Gradient Flows, Metastability and Interacting Particle Systems

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Kaveh Bashiri

aus

Bonn

Bonn, 2020

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

1. Gutachter: Prof. Dr. Anton Bovier
2. Gutachter: Prof. Dr. Matthias Erbar

Tag der Promotion: 24.04.2020
Erscheinungsjahr: 2020

Es ist nicht das Wissen, sondern das Lernen,
nicht das Besitzen, sondern das Erwerben,
nicht das Dasein, sondern das Hinkommen,
was den größten Genuß gewährt.

– Carl Friedrich Gauß, (1777 - 1855)

Acknowledgements

Don't take for granted the love this life gives you.
When you get where you're going, don't forget to turn back around.
– Lori McKenna, 2016

There are no words that can describe my gratitude towards my supervisor, my mentor and my friend Prof. Dr. Anton Bovier. Ever since I met him as a tutor for his course "Einführung in die Wahrscheinlichkeitstheorie" in the winter term 2012-2013, I am blessed by his infinite support and his perfect guidance. I was immediately impressed by his strong mathematical intuition, his unlimited generosity and his great sense of humour. I am and I always will be grateful to have been a part of his group for the last years.

Moreover, I would like to thank my dear friend Prof. Dr. Georg Menz for his guidance towards the third part of my thesis. I benefit a lot from his great intuition and his expertise.

I also would like to thank Prof. Dr. Muhittin Mungan for always sharing with me his invaluable advices and experiences. I am grateful to have found a true friend in him.

Numerous thanks go out to

- Dr. Matthias Erbar for many useful discussions and for taking part in the Ph.D. defence committee.
- Prof. Dr. Martin Rumpf for being my mentor and for taking part in the Ph.D. defence committee.
- Prof. Dr. Waldemar Kolanus for taking part in the Ph.D. defence committee.

Throughout my Ph.D. program I had the chance to discuss mathematical (and also non-mathematical) content with numerous brilliant minds, which helped me to proceed and to grow further. Among these persons, I would like to thank Prof. Dr. Sergio Albeverio, Prof. Dr. Patrik Ferrari, Prof. Dr. Frank den Hollander, Prof. Dr. Dmitry Ioffe, Prof. Dr. Constanza Rojas-Molina, Dr. Sebastian Andres, Dr. Lorenzo Dello Schiavo, Dr. Max Fathi, Dr. Elena Pulvirenti, Dr. André Schlichting and Dr. Martin Slowik.

Furthermore, I would like to thank my current and my former colleagues from the Abteilung für Wahrscheinlichkeitstheorie, the so-called "W-Theorie Familie". Especially, I would like to thank Mei-Ling Wang for her unlimited kindness and support. Moreover, I am grateful for having been a participant in the unforgettable "Klassenfahrt" to Villa la Collina at Lago di Como.

I also wish to express my gratitude to the Collaborative Research Center 1060, the Hausdorff Center for Mathematics and the Bonn International Graduate School in Mathematics for their overwhelming support at every stage of my Ph.D. program.

Special thanks go out to my family and my friends. Their love, patience and support is an invaluable gift. Especially I would like to thank "her", Maedeh, my soulmate, my wife, the most beautiful soul on this planet, for being my inspiration and my motivation in every single step I take.

Finally, as my faith guides me throughout my whole life, I would like to thank God for blessing me with his light and his infinite love in every second of my life.

Summary

Many stochastic models exhibit a phenomenon called *metastability*. The first goal of this thesis is to study this phenomenon for certain classes of interacting particle systems. The second goal of this thesis is the following. Many models that are expected to exhibit metastable behaviour consist of a large number of particles. Thus, their dynamics takes place in a high-dimensional configuration space. It is then a typical idea to describe the system on the *macroscopic level* by introducing a *macroscopic order parameter*. In the case of *high-dimensional diffusion systems*, the *empirical distribution* turns out to be a suitable order parameter. The reason is that, under this mapping, the Markov property of the system is preserved. Hence, the macroscopic level is given by the infinite-dimensional space of probability measures. Therefore, in order to study the macroscopic behaviour, it is useful to have the structure of a *Riemannian manifold* on the space of probability measure. In the seminal papers [83] and [111], it is shown that the so-called *Wasserstein formalism* provides such a structure. The second goal of this thesis is to extend this Wasserstein formalism to a certain class of diffusion equations, and to use this formalism to build a rigorous bridge between the microscopic and the macroscopic level in the case of *local mean-field interacting diffusions*. It is left for future research to apply these results to study the metastable behaviour of the system on the macroscopic level.

The outline of this thesis is as follows. In Chapter I we provide a brief introduction to the main topics of this thesis. We briefly describe the phenomenon of *metastability*, explain the main steps in the construction of *Wasserstein gradient flows*, and illustrate the *Fathi-Sandier-Serfaty approach* by a simple example. Moreover, we provide a first formulation of the main results of this thesis.

In Chapter II we study the metastable behaviour of three modifications of the standard, two-dimensional Ising model. The first model is an anisotropic version of the Ising model, where the interaction energy takes different values on vertical and horizontal bonds. The second model adds next-nearest-neighbour attraction to the standard Ising model. In the third model, the magnetic field is assumed to have different alternating signs on even and on odd rows. The results of Chapter II were published as the paper [11].

In Chapter III we first establish a *gradient flow representation* for evolution equations that depend on a non-evolving parameter. These equations are connected to a *local mean-field interacting spin system*. We then use the gradient flow representation to prove a *large deviation principle* and a *law of large numbers* for the empirical process associated to this system. This is done by using the Fathi-Sandier-Serfaty approach. The results of Chapter III were published as the paper [13].

In Chapter IV we consider a system of N mean-field interacting diffusions that are driven by a single-site potential of the form $z \mapsto z^4/4 - z^2/2$. The strength of the noise is measured by $\varepsilon > 0$, and the strength of the interaction by $J > 1$. Choosing the *empirical mean*, $P : \mathbb{R}^N \to \mathbb{R}$, $Px = 1/N \sum_i x_i$, as the macroscopic order parameter, we show that the resulting macroscopic Hamiltonian admits two global minima, one at $-m_\varepsilon^\star \in (-\infty, 0)$, and one at $m_\varepsilon^\star \in (0, \infty)$. We are interested in the transition time to the hyperplane $P^{-1}(m_\varepsilon^\star)$, when the initial configuration is close to $P^{-1}(-m_\varepsilon^\star)$. The main result is a formula for this transition time, which is reminiscent of the celebrated *Eyring-Kramers formula* up to a multiplicative error term that tends to 1 as $N \uparrow \infty$ and $\varepsilon \downarrow 0$. Finally, we add estimates on this transition time in the case $\varepsilon = 1$ and for a large class of single-site potentials. The results of Chapter

IV are contained in the preprint [14] and are the result of a collaboration with Georg Menz (UCLA).

In Chapter V we again consider the system of Chapter IV in the case $\varepsilon = 1$ and for a large class of single-site potentials. This time, instead of the empirical *mean*, we choose the empirical *distribution* as the order parameter. We then prove some results about the basins of attraction in the macroscopic energy landscape. These results provide a first step towards the investigation of the metastable behaviour of the empirical process associated to (local) mean-field interacting diffusions, which we motivated above. The results of Chapter V are contained in the preprint [12].

Contents

Chapter I

Introduction

The goal of this introduction is to provide a motivation and a background for the main results in this thesis. The main three topics in this thesis are *metastability*, *Wasserstein gradient flows* and the *(Fathi-)Sandier-Serfaty approach*. In this introduction, we discuss the main ideas behind these topics, briefly comment on their historical background and introduce simple examples to illustrate the main ideas. Moreover, we provide a first formulation of the main results of the Chapters II–V.

This chapter is organized as follows. In Section I.1 we introduce the concept of *metastability*. We introduce its main elements, briefly discuss the most common mathematical approaches, and provide two simple examples to illustrate the so-called *potential-theoretic approach to metastability*, which is the basis of Chapter II and Chapter IV.

In Section I.2 we provide a brief introduction to the theory of *Wasserstein gradient flows*. As a motivation, we start with the construction of gradient flows in Euclidean spaces. Then we introduce the main elements of the construction of Wasserstein gradient flows, i.e., of gradient flows in the space of probability measures with finite second moment equipped with the so-called *Wasserstein distance*. The main observation is the close relation between Wasserstein gradient flows and solutions to *diffusion equations*. This section is the basis of our construction of *Wasserstein-like gradient flows* in Chapter III. We also comment on the possible application of the Wasserstein formalism in the study of metastability, which is aimed for future research.

In Section I.3 we use a simple example in the setting of the Wasserstein space to introduce the main ideas of the so-called *Sandier-Serfaty approach*. Indeed, it turns out that this example already contains many crucial ideas that are used in Chapter III, where we apply a slight extension of the Sandier-Serfaty approach to prove a *law of large numbers* and a *large deviation principle* for a sequence of *local mean-field interacting diffusions*.

In the Sections I.4–I.7 we introduce the setting of the Chapters II–V, respectively. Moreover, we provide a first formulation of the main results in these chapters.

In Section I.8 we list some open questions, related to the results in this thesis, that are aimed for future research.

Finally, at the end of this chapter, we introduce some notational conventions that we use throughout this thesis.

In this introductory treatment we only focus on the main ideas and omit most of the technical details.

I.1 Metastability

In this section we provide a brief introduction to the phenomenon of metastability. The main goal is to introduce the main elements of this phenomenon, and to give a first description of the rigorous study of metastability. Special emphasis is made on the so-called *potential-theoretic approach to metastability*, which is the basis of Chapter II and Chapter IV.

This section is organized as follows. We start in Subsection I.1.1 by introducing a simple thought experiment, from which we deduce a paradigmatic description of metastability. This description should act as a guiding rule throughout the whole thesis. Then, in Subsection I.1.2, we state the main goals in the rigorous mathematical study of metastability, and briefly explain the most common approaches to tackle metastability. In order to exemplify these ideas we consider in Subsection I.1.3 two specific models, where the main elements of metastability can be observed very easily. The first model is a *one-dimensional diffusion in a double-well landscape at low temperature*, and the second one is *the Curie-Weiss model*. In order to analyse the metastable behaviour of these models, we apply the so-called *potential-theoretic approach to metastability*. This should provide a motivation for the application of this approach in the Chapters II and IV.

I.1.1 Paradigmatic description

In many physical, biological or chemical systems, one can observe a universal phenomenon called *metastability*. In the following, we first describe this phenomenon in a very simple thought experiment, which, although it might seem trivial, already provides many insights into the rigorous study of metastability. We then use this example to formulate a general paradigmatic description of metastability.

Suppose that, in a two-dimensional world, two valleys are separated by a mountain, and a ball is located in the base of the left valley as in Figure I.1 a). Due to thermal fluctuations, such as strong wind, every now and then, the ball is moved to the left and to the right (Figure I.1 b)). However, gravitational force constantly pushes the ball to the base of the valley. But eventually, after a very long time, the ball will be moved so much to the right (for example due to a hurricane) that it reaches the peak of the mountain (Figure I.1 c)), and falls into the right valley and reaches its base very fast (Figure I.1 d)).

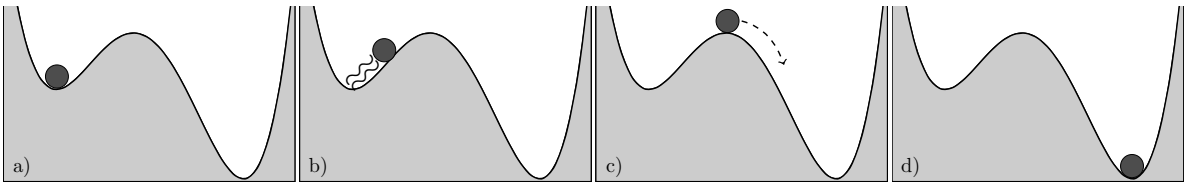


Figure I.1: A paradigmatic example of metastability.

This basic picture leads to a first description of metastability as follows. Suppose that the states of a system are associated to an energy functional $E : S \to \mathbb{R}$, where S denotes the state space of the system. For simplicity, we assume that S is connected and that E admits exactly one global minimum at $s \in S$ and exactly one local minima at some point $m \in S$ such that $E(m) < E(s)$. In the situation of Figure I.1, the role of the energy functional is played by the mountain landscape. Moreover, suppose that there is some source of noise in the system. In Figure I.1 the noise was given by meteorological events, such as winds and hurricanes.

Then we say that the system persists in a *metastable state* if it is trapped in a neighbourhood of m, that is, it is trapped around a state that is associated to a local minimum of E. In order to leave this valley around the local minimum, the energy of the system has to be increased. Consequently, the system resides around this metastable state for a relatively long time. However, due to the presence of noise, after many unsuccessful attempts, the system is finally able to free itself from this valley, and to make the crossover to the state s, i.e., it reaches a state which is associated to a global minimum of E. This state is called a *stable state* of the system. Often, this crossover is triggered by the fact that the system reaches a *critical state*. In Figure I.1 this critical state was given by the mountain peak.

Moreover, provided that the dynamical system is of Markovian nature, the (appropriately rescaled) transition time to the stable state is often shown to be (approximately) exponentially distributed. This comes from the fact that the system returns to the metastable state many times before it eventually makes the crossover to the stable state.

Another way to understand the above description is to look at metastability as a "dynamical signature of a first-order phase transition"[1]. More precisely, suppose that a phase diagram is separated into two areas corresponding to the phases associated to the states m and s, respectively. Suppose that, starting from the phase associated to m, a parameter is varied across the phase transition curve. Then, the system resides for a relatively long and random time in the phase associated to m before it makes the transition to the phase associated to s. The dynamical description of this situation is the same as the one we gave after Figure I.1.

Of course, in almost all metastable systems of practical relevance, the energy functional E is far more complex than in the paradigmatic descriptions we provided so far. For example, the system may possess several metastable and stable states, and there could be many different critical states in-between these states. Moreover, it may be that these states are given by submanifolds instead of single points, and that, as in the example of Chapter II, any path connecting these states has to pass other valleys in the energy landscape of smaller depth.

A standard example from physics, where a metastable behaviour can be observed, is the case of *over-saturated water vapour*. Here, below the critical temperature, the formation of a water droplet of critical length is needed in order to achieve the transition from the gas-phase to the liquid-phase. An analogous situation holds for *over-cooled liquids* and for *magnetic hysteresis*.

I.1.2 Mathematical approaches to metastability

Throughout the last decades a vast literature has been written in order to study the phenomenon of metastability in a mathematically rigorous way. The main goals in this field of research are

(i) to compute the average transition time from the metastable to the stable state,

(ii) to estimate this transition time in probability,

(iii) to show that the transition time normalized with its average is exponentially distributed,

(iv) to identify the typical paths for the transition from the metastable to the stable state, and

[1] [29, p. 5]

(v) to show that, in order to make the transition from the metastable to the stable state, the system has to pass some critical states.

Mainly three methods have been crystallized to be very powerful to tackle these problems. We briefly introduce these approaches in the following.

The path-wise approach. The first method is called the *path-wise approach* to metastability and was initiated by Cassandro, Galves, Olivieri and Vares in [38]. Motivated by the *Freidlin-Wentzell theory* (see [74]), one uses large deviation estimates on the path space to identify the most likely paths of the system for the transition from the metastable to the stable state. More precisely, the large deviation principle yields that, with high probability, the most typical paths for this transition are close to the unique minimizer of the corresponding rate functional. In many models this minimizer is given by the time-reverse of the gradient flow for the associated free energy functional.[2] Consequently, the path-wise approach leads to a very detailed description of the typical paths that are realized by the system in the transition from the metastable to the stable state. However, a drawback of this approach is that the average transition time can only be computed up to logarithmic equivalence. For an extensive treatment on the path-wise approach to metastability, the reader is referred to [41], [73], [99] or [109].

The spectral approach. The second method is known as the *spectral approach* to metastability, and was initiated by Davies in the papers [42],[43], [44] and [45]. It is based on a detailed analysis of the spectrum of the generators of reversible Markov processes. The main observation is that the metastable behaviour of such processes is closely related to a certain decomposition of the spectrum of the generator into clusters. We refer to [76] and [77] for more details and further developments on this approach.

The potential-theoretic approach. The third method is the *potential-theoretic approach* to metastability, which was initiated by Bovier, Eckhoff, Klein and Gayrard in the seminal papers [30], [31] and [32]. The main idea in this approach is to translate the problem into the language of *electric networks*, and then to use potential theory to obtain useful representations for the quantities of interest. In particular, one obtains that the average transition time from the metastable to the stable state can be expressed in terms of *capacities*, for which powerful variational principles are known. Hence, the computation of sharp estimates basically reduces to an appropriate choice of test functions in those variational principles. This method is the basis of the Chapters II and IV in this thesis, and will be explained in further details in these chapters and in the examples from Subsection I.1.3. Moreover, the reader is referred to the monograph [29] by Bovier and den Hollander for an comprehensive treatment of this approach.

We also mention that there are two relatively new methods to tackle metastability that are derivations from the potential-theoretic approach. The first one is known in the literature as *the martingale approach* to metastability and was initiated in [16]. Here, one uses the quantities from the potential-theoretic approach to introduce a new definition of metastability, which is based on the fact that Markov processes are characterized as unique solutions of

[2]The relation between gradient flows and large deviation rate functionals will be investigated in detail in Chapter III in the infinite-dimensional setting of the so-called *Wasserstein space*.

martingale problems; see [92] for an introduction to this approach. The second method is called *the mean-difference approach* to metastability and was initiated in [102]. The main idea in this approach is to obtain a lower bound on the capacity in terms of the so-called *weighted transport distance*. The latter object is inspired by the theory of *optimal transportation*, and describes the cost between two measures in terms of their interpolation; see [102, 4.1].

I.1.3 Two simple examples

One-dimensional diffusion in a double-well landscape at low temperature. The classic (and probably also the easiest) example of a mathematical model, which possesses metastable behaviour is given by the one-dimensional stochastic differential equation

$$dx_t \;=\; -\psi'(x_t)\,dt \;+\; \sqrt{2\varepsilon}\,dB_t, \tag{I.1.1}$$

where B is a one-dimensional Brownian motion, $\varepsilon > 0$, and $\psi \in C^2(\mathbb{R})$ is a typical *double-well potential*, i.e., $\lim_{x\to\pm\infty}\psi(x) = \infty$ and ψ admits three critical points at $-\infty < m < z^* < s < \infty$ such that $\psi''(m), \psi''(s) > 0$ and $\psi''(z^*) < 0$. That is, ψ is of the form given in Figure I.2. In the paradigmatic description of Subsection I.1.1, ψ plays the role of the energy functional E, and the Brownian motion, B, plays the role of the noise. We interpret the parameter ε as the *temperature* of the system, since it measures the strength of the Brownian noise.

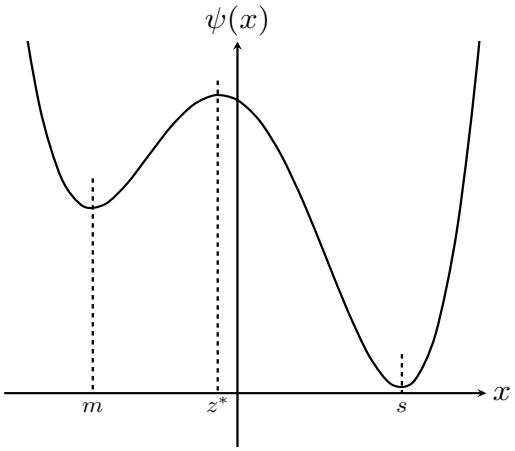


Figure I.2: A typical double-well potential.

We are interested in the average transition time of the system from the state m to the state s in the low-temperature regime. That is, we want to compute the asymptotic value of $\mathbb{E}_m[\tau_s]$ in the limit as $\varepsilon \downarrow 0$, where τ_s denotes the first hitting time of the state s. To do this, we apply the potential-theoretic approach as it was done in [32] (in a more general setting than here). However, we only sketch the main steps in the computations. The omitted details can be found in [29, Chapter 7 and Chapter 11] or [32]. See also Chapter IV in this thesis, where this method is used in a similar way.

Consider the Dirichlet problem given by

$$\begin{aligned} \mathcal{L}_\varepsilon h(x) \;&=\; 0 \qquad \text{for } x \in (m,s), \\ h(m) \;&=\; 1, \\ h(s) \;&=\; 0, \end{aligned} \tag{I.1.2}$$

where $\mathcal{L}_\varepsilon$ is the probability generator corresponding to the diffusion (I.1.1). It is a well-known fact in potential theory that this Dirichlet problem admits a unique solution, $h^\star_{m,s}$, which is called *equilibrium potential* of the *capacitor* (m, s). Moreover, $h^\star_{m,s}$ admits the probabilistic interpretation that for each $x \in (m, s)$, it is equal to the probability that the system returns to the metastable state m before it makes the transition to the stable state s; see [29, 7.15]. In the language of electrostatics, $h^\star_{m,s}$ can be seen as the electrostatic potential corresponding to the electric field between the plates m and s. Furthermore, in the particular case of one-dimensional reversible diffusions, we have an explicit representation formula for the equilibrium potential given by

$$h^\star_{m,s}(x) \;=\; \frac{\int_x^s \mathrm{e}^{\frac{1}{\varepsilon}\psi(z)}\,dz}{\int_m^s \mathrm{e}^{\frac{1}{\varepsilon}\psi(z)}\,dz} \qquad \text{for } x \in (m, s) \tag{I.1.3}$$

(cf. [29, (7.2.88)]). In a similar way, using that the function $x \mapsto \mathbb{E}_x[\tau_s]$ is also a solution of a certain Dirichlet problem (see [29, 7.30]), we can show that the expected transition time $\mathbb{E}_m[\tau_s]$ can be represented as

$$\mathbb{E}_m[\tau_s] \;=\; \frac{\int_m^s h^\star_{m,s}(z)\ \mathrm{e}^{-\frac{1}{\varepsilon}\psi(z)}\,dz}{\varepsilon \int_m^s (h^\star_{m,s})'(z)^2\ \mathrm{e}^{-\frac{1}{\varepsilon}\psi(z)}\,dz}. \tag{I.1.4}$$

Then, in view of (I.1.3) and (I.1.4), standard Laplace asymptotics yields that

$$\mathbb{E}_m[\tau_s] \;=\; \frac{2\pi}{\sqrt{|\psi''(m)|\psi''(z^*)}}\ \mathrm{e}^{\frac{1}{\varepsilon}(\psi(z^*)-\psi(m))}\ (1 \;+\; o_\varepsilon(1))\,, \tag{I.1.5}$$

where $o(1)$ stands for a term, which converges to 0 as $\varepsilon \downarrow 0$. Equation (I.1.5) is known in the literature as *Kramers formula*. Its multi-dimensional generalizations are called *Eyring-Kramers formula*. Such results are also known in the literature as *Kramers' law*.

We now provide some remarks on the historical background on the derivation of the Eyring-Kramers formula. First, based on chemical experiments, Arrhenius found out in [6] that the logarithmic asymptotics of the average transition time is given by the energy barrier that the system has to overcome to make the crossover to the valley corresponding to the global minimum, i.e.,

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}_m[\tau_s] \;=\; \psi(z^*) - \psi(m). \tag{I.1.6}$$

A first rigorous proof for this claim (in the multi-dimensional setting) was given in [126] by using the path-wise approach to metastability. We refer to [109] for more details on the path-wise approach to metastability for diffusion models at low temperature.

The system (I.1.1) has also been the object of study in the groundbreaking paper [90] by Kramers in the context of chemical reactions. Among other results, Kramers derived the Kramers formula, (I.1.5), for the one-dimensional model. That is, he improved (in dimension 1) Arrhenius' conjecture (equation (I.1.6)) by identifying the prefactor in front of the exponential term $\mathrm{e}^{\frac{1}{\varepsilon}(\psi(z^*)-\psi(m))}$.

In the multi-dimensional case, the Eyring-Kramers formula was first conjectured in [69] and [78] in the context of quantum statistical mechanics. The first rigorous proof was given by Sugiura in the papers [123] and [124] for the special case that all local minima of the potential function are of the same height (i.e., in the one-dimensional case of (I.1.1), the proof was

given under the assumption that $\psi(m) = \psi(s)$). The proofs in [123] and [124] are based on studying the asymptotics of the principal eigenvalue of the generator $\mathcal{L}_\varepsilon$. The first proof of the Eyring-Kramers formula in the full generality as it was conjectured in [78], is given in [32] via the potential-theoretic approach. Since then, the results in [32] have been generalized in many directions, including the infinite-dimensional case of stochastic differential equations (see [9], [10], [19] and [22]), the case when the saddle points are not quadratic (see [21]) or the case of non-reversible diffusions (see [27], [93], and [94]).

Due to the fact that there is by now a vast literature devoted to the study of metastability for (finite or infinite-dimensional) diffusion models at low temperature, the previous review is far from complete, and we refer to [18], [29] and [109] for a more detailed historical background. The main goal here was to list the main contributions for the three approaches listed in Subsection I.1.2, and to emphasize the usefulness of the potential-theoretic approach for the derivation of sharp asymptotics of the average transition time between metastable and stable states.

The Curie-Weiss model. A fundamental idea of statistical mechanics is the reduction of a high-dimensional, microscopic system to a low-dimensional state via a suitable mapping. This map is often called the *macroscopic order parameter*, and the whole procedure is called *coarse-graining*. Probably the easiest example for coarse-graining in a mathematical model is the *Curie-Weiss model of a ferromagnet*. In the following we first define the microscopic model and introduce the macroscopic order parameter. Then we analyse the metastable behaviour of the coarse-grained process by applying the potential-theoretic approach. As in the previous example, we only provide a sketch of the computations here. More details can be found in [29, Part V] and [30].

The state space of the Curie-Weiss model is given by $S_N = \{-1, +1\}^N$, and the *energy* (or *Hamiltonian*) of the system is given by

$$\mathrm{H}_N(\sigma) \;=\; -\frac{1}{2N}\sum_{i,j=1}^{N}\sigma_i\sigma_j \;-\; h\sum_{i=1}^{N}\sigma_i \quad \text{for } \sigma \in S_N, \tag{I.1.7}$$

where $h \in \mathbb{R}$. We consider a discrete-time Markov chain, $(\sigma(n))_{n\in\mathbb{N}}$, on S_N defined via the Metropolis transition probabilities given by

$$p_{\beta,N}(\sigma,\sigma') \;=\; \mathbb{1}_{\|\sigma-\sigma'\|_1=2}\,\frac{1}{N}\,\mathrm{e}^{-\beta\,[\mathrm{H}_N(\sigma')-\mathrm{H}_N(\sigma)]_+} \quad \text{for } \sigma \neq \sigma', \tag{I.1.8}$$

where $\beta > 0$ and $\|\cdot\|_1$ denotes the ℓ^1-norm on S_N. Consequently, the unique reversible measure for this Markov chain is given by the *Gibbs measure*

$$\mu_{\beta,N}(\sigma) \;=\; \frac{1}{Z_{\beta,N}}\,\mathrm{e}^{-\beta\,\mathrm{H}_N(\sigma)} \quad \text{for } \sigma \in S_N, \tag{I.1.9}$$

for some normalization constant $Z_{\beta,N}$.

This model is one of the simplest examples of a *mean-field-interacting* model, i.e., the interaction in this model is a function of the *empirical mean* defined by

$$m_N(\sigma) \;=\; \frac{1}{N}\sum_{i=1}^{N}\sigma_i \quad \text{for } \sigma \in S_N. \tag{I.1.10}$$

Indeed, the Hamiltonian H_N can be rewritten as

$$\mathrm{H}_N(\sigma) \;=\; -N\left(\frac{1}{2}m_N(\sigma)^2 \;+\; h\, m_N(\sigma)\right) \;=:\; N\, E(m_N(\sigma)). \tag{I.1.11}$$

This suggests to choose the map m_N as the macroscopic order parameter. It turns out that the pushed process, $(m_N(\sigma(n)))_{n\in\mathbb{N}}$, is a discrete-time Markov chain with state space

$$\Gamma_N \;=\; \{-1, -1+2N^{-1}, \dots, 1-2N^{-1}, 1\} \;\subset\; [-1,1], \tag{I.1.12}$$

and with transition probabilities given by

$$r_{\beta,N}(m,m') \;=\; \mathrm{e}^{-\beta N\,[E(m')-E(m)]_+}\;\left(\frac{1-m}{2}\mathbb{1}_{m'=m+2N^{-1}} + \frac{1+m}{2}\mathbb{1}_{m'=m-2N^{-1}}\right) \tag{I.1.13}$$

for $m \neq m'$. The unique reversible measure for $(m_N(\sigma(n)))_{n\in\mathbb{N}}$ is given by

$$\nu_{\beta,N}(m) \;=\; \frac{1}{Z_{\beta,N}}\,\mathrm{e}^{-\beta\, N\, f_{\beta,N}(m)} \qquad \text{for } m \in \Gamma_N, \tag{I.1.14}$$

where, for $m \in [-1,1]$,

$$f_{\beta,N}(m) \;=\; f_\beta(m) \;+\; \frac{1}{2N\beta}\log\left(\frac{\pi N(1-m^2)}{2}\right)(1+o_N(1))\,, \tag{I.1.15}$$

for some function $f_\beta : \mathbb{R} \to \mathbb{R}$ and where $o_N(1)$ stands for a term that converges to 0 as $N \to \infty$. If $\beta > 1$ and $|h|$ is small enough, it can be shown that f_β is a double-well potential (as in Figure I.2) with two local minima at some points $-1 \le m_-^\star < m_+^\star \le 1$. This indicates that the process $(m_N(\sigma(n)))_{n\in\mathbb{N}}$ admits metastable behaviour, and we are interested in the average transition time, $\mathbb{E}_{m_+^\star(N)}[\tau_{m_-^\star(N)}]$, of the process from the state $m_+^\star(N)$ to the state $m_-^\star(N)$, where $m_+^\star(N)$ and $m_-^\star(N)$ are the points in Γ_N that are closest to $m_+^\star$ and $m_-^\star$, respectively.

In order to compute this average transition time, as in the example of the one-dimensional diffusion in a double-well landscape, we apply the potential-theoretic approach. Note that $(m_N(\sigma(n)))_{n\in\mathbb{N}}$ is a one-dimensional nearest-neighbour random walk. Therefore, proceeding as in [29, Section 7.1.4], the potential-theoretic approach provides an explicit representation of $\mathbb{E}_{m_+^\star(N)}[\tau_{m_-^\star(N)}]$ given by

$$\mathbb{E}_{m_+^\star(N)}[\tau_{m_-^\star(N)}] \;=\; \sum_{\substack{m,m'\in\Gamma_N\,:\,m\le m',\\ m_-^\star(N)<m\le m_+^\star(N)}} \frac{\nu_{\beta,N}(m')}{\nu_{\beta,N}(m)\, r_{\beta,N}(m, m-2N^{-1})}. \tag{I.1.16}$$

Using standard techniques (see [29, Chapter 13]), we can compute the asymptotic value of the sum, and obtain that

$$\begin{aligned}\mathbb{E}_{m_+^\star(N)}[\tau_{m_-^\star(N)}] \;&=\; \mathrm{e}^{\beta\, N\,[f_\beta(z^\star)-f_\beta(m_+^\star)]} \\ &\quad\times\; \frac{1}{1-z^\star}\sqrt{\frac{1-(z^\star)^2}{1-(m_+^\star)^2}}\;\frac{\pi\, N}{\beta\sqrt{|f_\beta''(z^\star)|\, f_\beta''(m_+^\star)}}\;(1+o_N(1))\,,\end{aligned} \tag{I.1.17}$$

where $z^\star \in (m_-^\star, m_+^\star)$ denotes the saddle point as in Figure I.2.

This result has been generalized in many ways. For example, in [24], [25] and [30] the magnetic field h is replaced by certain random variables, and in [35] and [53] the underlying graph in (I.1.7) (which is the complete graph) is replaced by the Erdös-Réyni random graph. Moreover, in [119] the metastable behaviour of the Potts version of the Curie-Weiss model is studied. We refer to [29, Part V] for further references. The analogous situation in a continuous setting, i.e., a system of mean-field interacting diffusions, is studied in Chapter IV of this thesis.

I.2 Wasserstein gradient flows

It is well-known that many classes of diffusion equations can be represented as so-called *Wasserstein gradient flows*, i.e. as *gradient flows* in the space of probability measures equipped with the (L^2-) *Wasserstein distance*. This fact was first discovered in the seminal works [83] and [111], and has been formalized and extended to a large class of diffusion equations in [3].

There are mainly five arguments that speak in favour of the Wasserstein gradient flow representation for diffusion equations.

- The first one is that it entails a lot of useful properties such as contraction estimates (see Lemma I.6 or Theorem III.27), stability with respect to gamma-convergence (see [3, 11.2.1]), regularization estimates (see Theorem III.27), and a variational characterization as a minimum of an "energy-dissipation functional" (see Lemma I.8 or Theorem III.40).

- The second argument is that this formalism is strongly connected to certain functional inequalities such as the *HWI inequality*, the *log-Sobolev inequality*, the *transport inequality* or the *Poincaré inequality*; see, for instance, [2], [62] [64], [97], [112], [121] or [122]. There is by now a vast literature on these inequalities and it is known that they can be applied in many different fields. We refer to [79], [80], [95], [102] or [120] and references therein for more information on that. In this thesis, we do not consider functional inequalities.

- The third argument is that these representations can be used to study convergence (and large deviation principles) of sequences of evolution systems by using the so-called *(Fathi-)Sandier-Serfaty approach*. We explain this approach and its advantages in Section I.3.

- The fourth argument is that the Wasserstein formalism appears naturally in the setting of the *empirical distribution process* corresponding to *mean-field interacting diffusions*. We explain this in more detail at the beginning of Subsection I.6.4.

- The fifth argument is that the Wasserstein gradient flow is known to be "a natural and physically meaningful structure"[3] for certain diffusion equations. We provide an intuitive explanation of this in Remark I.10.

We now explain the main goal of this section. In Chapter III of this thesis we extend certain results from [3], and establish a gradient flow representation for evolution equations that depend on a non-evolving parameter. This is done by considering a slightly *modified Wasserstein distance*. The main ideas in Chapter III are the same as those in [3]. Therefore,

[3] [5, p. 421]

we provide in this section a brief introduction into the theory of Wasserstein gradient flows developed in [3]. In this way, we motivate the main ideas of Chapter III by the (simpler) classical setting of the Wasserstein space. Hence, this section should act as a guide for the proofs and the results from Chapter III.

The construction of Wasserstein gradient flows from [3] is introduced in Subsection I.2.2. In order to motivate it, we consider in Subsection I.2.1 the simple and well-known case of *gradient flows in Euclidean spaces*. Indeed, the construction of gradient flows in the purely metric framework of the Wasserstein space is inspired by the construction of gradient flows in Euclidean spaces. This should provide an intuition for the abstract metric objects defined in Subsection I.2.2.

In this introductory treatment, we only state the main results and omit most of the proofs. For more details, we refer to [3] and also to Chapter III, where, as we already mentioned, the main ideas are the same.

In this section we fix $d \in \mathbb{N}$, $\lambda \in \mathbb{R}$ and $T \in (0, \infty)$.

I.2.1 Motivation: The Euclidean case

This subsection is organized as follows. We first define *(Euclidean) gradient flows* (in (I.2.3)) and infer some immediate monotonicity property of the flows along the driving functional (in (I.2.4)). Then we study the question of existence in Lemma I.1, and show some contraction estimate (see (I.2.5)) which ensures uniqueness of gradient flows. Finally, we state a variational characterization of gradient flows as the unique minimum of an "energy-dissipation functional". This is known in the literature as the characterization of gradient flows as *curves of maximal slope*.

Euclidean gradient flows. Let $\phi \in C^1(\mathbb{R}^d)$ be λ-*convex*, i.e., for all $x, y \in \mathbb{R}^d$,

$$\phi(tx + (1-t)y) \;\leq\; (1-t)\phi(y) + t\phi(x) - t(1-t)\,\frac{\lambda}{2}\,|x-y|^2 \quad \text{for all } t \in [0,1]. \tag{I.2.1}$$

We say that $z : [0,T] \to \mathbb{R}^d$ is an *absolutely continuous curve* if there exists some function $m \in \mathrm{L}^2((0,T))$ such that

$$|z_s - z_t| \;\leq\; \int_s^t m(r)\,dr \quad \text{for all } 0 < s < t < T. \tag{I.2.2}$$

Consequently, we have that z is differentiable almost everywhere in $[0,T]$ (see [7, 4.4.1]). Then, the curve z is called *(Euclidean) gradient flow for the functional* ϕ if for all $t \in [0,T]$,

$$-\nabla\phi(z_t) \;=\; \dot{z}_t. \tag{I.2.3}$$

As a simple consequence of this definition we obtain that by the chain rule,

$$\frac{d}{dt}\phi(z_t) \;=\; \dot{z}_t\,\nabla\phi(z_t) \;=\; -|\nabla\phi(z_t)|^2. \tag{I.2.4}$$

Hence, the functional ϕ is non-increasing along the gradient flow curve.

Existence and uniqueness of Euclidean gradient flows. In the following lemma we study the question of existence and uniqueness of gradient flows for ϕ.

Lemma I.1 (Existence and uniqueness of Euclidean gradient flows)
Let $\phi \in C^1(\mathbb{R}^d)$ be λ-convex. Then, the following statements hold true.

(i) For $i = 1, 2$, let z^i be a gradient flow for ϕ with initial value $z_0^i \in \mathbb{R}^d$, i.e., $\lim_{t \downarrow 0} z_t^i = z_0^i$. Then, for all $t \in [0, T]$,

$$|z_t^1 - z_t^2| \;\leq\; \mathrm{e}^{-\lambda t}\, |z_0^1 - z_0^2|. \tag{I.2.5}$$

(ii) For all $z_0 \in \mathbb{R}^d$, there exists a unique gradient flow for ϕ with initial value z_0.

Proof. We first show part *(i)*. Since z^1 and z^2 are gradient flows for ϕ, and since ϕ is λ-convex, we have that for all $t \in (0, T]$,

$$\frac{d}{dt}|z_t^1 - z_t^2|^2 \;=\; -2(\nabla\phi(z_t^1) - \nabla\phi(z_t^2), z_t^1 - z_t^2) \;\leq\; -2\lambda\, |z_t^1 - z_t^2|^2. \tag{I.2.6}$$

Then, Gronwall's lemma yields part *(i)*.

To show part *(ii)*, note that the uniqueness claim immediately follows from part *(i)*, and that the existence is a consequence of a standard Picard-Lindelöf-iteration argument by using that ϕ is locally Lipschitz (see [8, 17.1.1]).

However, there is an alternative way to prove the existence claim, which was used in [49] in a purely metric setting; see also [3, p. 41] for more references. We now briefly introduce this method in the Euclidean setting and indicate that it leads to the existence of gradient flows. This should provide an intuitive reason why this method also leads to the existence of gradient flows in the purely metric framework of the so-called *Wasserstein space* that is introduced in Subsection I.2.2.

Fix a step size $\tau > 0$, and consider the *implicit Euler scheme* given by

$$z_n^\tau \;:=\; \operatorname*{argmin}_{y \in \mathbb{R}^d} \left(\phi(y) + \frac{1}{2\tau}|z_{n-1}^\tau - y|^2 \right) \;=:\; \operatorname*{argmin}_{y \in \mathbb{R}^d} \Upsilon^{n-1}(y) \tag{I.2.7}$$

for all $n \in \mathbb{N}$ such that $n\tau < T$, and with the piecewise constant interpolation

$$z_t^\tau := z_n^\tau \quad \text{for } t \in ((n-1)\tau, n\tau]. \tag{I.2.8}$$

Then, by computing the *Euler-Lagrange equation*, we observe that for all $t \in ((n-1)\tau, n\tau]$,

$$0 \;=\; \frac{d}{d\delta}\Big|_{\delta=0} \Upsilon^{n-1}(z_t^\tau + \delta y) \;=\; \left(\nabla\phi(z_t) + \frac{z_t^\tau - z_{t-\tau}^\tau}{\tau}, y \right) \quad \text{for all } y \in \mathbb{R}^d, \tag{I.2.9}$$

and hence,

$$\frac{z_t^\tau - z_{t-\tau}^\tau}{\tau} = -\nabla\phi(z_t^\tau). \tag{I.2.10}$$

Equation (I.2.10) is the *implicit time discretization* of (I.2.3), and therefore indicates that, as $\tau \downarrow 0$, the scheme defined by (I.2.7) and (I.2.8) converges to the solution of (I.2.3). □

The scheme defined by (I.2.7) and (I.2.8) was also used in [83] to show the existence of gradient flows in the purely metric framework of the *Wasserstein space*; see Lemma I.6. As in [49] and [83], this scheme can be used to *define* gradient flows as the limit of the scheme (provided that it converges). The advantage of this definition is that it requires both less assumptions on the ambient space (a purely metric framework is sufficient) and less assumptions on the regularity of the driving functional. This is known in the literature as the definition of gradient flows as *generalized minimizing movements* (see [3, 2.0.6]).

Characterization as curves of maximal slopes. There is also a third way to define gradient flows, which is based on the characterization given in the following lemma.

Lemma I.2 (Characterization as curves of maximal slopes)
Let $\phi \in C^1(\mathbb{R}^d)$ be λ-convex, and let $\mathcal{AC}((0,T);\mathbb{R}^d)$ denote the set of all absolutely continuous curves in $\mathbb{R}^d$. Let $\mathscr{J}_{\phi,T} : C((0,T);\mathbb{R}^d) \to [0,\infty]$ be defined by

$$\mathscr{J}_{\phi,T}[z] \;=\; \begin{cases} \phi(z_T) - \phi(z_0) + \frac{1}{2}\int_0^T \left(|\nabla\phi|^2(z_t) + |\dot{z}_t|^2\right) dt & \text{if } z \in \mathcal{AC}((0,T);\mathbb{R}^d), \\ \infty & \text{else.} \end{cases} \tag{I.2.11}$$

Let $z_0 \in \mathbb{R}^d$. For any curve $z \in \mathcal{AC}((0,T);\mathbb{R}^d)$ such that $\lim_{t\downarrow 0} z_t = z_0$, we have that $\mathscr{J}_{\phi,T}[z] \geq 0$. Equality holds if and only if z is the gradient flow for ϕ with initial value z_0.

Proof. Let $z \in \mathcal{AC}((0,T);\mathbb{R}^d)$. Then, by using the chain rule and Young's inequality, we have that

$$\phi(z_T) - \phi(z_0) \;=\; \int_0^T \nabla\phi(z_t)\dot{z}_t\, dt \;\geq\; -\frac{1}{2}\int_0^T \left(|\nabla\phi|^2(z_t) + |\dot{z}_t|^2\right) dt. \tag{I.2.12}$$

This shows that $\mathscr{J}_{\phi,T}$ is well-defined. Finally, equality holds in (I.2.12) if and only if z is the gradient flow for the functional ϕ. □

In Lemma I.2, we have seen that in Euclidean spaces and for sufficiently regular ϕ, the unique minimizer of an *energy-dissipation functional* is given by the gradient flow for ϕ. This fact is known for a more abstract setting than in this subsection (see [3, 2.3.1 and 2.3.3]). Therefore, we can alternatively *define* gradient flows as the minimizer (if it exists) of these functionals. This is known in the literature as the definition of gradient flows as *curves of maximal slope* (see [3, 1.3.2]).

Another advantage of this definition is its stability under the so-called *gamma-liminf-inequalities*. This was observed for the first time in [115] and [118] by Sandier and Serfaty in a general setting, and will be used in Chapter III of this thesis. In Section I.3 we show this stability result for a simple example in the Wasserstein space.

I.2.2 Gradient flows in the Wasserstein space

In this subsection, we translate the concepts and the results from the Euclidean setting in Subsection I.2.1 to the metric framework of the so-called *Wasserstein space*. The main goal is to introduce the main elements of the construction of gradient flows in the Wasserstein space, and to show their connection to weak solutions of the *Fokker-Planck equations*. More precisely, we show that gradient flows in the Wasserstein space for certain functionals are the unique weak solutions of diffusion equations of the form

$$\partial_t \rho_t \;=\; \Delta\rho_t + \operatorname{div}_x\left(\nabla V\, \rho_t\right), \tag{I.2.13}$$

where $V \in C^2(\mathbb{R}^d)$ is λ-convex (recall (I.2.1)) and bounded from below.

This subsection is organized as follows. We start by defining the *Wasserstein distance* and the *Wasserstein space* in (I.2.14) and (I.2.15), respectively. Then, we introduce *absolutely continuous curves* in the Wasserstein space and state their connection to solutions of the *continuity equation* in Lemma I.3. This will be a key element to build the bridge to (I.2.13).

Afterwards, we define the notion of *Wasserstein gradient flows* in Definition I.5 and state their existence and uniqueness in Lemma I.6. The latter is a consequence of the same type of contraction estimate as in (I.2.5). In Lemma I.8 we state the characterization as curves of maximal slopes, and in Lemma I.9 we build the bridge to (I.2.13). Then, we mention recently developed extensions of the previous results to other evolution equations and stochastic processes. Finally, we discuss possible applications of the Wasserstein formalism to study the metastable behaviour of stochastic processes.

The Wasserstein space. Initiated in [84], [85] and [104], the theory of optimal transportation has become a useful tool in numerous fields such as physics, partial differential equations or geometry; see [127] for more details on applications. In this thesis, we are interested in the particular case of the *Wasserstein distance*, where the cost function in the *Monge-Kantorovich formulation of optimal transportation* (see [127, Chapters 4 and 5]) is given by a distance. More precisely, the Wasserstein distance W_2 on the space of probability measures on $\mathbb{R}^d$, $\mathcal{M}_1(\mathbb{R}^d)$, is defined by

$$W_2^2(\mu,\nu) \;=\; \inf_{\gamma\in \mathrm{Cpl}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \, d\gamma(x,y) \quad \text{for } \mu,\nu\in\mathcal{M}_1(\mathbb{R}^d), \tag{I.2.14}$$

where $\mathrm{Cpl}(\mu,\nu)$ denotes the space of all probability measures on $(\mathbb{R}^d)^2$ that have μ and ν as marginals. We denote by $\mathrm{Opt}(\mu,\nu)\subset\mathrm{Cpl}(\mu,\nu)$ the set of all measures that realize the infimum in (I.2.14), and call these measures *optimal plans*; see [127, 4.1] for the existence of optimal plans.

It turns out that, restricted to the *Wasserstein space* $\mathcal{P}_2(\mathbb{R}^d)\subset\mathcal{M}_1(\mathbb{R}^d)$ defined by

$$\mathcal{P}_2(\mathbb{R}^d) := \left\{ \mu\in\mathcal{M}_1(\mathbb{R}^d) \,\middle|\, \int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty \right\}, \tag{I.2.15}$$

the Wasserstein distance satisfies the axioms of a metric. Moreover, it is shown in [127, 6.18] that the space $\mathcal{P}_2(\mathbb{R}^d)$ equipped with the Wasserstein distance is even a *Polish space*.

Another useful fact is the following characterization of convergence in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ (cf. [127, 6.8]). Let $(\mu^n)_{n\in\mathbb{N}}\subset\mathcal{P}_2(\mathbb{R}^d)$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$. Then we have that $\lim_{n\to\infty} W_2^2(\mu^n,\mu)=0$ if and only if

$$\mu^n \rightharpoonup \mu \quad \text{and} \quad \lim_{n\to\infty}\int_{\mathbb{R}^d} |x|^2 \, d\mu^n \;=\; \int_{\mathbb{R}^d} |x|^2 \, d\mu, \tag{I.2.16}$$

where we write $\mu^n\rightharpoonup\mu$ and say that μ^n *converges weakly to* μ *in* $\mathcal{M}_1(\mathbb{R}^d)$ if

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} f\, d\mu^n \;=\; \int_{\mathbb{R}^d} f\, d\mu \quad \text{for all continuous and bounded } f:\mathbb{R}\to\mathbb{R}. \tag{I.2.17}$$

In particular, for all $c\in(0,\infty)$, the set

$$\left\{ \mu\in\mathcal{P}_2(\mathbb{R}^d) \;\middle|\; \int_{\mathbb{R}^d} |x|^4 \, d\mu \le c \right\} \text{ is compact in } (\mathcal{P}_2(\mathbb{R}^d), W_2). \tag{I.2.18}$$

Absolutely continuous curves. Analogously to Subsection I.2.1, gradient flows in the Wasserstein space are required to have enough regularity, namely to be *absolutely continuous curves*. In the Wasserstein space, we say that a curve $(\mu_t)_{t\in[0,T]} \subset \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous if there exists some function $m \in \mathrm{L}^2((0,T))$ such that

$$W_2(\mu_s,\mu_t) \;\le\; \int_s^t m(r)\,dr \quad \text{for all } 0<s<t<T. \tag{I.2.19}$$

We denote the set of all absolutely continuous curves in $(\mathcal{P}_2(\mathbb{R}^d),W_2)$ by $\mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}^d))$. It is shown in [3, 1.1.2] that for all $\mu \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}^d))$, there exists $|\mu'| \in \mathrm{L}^2((0,T))$, called the *metric derivative* of $(\mu_t)_{t\in[0,T]}$, such that

$$|\mu'|(t) \;=\; \lim_{s\to t} \frac{W_2(\mu_s,\mu_t)}{|s-t|} \quad \text{for almost every } t\in(0,T). \tag{I.2.20}$$

An important observation is that absolutely continuous curves in $(\mathcal{P}_2(\mathbb{R}^d),W_2)$ are characterized as distributional solutions of the so-called *continuity equation*. This characterization is the key fact to build the bridge to the diffusion equation (I.2.13), and is given in the following lemma. The proof of this result is given in [3, Chapter 8].

Lemma I.3 (Absolutely continuous curves and the continuity equation)
The curve $(\mu_t)_{t\in(0,T)} \subset \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous in $(\mathcal{P}_2(\mathbb{R}^d),W_2)$ if and only if there exists a vector field $v:(0,T)\times\mathbb{R}^d\to\mathbb{R}^d$ such that

- $t \mapsto \|v_t\|_{\mathrm{L}^2(\mu_t)} \;\in\; \mathrm{L}^2((0,T))$,
- $\partial_t\mu_t + \mathrm{div}_x(\mu_t\, v_t) \;=\; 0$ *in* $(0,T)\times\mathbb{R}^d$ *in the sense of distributions, i.e., for all* $\varphi \in C_c^\infty((0,T)\times\mathbb{R}^d)$,

$$\int_{(0,T)\times\mathbb{R}^d} \Big(\partial_t\varphi_t(x) + \langle \nabla_x\varphi_t(x), v_t(x)\rangle\Big) d\mu_t(x)dt \;=\; 0, \tag{I.2.21}$$

  where div_x *and* ∇_x *denote the divergence and the gradient operator with respect to the space variable* x*, respectively, and*

- $v_t \;\in\; \overline{\{\nabla_x\varphi \,|\, \varphi\in C_c^\infty(\mathbb{R}^d)\}}^{\mathrm{L}^2(\mu_t)}$ *for almost every* t.

Moreover, $\|v_t\|_{\mathrm{L}^2(\mu_t)} = |\mu'|(t)$ *for almost every* t *and* v *is uniquely determined almost everywhere with respect to the Lebesgue measure on* $(0,T)$*. This vector field* v *is called* tangent velocity field*, and for* $t\in[0,T]$*, the space*

$$\mathrm{Tan}_{\mu_t}\mathcal{P}_2(\mathbb{R}^d) \;:=\; \overline{\{\nabla_x\varphi \,|\, \varphi\in C_c^\infty(\mathbb{R}^d)\}}^{\mathrm{L}^2(\mu_t)} \tag{I.2.22}$$

is called the tangent space *at* μ_t.

An intuitive picture for this result is given as follows. Suppose that μ_t describes the density of a cloud of gas at time t. Then, among all vector fields that describe the velocity of the particles, the tangent velocity field v_t from Lemma I.3 is the one with minimal total kinetic energy $\int_0^T \|v_t\|^2_{\mathrm{L}^2(\mu_t)}\,dt$ (cf. [59, p. 5]).

Wasserstein gradient flows. We would like to translate the definition, (I.2.3), of a gradient flow in the Euclidean setting into the present metric framework of the Wasserstein space. However, it is a priori not clear how to introduce a differentiable structure here. The main idea in the groundbreaking paper [111] by Felix Otto is to solve this problem by inducing a formal *Riemannian structure* on the space $\mathcal{P}_2(\mathbb{R}^d)$. More precisely, formally, by using the notion of the tangent space from Lemma I.3, he introduced a *metric tensor* in order to define the *gradient* of a functional on $\mathcal{P}_2(\mathbb{R}^d)$ as it is done in Riemannian geometry. From this notion of gradient, the notion of *gradient flows* is defined analogously to (I.2.3). His fundamental observation was that, as a consequence of this construction, for certain type of functionals (such as the *relative entropy* defined in (I.2.34)) these gradient flows are the solutions of diffusion equations such as (I.2.13).

Inspired by Otto's formal point of view, the corresponding rigorous construction was later introduced in the monograph [3] by Ambrosio, Gigli and Savaré. However, instead of defining a gradient on the Wasserstein space, they relied on the notion of *subdifferentials*. The reason is that, on the one hand, its conditions are easier to verify (since it only demands lower bounds instead of equalities), and on the other hand, it requires less regularity assumptions on the corresponding functional so that it is possible to consider a larger class of gradient flows.

In this introductory treatment, we roughly sketch the construction introduced in [3]. We only provide the main ideas here. For more details, we refer to [3] and Chapter III in this thesis, where we adapt the notions from [3] in order to introduce a differentiable structure on a modified Wasserstein space.

First we define the notion of *subdifferentials in the Wasserstein space* (cf. [3, Chapter 10] and Definition III.21).

Definition I.4 (Subdifferentials in the Wasserstein space)
Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be proper[4] *and lower semi-continuous with respect to W_2. Let $\mu \in D(\phi) \cap \mathcal{P}_2(\mathbb{R}^d)$ (i.e. $\phi(\mu) < \infty$) and let $\xi \in \mathrm{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$, where*

$$\mathrm{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) \;=\; \overline{\{\nabla_x \varphi \,|\, \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{\mathrm{L}^2(\mu)}. \tag{I.2.23}$$

Then we say that ξ belongs to the set of (strong) subdifferentials of ϕ at μ, *and write $\xi \in \partial\phi(\mu)$, if*

$$\phi(\mathrm{T}_\# \mu) - \phi(\mu) \;\geq\; \int_{\mathbb{R}^d} \xi\,(\mathrm{T} - \mathrm{Id})\; d\mu \;+\; o(\|\mathrm{T} - \mathrm{Id}\|_{\mathrm{L}^2(\mu)}) \quad \text{as } \|\mathrm{T} - \mathrm{Id}\|_{\mathrm{L}^2(\mu)} \to 0, \tag{I.2.24}$$

where $\mathrm{T}_\#\mu$ denotes the image measure of μ under the map $\mathrm{T} \in \mathrm{L}^2(\mu)$.

From this notion of subdifferentials, the definition of *Wasserstein gradient flows* is an easy adaptation of (I.2.3), and is given as follows.

Definition I.5 (Gradient flows in the Wasserstein space) *Let $(\mu_t)_{t\in[0,T]}$ be absolutely continuous in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ with corresponding tangent velocity field v. Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be proper and lower semi-continuous with respect to W_2. Then $(\mu_t)_{t\in[0,T]}$ is called* (Wasserstein) gradient flow for ϕ *with* initial value $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *if*

$$-v_t \;\in\; \partial\phi(\mu_t) \;\; \text{for a.e. } t \in (0,T) \qquad \text{and} \qquad \lim_{t\downarrow 0} W_2(\mu_t, \mu_0) \;=\; 0. \tag{I.2.25}$$

[4] We say that a functional $\phi : X \to (-\infty, \infty]$ on a Polish space (X, d) is *proper* if $\phi(\mu) > -\infty$ for all $\mu \in X$ and there exists $\mu \in X$ such that $\phi(\mu) < \infty$.

Existence and uniqueness of Wasserstein gradient flows. In this paragraph we translate the results (I.2.4) and Lemma I.1 from the Euclidean setting to the present metric framework of the Wasserstein space.

Analogously to the condition (I.2.1) in the Euclidean setting, the driving functionals of the Wasserstein gradient flows are required to satisfy some convexity property. In this framework this property is called *λ-convexity along (generalized) geodesics* or *strong λ-convexity*, where $\lambda \in \mathbb{R}$. In order to avoid too much terminology in this introductory treatment, we omit the precise definition of this property, and refer to [3, 9.1.4 and 9.2.4] and Definition III.19 in this thesis.

We are now in the position to state the existence and uniqueness of gradient flows in the Wasserstein space.

Lemma I.6 (Existence and uniqueness of Wasserstein gradient flows)
Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be proper, strongly λ-convex, lower semi-continuous with respect to W_2 and coercive[5]*. Then the following statements hold true.*

(i) (Existence) *For each $\mu_0 \in \overline{D(\phi)}$, there exists a gradient flow for ϕ with initial value μ_0.*

(ii) (λ-contraction and uniqueness) *Let $(\mu_t)_{t\in(0,T)}$ and $(\nu_t)_{t\in(0,T)}$ be gradient flows for ϕ with initial values $\mu_0 \in \overline{D(\phi)}$ and $\nu_0 \in \overline{D(\phi)}$, respectively. Then, for all $t \in (0,T)$,*

$$W_2(\mu_t, \nu_t) \;\leq\; \mathrm{e}^{-\lambda t} W_2(\mu_0, \nu_0). \tag{I.2.27}$$

In particular, for each $\mu_0 \in \overline{D(\phi)}$, the gradient flow for ϕ with initial value μ_0 is unique.

(iii) (Monotonicity along gradient flows) *Let $(\mu_t)_{t\in(0,T)}$ be the gradient flow for ϕ with initial value $\mu_0 \in \overline{D(\phi)}$. Then, for almost every $t \in (0,T)$*

$$\frac{d}{dt}\phi(\mu_t) \;=\; -\|v_t\|^2_{\mathrm{L}^2(\mu_t)}. \tag{I.2.28}$$

Proof. The proof is given in [3, 11.2.1]. We only note that it is based on the following *implicit Euler scheme*, which we already motivated in the proof of Lemma I.1 in the Euclidean setting. Let $\mu_0 \in \overline{D(\phi)}$ and let $\tau > 0$. Define recursively:

$$\begin{cases} \mu_0^\tau \;:=\; \mu_0, \\ \mu_n^\tau \;\in\; \operatorname*{argmin}_{\nu\in\mathcal{P}_2(\mathbb{R}^d)} \left(\phi(\nu) + \frac{1}{2\tau} W_2(\mu_{n-1}^\tau, \nu)^2\right) \text{ for } n \in \mathbb{N}, \end{cases} \tag{I.2.29}$$

and define the piecewise constant interpolating trajectory $(\bar{\mu}_t^\tau)_{t\in[0,T]}$ by

$$\begin{cases} \bar{\mu}_0^\tau \;:=\; \mu_0, \\ \bar{\mu}_t^\tau \;:=\; \mu_n^\tau \quad \text{for } t \in ((n-1)\tau, n\tau] \text{ for all } n \in \mathbb{N} \text{ such that } n\tau \leq T. \end{cases} \tag{I.2.30}$$

Then [3, 11.1.4 and 11.2.1] yields the convergence of this scheme with respect to W_2 towards a curve $(\mu_t)_{t\in(0,T)} \in \mathcal{AC}((0,T); \mathcal{P}_2(\mathbb{R}^d))$ which satisfies (I.2.25) and the claims *(ii)* and *(iii)*. □

[5] We say that a functional $\phi : X \to (-\infty, \infty]$ on a Polish space (X, d) is *coercive* if there exists $\mu^* \in X$ and $r^* > 0$ such that

$$\inf\{\phi(\nu) \,|\, \nu \in X\,,\, \mathrm{d}(\nu, \mu^*) \leq r^*\} > -\infty \quad \text{(cf. [3, (2.4.10)])}. \tag{I.2.26}$$

Characterization as curves of maximal slopes. In this paragraph we show that the results from Lemma I.2 can be translated to the setting of the Wasserstein space. That is, we characterize Wasserstein gradient flows as *curves of maximal slopes*. This result will be a key ingredient in Chapter III for the application of the so-called *Sandier-Serfaty approach*. We motivate this approach in Section I.3.

Before we state the result, we need to define the *metric slope* of a functional on $\mathcal{P}_2(\mathbb{R}^d)$, which plays the role of the modulus of the gradient in (I.2.11).

Definition I.7 (Metric slope) *Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be proper and lower semi-continuous with respect to W_2. Then the* metric slope $|\partial\phi| : D(\phi) \to [0, \infty]$ *is defined by*

$$|\partial\phi|(\mu) \;=\; \limsup_{\nu \to \mu} \frac{(\phi(\mu) - \phi(\nu))^+}{W_2(\mu, \nu)}. \tag{I.2.31}$$

This definition is consistent with Definition I.4 in the following sense. Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be proper, strongly λ-convex, lower semi-continuous with respect to W_2 and coercive. Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and suppose that the set $\partial\phi(\mu)$ is not empty. Then, in [4, 4.7, 4.8 and 4.10] it is shown that

$$|\partial\phi|(\mu) \;=\; \min\left\{ \|\xi\|_{\mathrm{L}^2(\mu)} \;\middle|\; \xi \in \partial\phi(\mu) \right\}. \tag{I.2.32}$$

This suggests that, intuitively, $|\partial\phi|(\mu)$ can be seen as a length of the gradient in the Wasserstein space with respect to the L^2-norm $\|\cdot\|_{\mathrm{L}^2(\mu)}$. Hence, in comparison to (I.2.11), $|\partial\phi|$ should be the metric analogue of the modulus of the Euclidean gradient given by $|\nabla\phi|$. In the following lemma we see another indication that this intuition is correct.

Lemma I.8 (Characterization as curves of maximal slopes)
Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be proper, strongly λ-convex, lower semi-continuous with respect to W_2 and coercive. Define $\mathscr{J}_{\phi,T} : C([0,T]\,;\mathcal{P}_2(\mathbb{R}^d)) \to [0,\infty]$ by

$$\mathscr{J}_{\phi,T}[(\nu_t)_{t\in(0,T)}] := \phi(\nu_T) - \phi(\nu_0) + \frac{1}{2}\int_0^T \left(|\partial\phi|^2(\nu_t) + |\nu'|^2(t)\right) dt, \tag{I.2.33}$$

if $(\nu_t)_{t\in(0,T)} \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}^d))$ and $\mathscr{J}_{\phi,T}[(\nu_t)_{t\in(0,T)}] = \infty$ else. Let $\mu_0 \in D(\phi)$. For any curve $(\mu_t)_{t\in(0,T)} \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}^d))$ such that $\lim_{t\to 0} W_2(\mu_t, \mu_0) = 0$ we have that $\mathscr{J}_{\phi,T}[(\mu_t)_{t\in(0,T)}] \geq 0$. Equality holds if and only if $(\mu_t)_{t\in(0,T)}$ is the gradient flow for ϕ with initial value μ_0.

The functional $\mathscr{J}_{\phi,T}$ is sometimes called the energy-dissipation functional *corresponding to the gradient flow for ϕ.*

Proof. The proof is given in [3, 11.1.3 and 11.2.1]. □

Connection to Fokker-Planck equations and reversible diffusion processes. We now show that gradient flows for the so-called *relative entropy* are the unique weak solutions to the *Fokker-Planck equation* (I.2.13).

Let $V \in C^2(\mathbb{R}^d)$ be λ-convex (recall (I.2.1)). For simplicity, we suppose that V is bounded from below. Let $\nu(dx) = \mathrm{e}^{-V(x)}dx$, and for $\mu \in \mathcal{M}_1(\mathbb{R}^d)$, define the *relative entropy* between μ and ν by

$$\mathcal{H}_\nu(\mu) \;:=\; \mathcal{H}(\mu\,|\,\nu) \;:=\; \begin{cases} \int_{\mathbb{R}^d} \log\left(\frac{d\mu}{d\nu}\right) d\mu & : \mu \ll \nu, \\ \infty & : \text{else.} \end{cases} \tag{I.2.34}$$

It is shown in [3, Chapter 9] that the functional $\mathcal{H}_\nu$ satisfies the assumptions of the Lemmas I.6 and I.8. Hence, there exists a unique gradient flow for $\mathcal{H}_\nu$, and it is characterized as a curve of maximal slope. Moreover, by combining Lemma I.3 and Definition I.5, we can show that this gradient flow is a weak solution to the Fokker-Planck equation (I.2.13). The precise result is given in the following lemma.

Lemma I.9 *Let $\mu_0 \in D(\mathcal{H}_\nu)$, and let $(\mu_t)_{t\in[0,T]} \subset \mathcal{P}_2(\mathbb{R}^d)$ be such that $\lim_{t\to 0} W_2(\mu_t, \mu_0) = 0$. Then $(\mu_t)_{t\in[0,T]}$ is the gradient flow for $\mathcal{H}_\nu$ if and only if*

(i) $\mu_t(dx) \;=\; \rho_t(x)\,dx$ *for all $t \in [0,T]$ for some density function ρ_t,*

(ii) the curve of densities $(\rho_t)_{t\in[0,T]}$ is a weak solution to

$$\partial_t \rho_t \;=\; \Delta\rho_t + \operatorname{div}\left(\nabla V\, \rho_t\right), \tag{I.2.35}$$

where Δ and div *denote the Laplacian and the divergence with respect to the space variable x.*

(iii) $\int_0^T |\partial\mathcal{H}_\nu|^2(\mu_t)\,dt \;<\; \infty$.

Proof. The proof is given in [3, 11.2.8]. We only sketch the "only if"-part here. Let $(\mu_t)_{t\in[0,T]}$ be the gradient flow for $\mathcal{H}_\nu$, and let $(v_t)_{t\in[0,T]}$ denote the corresponding tangent velocity field from Lemma I.3. Note that, by [59, 4.6], the unique strong subdifferential of $\mathcal{H}_\nu$ at some measure μ with density ρ is given by $\xi = \nabla\rho/\rho + \nabla V$. Then, the assertion of this lemma is an easy consequence of the definition of gradient flows (see (I.2.25)) and the fact that the curve $(\mu_t)_{t\in[0,T]}$ satisfies the continuity equation (see Lemma I.3). □

Using Lemma I.9, we immediately find the link between (Wasserstein) gradient flows for the relative entropy and reversible diffusion processes of a certain type. Let $(\mu_t)_{t\in[0,T]}$ be the gradient flow for $\mathcal{H}_\nu$, and let $(\rho_t)_{t\in[0,T]}$ be the flow of its probability densities. In Lemma I.9, we have seen that $(\rho_t)_{t\in[0,T]}$ is a weak solution to (I.2.35). Then, it is shown in [110, p. 111] that $(\mu_t)_{t\in[0,T]}$ is the flow of marginal laws of the reversible diffusion process $(x_t)_{t\in[0,T]}$ given by

$$dx_t \;=\; -\nabla V(x_t)\,dt \;+\; \sqrt{2}\,dB_t \quad \text{for } t \in (0,T], \tag{I.2.36}$$

and with x_0 being a random variable distributed according to μ_0.

Remark I.10 *We now provide a first intuitive explanation why the Wasserstein gradient flow formalism is seen as "a natural and physically meaningful structure"*[6] *to represent* (I.2.35).

Take $N \in \mathbb{N}$ independent copies, $(x_t^0)_{t\in[0,T]}, \dots, (x_t^{N-1})_{t\in[0,T]}$, of the diffusion (I.2.36) *with initial value being distributed according to μ_0. We now show that this system can be seen as the* microscopic origin *of the system* (I.2.35) *with respect to the Wasserstein formalism.*

[6] [5, p. 421]

Let the space $\mathcal{M}_1(\mathbb{R}^d)$ *be equipped with the topology induced by the notion of weak convergence, which we defined in* (I.2.17)*. Define the* empirical process $(K^N(t))_t \in C([0,T];\mathcal{M}_1(\mathbb{R}^d))$ *by*

$$K^N(t) \;=\; \frac{1}{N}\sum_{i=0}^{N-1}\delta_{x_t^i} \qquad \text{for } t\in[0,T]. \tag{I.2.37}$$

Then, it is well-known (see for instance Chapter III) that the sequence $(K^N(t))_t$ *satisfies a* law of large numbers*. That is, it converges, with respect to the weak topology, to the deterministic limit given by the solution* ρ *of* (I.2.35)*. Moreover, again by Chapter III, it can even be shown that* $(K^N(t))_t$ *satisfies a* large deviation principle *with rate function*

$$I[(\nu_t)_t] := \frac{1}{2}\mathscr{J}_{\mathcal{H}_\nu,T}[(\nu_t)_t] + \mathcal{H}(\nu_0|\mu_0) \qquad \text{for } (\nu_t)_t \in C([0,T];\mathcal{M}_1(\mathbb{R}^d)), \tag{I.2.38}$$

where $\mathscr{J}_{\mathcal{H}_\nu,T}$ *is the entropy-dissipation functional from Lemma I.8. This observation shows that the Wasserstein formalism does not only represent the solution of* (I.2.35)*, but also describes the* fluctuations *of the* most natural *microscopic particle system, which approximates the solution of* (I.2.35)*.*

This is a new perspective in comparison to other gradient flow representations of the solution of (I.2.35) *(such as, for example, the Hilbertian gradient flow representation introduced in [5, Section 1.4]). Of course, other gradient flow representations may describe the fluctuations of other microscopic particle systems than the one defined by* $(x_t^0)_{t\in[0,T]},\dots,(x_t^{N-1})_{t\in[0,T]}$*. But the system* $(x_t^0)_{t\in[0,T]},\dots,(x_t^{N-1})_{t\in[0,T]}$ *is the simplest and the canonical particle system that describes the solution of* (I.2.35)*. Hence, we see the Wasserstein gradient flow formalism as the canonical representation of the solution of* (I.2.35)*, since it describes the fluctuations of the canonical particle system* $(x_t^0)_{t\in[0,T]},\dots,(x_t^{N-1})_{t\in[0,T]}$*.*

This relation between large deviation principles of particle systems and Wasserstein gradient flows is known for a much larger class of diffusion equations, and goes much deeper than the relation we stated here; see [1], [55], [67], [70], [113] or Chapter III for more details.

Finally, we refer the reader also to Subsection I.6.4, where, by considering a system of mean-field interacting diffusions*, we provide another explanation why the Wasserstein formalism is seen as the "natural framework".*

Extension to other evolution systems. The link between Wasserstein gradient flows for relative entropy functionals and reversible diffusion processes of the form (I.2.36) was first discovered in the seminal papers [83] and [111]. In recent years, this gradient flow representation has been translated to other evolution systems. For instance, it is known that there are many more reversible diffusion processes that can be described by the Wasserstein gradient flow formalism from this section. These include the class of *McKean-Vlasov equations*; see [3, Section 11.2.1] or [37]. In Chapter III we extend this connection to the class of the so-called *local McKean-Vlasov equations*. This is done by modifying the Wasserstein distance such that the dependence on a non-evolving parameter in these equations is taken into account. As a consequence of this modification, we have to rebuild the whole gradient flow framework of this section for this modified Wasserstein distance.

Moreover, by an appropriate manipulation of the dynamical formulation of the Wasserstein distance (the so-called *Benamou-Brenier formula*[7]), this connection has been extended

[7] See [17, Proposition 1.1] or [3, Chapter 8]

to other systems than diffusion processes. For instance, Wasserstein-like gradient flow representations has been shown for *Markov chains* (see [63], [65], [66], [98] or [103]), the *Boltzmann equation* (see [15] or [61]) or *jump processes* (see [60]).

Connection to metastability. We have seen in the example of the Curie-Weiss model in Subsection I.1.3 that *coarse-graining* is a useful tool to study the metastable behaviour of models that consist of a large number of particles. In the special case of *high-dimensional diffusion systems* the *macroscopic order parameter* is often given by the *empirical distribution*. The reason is that this map preserves the Markov property of the system. This fact can be verified easily by applying [57, Vol I, p. 325, Theorem 10.13]. Hence, the macroscopic level is given by the space of probability measures, and in order to study the macroscopic behaviour of the system, it is necessary to have a Riemannian structure on this space. The latter is provided by the Wasserstein formalism that we introduced in this section. Hence, the Wasserstein formalism should pave the way for the rigorous investigation of the metastable behaviour of empirical distribution processes.

A first indication that this idea should lead to the desired metastability results is given by combining the results of Chapter III and Chapter V. Indeed, in Chapter III we show that the empirical distribution process associated to *(local) mean-field interacting diffusions* can be approximated by Wasserstein-like gradient flows for a functional $\mathcal{F}$. Hence, in order to study the metastable behaviour of the empirical distribution process, it is useful to analyse the long-time behaviour of the gradient flows for $\mathcal{F}$. This is the content of Chapter V, where we study, in a simplified context, the ergodic behaviour and the basins of attraction of the gradient flows for $\mathcal{F}$. Moreover, another indication for the connection between metastability of empirical distribution processes and the Wasserstein formalism is discussed at the beginning of Subsection I.6.4.

We finally note that there is also another approach to study metastability by using gradient flow representations (and the so-called *Sandier-Serfaty approach* that we introduce in Section I.3). This approach was used in [5], [81], [114] and [117]. Here, the authors show the convergence of the upscaled dynamics to a finite-state Markov chain, where each state corresponds to a metastable state. Then, the rates of the transitions of the Markov chain between these states are given by the Eyring-Kramers formula, which we motivated in the first example of Subsection I.1.3. It is left for future research to apply this approach for the system of mean-field interacting diffusions from Subsection I.6.4.

I.3 The Sandier-Serfaty approach

In Section I.2 we have seen that many evolution systems can be represented as gradient flows with respect to Wasserstein (or Wasserstein-like) distances. The goal of this section is to show that these representations can be used to study the convergence of sequences of evolution systems. This fact was first discovered in the paper [115] in the context of gradient flows in general Hilbert spaces, and is known in the literature as the *Sandier-Serfaty approach*. This approach relies on the so-called *gamma-liminf inequalities* for the "energy-dissipation functional", which appears in the variational characterization of the respective gradient flows (see Lemma I.2 or Lemma I.8). Successful applications of the Sandier-Serfaty approach are given, for example, in [5], [36], [61], [63], [70], [71], [118] or Chapter III of this thesis, where we apply this approach to prove a *law of large numbers* for the empirical distribution process

associated to a *local mean-field interacting spin system*.

The Sandier-Serfaty approach has the following advantages.

- The first one is that it provides an elegant and simple way to prove the convergence of evolution systems. Indeed, the main ideas in this approach are model-independent and have been used successfully in many different settings; again, see [5], [36], [61], [63], [70], [71] or Chapter III. More precisely, as we already mentioned, the main step is to show the gamma-liminf inequalities. And in order to show these inequalities, one typically makes use of certain duality representation formulas and lower semi-continuity properties of the objects appearing in the energy-dissipation functional. It is known that the latter two facts hold true for many classes of evolution systems. This makes the Sandier-Serfaty approach applicable for many different settings.

- The second advantage is that a successful application of the Sandier-Serfaty approach does not only show the convergence of evolution systems. It rather shows the *convergence of the gradient flow structures* of the respective systems. We have seen in Section I.2 that this gradient flow structure encodes many dynamic properties of the system, such as the *free energy landscape*, the *large deviation principle of the microscopic origin* (see Remark I.10) or the *stationary states* (see Lemma V.4). Therefore, the Sandier-Serfaty approach yields a rigorous connection between the level of the sequences and the limiting object with regard to these properties. In particular, one obtains the convergence of the objects in the energy-dissipation functionals corresponding to the gradient flows (cf. Step 5 of the proof of Theorem III.62). Especially, in certain cases, from the convergence of the free energies (cf. (III.3.2)), one can deduce the so-called *propagation of chaos*; see for example the comments after [61, 1.2].

- Another advantage is that this approach can be used to study the metastable behaviour of stochastic processes. This is the content of the papers [5], [81], [114] and [117]. We already mentioned this relation at the end of Section I.2.

In this section we apply the Sandier-Serfaty approach for a simple example in the context of the Wasserstein space. Namely, in the setting of a sequence of reversible diffusions in the limit of vanishing noise, i.e., in a similar setting as in (I.1.1). It turns out that this example already contains many ideas for the application of this approach in Chapter III. Intuitively, the setting of this section can even be seen as a finite-dimensional version of the setting of Chapter III. Indeed, in this example we show that the diffusion process converges, in the limit of vanishing noise, to a deterministic process given by an Euclidean gradient flow. In Chapter III we show the analogous result for the empirical distribution process associated to a system of $N \in \mathbb{N}$ interacting spins. More precisely, we show that, as $N \to \infty$, the empirical distribution process converges to a deterministic process given by a Wasserstein-like gradient flow. Therefore, the example in this section should act as a guide and a motivation for the results and the proofs of Chapter III.

We finally note that, in the context of reversible diffusion processes, Fathi shows in [70] that a slight extension of the Sandier-Serfaty approach yields the *large deviation principle* for the sequence. This extended scheme is sometimes known in the literature as the *Fathi-Sandier-Serfaty approach*. We use this approach in Chapter III to show that the spin system also satisfies a large deviation principle. In the example of this section, the Fathi-Sandier-Serfaty approach is also applicable. That is, with a little bit of additional work, we could

also prove the so-called *Schilder theorem* (see [52, Chapter 5]). However, in this introductory treatment we only focus on the main ideas, which are already present in the application of the Sandier-Serfaty approach. We postpone the details for the extension by Fathi to Chapter III.

We now introduce the model that we consider in this section. Let, for each $\varepsilon > 0$,

- x_0^ε be a random variable distributed according to some $\mu_0^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$, and
- $(x_t^\varepsilon)_{t\in(0,T]}$ be the solution of the stochastic differential equation

$$dx_t^\varepsilon \;=\; -\nabla V(x_t^\varepsilon)\,dt \;+\; \sqrt{2\varepsilon}\,dB_t \tag{I.3.1}$$

with initial condition x_0^ε, where B is a d-dimensional Brownian motion and $V \in C^2(\mathbb{R}^d)$ is assumed to be λ-convex (see (I.2.1)) and such that, for some $c > 0$,

$$|V(x)| \;\geq\; c(|x|^4 + |x|^2 - 1) \quad \text{for all } x \in \mathbb{R}^d. \tag{I.3.2}$$

The goal is to study the sequence $(x_t^\varepsilon)_{t\in[0,T]}$ in the limit as $\varepsilon \downarrow 0$. Of course, in this example, there are standard probabilistic tools to show that, for suitable initial conditions and as $\varepsilon \downarrow 0$, the sequence $(x_t^\varepsilon)_{t\in[0,T]}$ converges (in some sense that we specify later) to the Euclidean gradient flow (see Subsection I.2.1) for the functional V. However, here we would like to prove this claim by applying the Sandier-Serfaty approach. In this way, we already introduce many ideas that are used in Chapter III in the more complex setting of local mean-field interacting spin systems.

The two most important ingredients for the Sandier-Serfaty approach are *the gradient flow representation on the level of the sequences* and *the gradient flow representation for the limiting object*. In Section I.2 we introduced the gradient flow representation on the level of the sequences. Indeed, arguing in the same way as in the comment after Lemma I.9, we can show that the flow of marginal laws $(\mu_t^\varepsilon)_{t\in[0,T]}$ corresponding to the stochastic process $(x_t^\varepsilon)_{t\in[0,T]}$ is given by the unique Wasserstein gradient flow for the functional

$$\mathcal{H}_\varepsilon(\cdot) \;:=\; \varepsilon\,\mathcal{H}(\cdot \,|\, \mathrm{e}^{-\frac{1}{\varepsilon}V(x)}dx) \tag{I.3.3}$$

with initial law μ_0^ε. Moreover, as we already mentioned, the gradient flow representation for the limiting object is given by the Euclidean gradient flow for the functional V, which we studied in Subsection I.2.1.

We have now collected the main ingredients to state and prove the following result.

Lemma I.11 *Let $z_0 \in \mathbb{R}^d$ and let $z \in \mathcal{AC}((0,T);\mathbb{R}^d)$ be the unique Euclidean gradient flow for the functional V with initial value z_0. Suppose that the sequence $(\mu_0^\varepsilon)_{\varepsilon>0}$ of initial values is* well-prepared*, i.e.,*

$$\lim_{\varepsilon\downarrow 0} W_2(\mu_0^\varepsilon, \delta_{z_0}) \;=\; 0 \quad \textit{and} \quad \lim_{\varepsilon\downarrow 0} \mathcal{H}_\varepsilon(\mu_0^\varepsilon) \;=\; V(z_0). \tag{I.3.4}$$

Then, for all $t \in [0,T]$,

$$\lim_{\varepsilon\downarrow 0} W_2(\mu_t^\varepsilon, \delta_{z_t}) \;=\; 0 \quad \textit{and} \quad \lim_{\varepsilon\downarrow 0} \mathcal{H}_\varepsilon(\mu_t^\varepsilon) \;=\; V(z_t). \tag{I.3.5}$$

Proof. First notice that, by (I.3.2) and [83, p. 9], there exist $c_0, \varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$,

$$\mathcal{H}_\varepsilon(\mu) \;\geq\; c_0 \left(\int_{\mathbb{R}^d} (|x|^4 + |x|^2)\, d\mu - 1 \right) \;\geq\; -c_0 \quad \text{for all } \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{I.3.6}$$

Moreover, note that by Lemma I.8,

$$\mathscr{J}_{\mathcal{H}_\varepsilon, T}[(\mu_t^\varepsilon)_{t\in(0,T)}] \;=\; 0 \quad \text{for all } \varepsilon < \varepsilon_0. \tag{I.3.7}$$

Combining the monotonicity property (I.2.28) with (I.3.4), (I.3.6) and (I.3.7), yields that for some $0 < \varepsilon_1 \leq \varepsilon_0$

$$\sup_{\varepsilon<\varepsilon_1} \int_0^T |(\mu^\varepsilon)'|^2(t)\, dt \;<\; \infty \quad \text{and} \quad \sup_{\varepsilon<\varepsilon_1} \sup_{t\in[0,T]} \mathcal{H}_\varepsilon(\mu_t^\varepsilon) \;<\; \infty. \tag{I.3.8}$$

In particular, in view of (I.3.6),

$$\sup_{\varepsilon<\varepsilon_1} \sup_{t\in[0,T]} \int_{\mathbb{R}^d} (|x|^4 + |x|^2)\, d\mu_t^\varepsilon \;<\; \infty. \tag{I.3.9}$$

Step 1. [Compactness.]
As in most of the applications of the Sandier-Serfaty approach, the compactness of the sequence $\{(\mu_t^\varepsilon)_t\}_\varepsilon := \{(\mu_t^\varepsilon)_{t\in[0,T]}\}_{\varepsilon<\varepsilon_1}$ is shown by applying the Arzelá-Ascoli theorem (see [86, Chapter 7, Theorem 17]). That is, we have to show that, provided that the space $C([0,T]\,;\,\mathcal{P}_2(\mathbb{R}^d))$ is equipped with the uniform topology with respect to W_2, the sequence $\{(\mu_t^\varepsilon)_t\}_{\varepsilon<\varepsilon_1}$ is equi-continuous and that $\{\mu_t^\varepsilon\}_{\varepsilon<\varepsilon_1}$ is compact for all $t \in [0,T]$. The latter is a straightforward consequence of the uniform bound on the fourth moment given in (I.3.9) and (I.2.18). To show the former note that by the absolute continuity of $(\mu_t^\varepsilon)_{t\in[0,T]}$,

$$W_2(\mu_t^\varepsilon, \mu_s^\varepsilon)^2 \;\leq\; \left(\int_s^t |(\mu^\varepsilon)'|(t)\, dt \right)^2 \;\leq\; (t-s) \int_0^T |(\mu^\varepsilon)'|^2(t)\, dt \tag{I.3.10}$$

for all $0 < s < t < T$ and $\varepsilon < \varepsilon_1$. Combining this with (I.3.8) yields the equi-continuity of $\{(\mu_t^\varepsilon)_t\}_\varepsilon$. Hence, we obtain the existence of a subsequence $\{(\mu_t^{\varepsilon_n})_t\}_{n\in\mathbb{N}}$ and a curve $(\mu_t)_t \in C([0,T]\,;\,\mathcal{P}_2(\mathbb{R}^d))$ such that $\lim_{n\to\infty} \varepsilon_n = 0$ and

$$\lim_{n\to\infty} \sup_{t\in[0,T]} W_2(\mu_t^{\varepsilon_n}, \mu_t) \;=\; 0. \tag{I.3.11}$$

Step 2. [$(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}^d))$.]
According to [96, Lemma 1], it suffices to show that

$$\sup_{0<h<T} \int_0^{T-h} \frac{1}{h^2} W_2(\mu_t, \mu_{t+h})^2\, dt \;<\; \infty \qquad \text{and} \qquad \int_0^T W_2(\mu_t, \delta_0)^2\, dt \;<\; \infty. \tag{I.3.12}$$

The second claim is a consequence of (I.3.9) and (I.3.11). To show the first claim, note that by (I.3.11), Fatou's lemma, Fubini's theorem and (I.3.8)

$$\begin{aligned}
\sup_{0<h<T} \int_0^{T-h} \frac{1}{h^2} W_2(\mu_t, \mu_{t+h})^2\, dt \;&\leq\; \sup_{0<h<T} \liminf_{n\to\infty} \int_0^{T-h} \frac{1}{h^2} W_2(\mu_t^{\varepsilon_n}, \mu_{t+h}^{\varepsilon_n})^2\, dt \\
&\leq\; \sup_{0<h<T} \sup_{\varepsilon<\varepsilon_1} \int_0^{T-h} \frac{1}{h} \int_t^{t+h} |(\mu^\varepsilon)'|^2(r)\, dr\, dt \\
&\leq\; \sup_{\varepsilon<\varepsilon_1} \int_0^T |(\mu^\varepsilon)'|^2(r)\, dr < \infty.
\end{aligned} \tag{I.3.13}$$

Step 3. [Gamma-liminf-inequalities.]
Let $\mathcal{V} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ be defined by

$$\mathcal{V}(\mu) \;:=\; \int_{\mathbb{R}^d} V \, d\mu \quad \text{ for } \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{I.3.14}$$

In this step we show the so-called *gamma-liminf-inequalities*, i.e., we show that for all $t \in [0,T]$,

$$\liminf_{n\to\infty} \mathcal{H}_{\varepsilon_n}(\mu_t^{\varepsilon_n}) \;\geq\; \mathcal{V}(\mu_t), \tag{I.3.15}$$

$$\liminf_{n\to\infty} |\partial \mathcal{H}_{\varepsilon_n}|(\mu_t^{\varepsilon_n}) \;\geq\; |\partial \mathcal{V}|(\mu_t), \quad \text{ and} \tag{I.3.16}$$

$$\liminf_{n\to\infty} \int_0^t |(\mu^{\varepsilon_n})'|^2(r)\, dr \;\geq\; \int_0^t |\mu'|^2(r)\, dr. \tag{I.3.17}$$

These gamma-liminf-inequalities are often shown by exploiting duality theorems that are available for the quantities on the left-hand sides of (I.3.15)–(I.3.17) for a large class of functionals.

We first show (I.3.15). Note that by [83, p. 9], for some $c' > 0$,

$$\mathcal{H}_{\varepsilon_n}(\mu_t^{\varepsilon_n}) \;\geq\; -\varepsilon_n c' \int_{\mathbb{R}^d} |x|^2 \, d\mu_t^{\varepsilon_n}(x) \;-\; \varepsilon_n c' \;+\; \int_{\mathbb{R}^d} V \, d\mu_t^{\varepsilon_n}. \tag{I.3.18}$$

Then, taking the limit as $n \to \infty$, we obtain (I.3.15) by using (I.3.9) for the first term and standard lower semi-continuity results for integrals (see [3, 5.1.7]) for the third term.

To show (I.3.16), we use that for all $t \in [0,T]$, (see Corollary III.39 in this thesis, [59, 4.3] or [3, 10.4.9])

$$|\partial \mathcal{H}_{\varepsilon_n}|(\mu_t^{\varepsilon_n}) \;=\; \sup_{\varphi \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d),\, \|\varphi\|_{\mathrm{L}^2(\mu_t^{\varepsilon_n})}>0} \frac{\left|\int_{\mathbb{R}^d} (\varphi \nabla V - \varepsilon_n \operatorname{div}\varphi)\, d\mu_t^{\varepsilon_n}\right|}{\|\varphi\|_{\mathrm{L}^2(\mu_t^{\varepsilon_n})}}. \tag{I.3.19}$$

Let $\varphi \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$ be such that $\|\varphi\|_{\mathrm{L}^2(\mu_t)} > 0$. Then, since $\lim_{n\to\infty} W_2(\mu_t^{\varepsilon_n}, \mu_t) = 0$, we have that for n large enough, $\|\varphi\|_{\mathrm{L}^2(\mu_t^{\varepsilon_n})} > 0$. Therefore,

$$\begin{aligned}
\liminf_{n\to\infty} |\partial \mathcal{H}_{\varepsilon_n}|(\mu_t^{\varepsilon_n}) \;&\geq\; \sup_{\varphi \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d),\, \|\varphi\|_{\mathrm{L}^2(\mu_t)}>0} \liminf_{n\to\infty} \frac{\left|\int_{\mathbb{R}^d} (\varphi \nabla V - \varepsilon_n \operatorname{div}\varphi)\, d\mu_t^{\varepsilon_n}\right|}{\|\varphi\|_{\mathrm{L}^2(\mu_t^{\varepsilon_n})}} \\
&=\; \sup_{\varphi \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d),\, \|\varphi\|_{\mathrm{L}^2(\mu_t)}>0} \frac{\left|\int_{\mathbb{R}^d} \varphi \nabla V \, d\mu_t\right|}{\|\varphi\|_{\mathrm{L}^2(\mu_t)}} \;=\; |\partial \mathcal{V}|(\mu_t),
\end{aligned} \tag{I.3.20}$$

where we use in the last equality that the dual representation (I.3.19) is also true in the case $\varepsilon_n = 0$.

It remains to show (I.3.17). Proceeding similarly as in Step 2 and using the definition of the metric derivative (see (I.2.20)), we have that for $\delta \in (0, t/2)$,

$$\begin{aligned}
\int_0^{t-\delta} |\mu'|^2(t)\, dt \;&\leq\; \liminf_{h\downarrow 0, h<\delta} \int_0^{t-\delta} \frac{1}{h^2} W_2(\mu_t, \mu_{t+h})^2 \, dt \\
&\leq\; \liminf_{h\downarrow 0, h<\delta} \int_0^{t-\delta} \liminf_{n\to\infty} \frac{1}{h^2} W_2(\mu_t^{\varepsilon_n}, \mu_{t+h}^{\varepsilon_n})^2 dt \\
&\leq\; \liminf_{n\to\infty} \int_0^t |(\mu^{\varepsilon_n})'|^2(r)\, dr.
\end{aligned} \tag{I.3.21}$$

Letting $\delta \downarrow 0$ concludes the proof of (I.3.17).

Step 4. [Proof of (I.3.5).]

Combining Step 3 and (I.3.4) shows that

$$\begin{aligned} 0 &= \liminf_{n\to\infty} \mathscr{J}_{\mathcal{H}_{\varepsilon_n},T}[(\mu_t^{\varepsilon_n})_{t\in[0,T]}] \;\geq\; \mathcal{V}(\mu_T) - V(z_0) + \frac{1}{2}\int_0^T \big(|\partial\mathcal{V}|^2(\nu_t) + |\mu'|^2(t)\big)\,dt \\ &= \mathscr{J}_{\mathcal{V},T}[(\mu_t)_{t\in[0,T]}] \;\geq\; 0, \end{aligned} \tag{I.3.22}$$

where we set $\mu_0 = \delta_{z_0}$. Hence, $\mathscr{J}_{\mathcal{V},T}[(\mu_t)_{t\in[0,T]}] = 0$. In view of Lemma I.8, this yields that $(\mu_t)_{t\in[0,T]}$ is the unique gradient flow for the functional $\mathcal{V}$ with initial value δ_{z_0}. However, by [3, 11.2.3], we have that this Wasserstein gradient flow is given by $\mu_t = \delta_{z_t}$, where $(z_t)_{t\in[0,T]}$ is the unique Euclidean gradient flow for the functional $\mathcal{V}$. This shows that every limit point of the sequence $\{(\mu_t^\varepsilon)_t\}_\varepsilon$ is given by $(\delta_{z_t})_{t\in[0,T]}$, which in turn implies the first claim in (I.3.5). Using Step 3, (I.3.4) and the fact that $\lim_{\varepsilon\downarrow 0} \mathscr{J}_{\mathcal{H}_\varepsilon,T}[(\mu_t^\varepsilon)_{t\in[0,T]}] = \mathscr{J}_{\mathcal{V},T}[(\mu_t)_{t\in[0,T]}]$ shows the second claim in (I.3.5). □

I.4 The results of Chapter II

One of the simplest models where one can rigorously study the phenomenon of metastability is the two-dimensional standard Ising model on a finite torus in the low temperature regime. Neves and Schonmann applied the path-wise approach[8] to this model in [108]. This was later rewritten in the Chapters 7.1–7.5 of [109]. The potential-theoretic approach was used to study the metastable behaviour of this model in [34] by Bovier and Manzo (see also Chapter 17 of [29]). Moreover, several other settings and regimes in the Ising model have been considered as well. For example, the Ising model on $\mathbb{Z}^d$ was considered in [50] (for $d = 2$) and in [39] (for $d \geq 3$), and the regime, where the magnetic field tends to zero was studied in [116].

In this section we formulate the results of Chapter II, where we study three modifications of the Ising model. Roughly speaking, the crucial difference between all three models and the standard Ising model is the fact that we lose the applicability of *isoperimetrical inequalities*. Namely, in the Ising case, for a given number of *up-spins*, the configurations with minimal energy are those droplets of up-spins whose shape is given by a square (or a quasi-square) with a possible bar of up-spins attached to one of its sides. Here we do not have this property. Instead we need to look at the *stability* of certain classes of configurations separately in order to specify the metastable and the critical state rigorously. The path-wise approach has already been applied to these models in [88], [89] and [107], respectively. In the Chapters 7.7–7.10 of [109], a brief overview of these three papers is given. In Chapter II we complement these results and apply the potential-theoretic approach.

This section is organized as follows. In Subsection I.4.1 we introduce a dynamical spin-flip model on the two-dimensional lattice, which is driven by a general *energy function* H. This setting is the basic framework for all three models that we consider in Chapter II. These models only differ in the precise form of the energy function H. Then, in the Subsections I.4.2, I.4.3 and I.4.4 we introduce these three models and state the main results. We also compare our results with those from the papers [88], [89] and [107]. The proofs are given in Chapter II, and rely on the so-called *metastability theorems* that are proven in [29, Chapter 16]. We explain this in more detail at the end of Section II.1.

[8] Recall Subsection I.1.2.

I.4.1 The abstract set-up

Let $\Lambda \subset \mathbb{Z}^2$ be a finite, square box with periodic boundary conditions, centred at the origin, and let $S = \{-1, 1\}^\Lambda$. S is called the *configuration space*, an element $\sigma \in S$ is called *configuration*, and at each *site* $x \in \Lambda$, $\sigma(x) \in \{-1, 1\}$ is called the *spin-value* at x.

The *energy* or *Hamiltonian* of the system is given by $\mathrm{H} : S \to \mathbb{R}$, and the *Gibbs measure* associated to H is given by

$$\mu_\beta(\sigma) = \frac{1}{Z_\beta} \mathrm{e}^{-\beta\,\mathrm{H}(\sigma)}, \qquad \text{for } \sigma \in S, \tag{I.4.1}$$

where $\beta > 0$ is called the *inverse temperature*, and Z_β is a normalization constant called *partition function*.

For $\sigma \in S$ and $x \in \Lambda$ we define $\sigma^x \in S$ by

$$\sigma^x(y) = \begin{cases} \sigma(y) & : y \neq x, \\ -\sigma(x) & : y = x. \end{cases} \tag{I.4.2}$$

For all $\sigma, \sigma' \in S$, we say that σ and σ' *communicate* and write $\sigma \sim \sigma'$ if there exists $x \in \Lambda$ such that $\sigma^x = \sigma'$. This induces a graph structure on S by defining an edge between each $\sigma, \sigma' \in S$ whenever $\sigma \sim \sigma'$.

The dynamics of the system is given by the continuous time Markov Chain $(\sigma_t)_{t\geq 0}$ on S, whose generator $\mathcal{L}_\beta$ is given by

$$(\mathcal{L}_\beta f)(\sigma) = \sum_{x\in\Lambda} c_\beta(\sigma, \sigma^x)(f(\sigma^x) - f(\sigma)), \tag{I.4.3}$$

where $f : S \to \mathbb{R}$ is a function and

$$c_\beta(\sigma, \sigma') = \begin{cases} \mathrm{e}^{-\beta \max\{0, \mathrm{H}(\sigma') - \mathrm{H}(\sigma)\}} & : \sigma \sim \sigma', \\ 0 & : \text{else.} \end{cases} \tag{I.4.4}$$

Notice that for $\beta = \infty$ only moves to configurations with lower or equal energy are permitted. Moreover, one can immediately see that the following *detailed balance condition* holds.

$$\mu_\beta(\sigma) c_\beta(\sigma, \sigma') = \mu_\beta(\sigma') c_\beta(\sigma', \sigma) \qquad \forall\, \sigma, \sigma' \in \mathcal{X}_\beta^{(n_\beta)} \tag{I.4.5}$$

Hence, the Markov chain is *reversible* with respect to the Gibbs measure. The law of $(\sigma_t)_{t\geq 0}$ given that $\sigma_0 = \sigma \in S$ is denoted by $\mathbb{P}_\sigma$, and for a set $A \subset S$, we denote its *first hitting time after the starting configuration has been left* by τ_{A}, i.e.

$$\tau_{\mathrm{A}} \;=\; \inf\{t > 0 \mid \sigma_t \in A, \exists\, 0 < s < t : \sigma_s \neq \sigma_0\}. \tag{I.4.6}$$

If $A = \{\sigma\}$ for some $\sigma \in S$, we write $\tau_\sigma = \tau_{\mathrm{A}}$.

Finally, we denote by $\boxminus \in S$ the configuration, where all spin values are equal to -1 and by $\boxplus \in S$ the configuration with all spin values being $+1$. In the three models that we study in Chapter II, the configuration $\boxminus$ serves as the *metastable state* of the system, and $\boxplus$ as the *stable state* of the system.

I.4.2 Anisotropic Ising model

The first model that we study in Chapter II is the same model as in [88]. In this model the interaction between neighbouring spins is anisotropic in the sense that the attraction on horizontal bonds is stronger than on vertical bonds. More precisely, the Hamiltonian is explicitly given by

$$\mathrm{H}_{\mathrm{A}}(\sigma) = -\frac{J_H}{2} \sum_{(x,y)\in\Lambda_H^\star} \sigma(x)\sigma(y) - \frac{J_V}{2} \sum_{(x,y)\in\Lambda_V^\star} \sigma(x)\sigma(y) - \frac{h}{2}\sum_{x\in\Lambda}\sigma(x), \tag{I.4.7}$$

where $\sigma \in S$, $J_H > J_V > 0$, $h > 0$, $\Lambda_H^\star$ is the set of *unordered horizontal nearest-neighbour bonds* in Λ and $\Lambda_V^\star$ is the set of *unordered vertical nearest-neighbour bonds* in Λ. Here and in the following the subscript A is added to remind that we are in the anisotropic case. The critical length in this model is given by

$$L_V^\star = \left\lceil \frac{2J_V}{h} \right\rceil. \tag{I.4.8}$$

We make the following assumptions in this model.

Assumption I.12 *(a)* $J_H > J_V$,

(b) $2J_V > h$,

(c) $\frac{2J_V}{h} \notin \mathbb{N}$,

(d) $|\Lambda|$ *is large enough.*

We discuss the reasons for these assumptions in Chapter II.

We now formulate the main result for this model. For a more precise formulation and the proof we refer to Chapter II.

Theorem I.13 (cf. Theorem II.8) *Let* $\mathcal{C}_{\mathrm{A}} \subset S$ *be the the set of all configurations consisting only of a rectangle with side lengths* $L_V^\star - 1$ *and* $L_V^\star$*, and with an additional protuberance attached to one of its longer sides; see Figure I.3 for an example. A more precise definition of this set is given in Chapter II. Suppose Assumption I.12. Then,*

(A) $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\tau_{\mathcal{C}_{\mathrm{A}}} < \tau_{\boxplus} \mid \tau_{\boxplus} < \tau_{\boxminus}] = 1$,

(B) for all $\chi \in \mathcal{C}_{\mathrm{A}}$*, we have that* $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\sigma_{\tau_{\mathcal{C}_{\mathrm{A}}}} = \chi] = \frac{1}{|\mathcal{C}_{\mathrm{A}}|}$,

(C) $\lim_{\beta\to\infty} \lambda_\beta \, \mathbb{E}_{\boxminus}[\tau_{\boxplus}] = 1$*, where* λ_β *is the second largest eigenvalue of* $-\mathcal{L}_\beta$,

(D) $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\tau_{\boxplus} > t\mathbb{E}_{\boxminus}[\tau_{\boxplus}]] = \mathrm{e}^{-t}$ *for all* $t \geq 0$*, and*

(E) $\lim_{\beta\to\infty} \mathrm{e}^{-\beta\Gamma_{\mathrm{A}}^\star} \mathbb{E}_{\boxminus}[\tau_{\boxplus}] = K$*, where*

$$\begin{aligned} \Gamma_{\mathrm{A}}^\star &= \mathrm{H}_{\mathrm{A}}(\chi) - \mathrm{H}_{\mathrm{A}}(\boxminus) \quad \text{for all } \chi \in \mathcal{C}_{\mathrm{A}}, \text{ and} \\ K^{-1} &= \frac{4(2L_V^\star - 1)}{3}|\Lambda|. \end{aligned} \tag{I.4.9}$$

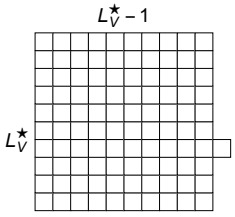


Figure I.3: An example of a configuration in $\mathcal{C}_{\mathrm{A}}$.

Part (A) of Theorem I.13 says that, in order to make the transition from the metastable to the stable state, the system has to pass the set $\mathcal{C}_{\mathrm{A}}$. Therefore, $\mathcal{C}_{\mathrm{A}}$ is seen as the set of *critical states* of the system. This provides a solution to problem (v) from Subsection I.1.2. Part (B) of Theorem I.13 says that the entrance into $\mathcal{C}_{\mathrm{A}}$ is uniformly distributed on $\mathcal{C}_{\mathrm{A}}$, and part (C) represents the average transition time of the system in terms of the spectrum of its generator. Part (D) of Theorem I.13 yields the asymptotic exponential distribution of $\tau_{\boxplus}$, and therefore provides an answer to problem (iii) from Subsection I.1.2. Finally, part (E) yields the precise asymptotics of the average transition time. This is the solution to problem (i) from Subsection I.1.2.

We now compare Theorem I.13 with the results that were already obtained in [88]. Part (A) of Theorem I.13 has already been shown in [88, Theorem 1]. Moreover, an estimate in probability of $\tau_{\boxplus}$ is proven in [88, Theorem 2], and the typical paths for the transition from $\boxminus$ to $\boxplus$ are identified in [88, Theorem 3]. Then, using standard techniques (cf. [109, Theorem 6.30 and (6.171)]), one can use these results to obtain both part (D) of Theorem I.13 and the asymptotics of $\mathbb{E}_{\boxminus}[\tau_{\boxplus}]$ up to logarithmic equivalence (i.e. the asymptotics without the pre-factor K). Hence, we provide here a new approach to prove part (A) and part (D) of Theorem I.13, and we provide a more precise estimate for $\mathbb{E}_{\boxminus}[\tau_{\boxplus}]$ than in [88].

I.4.3 Ising model with next-nearest-neighbour attraction

In the second model that we consider in Chapter II, we allow next-nearest-neighbour attraction, i.e. two spins that have Euclidean distance of $\sqrt{2}$ feel an interaction force. This next-nearest-neighbour attraction is assumed to be strictly weaker than the attraction between nearest-neighbour bonds. This has the physical intuition that next-nearest-neighbour attraction is seen as a perturbation of nearest-neighbour attraction. An interesting fact is that the local minima of the energy landscape are given by droplets of *octagonal shape*; see Figure I.4 for an example. For the path-wise approach to this model we refer to [89].

Here the Hamiltonian is given by

$$\mathrm{H}_{\mathrm{NN}}(\sigma) = -\frac{\tilde{J}}{2} \sum_{(x,y)\in\Lambda^\star} \sigma(x)\sigma(y) - \frac{K}{2} \sum_{(x,y)\in\Lambda^{\star\star}} \sigma(x)\sigma(y) - \frac{h}{2} \sum_{x\in\Lambda} \sigma(x), \tag{I.4.10}$$

where $\sigma \in S$, $\tilde{J} > K$, $h > 0$, $\Lambda^\star$ is the set of *unordered nearest-neighbour bonds* in Λ and $\Lambda^{\star\star}$ is the set of *unordered next-nearest-neighbour bonds* in Λ, i.e

$$\Lambda^{\star\star} = \big\{\{x,y\} \in \Lambda^2 \;\big|\; |x-y| = \sqrt{2}\big\}. \tag{I.4.11}$$

Here, the subscript NN is added to remind that we are in the case with next-nearest-neighbour

attraction. Set $J = \tilde{J} + 2K$. The critical lengths in this model are given by

$$\ell^\star = \left\lceil \frac{2K}{h} \right\rceil \qquad \text{and} \qquad D^\star = \left\lceil \frac{2J}{h} \right\rceil \qquad \text{and} \qquad L^\star = D^\star - 2(\ell^\star - 1). \tag{I.4.12}$$

We make the following assumptions in this model.

Assumption I.14 *(a)* $K > h$,

(b) $\tilde{J} \geq 2K + h$,

(c) $\frac{2J}{h} \notin \mathbb{N}$, $\frac{2K}{h} \notin \mathbb{N}$,

(d) $|\Lambda|$ *is large enough.*

We discuss the reasons for these assumptions in Chapter II.

We now formulate the main result for this model. For a more precise formulation and the proof we refer to Chapter II.

Theorem I.15 (cf. Theorem II.16) *Let* $\mathcal{C}_{\text{NN}} \subset S$ *be the set of all configurations that consist only of a so-called* octagon *of critical side lengths and with an additional protuberance attached at the* interior *of one of its longer* coordinate edges*; see Figure I.4 for an example and see Chapter II for a more precise definition of this set and of the notions octagon, interior and coordinate edge. Suppose Assumption I.14. Then,*

(A) $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\tau_{\mathcal{C}_{\text{NN}}} < \tau_{\boxplus} \mid \tau_{\boxplus} < \tau_{\boxminus}] = 1$,

(B) for all $\chi \in \mathcal{C}_{\text{NN}}$, *we have that* $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\sigma_{\tau_{\mathcal{C}_{\text{NN}}}} = \chi] = \frac{1}{|\mathcal{C}_{\text{NN}}|}$,

(C) $\lim_{\beta\to\infty} \lambda_\beta \, \mathbb{E}_{\boxminus}[\tau_{\boxplus}] = 1$, *where* λ_β *is the second largest eigenvalue of* $-\mathcal{L}_\beta$,

(D) $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\tau_{\boxplus} > t\mathbb{E}_{\boxminus}[\tau_{\boxplus}]] = \mathrm{e}^{-t}$ *for all* $t \geq 0$, *and*

(E) $\lim_{\beta\to\infty} \mathrm{e}^{-\beta\Gamma^\star_{\text{NN}}} \, \mathbb{E}_{\boxminus}[\tau_{\boxplus}] = K$, *where*

$$\begin{aligned} \Gamma^\star_{\text{NN}} &= \mathrm{H}_{\text{NN}}(\chi) - \mathrm{H}_{\text{NN}}(\boxminus) \quad \text{for all } \chi \in \mathcal{C}_{\text{NN}}, \text{ and} \\ K^{-1} &= \frac{4(2L^\star - 5)}{3}|\Lambda|. \end{aligned} \tag{I.4.13}$$

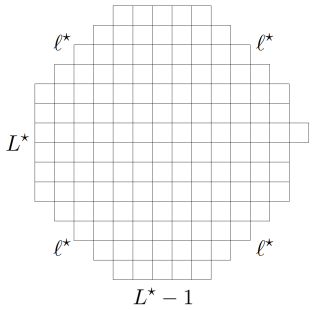


Figure I.4: An example of a configuration in $\mathcal{C}_{\text{NN}}$.

As in Theorem I.13, some of the results from Theorem I.15 are already known. For instance, part (A) and part (D) of Theorem I.15 and the asymptotics up to logarithmic equivalence of $\mathbb{E}_{\boxminus}[\tau_{\boxplus}]$ follow from [89, Theorem 1, 2 and 3]. The main contribution here is the sharp asymptotic expression of $\mathbb{E}_{\boxminus}[\tau_{\boxplus}]$ in Theorem I.15 (E).

I.4.4 Ising model with alternating magnetic field

In the third modification of the standard Ising model, the magnetic field is allowed to take alternating signs and absolute values on even and on odd rows. The path-wise approach has been applied to this model in [107]. The Hamiltonian here is given by

$$\mathrm{H}_{\pm}(\sigma) = -\frac{J}{2} \sum_{(x,y)\in\Lambda^\star} \sigma(x)\sigma(y) + \frac{h_2}{2} \sum_{x\in\Lambda_2} \sigma(x) - \frac{h_1}{2} \sum_{x\in\Lambda_1} \sigma(x), \tag{I.4.14}$$

where $\sigma \in S$, $J, h_2, h_1 > 0$, $\Lambda_2 = \{(x_1, x_2) \in \Lambda \,|\, x_2 \text{ is odd}\}$ are the *odd rows* in Λ, $\Lambda_1 = \Lambda \setminus \Lambda_2$ are the *even rows* and $\Lambda^\star$ is the set of *unordered nearest-neighbour bonds* in Λ. The critical lengths in this model are given by

$$l_b^\star = \left\lceil \frac{\mu}{\varepsilon} \right\rceil \qquad \text{and} \qquad l_h^\star = 2l_b^\star - 1, \tag{I.4.15}$$

where

$$\begin{aligned} &\varepsilon = h_1 - h_2, \quad \text{and} \\ &\mu = 2J - h_2. \end{aligned} \tag{I.4.16}$$

$l_b^\star$ will be the length of the basis of the critical droplet, and $l_h^\star$ will be its height. We make the following assumptions in this model.

Assumption I.16 *(a)* $h_1 > h_2$,

(b) $J > h_1$,

(c) $\frac{\mu}{\varepsilon} \notin \mathbb{N}$,

(d) $|\Lambda|$ *is large enough.*

We discuss the reasons for these assumptions in Chapter II.

We now formulate the main result for this model. For a more precise formulation and the proof we refer to Chapter II.

Theorem I.17 (cf. Theorem II.24) *Let $\mathcal{C}_\pm = \mathcal{C}_1 \cup \mathcal{C}_2 \subset S$, where the sets $\mathcal{C}_1$ and $\mathcal{C}_2$ are defined in Chapter II. Two examples of configurations in $\mathcal{C}_\pm$ are given in Figure I.5. Suppose Assumption I.16. Then,*

(A) $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\tau_{\mathcal{C}_\pm} < \tau_{\boxplus} \mid \tau_{\boxplus} < \tau_{\boxminus}] = 1$,

(B) for all $\chi \in \mathcal{C}_\pm$, we have that $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\sigma_{\tau_{\mathcal{C}_\pm}} = \chi] = \frac{1}{|\mathcal{C}_\pm|}$,

(C) $\lim_{\beta\to\infty} \lambda_\beta \, \mathbb{E}_{\boxminus}[\tau_{\boxplus}] = 1$, *where λ_β is the second largest eigenvalue of $-\mathcal{L}_\beta$,*

(D) $\lim_{\beta\to\infty} \mathbb{P}_{\boxminus}[\tau_{\boxplus} > t\mathbb{E}_{\boxminus}[\tau_{\boxplus}]] = \mathrm{e}^{-t}$ *for all $t \geq 0$, and*

(E) $\lim_{\beta\to\infty} \mathrm{e}^{-\beta\Gamma_\pm^\star} \, \mathbb{E}_{\boxminus}[\tau_{\boxplus}] = K$, *where*

$$\begin{aligned} \Gamma_\pm^\star &= \mathrm{H}_\pm(\chi) - \mathrm{H}_\pm(\boxminus) \quad \text{for all } \chi \in \mathcal{C}_\pm, \text{ and} \\ K^{-1} &= \frac{14\,(l_b^\star - 1)}{3} |\Lambda|. \end{aligned} \tag{I.4.17}$$

As in Theorem I.13 and Theorem I.17, the parts (A) and (D) of Theorem I.17 and the asymptotics up to logarithmic equivalence of $\mathbb{E}_{\boxminus}[\tau_{\boxplus}]$ follow from [107, Theorem 1, Section 4 and Section 5]. The main contribution here is the sharp asymptotic expression of $\mathbb{E}_{\boxminus}[\tau_{\boxplus}]$ in Theorem I.17 (E).

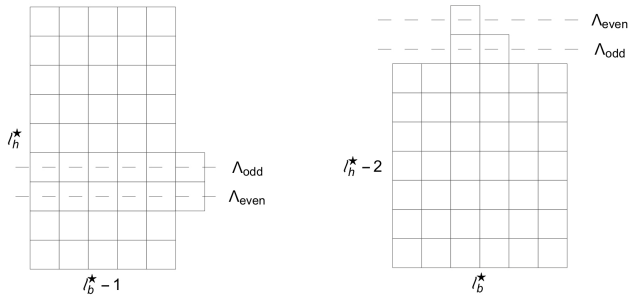


Figure I.5: Examples of configurations in $\mathcal{C}_{\pm}$.

I.5 The results of Chapter III

One of the biggest challenges in statistical mechanics is to deal with disordered systems that consist of a large number of particles. As we already mentioned in the example of the *Curie-Weiss model* in Subsection I.1.3, except of studying the behaviour of the system on the *microscopic level*, it is useful to study its behaviour under a suitable mapping. This mapping is often called the *macroscopic order parameter*. The goal is then to study the behaviour of the system on the *macroscopic level*.

In many examples, where the interaction between the particles is of *mean-field* type, this procedure can be applied successfully. For instance, in the papers [47] and [75], a system of $N \in \mathbb{N}$ *mean-field interacting diffusions* is considered. It is shown that, by taking the *empirical distribution* as the macroscopic order parameter, the macroscopic behaviour of the system can be described by the solution of the so-called *McKean-Vlasov equation*. Another example is the Curie-Weiss model, which we already introduced in Subsection I.1.3. We mentioned its macroscopic behaviour tacitly in equation (I.1.15), where we introduced the object f_β. More precisely, (I.1.15) implies that, by taking the *empirical mean* as the macroscopic order parameter, f_β serves as the *macroscopic free energy function* (or *macroscopic Hamiltonian*) of the system. Another mean-field setting was considered in the paper [63]. Here the authors use the Sandier-Serfaty approach, which we motivated in Section I.3, to study the macroscopic behaviour of a discrete mean-field interacting particle system.

We have two main goals in Chapter III. The first one is the extension of the results of [47] and [75] to the case, where the interaction is of *local* mean-field type instead of mean-field type. The second goal is to introduce a Wasserstein-like gradient flow structure on the macroscopic level, and to apply the (Fathi-)Sandier-Serfaty approach. The motivation behind the second goal is to make use of the advantages of the Wasserstein formalism and the (Fathi-)Sandier-Serfaty approach, which we explained in Section I.2 and Section I.3, respectively.

More precisely, we establish the following results in Chapter III.

- In Section III.1 we *modify the Wasserstein distance* and establish a *gradient flow formalism* with respect to the resulting metric. Then we show that gradient flows in this modified Wasserstein space correspond to partial differential equations, which depend on a non-evolving parameter. In particular, we investigate a special example, which will represent the macroscopic behaviour of a local mean-field interacting spin system, which is introduced in Section I.5.1 and more rigorously in Section III.2.1.

- In Section III.2 we use the Fathi-Sandier-Serfaty approach and the results of Section

III.1 to prove a *large deviation principle* for the system in Section I.5.1.

- In Section III.3 we use the Sandier-Serfaty approach to prove a *law of large numbers* for the system in Section I.5.1. Although this result already follows from the large deviation principle from Section III.2, we reprove the statement in order to obtain the law of large numbers for a slightly larger class of initial values and with respect to the stronger topology of the Wasserstein distance.

The results of Section III.1 are new, whereas some of the results of Section III.2 and III.3 have already been proven in [33] and [105], respectively, via different approaches. For instance, the large deviation principle was proven via the approach of the paper [47], and the law of large numbers was proven via the so-called *relative entropy method* (see [87] or [129] for more informations on this method). The main purpose of the Sections III.2 and III.3 is to use the Wasserstein formalism and the (Fathi-)Sandier-Serfaty approach to prove these claims. We motivated the advantages of these in the Sections I.2 and I.3.

There are three further important differences between the results from the Sections III.2 and III.3 and the results from [33] and [105]. The first one is that the rate function in Section III.2 differs from the one in [105]. The second difference is that, by using the gradient flow formalism from Section III.1, we also show here that the rate function admits a unique minimum point. This fact is not shown in [105]. The third difference is that in Section III.3 we also establish the convergence in the stronger topology of the Wasserstein distance. We believe that these differences and the advantages coming from the Wasserstein formalism and the (Fathi-)Sandier-Serfaty approach are useful ingredients for the study of the metastable behaviour of this model. This is planned for future research.

We finally note that, after the results of Chapter III have been published, the same methods as in Chapter III are used in the paper [36] to show the law of large numbers for a system of mean-field interacting diffusions. The setting in [36] differs from the one in Chapter III only in the precise form of the interaction part in the dynamics.

This section is organized as follows. In Subsection I.5.1 we introduce the microscopic spin system. In Subsection I.5.2 we define the macroscopic object and show how to modify the Wasserstein distance in order to obtain a gradient flow representation for this system. In Subsection I.5.3 we provide a first formulation of the main results of Chapter III and sketch the main ideas of the proofs.

I.5.1 The microscopic spin system

Let $T \in (0,\infty)$ and $N \in \mathbb{N}$. We denote by $\mathbb{T}$ the one-dimensional unit torus. Let $\Psi : \mathbb{R} \to \mathbb{R}$ and $J : \mathbb{T} \to \mathbb{R}$ be two functions that satisfy Assumption III.33 below. Moreover, let $B = (B^i)_{i=0,\dots,N-1}$ be an N-dimensional Brownian motion and $\mu_0^N \in \mathcal{M}_1(\mathbb{R}^N)$. In Chapter III we consider a system of N coupled stochastic differential equations given by

$$d\theta_t^{i,N} = -\Psi'\left(\theta_t^{i,N}\right) dt + \frac{1}{N}\sum_{j=0}^{N-1} J\left(\frac{i-j}{N}\right)\theta_t^{j,N}\, dt + \sqrt{2}\, dB_t^i, \quad t \in (0,T],\ 0 \le i \le N-1,$$
$$(\theta_0^{0,N},\dots\theta_0^{N-1,N}) \sim \mu_0^N. \tag{I.5.1}$$

For each $i = 0,\dots,N-1$ and $t \in [0,T]$, we call $\theta_t^{i,N}$ the *spin value* at time t of a particle, which is *located* at $i/N \in \mathbb{T}$. For a detailed historical review on such models we refer to [105, Subsection 1.1].

Define the *microscopic Hamiltonian* $\mathrm{H}^N : \mathbb{R}^N \to \mathbb{R}$ by

$$\mathrm{H}^N(\Theta) = \sum_{i=0}^{N-1} \left(\Psi\left(\theta^i\right) - \frac{1}{2N} \sum_{j=0}^{N-1} J\left(\frac{i-j}{N}\right) \theta^i \theta^j \right). \tag{I.5.2}$$

Let $\Theta_t^N := (\theta_t^{i,N})_{i=0,\dots,N-1}$ denote the vector of all N spins. Then we observe that

$$d\Theta_t^N = -\nabla \mathrm{H}^N(\Theta_t^N)\, dt + \sqrt{2}\, dB_t^N \quad \text{and} \quad \Theta_0^N \sim \mu_0^N. \tag{I.5.3}$$

Let μ_t^N denote the law of Θ_t^N for each $t \in [0,T]$. We have seen in Section I.2 that $(\mu_t^N)_{t\in[0,T]}$ can be represented as a *Wasserstein gradient flow* for the relative entropy

$$\mathcal{H}^N(\cdot) \;:=\; \mathcal{H}(\cdot \,|\, \mathrm{e}^{-\mathrm{H}^N(x)} dx). \tag{I.5.4}$$

Moreover, as we have shown in Lemma I.9, for each t, μ_t^N has a density ρ_t^N with respect to the Lebesgue measure on $\mathbb{R}^N$ and $(\rho_t^N)_{t\in[0,T]}$ is a weak solution to the *Fokker-Planck equation*

$$\partial_t \rho_t^N = \Delta \rho_t^N + \operatorname{div}\left(\nabla \mathrm{H}^N \rho_t^N\right). \tag{I.5.5}$$

In this thesis we focus on curves of laws rather than on the path-wise solutions of systems of stochastic differential equations. Hence, instead of the systems (I.5.1) and (I.5.3) we study $(\mu_t^N)_{t\in[0,T]}$ and $(\rho_t^N)_{t\in[0,T]}$. However, it is also possible to specify the roles of (I.5.1) and (I.5.3) in the results of this thesis; see [33].

In order to analyse the curves $(\mu_t^N)_{t\in[0,T]}$ as $N \to \infty$, we push all measures into the same space via the map K^N that sends a vector to the corresponding *empirical pair measure*, i.e.,

$$\begin{aligned} K^N : \mathbb{R}^N &\to \mathcal{M}_1(\mathbb{T}\times\mathbb{R}) \\ \Theta = (\theta^k)_{k=0}^{N-1} &\mapsto \frac{1}{N} \sum_{k=0}^{N-1} \delta_{\left(\frac{k}{N}, \theta^k\right)}. \end{aligned} \tag{I.5.6}$$

The goal is to state a law of large numbers and a large deviation principle for the sequence $\{((K^N)_{\#}\mu_t^N)_{t\in[0,T]}\}_N$, where $(K^N)_{\#}\mu_t^N$ denotes the image measure of μ_t^N under K^N.

I.5.2 The macroscopic object

We first explain intuitively what the limiting system should be. Note that (I.5.1) is of the form

$$d\theta_t^{i,N} = b\left(\frac{i}{N}\,,\, \theta_t^{i,N}\,;\, K^N(\Theta_t^N)\right) dt + \sqrt{2}\, dB_t^{i,N}, \tag{I.5.7}$$

where $b : \mathbb{T}\times\mathbb{R}\times\mathcal{M}_1(\mathbb{T}\times\mathbb{R}) \to \mathbb{R}$ is given by

$$b(x,\theta;\nu) = -\Psi'(\theta) + \int_{\mathbb{T}\times\mathbb{R}} J(x-x')\theta' d\nu(x',\theta'). \tag{I.5.8}$$

This suggests that the limiting system should be

$$d\hat{\theta}_t^x = b\left(x\,,\, \hat{\theta}_t^x\,;\, \mu_t\right) dt + \sqrt{2}\, dB_t^x, \qquad x \in \mathbb{T}, \tag{I.5.9}$$

where $\mu_t \in \mathcal{M}_1(\mathbb{T}\times\mathbb{R})$ is of the form $\mu_t = \mu_t^x\,dx$ and such that μ_t^x is the law of $\hat{\theta}_t^x$ for all t and x. However, this in turn suggests that μ_t should have a density ρ_t with respect to the Lebesgue measure on $\mathbb{T}\times\mathbb{R}$ for all $t\in(0,T]$ and $(\rho_t)_{t\in[0,T]}$ should be a weak solution of a partial differential equation of the form

$$\partial_t\rho_t(x,\theta) = \partial^2_{\theta\theta}\rho_t(x,\theta) + \partial_\theta\left(\rho_t(x,\theta)\left(\Psi'(\theta) - \int J(x-\bar{x})\bar{\theta}\rho_t(\bar{x},\bar{\theta})\,d\bar{\theta}d\bar{x}\right)\right). \tag{I.5.10}$$

It is not possible to find a representation of this partial differential equation in the usual Wasserstein setting, since there are no partial derivatives with respect to x. Hence, we have to modify the Wasserstein distance in such a way that the new metric takes into account that there is no evolution in this parameter. It turns out that the correct distance is given by

$$\mathrm{W}^{\mathrm{L}}(\mu,\nu)^2 := \int_{\mathbb{T}} W_2(\mu^x,\nu^x)^2\,dx, \tag{I.5.11}$$

where $\mu = \mu^x\,dx \in \mathcal{M}_1(\mathbb{T}\times\mathbb{R})$ and $\nu = \nu^x\,dx \in \mathcal{M}_1(\mathbb{T}\times\mathbb{R})$ are suitable, and W_2 is the Wasserstein distance on $\mathcal{P}_2(\mathbb{R})$, which we introduced in (I.2.14); see Section III.1 for the details. Now we have to rebuild the whole gradient flow theory from Section I.2 for this new metric in order to show that we can represent (I.5.10) in this new framework. This is the content of Section III.1.

I.5.3 Results

In this section, we state our main results and sketch the ideas of the corresponding proofs. The first result is the gradient flow formulation of (I.5.10). More precisely, we show that the results of Section I.2 also hold for the modified Wasserstein distance W^{L} defined in (I.5.11).

Theorem I.18 (Gradient flow formulation, cf. Theorems III.35, III.40 and III.41) *Define $\mathcal{F}:\mathcal{M}_1(\mathbb{T}\times\mathbb{R})\to(-\infty,\infty]$ by*

$$\mathcal{F}(\mu) := \mathcal{H}(\mu|\mathrm{e}^{-\Psi(\theta)}dxd\theta) - \frac{1}{2}\int_{(\mathbb{T}\times\mathbb{R})^2} J(x-x')\theta\theta' d\mu(x,\theta)d\mu(x',\theta'), \tag{I.5.12}$$

where $\mathcal{H}$ is the relative entropy functional (see (I.2.34)*). Let $\mu_0\in D(\mathcal{F})$. Then there exists a unique* W^{L}*-gradient flow $(\mu_t)_{t\in[0,T]}$ for $\mathcal{F}$ with initial value μ_0. Moreover, for all $t\in[0,T]$, μ_t has a density ρ_t with respect to the Lebesgue measure on $\mathbb{T}\times\mathbb{R}$ and $(\rho_t)_{t\in[0,T]}$ is a weak solution to* (I.5.10)*. Finally, $(\mu_t)_{t\in[0,T]}$ is the unique* W^{L}*-continuous curve such that $\lim_{t\downarrow 0}\mathrm{W}^{\mathrm{L}}(\mu_t,\mu_0)=0$ and $\mathscr{J}[(\mu_t)_{t\in[0,T]}]=0$, where, for smooth curves $(\nu_t)_{t\in[0,T]}$, $\mathscr{J}[(\nu_t)_{t\in[0,T]}]$ is defined by*

$$\mathscr{J}[(\nu_t)_{t\in[0,T]}] := \mathcal{F}(\nu_T) - \mathcal{F}(\nu_0) + \frac{1}{2}\int_0^T \left(|\partial\mathcal{F}|^2(\nu_t) + |\nu'|^2(t)\right)dt, \tag{I.5.13}$$

where the objects $|\partial\mathcal{F}|$ and $|\nu'|$ are introduced in (III.1.70) *and* (III.1.40)*, respectively, and are defined analogously to* (I.2.31) *and* (I.2.20)*.*

To prove this result, we have to develop the same theory for W^{L} as in Section I.2 in the Wasserstein space. To this end, we first show that $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}),\mathrm{W}^{\mathrm{L}})$ is a Polish space (Lemma III.6, Lemma III.7 and Lemma III.8), where $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})$ is defined in (III.1.1) below.

Then we analyse curves in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$ and characterize W^{L}-absolutely continuous curve via distributional solutions of certain partial differential equations (Proposition III.10). This characterisation will later be the key fact to build the bridge to (I.5.10). In Subsection III.1.3, we introduce a subdifferential calculus in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$ and define the notion of gradient flows with it. Then we apply the abstract theory of Part I of the book [3] to show existence, uniqueness and further properties of W^{L}-gradient flows in Theorem III.27. In Subsection III.1.4, we finally consider the special case of the functional $\mathcal{F}$ and apply the previous results for this case and arrive at Theorem I.18.

The second result of Chapter III is the following large deviation principle.

Theorem I.19 (Large deviation principle, cf. Theorem III.47)
For all $N\in\mathbb{N}$, let $(\mu_t^N)_{t\in[0,T]}$ be defined as in Subsection I.5.1. Let $(\mu_0^N)_N$ satisfy Assumption III.43. Then $(\{(K^N)_\#\mu_t^N\}_{t\in[0,T]})_N$ satisfies a large deviation principle in $C([0,T];\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ with rate function

$$(\nu_t)_t \mapsto I[(\nu_t)_t] := \frac{1}{2}\mathscr{J}[(\nu_t)_t] + \mathcal{H}(\nu_0|\mu_0) \tag{I.5.14}$$

for some $\mu_0\in D(\mathcal{F})$ (see Theorem III.47 for details).

The proof is based on the paper [70] in the following way. For each N, let $\mathscr{J}^N := \mathscr{J}_{\mathcal{H}^N,T}$ be the energy-dissipation functional from Lemma I.8. That is, $(\mu_t^N)_{t\in[0,T]}$ is the unique W_2-continuous curve such that $\lim_{t\downarrow 0} W_2(\mu_t^N,\mu_0^N)=0$ and $\mathscr{J}^N[(\mu_t^N)_{t\in[0,T]}]=0$. Then, the results in [70] (combined with some additional arguments that we provide in the proof of Theorem III.47) show that in order to prove the large deviation principle for $(\{(K^N)_\#\mu_t^N\}_{t\in[0,T]})_N$ it is equivalent to show that the following two claims hold:

- If $(\nu_t)_{t\in[0,T]}\in C([0,T];\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ and $(\nu_t^N)_{t\in[0,T]}\in C([0,T];\mathcal{M}_1(\mathbb{R}^N))$ for all $N\in\mathbb{N}$ are such that $(K^N)_\#\nu_t^N \rightharpoonup \delta_{\nu_t}$ for all $t\in[0,T]$, then

$$\liminf_{N\to\infty}\frac{1}{N}\left(\frac{1}{2}\mathscr{J}^N[(\nu_t^N)_{t\in[0,T]}] + \mathcal{H}(\nu_0^N\,|\,\mu_0^N)\right) \geq I[(\nu_t)_{t\in[0,T]}]. \tag{I.5.15}$$

- For all $(\nu_t)_{t\in[0,T]}\in C([0,T];\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ there exists $(\nu_t^N)_{t\in[0,T]}\in C([0,T];\mathcal{M}_1(\mathbb{R}^N))$ for all $N\in\mathbb{N}$ such that $(K^N)_\#\nu_t^N \rightharpoonup \delta_{\nu_t}$ for all $t\in[0,T]$, and

$$\limsup_{N\to\infty}\frac{1}{N}\left(\frac{1}{2}\mathscr{J}^N[(\nu_t^N)_{t\in[0,T]}] + \mathcal{H}(\nu_0^N\,|\,\mu_0^N)\right) \leq I[(\nu_t)_{t\in[0,T]}]. \tag{I.5.16}$$

These two claims are shown in Subsection III.2.4 and III.2.5, respectively. Therefore, the large deviation principle is related to a (variant of) gamma-convergence result of the functionals $(\nu_t^N)_{t\in[0,T]}\mapsto \frac{1}{2}\mathscr{J}^N[(\nu_t^N)_t] + \mathcal{H}(\nu_0^N\,|\,\mu_0^N)$. We explain this in more detail in Section III.2.

The third result of Chapter III is the following law of large numbers.

Theorem I.20 (Law of large numbers; cf. Theorem III.62)
Let $(\mu_t)_{t\in[0,T]}$ be the W^{L}-gradient flow for $\mathcal{F}$ with initial value $\mu_0\in D(\mathcal{F})$. For all $N\in\mathbb{N}$, let $(\mu_t^N)_{t\in[0,T]}$ be defined as in Subsection I.5.1. Suppose that the sequence of initial conditions $(\mu_0^N)_N$ is such that $((K^N)_\#\mu_0^N)_N$ converges to δ_{μ_0} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ and

$$\lim_{N\to\infty}\frac{1}{N}\mathcal{H}(\mu_0^N\,|\,\mathrm{e}^{-\mathrm{H}^N}\mathrm{Leb}_{\mathbb{R}^N}) = \mathcal{F}(\mu_0). \tag{I.5.17}$$

Then $((K^N)_{\#}\mu_t^N)_N$ *converges to* δ_{μ_t} *weakly in* $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ *for all* $t\in[0,T]$ *and*

$$\lim_{N\to\infty}\frac{1}{N}\mathcal{H}(\mu_t^N\,|\,\mathrm{e}^{-\mathrm{H}^N}\mathrm{Leb}_{\mathbb{R}^N})=\mathcal{F}(\mu_t)\qquad\text{for all } t\in[0,T]. \tag{I.5.18}$$

Moreover, under some additional assumption on Ψ*, the convergence holds even in a stronger topology, which is induced by the Wasserstein topology on* $\mathcal{M}_1(\mathbb{T}\times\mathbb{R})$*.*

The assumption on the initial configurations here is weaker than in Assumption III.43. The proof uses the Sandier-Serfaty approach, which we introduced in Section I.3. The main ideas of the proof of Theorem I.20 are the same as in the proof of Lemma I.11.

Remark I.21 *Most of the statements that we prove in Chapter III can be extended easily. For instance, it is possible to add a random environment, which is drawn according to some* $\varsigma\in\mathcal{M}_1(\mathbb{R})$*, or to replace* $\mathbb{T}$ *by a compact Riemannian manifold* M*, or to allow the spins to take values in* $\mathbb{R}^d$ *for some* $d>1$*. The corresponding metric should then be of the form*

$$\mathrm{W}^{M,\varsigma}(\mu,\nu)^2:=\int_M\int_{\mathbb{R}}W_2(\mu^{m,\omega},\nu^{m,\omega})^2\,d\varsigma(\omega)\,d\mathrm{vol}(m). \tag{I.5.19}$$

Moreover, it is possible to generalize (I.5.10) *in various ways without much additional work. For instance, we could add a term of the form* $\partial^2_{\theta\theta}L_F(\rho_t(x,\theta))$ *for some function* $L_F:[0,\infty)\to[0,\infty)$ *as in [3, Example 9.3.6 and Subsection 10.4.3], or we could include a diffusion coefficient as in [70]. It is also straightforward to see that the single-site potentials* Ψ *could also be dependent on the space parameter* x*, and the quadratic interaction (given by the factor* $-\theta\theta'$ *in* (I.5.12)*) could be replaced by a more general class of interactions. However, we try to keep the notation as simple as possible and did not try to optimize our results.*

I.6 The results of Chapter IV

Already in the paper [46] it was conjectured that mean-field interacting diffusion systems exhibit metastable behaviour on the macroscopic level. More precisely, they consider the system (I.5.1) from Section I.5 in the special case that the function J is constant. Next they observe that the *macroscopic free energy* $\mathcal{F}$ (cf. (I.5.12)) admits two global minima[9]. Then they conjecture the exponential asymptotics of the average transition time between these minima for the empirical process. It is a long outstanding problem to verify this conjecture from [46] rigorously, and, in the next step, to compute this average transition time beyond the exponential asymptotics by using the potential-theoretic approach[10].

In Chapter IV we provide first progress towards these goals. In order to do this we simplify and modify the setting from Section I.5 in three ways.

- The first simplification is the following. In Section I.5 the system has two characteristics, a *fixed space variable* and a *spin value*. In Chapter IV we omit the fixed space variable. Therefore, there is only one characteristic left, and in Chapter IV we consider a system of *mean-field interacting diffusions*. Consequently, we only consider the special case, where the function J in (I.5.1) is constant.

[9] We reprove this statement in Lemma V.2. Note that it is shown in [106, Section IV.2] that this is also true in the *local mean-field* case.

[10] Recall Subsection I.1.2.

- The second simplification is that we switch, in the bulk part of Chapter IV, to the so-called *low-temperature regime*. That is, we introduce a parameter $\varepsilon > 0$, which measures the strength of the Brownian noise, and consider the regime $\varepsilon \ll 1$.

- The third modification is the choice of the macroscopic order parameter. In Section I.5 the macroscopic order parameter is given by the *empirical distribution*, whereas in Chapter IV it is given by the *empirical mean*. The advantage of this choice is the availability of the so-called *local Cramér theorem*, which is the fundamental tool in Chapter IV to pass from the microscopic variables to the macroscopic ones.

These simplifications lead to **important changes in the interpretation and in the notation** for Chapter IV. These changes are given as follows. In order to be consistent with many references that we use in Chapter IV (especially with [32]), we interpret the mean-field interacting system of diffusions in Chapter IV as *time-evolving space variables*, and denote it by x. Of course, mathematically, the analogue of this object in Chapter III is the spin value. But we decided to change this interpretation and notation in order to be consistent with the literature that we use. Moreover, in this way, we are consistent in the sense that *space variables are always denoted by x in this thesis*. However, the price is that it may lead to confusions, since the analogous objects in Chapter III and in Chapter IV are denoted differently.

Hence, in Chapter IV we are interested in the metastable behaviour of a system of $N \in \mathbb{N}$ mean-field interacting stochastic differential equations given by

$$dx_i^{N,\varepsilon}(t) \;=\; -\psi'\left(x_i^{N,\varepsilon}(t)\right)\,dt - \frac{J}{N}\sum_{j=0}^{N-1}\left(x_i^{N,\varepsilon}(t) - x_j^{N,\varepsilon}(t)\right)\,dt + \sqrt{2\varepsilon}\,dB_i(t), \tag{I.6.1}$$

where $t \in (0,\infty)$, $0 \le i \le N-1$, $\varepsilon > 0$, $B^N = (B_i)_{i=0,\dots,N-1}$ is an N-dimensional Brownian motion, $J > 0$ and the *single-site potential* $\psi : \mathbb{R} \to \mathbb{R}$ is given by $\psi(z) = \frac{1}{4}z^4 - \frac{1}{2}z^2$. We consider the strength ε of the Brownian noise as the *temperature* of the system.

We proceed as follows. First, in order to analyse the system for large N, we choose the *empirical mean*, $P : \mathbb{R}^N \to \mathbb{R}$, $Px = 1/N\sum_{i=0}^{N-1} x_i$, as the macroscopic order parameter. That is, we consider the image of the system under the map P. Then, as a result of an improvement of the well-known Cramér theorem for this setting, which we call *local Cramér theorem* (see Section I.6.2), we obtain a function $\bar{H}_\varepsilon : \mathbb{R} \to \mathbb{R}$, which we interpret as the *macroscopic Hamiltonian* of the system. A simple analysis shows that $\bar{H}_\varepsilon$ admits exactly two global minima at $-m_\varepsilon^\star < 0$ and $m_\varepsilon^\star > 0$, and that $\bar{H}_\varepsilon$ admits a unique local maximum at 0. This fact indicates that our model exhibits metastable behaviour with the two metastable states being the hyperplanes $P^{-1}(m_\varepsilon^\star)$ and $P^{-1}(-m_\varepsilon^\star)$. The goal of Chapter IV is to compute the average transition time to a region around $P^{-1}(m_\varepsilon^\star)$, when the system is initially close to $P^{-1}(-m_\varepsilon^\star)$.

We tackle this goal in two different regimes, the first one being the *low-temperature regime*, where the strength ε of the Brownian noise tends to zero, and the second one being the *high-temperature regime*, where we set $\varepsilon = 1$. We obtain the following results in Chapter IV.

- In Section IV.1 we show that in the low-temperature regime and under the assumption that $J > 1$, the average transition time is asymptotically given by a formula, which is of a similar form as the well-known Eyring-Kramers formula (see [32] or Subsection I.1.3) up to a multiplicative error term that tends to 1 as $N \to \infty$ and $\varepsilon \downarrow 0$. Such a result is often known as *Kramers' law* in the literature; see Subsection I.1.3 and see [18] for a review on such results.

- In Section IV.2 we consider the high-temperature regime, where we only show that, as $N \to \infty$, the average transition time is confined to an interval $[\alpha\, \mathrm{e}^{N\Delta}, \beta\, \mathrm{e}^{N\Delta}]$, where $\Delta = \bar{H}_1(0) - \bar{H}_1(-m_1^\star)$, and $0 < \alpha < \beta < \infty$ are independent of N. This result still holds true if we replace ψ by a large class of single-site potentials.

We now provide a short remark on the historical background of metastability results in high-dimensional diffusion models. In the papers [9], [10], [19], and [22], Kramers' law has been shown for systems of N nearest-neighbour interacting stochastic differential equations in low temperature. These models are considered as N-dimensional approximations of stochastic partial differential equations. A similar setting was studied in [20], where, instead of the potential-theoretic approach, the path-wise approach to metastability was used. As we already mentioned, for mean-field interacting systems in the high-temperature regime (i.e. for exactly the same setting as in Section IV.2), the asymptotic behaviour, up to logarithmic equivalence, of the average transition time has been stated without proof in [46, Theorem 4]. The rough estimates from Section IV.2 provide a slightly improved version of this conjecture under different initial conditions (see Section I.6.4 for more details).

This section is organized as follows. In Subsection I.6.1 we define the microscopic model. Then, in Subsection I.6.2 we introduce the macroscopic order parameter, and collect some result on the energy landscape of the model under this order parameter. In Subsection I.6.3 and I.6.4 we provide a first formulation of the two main results of Chapter IV.

I.6.1 The microscopic model

We consider a system of N stochastic differential equations defined by

$$dx_i^{N,\varepsilon}(t) \;=\; -\psi'\left(x_i^{N,\varepsilon}(t)\right)\, dt - \frac{J}{N}\sum_{j=0}^{N-1}\left(x_i^{N,\varepsilon}(t) - x_j^{N,\varepsilon}(t)\right)\, dt + \sqrt{2\varepsilon}\, dB_i(t), \tag{I.6.2}$$

where $t \in (0,\infty)$, $0 \le i \le N-1$, $\varepsilon > 0$, $B^N = (B_i)_{i=0,\dots,N-1}$ is an N-dimensional Brownian motion, $J > 0$ and the *single-site potential* $\psi : \mathbb{R} \to \mathbb{R}$ is given by

$$\psi(z) = \frac{1}{4}z^4 - \frac{1}{2}z^2. \tag{I.6.3}$$

This model has already been studied extensively in the literature; see for instance [46], [47], [48] and [75].

The *Gibbs measure* $\mu^{N,\varepsilon} \in \mathcal{P}(\mathbb{R}^N)$ corresponding to this model has the form

$$\mu^{N,\varepsilon}(dx) = \frac{1}{Z_{\mu^{N,\varepsilon}}}\, \mathrm{e}^{-\mathrm{H}^{N,\varepsilon}(x)} dx, \tag{I.6.4}$$

where $Z_{\mu^{N,\varepsilon}}$ is a normalization constant, and, for $x = (x_i)_{i=0,\dots,N-1} \in \mathbb{R}^N$, the *microscopic Hamiltonian* $\mathrm{H}^{N,\varepsilon} : \mathbb{R}^N \to \mathbb{R}$ is defined by

$$\mathrm{H}^{N,\varepsilon}(x) = \frac{1}{\varepsilon}\sum_{i=0}^{N-1}\psi(x_i) + \frac{1}{\varepsilon}\frac{J}{4N}\sum_{i,j=0}^{N-1}(x_i - x_j)^2. \tag{I.6.5}$$

For $t \in (0,\infty)$, let $x^{N,\varepsilon}(t) = (x_0^{N,\varepsilon}(t), \dots, x_{N-1}^{N,\varepsilon}(t))$. It is well-known that $\mu^{N,\varepsilon}$ is the unique stationary measure of the process $(x^{N,\varepsilon}(t))_{t\in(0,\infty)}$.

I.6.2 The macroscopic variables and the macroscopic energy landscape

The *empirical mean* $P : \mathbb{R}^N \to \mathbb{R}$ is defined by

$$Px \;=\; \frac{1}{N}\sum_{i=0}^{N-1} x_i. \tag{I.6.6}$$

This operator will act as the macroscopic order parameter for our microscopic system. That is, in order to analyse the process $(x^{N,\varepsilon}(t))_{t\in(0,\infty)}$ for large N, we study the image of this process under the map P. Therefore, intuitively, $\bar{\mu}^{N,\varepsilon} := P_{\#}\mu^{N,\varepsilon}$ describes the (long-time) macroscopic behaviour of the model, and it will be crucial to study the asymptotic behaviour of this measure.

In fact, in Proposition IV.1, we show that for any compact set $K \subset \mathbb{R}$, there exists a function $R_K : K \times [0,1] \times \mathbb{N} \to [0,\infty)$ and a constant $C_K > 0$ such that $|R_K(m,\varepsilon,N)| \;\leq\; C_K/\sqrt{N}$ for all $m \in K$, for ε small enough and N large enough, and such that

$$\bar{\mu}^{N,\varepsilon}(dm) \;=\; \mathrm{e}^{-N\bar{H}_\varepsilon(m)}\sqrt{\frac{\varphi_\varepsilon''(m)}{2\pi}}\,dm\;\;(1+R_K(\varepsilon,N,m))\,. \tag{I.6.7}$$

Here, $\varphi_\varepsilon : \mathbb{R} \to \mathbb{R}$ is the so-called *Cramér transform* of the Gibbs measure with respect to the single-site potential (or more precisely with respect to the effective single-site potential defined in (IV.1.2)) and is defined in (IV.1.9), and $\bar{H}_\varepsilon : \mathbb{R} \to \mathbb{R}$ is defined by

$$\bar{H}_\varepsilon(z) \;=\; \varphi_\varepsilon(z) \;-\; \frac{1}{\varepsilon}\frac{J}{2}z^2. \tag{I.6.8}$$

Since $\bar{\mu}^{N,\varepsilon}$ is the law of the empirical mean of a sequence of random variables, (I.6.7) can be seen as an improvement of the well-known *Cramér theorem* (cf. [52, 6.1.3]) for this setting. This explains, why we call this result *local Cramér theorem*.

Equation (I.6.7) shows that, for large N and for ε small enough, $\bar{\mu}^{N,\varepsilon}$ is very similar to a Gibbs measure with $\bar{H}_\varepsilon$ playing the role of the energy function. Therefore, we consider $\bar{H}_\varepsilon$ as the *macroscopic Hamiltonian* of the system. This suggests to study the analytic properties of the function $\bar{H}_\varepsilon$. We do this in Lemma IV.2, where we show that, for ε small enough. $\bar{H}_\varepsilon$ is a symmetric double-well function with two global minima at $-m_\varepsilon^\star < 0$ and $m_\varepsilon^\star > 0$, and with a local maximum at 0. That is, $\bar{H}_\varepsilon$ is of the form given in Figure I.6.

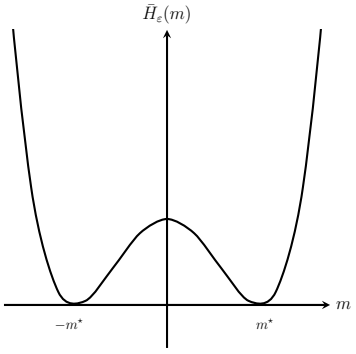


Figure I.6: Form of the graph of the function $\bar{H}_\varepsilon$.

I.6.3 The Eyring-Kramers formula at low temperature

The fact that the macroscopic Hamiltonian $\bar{H}_\varepsilon$ has two global minima at $-m_\varepsilon^\star$ and $m_\varepsilon^\star$ suggests that our system exhibits metastable behaviour in the following sense. Suppose that the initial condition of our system is concentrated in a small region around the hyperplane $P^{-1}(-m_\varepsilon^\star)$. Then, we expect that the average transition time to hit a small region around $P^{-1}(m_\varepsilon^\star)$ fulfils Kramers' law. This is the content of the main result of Chapter IV, which is formulated in Theorem I.22. (For a more detailed formulation of the result, we refer to Section IV.1.2.) In this theorem, we suppose that

$$J > 1. \tag{I.6.9}$$

The reason for this assumption is that, in this regime, we are able to control the microscopic fluctuations via functional inequalities. We explain this in further detail in Remark IV.10. To show the metastable behaviour for the case $J \leq 1$ is the content of future research.

Theorem I.22 (cf. Theorem IV.7) *Suppose* (I.6.9). *Let*

$$\mathcal{T} \;=\; \inf\{t > 0 \,|\, Px^{N,\varepsilon}(t) \;\geq\; m_\varepsilon^\star - \eta\} \tag{I.6.10}$$

for some specific $\eta \;=\; \Omega(\sqrt{\log(N)/N}\,\sqrt{\varepsilon\log(\varepsilon^{-1})})$ *(see* (IV.1.26)*). Then, for* ε *small enough, and for* N *large enough,*

$$\mathbb{E}_{\nu_{B^-,B^+}}[\mathcal{T}] \;=\; \frac{2\pi\,\sqrt{\varphi_\varepsilon''(-m_\varepsilon^\star)}\;\mathrm{e}^{N(\bar{H}_\varepsilon(0)-\bar{H}_\varepsilon(-m_\varepsilon^\star))}}{\varepsilon\,\sqrt{\bar{H}_\varepsilon''(-m_\varepsilon^\star)\,|\bar{H}_\varepsilon''(0)|\,\varphi_\varepsilon''(0)}}\,\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)+O\left(\varepsilon^2\right)\right), \tag{I.6.11}$$

where ν_{B^-,B^+} *is a probability measure, which is concentrated on the set* $\{\,Px \;=\; -m_\varepsilon^\star + \eta\,\}$ *and is called* last-exit biased distribution on B^- *(see* (IV.1.21) *for the definition of* ν_{B^-,B^+} *and see* (IV.1.27) *for the definition of the sets* B^- *and* B^+*), and where* $\mathbb{E}_{\nu_{B^-,B^+}}[\mathcal{T}] := \int \mathbb{E}_x[\mathcal{T}]\,d\nu_{B^-,B^+}(x)$.

To prove this result in Section IV.1, we proceed as follows.

In Subsection IV.1.1 we collect three important ingredients. More precisely, we first state the local Cramér theorem (i.e. (I.6.7)), which is the key tool in our proof to go from the microscopic variables to the macroscopic ones. Then, we study the analytic properties of the macroscopic Hamiltonian $\bar{H}_\varepsilon$ and show that its graph is of the form given in Figure I.6. And as the third ingredient, we collect the key elements from potential theory that allow us to rewrite the average transition time, $\mathbb{E}_{\nu_{B^-,B^+}}[\mathcal{T}]$, in terms of quantities from electric networks. Namely, we show that $\mathbb{E}_{\nu_{B^-,B^+}}[\mathcal{T}]$ is equal to the quotient of the *mass of the equilibrium potential* and the *capacity*; see Lemma IV.5 below or the first example of Subsection I.1.3, where we motivated the potential-theoretic approach to metastability in a simple setting.

After we collect these ingredients, we formulate the main result of Chapter IV in Subsection IV.1.2. The proof of this result is divided into three steps. The first step consists of showing the correct upper bound for the capacity in Subsection IV.1.3. This is done by using the so-called *Dirichlet principle* (see Lemma IV.6). Here we have to choose an appropriate test function and compute the asymptotic value of the corresponding *Dirichlet form*. In the second step, we compute in Subsection IV.1.4 the lower bound on the capacity by an adaptation of the so-called *two-scale approach*, which was initiated in the paper [80]. This is the main

point, where we use the assumption (I.6.9). We explain this is further detail in Remark IV.10. Finally, we compute in Subsection IV.1.5 the asymptotic value of the mass of the equilibrium potential. This follows from applying standard Laplace asymptotics, and by exploiting that the graph of $\bar{H}_\varepsilon$ has the form of a double-well function (cf. Figure I.6).

I.6.4 Rough estimates at high temperature

We also consider in Chapter IV the situation where the microscopic fluctuations in the system do not become negligible. That is, we study the system $(x^{N,1}(t))_{t\in(0,\infty)}$ given by (I.6.2) with $\varepsilon = 1$. It is not surprising that the methods that we use for the setting in Subsection I.6.3 do not yield the precise Eyring-Kramers formula in the present case. The reason is that in this case, the *entropy* of the paths matters substantially, i.e. the microscopic fluctuations do not allow to restrict solely to the macroscopic variables under the order parameter P. We believe that, in order to obtain the Eyring-Kramers formula, we need to consider, as in Section I.5, the *empirical distribution* $K^N : \mathbb{R}^N \to \mathcal{P}(\mathbb{R})$,

$$K^N x \;=\; \frac{1}{N}\sum_{i=0}^{N-1} \delta_{x_i} \tag{I.6.12}$$

as the order parameter instead of P. An heuristic argument for that is the following.

For all $N \in \mathbb{N}$ and $t \in (0,\infty)$, let $\gamma_N(t) = K^N(x^{N,1}(t))$. Already Dawson and Gärtner [48, (1.8)] obtained a diffusion-like equation for the evolution of $(\gamma_N(t))_{t\in(0,\infty)}$ of the form

$$d\langle \gamma_N(t), f\rangle = \langle \gamma_N(t), L(\gamma_N(t))f\rangle dt + \frac{1}{\sqrt{N}} dM_t^f \qquad \text{for all smooth functions } f, \tag{I.6.13}$$

where L is the generator of the McKean-Vlasov equation and M_t^f is a martingale for all such f with quadratic variation process $[M^f]_t$ given by (cf. [48, (1.9)])

$$d[M^f]_t = 2\langle \gamma_N(t), |f'|^2\rangle dt. \tag{I.6.14}$$

Moreover, the first term on the right-hand side of (I.6.13) can be interpreted in the Wasserstein formalism as follows. Let $\mathcal{F}$ be the *macroscopic free energy* functional on the Wasserstein space corresponding to $(\gamma_N(t))_{t\in(0,\infty)}$; see (I.5.12). Then, by using [72, Theorem D.28] we have that

$$\langle \gamma_N(t), L(\gamma_N(t))f\rangle = \langle \mathrm{Grad}_{\mathrm{Wass}}\mathcal{F}(\gamma_N(t)), f\rangle, \tag{I.6.15}$$

where $\mathrm{Grad}_{\mathrm{Wass}}\mathcal{F}$ is the gradient of $\mathcal{F}$ in the Wasserstein space interpreted in the sense of distributions as in [72, Definition 9.36]. Here we used the formal Riemannian setting on the Wasserstein space introduced in [111], and which we motivated in Section I.2. Thus, combining (I.6.13), (I.6.14), (I.6.15) and Theorem I.20, suggests that the random perturbations of the process $(\gamma_N(t))_{t\in(0,\infty)}$ are of order $1/\sqrt{N}$, and that the potential landscape for this process is given by the free energy in the Wasserstein space given by $\mathcal{F}$. This provides an intuitive justification that, in the limit as $N \uparrow \infty$, one is in a weak noise setting analogously to [32] (or the first example in Subsection I.1.3) but in the infinite dimensional Wasserstein space. Moreover, it justifies the choice K^N as the macroscopic order parameter, and shows that the Wasserstein setting, which we introduced in Section I.2, is the natural framework.

Following these observations, we should be able to follow the same strategy as in [32] (or [29, Chapter 11]) to study the metastable behaviour of the process $(\gamma_N(t))_{t\in(0,\infty)}$. In order to

do that, we plan to proceed as follows. We first use standard results from potential theory to represent expected transition times between the metastable states associated to $(\gamma_N(t))_{t\in(0,\infty)}$ in terms of Dirichlet forms on the Wasserstein space. The next goal is then to derive sharp asymptotics of these Dirichlet forms in the limit as $N \to \infty$. At this point we should benefit both from the results obtained in [51] and [128], where a Malliavin calculus is constructed on the Wasserstein space, and from the results of Chapter V, where we study the ergodic behaviour of the gradient flows for $\mathcal{F}$. The latter is an important ingredient in the study of the metastable behaviour of $(\gamma_N(t))_{t\in(0,\infty)}$, since we know from Chapter III that the process $(\gamma_N(t))_{t\in(0,\infty)}$ can be approximated by these gradient flows. The rigorous implementation of these thoughts is left for future research.

However, we can still obtain estimates for the mean transition time under the order parameter P. Here we replace ψ by single-site potentials of the form $z \mapsto \Psi(z) - \frac{J}{2}z^2$, where $\Psi : \mathbb{R} \to \mathbb{R}$ is a symmetric and bounded perturbation of a strictly convex function (cf. Assumption IV.13). Moreover, we have to assume that $J > \int_{\mathbb{R}} \mathrm{e}^{-\Psi(z)}\,dz/(\int z^2\,\mathrm{e}^{-\Psi(z)}\,dz)$. This condition is necessary for $\bar{H}_1$ to be of the form of a double-well function. (Note that the objects $\varphi_1, \bar{H}_1, \nu_{B_1^-,B_1^+}, \mathcal{T}$ are defined as in Theorem I.22 but with ψ replaced by $z \mapsto \Psi(z) - \frac{J}{2}z^2$ and with $\varepsilon = 1$.) That is, in the case $J \leq \int_{\mathbb{R}} \mathrm{e}^{-\Psi(z)}\,dz/(\int z^2\,\mathrm{e}^{-\Psi(z)}\,dz)$, we do not have a metastable behaviour for the system under the order parameter P. This is different than in Theorem I.22, where we can show that $\bar{H}_\varepsilon$ is a double-well function also in the case $J \leq 1$ (see Lemma IV.2). The main result is the following statement.

Theorem I.23 (cf. Theorem IV.17) *Suppose Assumption IV.13. Let $\pm m_1^\star$ be the two global minimisers of the macroscopic Hamiltonian $\bar{H}_1$. Then, for all N large enough and for some $a > 0$, which is independent of N,*

$$\begin{aligned}
\mathbb{E}_{\nu_{B_1^-,B_1^+}}[\mathcal{T}] &\geq \frac{2\pi\,\sqrt{\varphi_1''(-m_1^\star)}\,\mathrm{e}^{N(\bar{H}_1(0)-\bar{H}_1(-m_1^\star))}}{\sqrt{\bar{H}_1''(-m_1^\star)\,|\bar{H}_1''(0)|\,\varphi_1''(0)}}\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right), \quad \textit{and} \\
\mathbb{E}_{\nu_{B_1^-,B_1^+}}[\mathcal{T}] &\leq (1+a)\frac{2\pi\,\sqrt{\varphi_1''(-m_1^\star)}\,\mathrm{e}^{N(\bar{H}_1(0)-\bar{H}_1(-m_1^\star))}}{\sqrt{\bar{H}_1''(-m_1^\star)\,|\bar{H}_1''(0)|\,\varphi_1''(0)}}\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right).
\end{aligned} \tag{I.6.16}$$

The proof of this result is organized in the same way as the proof of Theorem I.22, and is given in Section IV.2.

Finally, we point out that Theorem I.23 provides a slight improvement of the conjecture given in [46, Theorem 4]. Indeed, the authors of [46] expect that for all $\delta > 0$, there exists $N_\delta \in \mathbb{N}$ such that for $N \geq N_\delta$ the expected transition time is confined to the interval $[\mathrm{e}^{N(\Delta-\delta)}, \mathrm{e}^{N(\Delta+\delta)}]$, where $\Delta = \bar{H}_1(0) - \bar{H}_1(-m_1^\star)$. Here we have used the simple fact that $\bar{H}_1(0) - \bar{H}_1(-m_1^\star)$ can be written in terms of the free energy functional $\mathcal{F}$ from (I.5.12); see Lemma V.2 for more details. However, we note that the initial condition in our setting is different than in the conjecture formulated in [46, Theorem 4].

I.7 The results of Chapter V

In Chapter IV we analyse the metastable behaviour of a system of *mean-field interacting diffusions* on the macroscopic level. The macroscopic order parameter is chosen to be the *empirical mean*. In the so-called high-temperature regime (see Subsection I.6.4), where the microscopic fluctuations do not become negligible, our results are limited to rough estimates

for the metastable transition time. We already mentioned that, in order to obtain sharp estimates on this transition time, the macroscopic order parameter should be the *empirical distribution*, and the correct framework to analyse the corresponding macroscopic objects should be the *Wasserstein setting*. Therefore, it is important to analyse the *macroscopic energy landscape* in the Wasserstein space and the *long-time behaviour* of the corresponding macroscopic objects. This is because these two ingredients are essential for the rigorous study of the metastable behaviour of a stochastic process.

In Chapter V we provide a first step towards this goal. Before we state the main results of Chapter V, we recall two facts that we know about the setting of Subsection I.6.4.

- If the macroscopic order parameter is chosen to be the *empirical mean*, then the corresponding macroscopic Hamiltonian $\bar{H}_1$ admits exactly three critical points, which are located at $-m_1^\star$, 0 and $m_1^\star$ for some $m_1^\star > 0$; see Section I.6.
- If the macroscopic order parameter is chosen to be the *empirical distribution*, then the corresponding macroscopic Hamiltonian is given by the functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}) \to (-\infty, \infty]$, which is defined by

$$\mathcal{F}(\mu) \;=\; \int_{\mathbb{R}} \log(\rho) d\mu + \int_{\mathbb{R}} \Psi \, d\mu - \frac{J}{2} \left(\int_{\mathbb{R}} z \, d\mu(z) \right)^2 \tag{I.7.1}$$

  if $\mu \in \mathcal{P}_2(\mathbb{R})$ has a Lebesgue density ρ, and $\mathcal{F}(\mu) = \infty$ otherwise. In other words, the macroscopic behaviour of the system is approximately given by the unique gradient flow for $\mathcal{F}$; see Section I.5.

Note that, as an immediate consequence of these two facts, we observe that $\mathcal{F}$ admits exactly three critical points as well. More precisely, we have that

$$|\partial \mathcal{F}|(\mu) = 0 \qquad \text{if and only if} \qquad \mu \in \{\mu^-, \mu^0, \mu^+\}, \tag{I.7.2}$$

where $\mu^-, \mu^0, \mu^+ \in \mathcal{P}_2(\mathbb{R})$ are defined through the objects φ_1^* and φ_1 from Section I.6; see (V.0.3). We explain this in more detail in Lemma V.2.

Having these facts in hand, the main goal of Chapter V is to study the *ergodic behaviour* of the gradient flows for $\mathcal{F}$, i.e., their possible convergence towards the stationary measures. Moreover, another goal is to obtain more informations on the *energy landscape determined by* $\mathcal{F}$. As we already mentioned, we believe that these goals are essential for the investigation of the metastable behaviour of the system in Subsection I.6.4 under the macroscopic order parameter K^N (see (I.6.12)). The following theorem and its by-products provide first progress in this direction. From now on, let $(S[\mu](t))_{t\in(0,\infty)}$ denote the unique Wasserstein gradient flow for $\mathcal{F}$ with initial value $\mu \in \overline{D(\mathcal{F})} = \mathcal{P}_2(\mathbb{R})$; see [3, 11.2.8] or Section I.2.

Theorem I.24 *Suppose Assumption V.1. Let $\mu \in \mathcal{P}_2(\mathbb{R})$. Then, there exists a measure $\mu^* \in \{\mu^-, \mu^0, \mu^+\}$ such that*

$$\lim_{t\to\infty} W_2(S[\mu](t), \mu^*) \;=\; 0 \quad \text{and} \quad \lim_{t\to\infty} \mathcal{F}(S[\mu](t)) = \mathcal{F}(\mu^*). \tag{I.7.3}$$

The proof of this result is given in Chapter V. As a by-product of this proof, we obtain the following two propositions, which are interesting on their own. The first one shows that inside the valleys of the set $\{\mu \in \mathcal{P}_2(\mathbb{R}) \,|\, \mathcal{F}(\mu) \le \mathcal{F}(\mu^0)\}$ the convergence of the gradient flows for $\mathcal{F}$ is determined by the sign of the mean of the initial value.

Proposition I.25 *Suppose Assumption V.1. Let $\mu \in \mathcal{P}_2(\mathbb{R})$ be such that $\int_{\mathbb{R}} z\, d\mu(z) \neq 0$ and $\mathcal{F}(\mu) \leq \mathcal{F}(\mu^0)$. Then,*

$$\lim_{t\to\infty} \mathcal{F}(S[\mu](t)) \;=\; \mathcal{F}(\mu^-) \;=\; \mathcal{F}(\mu^+), \tag{I.7.4}$$

and

$$\lim_{t\to\infty} W_2(S[\mu](t), \mu^-) \;=\; 0 \quad \text{if } \int_{\mathbb{R}} z\, d\mu(z) < 0 \quad \text{and} \tag{I.7.5}$$

$$\lim_{t\to\infty} W_2(S[\mu](t), \mu^+) \;=\; 0 \quad \text{if } \int_{\mathbb{R}} z\, d\mu(z) > 0. \tag{I.7.6}$$

The second by-product is the following proposition on the energy landscape determined by $\mathcal{F}$. This proposition provides useful informations about the topological properties of the basins of attraction of the stationary measures μ^-, μ^0 and μ^+. This is an important ingredient for the proof of Theorem I.24.

Proposition I.26 *Suppose Assumption V.1. Let $\mathcal{B}^-$, $\mathcal{B}^0$ and $\mathcal{B}^-$ be the* basins of attraction *of the stationary measures μ^-, μ^0 and μ^+, respectively. That is,*

$$\begin{aligned} \mathcal{B}^- &= \left\{\mu \in \mathcal{P}_2(\mathbb{R}) \;\middle|\; \lim_{t\to\infty} S[\mu](t) = \mu^-\right\}, \\ \mathcal{B}^+ &= \left\{\mu \in \mathcal{P}_2(\mathbb{R}) \;\middle|\; \lim_{t\to\infty} S[\mu](t) = \mu^+\right\}, \quad \text{and} \\ \mathcal{B}^0 &= \left\{\mu \in \mathcal{P}_2(\mathbb{R}) \;\middle|\; \lim_{t\to\infty} S[\mu](t) = \mu^0\right\}. \end{aligned} \tag{I.7.7}$$

Then, $\mathcal{B}^-$ and $\mathcal{B}^+$ are open subsets of $\mathcal{P}_2(\mathbb{R})$, and $\mathcal{B}^0$ is a closed subset of $\mathcal{P}_2(\mathbb{R})$.

The results of Chapter V are not completely new. Indeed, Theorem I.24 and Proposition I.25 are mild extensions of the results that have already been obtained in the paper [125]. The proofs in [125] are based on methods from the theory of partial differential equations. The main contributions of Chapter V are that we use the Wasserstein framework to prove these results (which provides shorter proofs than in [125]), and that the results hold in the stronger topology of the Wasserstein distance (whereas the results in [125] are formulated in terms of the weak topology). However, to our knowledge, Proposition I.26 is a new result. It is expected that this proposition will become useful in the study of the metastable behaviour of the system of Subsection I.6.4 via the Wasserstein framework. This is left for future research.

I.8 Future research

We have several aims for future research that are related to the results and the concepts of this thesis. In this section we briefly list some of them.

- The first aim is to extend the results of Chapter III to non-reversible settings. In order to find a gradient flow representation in such settings, the GENERIC framework introduced in [56] might be helpful.
- The main object of investigation for future research is to prove the sharp asymptotics of the transition time introduced in Subsection I.6.4. We have three ideas that might work out.

(i) As we motivated in Subsection I.6.4, extending the potential-theoretic approach to metastability to the Wasserstein space would be the canonical way to solve the problem. This approach relies on the results obtained in [51] and [128].

(ii) The proof of the lower bound of the capacity in the setting of Subsection I.6.3 is inspired by the so-called *two-scale approach* developed in [80]. This approach is restricted to Euclidean spaces. Hence, it is not applicable if the macroscopic order parameter is chosen to be the empirical distribution. It would be interesting to extend (or to find the analogue of) the two-scale approach in the case when the macroscopic order parameter does not map into an Euclidean space.

(iii) We already mentioned at the end of Section I.2 that there is an approach to study metastability based on gradient flow representations and the Sandier-Serfaty approach. This approach was used in [5], [81], [114] and [117]. It is interesting to see whether this approach is applicable in the setting of Subsection I.6.4. Here, the results of Chapter III and Chapter IV might be useful.

- Another object of investigation for future research is to extend the results of Subsection I.6.3 to the case $J \leq 1$.

Notation

We now list some notation that is used throughout this thesis. In the following let (Y, d) and $(\bar{Y}, \bar{\mathrm{d}})$ be Polish spaces. Note that, at the beginning of the Chapters II, III , IV and V, we introduce some notational conventions that are specific to the respective chapter.

- As it is usual in the literature, $C(Y)$ denotes the set of all continuous functions $f : Y \to \mathbb{R}$, and $C_b(Y)$ denotes the set of all continuous and bounded functions $f : Y \to \mathbb{R}$. Moreover, for a measure μ on Y and for $k \in \mathbb{N}$, $\mathrm{L}^k(\mu)$ denotes space of all measurable functions $f : Y \to \mathbb{R}$ such that $\int_Y |f|^k \, d\mu$ is finite.

- $\mathcal{M}_1(Y)$ denotes the space of all Borel probability measures on Y. We equip $\mathcal{M}_1(Y)$ with the topology of weak convergence, where we say that $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}_1(Y)$ *converges weakly in* $\mathcal{M}_1(Y)$ to $\mu \in \mathcal{M}_1(Y)$ (and write $\mu_n \rightharpoonup \mu$) if

$$\int_Y f d\mu_n \to \int_Y f d\mu \quad \text{for all } f \in C_b(Y). \tag{I.8.1}$$

  To emphasize the particular metric on Y, we sometimes say that $(\mu_n)_n$ *converges weakly in* $\mathcal{M}_1((Y, \mathrm{d}))$ to μ.

- For $\mu \in \mathcal{M}_1(Y)$ and a Borel map $f : Y \to \bar{Y}$, we denote by $f_{\#}\mu$ the image measure of μ by f.

Chapter II

Metastability in three modifications of the standard Ising model

The results of the present chapter have already been published as the paper [11].

Recall Section I.4, where we provide a motivation and a first formulation of the main results of this chapter. This chapter is organized as follows. In Section II.1 we introduce the setting and the results from [29, Chapter 16][1]. In particular, we state the so-called *metastability theorems*, which are the key results on which we rely in this chapter. Then, in the Sections II.2–II.4, we consider the three modifications of the standard Ising model, respectively, and provide the proofs of the results that we stated in Section I.4.

II.1 The abstract set-up and the metastability theorems

As in Section I.4, let $\Lambda \subset \mathbb{Z}^2$ be a finite torus centred at the origin and $S = \{-1, 1\}^\Lambda$ be the *configuration space* An element $\sigma \in S$ is called *configuration*, and at each *site* $x \in \Lambda$, $\sigma(x) \in \{-1, 1\}$ is called the *spin-value* at x. By abuse of notation, we often identify each configuration $\sigma \in S$ with the sites that have spin value $+1$, i.e.

$$\sigma \equiv \{x \in \Lambda \mid \sigma(x) = +1\}. \tag{II.1.1}$$

Moreover, we represent σ geometrically by identifying each $x \in \sigma$ with $\sigma(x) = +1$ with a closed unit square centered at x. See Figure II.1 for an example.

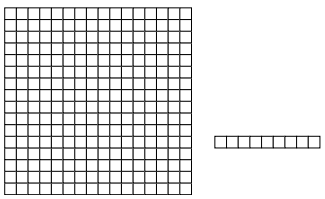


Figure II.1: Geometric representation of a configuration that assigns to each site in Λ the spin-value -1 except on a square of size 16×16 and a rectangle of size 1×8.

[1]The setting in [29] is more general. In order to keep the presentation here as simple as possible, we restrict to a dynamical spin-flip model on the two-dimensional lattice.

The *energy* or *Hamiltonian* of the system is given by $\mathrm{H} : S \to \mathbb{R}$. Let $\beta > 0$ be the *inverse temperature*. The *Gibbs measure* associated to H and β is given by

$$\mu_\beta(\sigma) = \frac{1}{Z_\beta} \mathrm{e}^{-\beta \mathrm{H}(\sigma)}, \qquad \text{for } \sigma \in S, \tag{II.1.2}$$

where Z_β is a normalization constant called *partition function*.

For $\sigma \in S$ and $x \in \Lambda$ we define $\sigma^x \in S$ by

$$\sigma^x(y) = \begin{cases} \sigma(y) & : y \neq x, \\ -\sigma(x) & : y = x. \end{cases} \tag{II.1.3}$$

For all $\sigma, \sigma' \in S$, we say that σ and σ' *communicate* and write $\sigma \sim \sigma'$ if there exists $x \in \Lambda$ such that $\sigma^x = \sigma'$. This induces a graph structure on S by defining an edge between each $\sigma, \sigma' \in S$ whenever $\sigma \sim \sigma'$.

The dynamics of the system is given by the continuous time Markov Chain $(\sigma_t)_{t \geq 0}$ on S, whose generator $\mathcal{L}_\beta$ is given by

$$(\mathcal{L}_\beta f)(\sigma) = \sum_{x \in \Lambda} c_\beta(\sigma, \sigma^x)(f(\sigma^x) - f(\sigma)), \tag{II.1.4}$$

where $f : S \to \mathbb{R}$ is a function and

$$c_\beta(\sigma, \sigma') = \begin{cases} \mathrm{e}^{-\beta \max\{0, \mathrm{H}(\sigma') - \mathrm{H}(\sigma)\}} & : \sigma \sim \sigma', \\ 0 & : \text{else.} \end{cases} \tag{II.1.5}$$

It is easy to see that the *detailed balance condition*,

$$\mu_\beta(\sigma) c_\beta(\sigma, \sigma') = \mu_\beta(\sigma') c_\beta(\sigma', \sigma) \qquad \forall\, \sigma, \sigma' \in \mathcal{X}_\beta^{(n_\beta)}, \tag{II.1.6}$$

holds. Hence, the dynamics is *reversible* with respect to the Gibbs measure. The law of $(\sigma_t)_{t \geq 0}$ given that $\sigma_0 = \sigma \in S$ will be denoted by $\mathbb{P}_\sigma$, and for any set $A \subset S$, let

$$\tau_{\mathrm{A}} = \inf\{t > 0 |\ \sigma_t \in A, \exists\, 0 < s < t : \sigma_s \neq \sigma_0\}. \tag{II.1.7}$$

If $A = \{\sigma\}$ for some $\sigma \in S$, then we write $\tau_\sigma = \tau_{\mathrm{A}}$.

Definition II.1 *i) Let $\sigma, \sigma' \in S$. The* communication height *between σ and σ' is defined by*

$$\Phi(\sigma, \sigma') = \min_{\gamma : \sigma \to \sigma'} \max_{\eta \in \gamma} \mathrm{H}(\eta), \tag{II.1.8}$$

where the minimum is taken over all finite paths γ of allowed moves in S going from σ to σ'.

ii) Let $\sigma, \sigma' \in S$. A finite path $\gamma : \sigma \to \sigma'$ is called optimal path *between σ and σ' if*

$$\Phi(\sigma, \sigma') = \max_{\eta \in \gamma} \mathrm{H}(\eta). \tag{II.1.9}$$

The set of all optimal paths between σ and σ' is denoted by $(\sigma \to \sigma')_{\mathrm{opt}}$.

iii) Let $\sigma \in S$. The stability level *of σ is defined by*

$$V_\sigma = \min_{\eta \in S : \mathrm{H}(\eta) < \mathrm{H}(\sigma)} \Phi(\sigma, \eta) - \mathrm{H}(\sigma). \tag{II.1.10}$$

Moreover, for $V \in \mathbb{R}$, we define

$$S_V = \{\sigma \in S \,|\, V_\sigma > V\}, \tag{II.1.11}$$

which is the set of all configurations, whose stability level is greater than V.

iv) The set of stable configurations *in S is defined by:*

$$S_{\mathrm{stab}} = \{\sigma \in S \mid \mathrm{H}(\sigma) = \min_{\eta \in S} \mathrm{H}(\eta)\}. \tag{II.1.12}$$

v) The set of metastable configurations *in S is defined by:*

$$S_{\mathrm{meta}} = \{\sigma \in S \mid V_\sigma = \max_{\eta \in S \setminus S_{\mathrm{stab}}} V_\eta \}. \tag{II.1.13}$$

vi) Let $(m, s) \in S_{\mathrm{meta}} \times S_{\mathrm{stab}}$. The energy barrier $\Gamma^\star(m, s)$ *between m and s is defined by*

$$\Gamma^\star(m, s) = \Phi(m, s) - \mathrm{H}(m). \tag{II.1.14}$$

Note that by [40, Theorem 2.4], we have that

$$\Gamma^\star(m, s) = \max_{\eta \in S \setminus S_{\mathrm{stab}}} V_\eta \quad \text{for all } (m, s) \in S_{\mathrm{meta}} \times S_{\mathrm{stab}}. \tag{II.1.15}$$

In the following definition we introduce the notion of a *critical configuration*. This resembles the idea of the critical state from Section I.1.

Definition II.2 (Definition 16.3 in [29])
Let $(m, s) \in S_{\mathrm{meta}} \times S_{\mathrm{stab}}$. Then $(\mathcal{P}^\star(m, s), \mathcal{C}^\star(m, s))$ is defined as the maximal subset of $S \times S$ such that

1.) $\forall \sigma \in \mathcal{P}^\star(m, s)\ \exists \sigma' \in \mathcal{C}^\star(m, s)\ :\ \sigma \sim \sigma'$, *and*
$\forall \sigma' \in \mathcal{C}^\star(m, s)\ \exists \sigma \in \mathcal{P}^\star(m, s)\ :\ \sigma \sim \sigma'$,

2.) $\forall \sigma \in \mathcal{P}^\star(m, s)\ :\ \Phi(m, \sigma) < \Phi(\sigma, s)$,

3.) $\forall \sigma' \in \mathcal{C}^\star(m, s)\ \exists \gamma : \sigma' \to s\ :\ \max_{\eta \in \gamma} \mathrm{H}(\eta) - \mathrm{H}(m) \le \Gamma^\star(m, s)\,,\ \Phi(m, \eta) \ge \Phi(\eta, s)\, \forall \eta \in \gamma$.

We call $\mathcal{P}^\star(m, s)$ the set of protocritical configurations *and $\mathcal{C}^\star(m, s)$ the* set of critical configurations.

The results from the Sections II.2–II.4 are based on the following metastability theorems that are taken from [29, Theorem 16.4–16.6]. These hold subject to the hypothesis

$$S_{\mathrm{meta}} = \{m\} \quad \text{and} \quad S_{\mathrm{stab}} = \{s\}, \tag{H1}$$

where $m, s \in S$. One challenge in the Sections II.2–II.4 is to verify this hypothesis for the three specific models. Under (H1), it would not lead to confusions if we abbreviate $\mathcal{P}^\star = \mathcal{P}^\star(m, s)$, $\mathcal{C}^\star = \mathcal{C}^\star(m, s)$ and $\Gamma^\star = \Gamma^\star(m, s)$.

Theorem II.3 (Theorem 16.4 in [29]) *Consider* $(\mathcal{P}^\star, \mathcal{C}^\star)$ *from Definition II.2. Suppose* (H1)*. Then,*

a) $\lim_{\beta\to\infty} \mathbb{P}_m[\tau_{\mathcal{C}^\star} < \tau_s \mid \tau_s < \tau_m] = 1$*, and*

b) if, moreover, the assumption

$$\sigma' \to |\{\sigma \in \mathcal{P}^\star \ : \ \sigma \sim \sigma'\}| \text{ is constant on } \mathcal{C}^\star \tag{H2}$$

holds, then for all $\chi \in \mathcal{C}^\star$, $\lim_{\beta\to\infty} \mathbb{P}_m[\sigma_{\tau_{\mathcal{C}^\star}} = \chi] = \frac{1}{|\mathcal{C}^\star|}$.

Theorem II.3 says that, in order to make the crossover from the metastable to the stable configuration, the system has to pass the set of critical configurations. If, in addition, assumption (H2) holds, then part b) of Theorem II.3 says that the entrance into $\mathcal{C}^\star$ is uniformly distributed on $\mathcal{C}^\star$.

Theorem II.4 (Theorem 16.6 in [29]) *Subject to* (H1)*, it holds that*

a) $\lim_{\beta\to\infty} \lambda_\beta \mathbb{E}_m[\tau_s] = 1$*, where* λ_β *is the second largest eigenvalue of* $-\mathcal{L}_\beta$*, and*

b) $\lim_{\beta\to\infty} \mathbb{P}_m[\tau_s > t \cdot \mathbb{E}_m[\tau_s]] = e^{-t}$ *for all* $t \geq 0$.

Theorem II.4 represents the average transition time of the system in terms of the spectrum of its generator and part b) yields the asymptotic exponential distribution of τ_s.

Theorem II.5 (Theorem 16.5 and Lemma 16.17 in [29]) *Suppose* (H1)*. Then,*

a) there exists a constant $K \in (0, \infty)$ *such that* $\lim_{\beta\to\infty} \mathrm{e}^{-\beta\Gamma^\star} \mathbb{E}_m[\tau_s] = K$*, and*

b) define

- $S^\star \subset S$ *be the subgraph obtained by removing all vertices* η *with* $H(\eta) > \Gamma^\star + H(m)$ *and all edges incident to these vertices,*
- $S^{\star\star} \subset S^\star$ *be the subgraph obtained by removing all vertices* η *with* $H(\eta) = \Gamma^\star + H(m)$ *and all edges incident to these vertices,*
- $S_m = \{\eta \in S \mid \Phi(m, \eta) < \Phi(\eta, s) = \Gamma^\star + H(m)\}$,
- $S_s = \{\eta \in S \mid \Phi(\eta, s) < \Phi(m, \eta) = \Gamma^\star + H(m)\}$,
- $S_1, \ldots, S_I \subset S^{\star\star}$ *be such that* $S^{\star\star} \setminus (S_m \cup S_s) = \cup_{i=1}^I S_i$ *and each* S_i *is a maximal set of communicating configurations,*

then,

$$\frac{1}{K} = \min_{C_1,\ldots,C_I \in [0,1]} \min_{\substack{h:S^\star \to [0,1] \\ h|_{S_m}=1, h|_{S_s}=0, h|_{S_i}=C_i \, \forall i}} \frac{1}{2} \sum_{\eta,\eta' \in S^\star} \mathbb{1}_{\{\eta\sim\eta'\}} [h(\eta) - h(\eta')]^2. \tag{II.1.16}$$

Note that the first minimum runs over all constants $C_1, \ldots, C_I \in [0, 1]$.

Theorem II.5 yields the precise asymptotics of the average transition time and provides a variational formula to compute the pre-factor.

The first goal in this chapter is to verify Theorem II.3, II.4 and II.5 for the three specific models we introduced in Section I.4 in the introduction. That is, we have to show that the conditions of those theorems are satisfied. In all these three models we have that

$$m = \boxminus, \qquad \text{and} \qquad s = \boxplus, \tag{II.1.17}$$

where $\boxminus \in S$ is the configuration where all spin values are equal to -1 and $\boxplus$ is the configuration with all spin values being $+1$. The second goal in this chapter is to compute the precise value of the pre-factor K from Theorem II.5 for each of these three models.

Hence, for each model we have to

- compute $\Gamma^\star = \Phi(\boxminus, \boxplus) - \mathrm{H}(\boxminus)$,
- identify the sets $\mathcal{P}^\star$ and $\mathcal{C}^\star$,
- verify hypothesis (H1) and if possible hypothesis (H2), and
- compute K.

These tasks are treated in the Sections II.2–II.4 for the respective models.

We conclude this section with some definitions hat are used throughout this chapter.

Further definitions

- For $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer greater than x.
- For $l_1, l_2 \in \mathbb{N}$, $R(l_1 \times l_2)$ denotes the set of all configurations consisting of a single rectangle with horizontal length l_1 and vertical length l_2 somewhere on the torus Λ. An element $\sigma \in R(l_1 \times l_2)$ is called *rectangle* and will often be denoted by $l_1 \times l_2$, since usually we can ignore the position of the rectangle in the torus. For this reason, by abuse of notation, we often identify the whole set $R(l_1 \times l_2)$ with $l_1 \times l_2$. We also define $R(l_1, l_2) = R(l_1 \times l_2) \cup R(l_2 \times l_1)$. If $|l_1 - l_2| = 1$ or $|l_1 - l_2| = 0$, then $l_1 \times l_2$ is called *quasi-square* or *square*, respectively. $1 \times l_2$ is called *vertical bar* or *column* and $l_1 \times 1$ is called *horizontal bar* or *row*.
- For a rectangle $R \in S$, we denote by $P_H R \in \mathbb{N}$ its *horizontal length*, and by $P_V R \in \mathbb{N}$ its *vertical length*.
- For $\sigma \in S$, let $|\sigma|$ be the *area* of σ, i.e. its number of $(+1)$–spins. Further, $\partial(\sigma)$ is the Euclidean boundary of σ in its geometric representation and $|\partial(\sigma)|$ denotes the *perimeter*, i.e. the length of $\partial(\sigma)$.
- Let $\sigma \in S$. We say that σ is *connected* if $\sigma \setminus \partial(\sigma)$ is connected in the Euclidean space $\mathbb{R}^2$.
- Let $\sigma \in S$. A *cluster* of σ is a maximally connected component of σ.
- Two droplets on the torus are called *isolated* if their Euclidean distance is greater or equal to $\sqrt{2}$.
- Let $\sigma \in S$ and $x \in \Lambda$ be such that $\sigma(x) = +1$. Then x is called *protuberance* if $\sum_{y \in \Lambda : |y-x|=1} \sigma(y) = -2$.

- Let $\sigma \in S$ be connected and l be either a vertical bar or a horizontal bar. Then l is called *attached* to σ if for all $x \in l$ there exists $y \in \Lambda \setminus l$ such that $|y - x| = 1$ and $z \in \sigma$ such that $|z - x| = 1$.

- If $\sigma \in S$ consists of a single, connected droplet, then $R(\sigma)$ is the smallest rectangle that contains σ.

- A *row* or a *column* of a connected configuration $\sigma \in S$ is defined as the intersection of a row or a column of Λ with σ.

- $\sigma \in S$ is called a *local minimum* of H if $\mathrm{H}(\sigma^x) > \mathrm{H}(\sigma)$ for all $x \in \Lambda$.

- For $A \subset S$, let $\partial^+ A = \{\sigma \in S \setminus A \,|\, \exists \sigma' \in S : \sigma \sim \sigma'\}$ denote the *outer boundary* A. We also define $A^+ = A \cup \partial^+ A$. Moreover, if $\eta \in S$, then $A \sim \eta \subset S$ is defined by $A \sim \eta = \{\sigma \in A \,|\, \sigma \sim \eta\}$.

II.2 Anisotropic Ising model

Recall the setting from Subsection I.4.2 and that the Hamiltonian for the anisotropic Ising model is given by

$$\mathrm{H_A}(\sigma) = -\frac{J_H}{2} \sum_{(x,y)\in\Lambda_H^\star} \sigma(x)\sigma(y) - \frac{J_V}{2} \sum_{(x,y)\in\Lambda_V^\star} \sigma(x)\sigma(y) - \frac{h}{2} \sum_{x\in\Lambda} \sigma(x), \tag{II.2.1}$$

where $\sigma \in S$, $J_H, J_V, h > 0$, $\Lambda_H^\star$ is the set of *unordered horizontal nearest-neighbour bonds* in Λ and $\Lambda_V^\star$ is the set of *unordered vertical nearest-neighbour bonds* in Λ.

Using the geometric representation of σ, one can rewrite $\mathrm{H_A}(\sigma)$ as

$$\mathrm{H_A}(\sigma) = \mathrm{H_A}(\boxminus) - h|\sigma| + J_H|\partial_V(\sigma)| + J_V|\partial_H(\sigma)|, \tag{II.2.2}$$

where $|\partial_V(\sigma)|$ is the length of the vertical part of $\partial(\sigma)$ and $|\partial_H(\sigma)|$ is the length of the horizontal part of $\partial(\sigma)$. In the example in Figure II.1 we have that $|\partial_V(\sigma)| = 34$ and $|\partial_H(\sigma)| = 40$.

Recall that the critical length in this model is given by

$$L_V^\star = \left\lceil \frac{2J_V}{h} \right\rceil. \tag{II.2.3}$$

We make the following assumptions in this section.

Assumption II.6 *a)* $J_H > J_V$,

b) $2J_V > h$,

c) $\frac{2J_V}{h} \notin \mathbb{N}$,

d) $|\Lambda| > \left(\max\{\frac{2J_H}{hL_V^\star - 2J_V}, \frac{2J_H(L_V^\star - 1)}{2J_V - h(L_V^\star - 1)} + L_V^\star\}\right)^2$.

By symmetry, Assumption II.6 a) could be chosen the other way around. Assumption II.6 b) implies that the dynamics prefers aligned neighbouring spins to (+1)–spins. This is essential to obtain the metastable behavior of the system. Indeed, if $2J_V \leq h$, then $L_V^\star = 1$ and therefore, each configuration with a single (+1)–spin somewhere in Λ is a critical configuration of the system. It follows from Assumption II.6 c) that

$$(L_V^\star - 1)h < 2J_V < L_V^\star h. \tag{II.2.4}$$

In Subsection II.2.2 and Subsection II.2.4 the importance of (II.2.4) will become clear. Assumption II.6 d) is made to avoid certain degenerate situations. For instance, if $|\Lambda|$ is small enough, all optimal paths between $\boxminus$ and $\boxplus$ contain a configuration, which consists of a single rectangle, where one side wraps around the torus and the other side is of length strictly smaller than $L_V^\star - 1$. For more details see (II.2.10) or the proof of Lemma II.11. Moreover, d) ensures that the torus is large enough to contain at least a critical droplet.

Recall the definition of $R(l_1, l_2)$ from the end of Section II.1. Before stating the main result of this section, we need the following definition.

Definition II.7 *We denote by $R(L_V^\star - 1, L_V^\star)^{\mathrm{1pr}}$ the set of all configurations consisting only of a rectangle from $R(L_V^\star - 1, L_V^\star)$ and with an additional protuberance attached to one of its longer sides. The right droplet in Figure II.2 provides an example.*

Moreover, we denote by $R(L_V^\star - 1, L_V^\star)^{\mathrm{2pr}}$ the set of all configurations that are obtained from a configuration in $R(L_V^\star - 1, L_V^\star)^{\mathrm{1pr}}$ by adding a second (+1)–spin, which is attached to the rectangle and adjacent to the protuberance.

We now formulate the main result of this section.

Theorem II.8 *Under Assumption II.6, the pair $(\boxminus, \boxplus)$ satisfies* (H1) *and* (H2) *so that Theorems II.3–II.5 hold for the anisotropic Ising model. Moreover,*

- $\mathcal{P}^\star = R(L_V^\star - 1, L_V^\star)$,
- $\mathcal{C}^\star = R(L_V^\star - 1, L_V^\star)^{\mathrm{1pr}}$,
- $\Phi(\boxminus, \boxplus) - \mathrm{H}_\Lambda(\boxminus) = 2L_V^\star(J_H + J_V) - h(1 + (L_V^\star - 1)L_V^\star) =: \Gamma_\Lambda^\star =: \mathrm{E}_\Lambda^\star - \mathrm{H}_\Lambda(\boxminus)$,
- $K^{-1} = \frac{4(2L_V^\star - 1)}{3}|\Lambda|$.

Proof. The proof is divided into the Subsections II.2.1–II.2.6. □

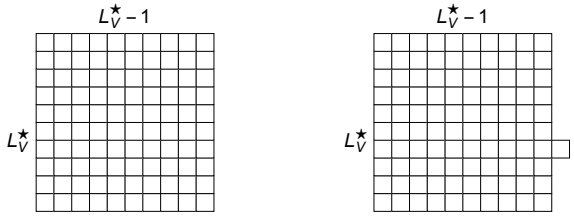


Figure II.2: The left object is an element in $\mathcal{P}^\star$ and the right object is an element in $\mathcal{C}^\star$.

II.2.1 Proof of $\Phi(\boxminus, \boxplus) - \mathrm{H_A}(\boxminus) \leq \Gamma^\star_{\mathrm{A}}$

It will be enough to construct a path $\gamma_{\mathrm{A}} = (\gamma_{\mathrm{A}}(n))_{n\geq 0} : \boxminus \to \boxplus$ such that

$$\max_{\eta\in\gamma_{\mathrm{A}}} \mathrm{H_A}(\eta) \leq \mathrm{H_A}(\boxminus) + \Gamma^\star_{\mathrm{A}} = \mathrm{E}^\star_{\mathrm{A}}. \tag{II.2.5}$$

This path will be called *reference path*.

Construction of γ_{A}. Let $\gamma_{\mathrm{A}}(0) = \boxminus$. In the first step an arbitrary (-1)–spin is flipped. Then γ_{A} first passes through a sequence of squares and quasi-squares as follows. If at some step i, $\gamma_{\mathrm{A}}(i)$ is a square, then a protuberance is added above the droplet. Afterwards, this row is filled by successively flipping in this row adjacent (-1)–spins until the droplet has the shape of a quasi-square. Next, a protuberance is added on the right of the droplet. Similarly as before, successively, adjacent (-1)–spins are flipped in this column until the droplet has the shape of a square again. This procedure is stopped, when $R((L^\star_V - 1) \times L^\star_V)$ is reached.

Now a protuberance is added on the right of the droplet and this column is filled until $R((L^\star_V - 1) \times (L^\star_V + 1))$ is reached. This adding structure is repeated until the droplet winds around the torus. Next, a protuberance is added above the droplet and the corresponding row is filled until this row also winds around the torus. This is repeated until $\boxplus$ is reached.

Inequality (II.2.5) *holds.* Let $k^\star$ be such that $\gamma_{\mathrm{A}}(k^\star) \in R((L^\star_V - 1) \times L^\star_V)$. Then $\mathrm{H_A}(\gamma_{\mathrm{A}}(k^\star)) = \mathrm{E}^\star_{\mathrm{A}} - 2J_V + h < \mathrm{E}^\star_{\mathrm{A}}$. If we go backwards in the path from that point on, then we will have to cut the top row of $R((L^\star_V - 1) \times L^\star_V)$, which has the length $L^\star_V - 1$. This is an increase of the energy in each step by h for $(L^\star_V - 2)$ times until the top row turns into a protuberance. At this point the energy equals to

$$\mathrm{H_A}(\gamma_{\mathrm{A}}(k^\star - (L^\star_V - 2))) = \mathrm{E}^\star_{\mathrm{A}} - 2J_V + (L^\star_V - 1)h < \mathrm{E}^\star_{\mathrm{A}} \tag{II.2.6}$$

by (II.2.4). Cutting the last protuberance decreases the energy by $2J_H - h$. By the same arguments, if we keep on going backwards in the path of γ_{A}, we will always stay below $\mathrm{E}^\star_{\mathrm{A}}$, since the size of the above and right bars of the droplets will be at most $L^\star_V - 1$. Hence, we get that

$$\max_{i=1,\dots,k^\star} \mathrm{H_A}(\gamma_{\mathrm{A}}(i)) < \mathrm{E}^\star_{\mathrm{A}}. \tag{II.2.7}$$

We now consider the remaining path of γ_{A} after the step $k^\star + 2$. It holds that $\mathrm{H_A}(\gamma_{\mathrm{A}}(k^\star + 2)) = \mathrm{E}^\star_{\mathrm{A}} - h < \mathrm{E}^\star_{\mathrm{A}}$. While filling the right column, the energy decreases by h at every step. After the right column is filled, a protuberance is added on the right side and the energy increases by $2J_V - h$. Again by (II.2.4), we get that

$$\mathrm{H_A}(\gamma_{\mathrm{A}}(k^\star + (L^\star_V + 1))) = \mathrm{E}^\star_{\mathrm{A}} + 2J_V - L^\star_V h < \mathrm{E}^\star_{\mathrm{A}}. \tag{II.2.8}$$

Repeating this until the droplet wraps around the torus, the following energy level is reached

$$\mathrm{E}^\star_{\mathrm{A}} - (hL^\star_V - 2J_V)(\sqrt{|\Lambda|} - L^\star_V) - h(L^\star_V - 1) - 2J_V L^\star_V. \tag{II.2.9}$$

Now we add a protuberance above the droplet and the energy increases by $2J_H - h$. Assumption II.6 d) and (II.2.4) imply that

$$\begin{aligned} &\mathrm{E}^\star_{\mathrm{A}} - (hL^\star_V - 2J_V)(\sqrt{|\Lambda|} - L^\star_V) - h(L^\star_V - 1) - 2J_V L^\star_V + 2J_H - h \\ &\quad \leq \mathrm{E}^\star_{\mathrm{A}} + (hL^\star_V - 2J_V)L^\star_V - hL^\star_V - 2J_V L^\star_V \\ &\quad < \mathrm{E}^\star_{\mathrm{A}} - 2J_V L^\star_V. \end{aligned} \tag{II.2.10}$$

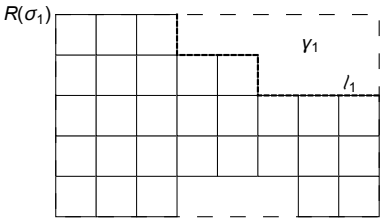
Filling this row, decreases the energy by $(\sqrt{|\Lambda|}-1)h+2J_V$. In the same way, one can show that the remaining part of the path stays below $\mathrm{E}_{\mathrm{A}}^{\star}$. Combining this with (II.2.7) and the fact that $\mathrm{H}_{\mathrm{A}}(\gamma_{\mathrm{A}}(k^{\star}+1))=\mathrm{E}_{\mathrm{A}}^{\star}$, we infer (II.2.5).

II.2.2 Proof of $\Phi(\boxminus,\boxplus)-\mathrm{H}_{\mathrm{A}}(\boxminus)\geq\Gamma_{\mathrm{A}}^{\star}$

It suffices to show that every optimal path from $\boxminus$ to $\boxplus$ has to pass through $R(L_V^{\star}-1,L_V^{\star})^{1\mathrm{pr}}$. We first list a few observations. Recall the definition of *local minimum* from the end of Section II.1.

Lemma II.9 *Let $\sigma\in S$ be a local minimum of H_{A}. Then σ is a union of isolated rectangles.*

Proof. Suppose that σ has a connected component σ_1 that is not a rectangle. Consider a connected component γ_1 of $R(\sigma_1)\cap(\mathbb{Z}^2\setminus\sigma_1)$. Let l_1 be the maximal component of the boundary of γ_1 that does not belong to the boundary of $R(\sigma_1)$. An example would be:



Then, since σ_1 is connected and l_1 lies inside $R(\sigma_1)$, l_1 has both a horizontal part and a vertical part. Let $x\in\gamma_1$ be a site, whose boundary intersects both a horizontal part and a vertical part of l_1. In particular, $\sigma(x)=-1$ and x has at least two nearest-neighbour $(+1)$–spins. It is easy to see that σ^x has strictly lower energy than σ. □

Corollary II.10 *Assume that $\sigma\in S$ consists of a unique cluster. Then,*

$$\mathrm{H}_{\mathrm{A}}(\sigma)\geq\mathrm{H}_{\mathrm{A}}(R(\sigma)), \tag{II.2.11}$$

and equality holds if and only if $\sigma=R(\sigma)$.

We first show that every optimal path has to cross $R(L_V^{\star}-1,L_V^{\star})$.

Lemma II.11 *Let $\gamma\in(\boxminus,\boxplus)_{\mathrm{opt}}$. Then γ has to cross $R(L_V^{\star}-1,L_V^{\star})$.*

Proof. Assume the contrary, i.e. $\gamma\cap R(L_V^{\star}-1,L_V^{\star})=\emptyset$. Let us first assume that throughout its whole path γ consists of a unique cluster. On its way to $\boxplus$, γ has to cross a configuration, whose rectangular envelope has both horizontal and vertical length greater or equal to $L_V^{\star}$. Let

$$\bar{t}=\min\{l\geq 0\,|\,P_H R(\gamma(l)),P_V R(\gamma(l))\geq L_V^{\star}\}. \tag{II.2.12}$$

Since γ is assumed to consist of a unique cluster, we have that either $P_H R(\gamma(\bar{t}-1))=L_V^{\star}-1$ holds or $P_V R(\gamma(\bar{t}-1))=L_V^{\star}-1$. In the following we analyze both cases and show that the assumption $\gamma\cap R(L_V^{\star}-1,L_V^{\star})=\emptyset$ leads to a contradiction.

Case 1. $[P_V R(\gamma(\bar{t}-1)) = L_V^\star - 1]$.
From the definition of $\bar{t}$, it is clear that $R(\gamma(\bar{t}-1)) \in R((L_V^\star + m) \times (L_V^\star - 1))$ for some $m \geq 0$.

Case 1.1. $[m = 0]$.
By hypothesis, γ does not cross $R(L_V^\star \times (L_V^\star - 1))$. Hence, Corollary II.10 yields that

$$\mathrm{H_A}(\gamma(\bar{t}-1)) > \mathrm{H_A}(L_V^\star \times (L_V^\star - 1)) = \mathrm{H_A}(\boxminus) + \Gamma_{\mathrm{A}}^\star - 2J_H + h = \mathrm{E}_{\mathrm{A}}^\star - 2J_H + h. \quad \text{(II.2.13)}$$

The minimal increase of energy to enlarge the vertical length of the rectangular envelope of a configuration is $2J_H - h$. Hence,

$$\mathrm{H_A}(\gamma(\bar{t})) \geq \mathrm{H_A}(\gamma(\bar{t}-1)) + 2J_H - h > \mathrm{E}_{\mathrm{A}}^\star. \quad \text{(II.2.14)}$$

This contradicts $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$, since we already know from Subsection II.2.1 that $\Phi(\boxminus, \boxplus) \leq E_{\mathrm{A}}^\star$.

Case 1.2. $[m \in [1, \sqrt{|\Lambda|} - L_V^\star)]$.
Again, by Corollary II.10 we have that

$$\begin{aligned} \mathrm{H_A}(\gamma(\bar{t}-1)) &\geq \mathrm{H_A}((L_V^\star + m) \times (L_V^\star - 1)) \\ &= \mathrm{H_A}(L_V^\star \times (L_V^\star - 1)) + m(2J_V - h(L_V^\star - 1)) \\ &> \mathrm{E}_{\mathrm{A}}^\star - 2J_H + h, \end{aligned} \quad \text{(II.2.15)}$$

where we used inequality (II.2.4) in the last step. As before, this leads to a contradiction, since

$$\mathrm{H_A}(\gamma(\bar{t})) \geq \mathrm{H_A}(\gamma(\bar{t}-1)) + 2J_H - h > \mathrm{E}_{\mathrm{A}}^\star. \quad \text{(II.2.16)}$$

Case 1.3. $[m = \sqrt{|\Lambda|} - L_V^\star]$.
In this case, $\gamma(\bar{t}-1)$ wraps around the torus. Using Assumption II.6 d), we infer that

$$\begin{aligned} &\mathrm{H_A}(\gamma(\bar{t}-1)) \geq \mathrm{H_A}(\sqrt{|\Lambda|} \times (L_V^\star - 1)) \\ &\quad = \mathrm{H_A}(L_V^\star \times (L_V^\star - 1)) + (\sqrt{|\Lambda|} - L_V^\star)(2J_V - h(L_V^\star - 1)) - 2J_V(L_V^\star - 1) \\ &\quad > \mathrm{H_A}(L_V^\star \times (L_V^\star - 1)) = \mathrm{E}_{\mathrm{A}}^\star - 2J_H + h. \end{aligned} \quad \text{(II.2.17)}$$

Finally,

$$\mathrm{H_A}(\gamma(\bar{t})) \geq \mathrm{H_A}(\gamma(\bar{t}-1)) + 2J_H - h > \mathrm{E}_{\mathrm{A}}^\star, \quad \text{(II.2.18)}$$

which is a contradiction.

Case 2. $[P_H R(\gamma(\bar{t}-1)) = L_V^\star - 1]$.
Here we have that $R(\gamma(\bar{t}-1)) \in R((L_V^\star - 1) \times (L_V^\star + m'))$ for some $m' \geq 0$.

Case 2.1. $[m' = 0]$.
Since γ does not cross $R((L_V^\star - 1) \times L_V^\star)$, we have by Corollary II.10 that

$$\mathrm{H_A}(\gamma(\bar{t}-1)) > \mathrm{H_A}((L_V^\star - 1) \times L_V^\star) = \mathrm{E}_{\mathrm{A}}^\star - 2J_V + h. \quad \text{(II.2.19)}$$

The minimal increase of energy to enlarge the horizontal length of the rectangular envelope of a configuration is $2J_V - h$. Hence,

$$\mathrm{H_A}(\gamma(\bar{t})) \geq \mathrm{H_A}(\gamma(\bar{t}-1)) + 2J_V - h > \mathrm{E}_{\mathrm{A}}^\star. \quad \text{(II.2.20)}$$

As before, this contradicts $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$.

Case 2.2. $[m' \in [1, \sqrt{|\Lambda|} - L_V^\star)]$.
This case also leads to a contradiction, since

$$\begin{aligned} \mathrm{H_A}(\gamma(\bar{t})) &\geq \mathrm{H_A}(\gamma(\bar{t}-1)) + 2J_V - h \geq \mathrm{H_A}((L_V^\star - 1) \times (L_V^\star + m')) + 2J_V - h \\ &= \mathrm{H_A}((L_V^\star - 1) \times L_V^\star) + m'(2J_H - h(L_V^\star - 1)) + 2J_V - h \\ &> \mathrm{E_A^\star}, \end{aligned} \tag{II.2.21}$$

where we have used inequality (II.2.4) and Assumption II.6 a) in the last step.

Case 2.3. $[m' = \sqrt{|\Lambda|} - L_V^\star]$.
Using Assumption II.6 d), we infer that

$$\begin{aligned} &\mathrm{H_A}(\gamma(\bar{t}-1)) \geq \mathrm{H_A}((L_V^\star - 1) \times \sqrt{|\Lambda|}) \\ &\quad = \mathrm{H_A}((L_V^\star - 1) \times L_V^\star) + (\sqrt{|\Lambda|} - L_V^\star)(2J_H - h(L_V^\star - 1)) - 2J_H(L_V^\star - 1) \\ &\quad > \mathrm{H_A}((L_V^\star - 1) \times L_V^\star) = \mathrm{E_A^\star} - 2J_V + h. \end{aligned} \tag{II.2.22}$$

Finally,

$$\mathrm{H_A}(\gamma(\bar{t})) \geq \mathrm{H_A}(\gamma(\bar{t}-1)) + 2J_V - h > \mathrm{E_A^\star}, \tag{II.2.23}$$

which is a contradiction.

Now suppose that γ can consist of several clusters, i.e. at each step $j \in \mathbb{N}$, $\gamma(j)$ consists of $n_j \in \mathbb{N}$ clusters, which are denoted by $\gamma^1(j), \ldots, \gamma^{n_j}(j)$. The proof follows from similar arguments as in the first part of the proof of this lemma. Thus, we only provide the main arguments and omit the details.

Using formula (II.2.2) and Corollary II.10, we infer that for all $j \in \mathbb{N}$,

$$\begin{aligned} \mathrm{H_A}(\,\gamma(j)\,) &= \sum_{k=1}^{n_j} \mathrm{H_A}(\,\gamma^k(j)\,) - (n_j - 1)\,\mathrm{H_A}(\boxminus) \\ &\geq \sum_{k=1}^{n_j} \mathrm{H_A}(\,R(\gamma^k(j))\,) - (n_j - 1)\,\mathrm{H_A}(\boxminus). \end{aligned} \tag{II.2.24}$$

For all $j \in \mathbb{N}$ and $k \leq n_j$, set $\ell_V^k(j) = P_V R(\gamma^k(j))$ and $\ell_H^k(j) = P_H R(\gamma^k(j))$. Then,

$$\begin{aligned} &\sum_{k=1}^{n_j} \mathrm{H_A}(\,R(\gamma^k(j))\,) - (n_j - 1)\,\mathrm{H_A}(\boxminus) \\ &\qquad = \mathrm{H_A}(\boxminus) + 2J_H \sum_{k=1}^{n_j} \ell_V^k(j) + 2J_V \sum_{k=1}^{n_j} \ell_H^k(j) - h \sum_{k=1}^{n_j} \ell_V^k(j)\ell_H^k(j). \end{aligned} \tag{II.2.25}$$

Set $\ell_V(j) = \sum_{k=1}^{n_j} \ell_V^k(j)$ and $\ell_H(j) = \sum_{k=1}^{n_j} \ell_H^k(j)$ and define

$$\tilde{t} = \min\{j \in \mathbb{N} \mid \ell_H(j), \ell_V(j) \geq L_V^\star\}. \tag{II.2.26}$$

We have that either $\ell_V(\tilde{t}-1) = L_V^\star - 1$ holds or $\ell_H(\tilde{t}-1) = L_V^\star - 1$. We only treat the case when $\ell_H(\tilde{t}-1) = L_V^\star - 1$, since the other case is a straightforward combination of Case 1 above and the following arguments.

By the definition of $\tilde{t}$, we have that $\ell_V(\tilde{t}-1) = L_V^\star + \tilde{m}$ for some $\tilde{m} \geq -1$. (If $n_{\tilde{t}-1} \geq n_{\tilde{t}}$, then $\tilde{m} \geq 0$, and if $n_{\tilde{t}-1} < n_{\tilde{t}}$, then $\tilde{m} \geq -1$). From (II.2.24) and (II.2.25), we infer that

$$\begin{aligned} \mathrm{H_A}(\,\gamma(\tilde{t}-1)\,) &\geq \mathrm{H_A}(\boxminus) + 2J_H\,(L_V^\star + \tilde{m}) + 2J_V(L_V^\star - 1) - h\sum_{k=1}^{n_{\tilde{t}-1}} \ell_V^k(\tilde{t}-1)\ell_H^k(\tilde{t}-1), \text{ and} \\ \mathrm{H_A}(\,\gamma(\tilde{t})\,) &\geq \mathrm{H_A}(\,\gamma(\tilde{t}-1)\,) + 2J_V - h \qquad \text{(II.2.27)} \\ &\geq \mathrm{H_A}(\boxminus) + 2J_H\,(L_V^\star + \tilde{m}) + 2J_V L_V^\star - h\left(\sum_{k=1}^{n_{\tilde{t}-1}} \ell_V^k(\tilde{t}-1)\ell_H^k(\tilde{t}-1) + 1\right). \end{aligned}$$

Notice the following estimate

$$\sum_{k=1}^{n_{\tilde{t}-1}} \ell_V^k(\tilde{t}-1)\ell_H^k(\tilde{t}-1) \leq \sum_{p=1}^{n_{\tilde{t}-1}} \ell_V^p(\tilde{t}-1)\sum_{k=1}^{n_{\tilde{t}-1}} \ell_H^k(\tilde{t}-1) = (L_V^\star + \tilde{m})(L_V^\star - 1), \qquad \text{(II.2.28)}$$

where the inequality is strict whenever $n_{\tilde{t}-1} > 1$. We now have to show that all possible values for $\tilde{m}$ and the hypothesis that $\gamma \cap R(L_V^\star - 1, L_V^\star) = \emptyset$ yield to the fact that $\mathrm{H_A}(\,\gamma(\tilde{t})\,) > \mathrm{E_A^\star}$, which is a contradiction. However, using (II.2.27) and (II.2.28), we can proceed as in Case 2.1–Case 2.3 above. The details are straightforward adaptations and are therefore omitted. This concludes the proof of this lemma. □

The following lemma concludes the proof of $\Phi(\boxminus, \boxplus) - \mathrm{H_A}(\boxminus) \geq \Gamma_{\mathrm{A}}^\star$.

Lemma II.12 *Let $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$. In order to cross a configuration whose rectangular envelope has both vertical and horizontal length greater or equal to $L_V^\star$, γ has to pass through $R(L_V^\star - 1, L_V^\star)$ and $R(L_V^\star - 1, L_V^\star)^{\mathrm{1pr}}$. In particular, each optimal path between $\boxminus$ and $\boxplus$ has to cross $R(L_V^\star - 1, L_V^\star)^{\mathrm{1pr}}$.*

Proof. Consider the time step $\bar{t}$ defined in (II.2.12). In the proof of Lemma II.11 we have seen that necessarily $\gamma(\bar{t}-1) \in R(L_V^\star - 1, L_V^\star)$. Note that $\min\{P_V(\gamma(\bar{t}-1)), P_H(\gamma(\bar{t}-1))\} = L_V^\star - 1$ and that $P_V R(\gamma(\bar{t})), P_H R(\gamma(\bar{t})) \geq L_V^\star$. Therefore, $\gamma(\bar{t})$ must be obtained from $\gamma(\bar{t}-1)$ by adding a protuberance at a longer side of the rectangle $\gamma(\bar{t}-1)$. This implies that $\gamma(\bar{t})$ needs to belong to $R(L_V^\star - 1, L_V^\star)^{\mathrm{1pr}}$. □

II.2.3 Identification of $\mathcal{P}^\star$ and $\mathcal{C}^\star$

From Subsection II.2.1, we get that $R(L_V^\star - 1, L_V^\star) \subset \mathcal{P}^\star$. Now let $\sigma \in \mathcal{P}^\star$ and $x \in \Lambda$ be such that $\sigma^x \in \mathcal{C}^\star$. If follows from the definition of $\mathcal{P}^\star$ and $\mathcal{C}^\star$ that there exists $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$ and $\ell \in \mathbb{N}$ such that

(i) $\gamma(\ell) = \sigma$ and $\gamma(\ell+1) = \sigma^x$,

(ii) $\mathrm{H_A}(\gamma(k)) < \mathrm{E_A^\star}$ for all $k \in \{0, \ldots, \ell\}$,

(iii) $\Phi(\boxminus, \gamma(k)) \geq \Phi(\gamma(k), \boxplus)$ for all $k \geq \ell + 1$.

By Lemma II.12, (ii) implies that $\min(P_H R(\sigma), P_V R(\sigma)) \leq L_V^\star - 1$, since otherwise the energy level $\mathrm{E_A^\star}$ would have been reached. There are two possible cases.

Case 1. $[P_H R(\sigma^x), P_V R(\sigma^x) \geq L_V^\star]$.
Lemma II.12 implies that we necessarily have that $\sigma \in R(L_V^\star - 1, L_V^\star)$ and $\sigma^x \in R(L_V^\star - 1, L_V^\star)^{1\text{pr}}$.

Case 2. $[\min(P_H R(\sigma^x), P_V R(\sigma^x)) \leq L_V^\star - 1]$.
Also by Lemma II.12, there must exist some $k^\star \geq \ell + 2$ such that $\gamma(k^\star) \in R(L_V^\star - 1, L_V^\star)$. But this contradicts (iii), since $\Phi(\boxminus, \gamma(k^\star)) < \Phi(\gamma(k^\star), \boxplus) = \mathrm{E}_{\mathrm{A}}^\star$.

Hence, only Case 1 can hold true. We conclude that $\mathcal{P}^\star = R(L_V^\star - 1, L_V^\star)$ and $\mathcal{C}^\star = R(L_V^\star - 1, L_V^\star)^{1\text{pr}}$.

II.2.4 Verification of (H1)

Obviously, $S_{\text{stab}} = \{\boxplus\}$, since $\boxplus$ minimizes all three sums in (II.2.1). It remains to show that $S_{\text{meta}} = \{\boxminus\}$.

Let $\sigma \in S \setminus \{\boxminus, \boxplus\}$. We have to show that $V_\sigma < \Gamma_{\mathrm{A}}^\star$, i.e. there exists $\sigma' \in S$ such that $\mathrm{H}_{\mathrm{A}}(\sigma') < \mathrm{H}_{\mathrm{A}}(\sigma)$ and $\Phi(\sigma, \sigma') - \mathrm{H}_{\mathrm{A}}(\sigma) < \Gamma_{\mathrm{A}}^\star$. There are four possible cases.

Case 1. [σ contains a cluster, which is not a rectangle].
Lemma II.9 implies that σ is not a local minimum, i.e. there exists $x \in \Lambda$ such that $\mathrm{H}_{\mathrm{A}}(\sigma^x) < \mathrm{H}_{\mathrm{A}}(\sigma)$. Moreover, $\Phi(\sigma, \sigma^x) - \mathrm{H}_{\mathrm{A}}(\sigma) = 0 < \Gamma_{\mathrm{A}}^\star$.

Case 2. [σ contains a cluster, which is a rectangle $R = l_1 \times l_2$ with $l_2 \geq L_V^\star$ and $l_1 < \sqrt{|\Lambda|}$].
Let σ' be obtained from σ by attaching on the right of R a new column of length l_2. Then,

$$\begin{aligned} \mathrm{H}_{\mathrm{A}}(\sigma') \leq \mathrm{H}_{\mathrm{A}}(\sigma) + 2J_V - l_2 h \leq \mathrm{H}_{\mathrm{A}}(\sigma) + 2J_V - L_V^\star h < \mathrm{H}_{\mathrm{A}}(\sigma), \quad \text{and} \\ \Phi(\sigma, \sigma') - \mathrm{H}_{\mathrm{A}}(\sigma) = 2J_V - h < \Gamma_{\mathrm{A}}^\star. \end{aligned} \tag{II.2.29}$$

Case 3. [σ contains a cluster, which is a rectangle $R = l_1 \times l_2$ with $l_2 < L_V^\star$ and $l_1 < \sqrt{|\Lambda|}$].
Let σ' be obtained from σ by cutting the right column of R. Then,

$$\begin{aligned} \mathrm{H}_{\mathrm{A}}(\sigma') = \mathrm{H}_{\mathrm{A}}(\sigma) - 2J_V + l_2 h \leq \mathrm{H}_{\mathrm{A}}(\sigma) - 2J_V + (L_V^\star - 1)h < \mathrm{H}_{\mathrm{A}}(\sigma), \quad \text{and} \\ \Phi(\sigma, \sigma') - \mathrm{H}_{\mathrm{A}}(\sigma) = (l_2 - 1)h < \Gamma_{\mathrm{A}}^\star. \end{aligned} \tag{II.2.30}$$

Case 4. [σ contains a cluster, which is a rectangle $R = l_1 \times l_2$ with $l_1 = \sqrt{|\Lambda|}$].
Let σ' be obtained from σ by attaching above R a row that also wraps around the torus. Then, by Assumption II.6 d),

$$\begin{aligned} \mathrm{H}_{\mathrm{A}}(\sigma') = \mathrm{H}_{\mathrm{A}}(\sigma) + 2J_H - l_1 h < \mathrm{H}_{\mathrm{A}}(\sigma), \quad \text{and} \\ \Phi(\sigma, \sigma') - \mathrm{H}_{\mathrm{A}}(\sigma) = 2J_H - h < \Gamma_{\mathrm{A}}^\star. \end{aligned} \tag{II.2.31}$$

We conclude that $S_{\text{meta}} = \{\boxminus\}$.

II.2.5 Verification of (H2)

Obviously, $|\{\sigma \in \mathcal{P}^\star \,|\, \sigma \sim \sigma'\}| = 1$ for all $\sigma' \in \mathcal{C}^\star$. Therefore, (H2) holds.

II.2.6 Computation of K

The starting point for the computation of K is the variational formula (II.1.16). Recall the definitions of $\partial^+ A$ and A^+ for a subset $A \subset S$ from the end of Section II.1.

Lower bound. Since the sum in (II.1.16) has only non-negative summands, we can bound K^{-1} from below by

$$\frac{1}{K} \geq \min_{C_1,\dots,C_I\in[0,1]} \min_{\substack{h:S^\star\to[0,1]\\ h|_{S_\boxminus}=1, h|_{S_\boxplus}=0, h|_{S_i}=C_i\,\forall i}} \frac{1}{2} \sum_{\eta,\eta'\in(\mathcal{C}^\star)^+} \mathbb{1}_{\{\eta\sim\eta'\}}[h(\eta)-h(\eta')]^2. \tag{II.2.32}$$

Obviously, $\partial^+\mathcal{C}^\star \cap S^\star = R(L_V^\star-1, L_V^\star) \cup R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}}$. Moreover, similar computations as in Subsection II.2.1 show that $R(L_V^\star-1, L_V^\star) \subset S_\boxminus$ and $R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}} \subset S_\boxplus$. This leads to

$$\begin{aligned}\frac{1}{K} &\geq \min_{h:\mathcal{C}^\star\to[0,1]} \sum_{\eta\in\mathcal{C}^\star} \left(\sum_{\eta'\in R(L_V^\star-1,L_V^\star),\eta'\sim\eta} [1-h(\eta)]^2 + \sum_{\eta'\in R(L_V^\star-1,L_V^\star)^{2\mathrm{pr}},\eta'\sim\eta} h(\eta)^2 \right) \\ &= \sum_{\eta\in\mathcal{C}^\star} \min_{h\in[0,1]} \Big(|R(L_V^\star-1, L_V^\star)\sim\eta|\ [1-h]^2 + |R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}}\sim\eta|\ h^2 \Big) \\ &= \sum_{\eta\in\mathcal{C}^\star} \frac{|R(L_V^\star-1, L_V^\star)\sim\eta| \cdot |R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}}\sim\eta|}{|R(L_V^\star-1, L_V^\star)\sim\eta| + |R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}}\sim\eta|}.\end{aligned} \tag{II.2.33}$$

For all $\eta\in\mathcal{C}^\star$ we have that $|R(L_V^\star-1, L_V^\star)\sim\eta| = 1$. If the protuberance in η is attached at a corner of $(L_V^\star-1)\times L_V^\star$, then $|R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}}\sim\eta| = 1$, otherwise $|R(L_V^\star-1, L_V^\star)^{2\mathrm{pr}}\sim\eta| = 2$. Taking into account that there are $|\Lambda|$ possible locations for each shape of a critical droplet and 2 possible rotations, we obtain that

$$\frac{1}{K} \geq \Big(2(L_V^\star-2)\frac{2}{3} + 4\frac{1}{2}\Big)2|\Lambda| = \frac{4(2L_V^\star-1)}{3}|\Lambda|. \tag{II.2.34}$$

Upper bound. Define

$$\begin{aligned}\mathscr{S}^- &= \{\sigma\in S^\star \mid \min(P_H R(\eta), P_V R(\eta)) \leq L_V^\star-1 \text{ for all clusters } \eta \text{ of } \sigma\}, \text{ and} \\ \mathscr{S}^+ &= \{\sigma\in S^\star \mid \text{ there exists a cluster } \eta \text{ of } \sigma \text{ such that } P_H R(\eta), P_V R(\eta) \geq L_V^\star\}.\end{aligned} \tag{II.2.35}$$

Note that $S^\star = \mathscr{S}^- \cup \mathscr{S}^+$, $\mathscr{S}^- \cap \mathscr{S}^+ = \emptyset$, $\mathcal{P}^\star \subset \mathscr{S}^-$ and $\mathcal{C}^\star \subset \mathscr{S}^+$. Using the same arguments as in Lemma II.11, we can show the following fact for transitions between $\mathscr{S}^-$ and $\mathscr{S}^+$.

Lemma II.13 *Let $\sigma\in\mathscr{S}^-$ and $\sigma'\in\mathscr{S}^+$. Then $\sigma\sim\sigma'$ if and only if $\sigma\in\mathcal{P}^\star$ and $\sigma'\in\mathcal{C}^\star$.*

Proof. We omit the details of this proof, since they walk along the same lines as the proof of Lemma II.11. □

Recall that the sets $S_1,\dots,S_I$ are assumed to be maximal sets of communicating configurations. Hence, for all $i=1,\dots,I$, we have that either $S_i\subset\mathscr{S}^-$ or $S_i\subset\mathscr{S}^+$, since $S^\star = \mathscr{S}^-\cup\mathscr{S}^+$ and $\mathscr{S}^-\cap\mathscr{S}^+ = \emptyset$. For the same reason and by Subsection II.2.1, we have that $S_\boxminus\subset\mathscr{S}^-$ and $S_\boxplus\subset\mathscr{S}^+$. Therefore, we can estimate K^{-1} from above by restricting the minimum in (II.1.16) only to those functions $h: S^\star\to[0,1]$ such that

$$\begin{aligned}&h(\eta) = 1 \text{ for all } \eta\in\mathscr{S}^-, \text{ and} \\ &h(\eta) = 0 \text{ for all } \eta\in\mathscr{S}^+\setminus\mathcal{C}^\star.\end{aligned} \tag{II.2.36}$$

The restriction to such functions is allowed, since we can choose $C_i = 1$ for all $i \in \{1, \dots, I\}$ such that $S_i \subset \mathscr{S}^-$ and $C_i = 0$ for all $i \in \{1, \dots, I\}$ such that $S_i \subset \mathscr{S}^+$. Thus,

$$\begin{aligned} \frac{1}{K} &\leq \min_{\substack{h:S^\star \to [0,1] \\ h|_{\mathscr{S}^-}=1, h|_{\mathscr{S}^+ \setminus \mathcal{C}^\star}=0}} \frac{1}{2} \sum_{\eta,\eta' \in S^\star} \mathbb{1}_{\{\eta \sim \eta'\}} [h(\eta) - h(\eta')]^2 \\ &= \min_{\substack{h:S^\star \to [0,1] \\ h|_{\mathscr{S}^-}=1, h|_{\mathscr{S}^+ \setminus \mathcal{C}^\star}=0}} \frac{1}{2} \sum_{\eta \in \mathscr{S}^-, \eta' \in \mathscr{S}^-} \mathbb{1}_{\{\eta \sim \eta'\}} [h(\eta) - h(\eta')]^2. \end{aligned} \tag{II.2.37}$$

Using Lemma II.13, the right-hand side is equal to

$$\begin{aligned} &\min_{\substack{h:(\mathcal{C}^\star)^+ \to [0,1] \\ h|_{\mathscr{S}^- \cap \partial^+ \mathcal{C}^\star}=1, h|_{\mathscr{S}^+ \cap \partial^+ \mathcal{C}^\star}=0}} \frac{1}{2} \sum_{\eta,\eta' \in (\mathcal{C}^\star)^+} \mathbb{1}_{\{\eta \sim \eta'\}} [h(\eta) - h(\eta')]^2 \\ &= \min_{h:\mathcal{C}^\star \to [0,1]} \sum_{\eta \in \mathcal{C}^\star} \left(\sum_{\eta' \in R(L_V^\star - 1, L_V^\star), \eta' \sim \eta} [1 - h(\eta)]^2 + \sum_{\eta' \in R(L_V^\star - 1, L_V^\star)^{2\mathrm{pr}}, \eta' \sim \eta} h(\eta)^2 \right) \\ &= \frac{4(2L_V^\star - 1)}{3} |\Lambda|. \end{aligned} \tag{II.2.38}$$

This concludes the proof.

II.3 Ising model with next-nearest-neighbour attraction

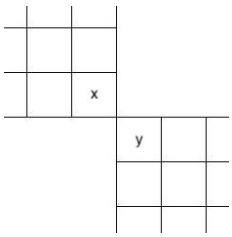
In this section (cf. Subsection I.4.3) the Hamiltonian is given by

$$\mathrm{H}_{\mathrm{NN}}(\sigma) = -\frac{\tilde{J}}{2} \sum_{(x,y) \in \Lambda^\star} \sigma(x)\sigma(y) - \frac{K}{2} \sum_{(x,y) \in \Lambda^{\star\star}} \sigma(x)\sigma(y) - \frac{h}{2} \sum_{x \in \Lambda} \sigma(x), \tag{II.3.1}$$

where $\sigma \in S$, $\tilde{J}, K, h > 0$, $\Lambda^\star$ is the set of *unordered nearest-neighbour bonds* in Λ and $\Lambda^{\star\star}$ is the set of *unordered next-nearest-neighbour bonds* in Λ (cf. (I.4.11)). We can rewrite $\mathrm{H}_{\mathrm{NN}}(\sigma)$ as

$$\mathrm{H}_{\mathrm{NN}}(\sigma) = \mathrm{H}_{\mathrm{NN}}(\boxminus) - h|\sigma| + J|\partial(\sigma)| - K|A(\sigma)|, \tag{II.3.2}$$

where $J = \tilde{J} + 2K$ and $|A(\sigma)|$ is the number of corners (or right angles) of σ. Indeed, a unit segment of $\partial(\sigma)$ breaks two next-nearest-neighbour-bonds. However, at each corner the same broken next-nearest-neighbour-bond is counted twice. This explains the term $-K|A(\sigma)|$ in (II.3.2). Moreover, in the situation



we count four corners, since the bond between x and y is not broken, but we have counted it as such due to the four unit segments surrounding this bond.

Recall that the critical lengths in this model are given by

$$\ell^\star = \left\lceil \frac{2K}{h} \right\rceil \qquad \text{and} \qquad D^\star = \left\lceil \frac{2J}{h} \right\rceil \qquad \text{and} \qquad L^\star = D^\star - 2(\ell^\star - 1). \tag{II.3.3}$$

We make the following assumptions for this section.

Assumption II.14 *a)* $K > h$,

b) $\tilde{J} \geq 2K + h$,

c) $\frac{2J}{h} \notin \mathbb{N}$, $\frac{2K}{h} \notin \mathbb{N}$,

d) $|\Lambda| > \left(\frac{2J(D^\star - 1)}{2J - h(D^\star - 1)} + D^\star \right)^2$.

Similarly as in Section II.2, a) and b) induce a hierarchy in the sense that for the system it is most important to align nearest-neighbours, then next-nearest-neighbours and then to align the spin values with the sign of the magnetic field. As in Section II.2, this assumption is essential to obtain the metastable behavior of the system. Moreover, a) respects a hypothesis made in [89] (but not every hypothesis in there). Assumption c) is made for non-degeneracy reasons. Assumption d) implies that it is not profitable to enlarge a droplet such that one side is subcritical and the other side wraps around the torus. This will become clear later in Lemma II.19. Moreover, d) ensures that the torus is large enough to contain at least a critical droplet. It immediately follows from Assumption II.14 c) that

$$(\ell^\star - 1)h < 2K < \ell^\star h \qquad \text{and} \qquad (D^\star - 1)h < 2J < D^\star h. \tag{II.3.4}$$

We need a few definitions that are mostly carried over from [89].

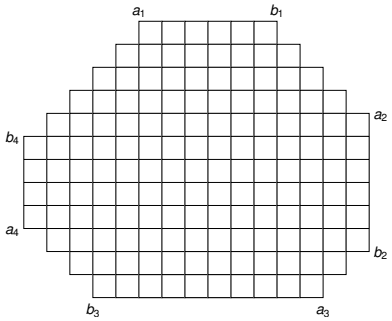
Definition II.15

- *$A \subset \mathbb{Z}^2$ is called an* oblique bar *if $A = \{x_1, \dots, x_n\}$ for some $n \in \mathbb{N}$ and it holds that either $x_i = x_{i-1} + (1,1)^T$ or $x_i = x_{i-1} + (1,-1)^T$ for all $2 \leq i \leq n$.*
- *We say that $\sigma \in S$ is an* octagon of side lengths $D_n, D_w \in \mathbb{N} \cap [1, \sqrt{|\Lambda|} - 1]$ and oblique edge lengths $\ell_{ne}, \ell_{nw}, \ell_{sw}, \ell_{se} \in \mathbb{N}$ *and write $\sigma \in Q(D_n, D_w; \ell_{ne}, \ell_{nw}, \ell_{sw}, \ell_{se})$ if the geometric representation of σ has the following form (cf. [89, Scheme 2.2]). σ is connected and inscribed in a rectangle from $R(D_n, D_w)$. Moreover, σ has four straight edges with endpoints a_i, b_i, $i = 1, \dots, 4$ and four oblique edges that have a local staircase structure with endpoints (b_1, a_2), (b_2, a_3), (b_3, a_4), (b_4, a_1). The* lengths of its oblique edges *are defined by*

$$\ell_{ne} = 1 + \frac{1}{\sqrt{2}}|b_1 - a_2|, \qquad \ell_{se} = 1 + \frac{1}{\sqrt{2}}|b_2 - a_3|, \tag{II.3.5}$$

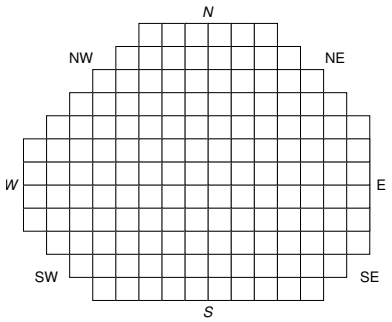
$$\ell_{sw} = 1 + \frac{1}{\sqrt{2}}|b_3 - a_4|, \qquad \ell_{nw} = 1 + \frac{1}{\sqrt{2}}|b_4 - a_1|. \tag{II.3.6}$$

An example with $D_n = 15$, $D_w = 12$, $\ell_{ne} = 5$, $\ell_{nw} = 6$, $\ell_{sw} = 4$, $\ell_{se} = 3$ is given by



We often abuse the notation by identifying $Q(D_n, D_w; \ell_{ne}, \ell_{nw}, \ell_{sw}, \ell_{se})$ *with configurations from this set.*

- *For* $Q \in Q(D_n, D_w; \ell_{ne}, \ell_{nw}, \ell_{sw}, \ell_{se})$*, the upper right edge of length* ℓ_{ne} *is called* NE-edge*, the upper left edge of length* ℓ_{nw} NW-edge*, the down left edge of length* ℓ_{sw} SW-edge *and the down right edge of length* ℓ_{se} *is called* SE-edge*. These four edges are also called* oblique edges*. The four remaining horizontal or vertical edges are called* coordinate edges*. We call the upper coordinate edge* N-edge*, the left one* W-edge*, the bottom one* S-edge *and the right coordinate edge* E-edge*.*



- $Q(D_n, D_w; \ell_{ne}, \ell_{nw}, \ell_{sw}, \ell_{se})$ *is called* stable octagon *if each of his eight edges has length greater or equal to 2.*
- *We abbreviate* $Q(D_n, D_w; \ell, \ell, \ell, \ell) = Q(D_n, D_w; \ell)$ *for all* $\ell \in \mathbb{N}$ *and* $Q(D_n, D_w; \ell^\star) = Q(D_n, D_w)$*. Moreover, we write* $Q(3\ell - 2, 3\ell - 2; \ell) = Q(\ell)$ *for all* $\ell \in \mathbb{N}$*, which corresponds to the case, where all eight edges have the same length given by* ℓ*.*
- $Q(D_n, D_w; \ell)^{\mathrm{1pr}}$ *denotes the set of all configurations that are obtained from a configuration in* $Q(D_n, D_w; \ell)$ *by adding a protuberance somewhere at the* interior *of one of its longest coordinate edges. Here the interior of the coordinate edge contains every site of the edge except for the two sites at the end of the edge. The right droplet in Figure II.3 provides an example.*
- $Q(D_n, D_w; \ell)^{\mathrm{2pr}}$ *denotes the set of all configurations that are obtained from a configuration in* $Q(D_n, D_w; \ell)^{\mathrm{1pr}}$ *by adding a second* $(+1)$*–spin adjacent to the protuberance at the* interior *of the coordinate edge.*

Note that the energy of an octagon $Q \in Q(D_n, D_w; \ell_{ne}, \ell_{nw}, \ell_{sw}, \ell_{se})$ is given by

$$\mathrm{H_{NN}}(Q) = \mathrm{H_{NN}}(\boxminus) - hD_nD_w + 2J(D_n + D_w) + \sum_{a \in \{ne,nw,sw,se\}} F(\ell_a), \tag{II.3.7}$$

where $F(\ell) = -K(2\ell - 1) + \frac{1}{2}h(\ell - 1)\ell$. Now we can formulate the main result of this section.

Theorem II.16 *Under Assumption II.14, the pair* $(\boxminus, \boxplus)$ *satisfies* (H1) *and* (H2) *so that Theorems II.3–II.5 hold for the Ising model with next-nearest-neighbour attraction. Moreover,*

- $\mathcal{P}^\star = Q(D^\star - 1, D^\star)$,
- $\mathcal{C}^\star = Q(D^\star - 1, D^\star)^{\mathrm{1pr}}$,
- $\Phi(\boxminus, \boxplus) - \mathrm{H}_{\mathrm{NN}}(\boxminus) = \mathrm{H}_{\mathrm{NN}}(Q(D^\star - 1, D^\star)) + 2J - 4K - h =: \Gamma^\star_{\mathrm{NN}} =: \mathrm{E}^\star_{\mathrm{NN}} - \mathrm{H}_{\mathrm{NN}}(\boxminus)$,
- $K^{-1} = \frac{4(2L^\star - 5)}{3}|\Lambda|$.

Proof. The proof is divided into the Subsection II.3.1–II.3.6. □

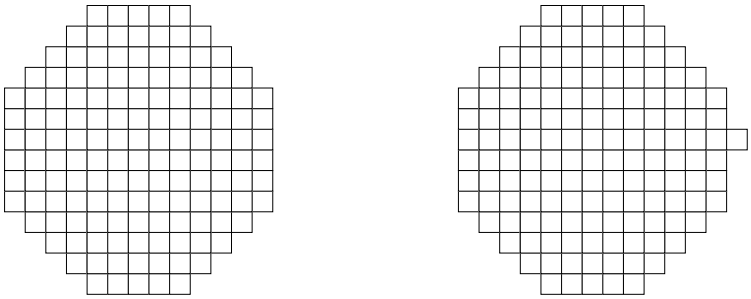


Figure II.3: The left object is an element in $\mathcal{P}^\star$ and the right object is an element in $\mathcal{C}^\star$.

II.3.1 Proof of $\Phi(\boxminus, \boxplus) - \mathrm{H}_{\mathrm{NN}}(\boxminus) \leq \Gamma^\star_{\mathrm{NN}}$

As in Subsection II.2.1, we need to construct a reference path $\gamma_{\mathrm{NN}} : \boxminus \to \boxplus$ such that

$$\max_{\eta \in \gamma_{\mathrm{NN}}} \mathrm{H}_{\mathrm{NN}}(\eta) \leq \mathrm{H}_{\mathrm{NN}}(\boxminus) + \Gamma^\star_{NN} = \mathrm{E}^\star_{\mathrm{NN}}. \tag{II.3.8}$$

Construction of γ_{NN}. We only sketch the construction of γ_{NN}, since we can rely on [89] and we are mainly interested in the part of the path around the critical configuration.

- [From $\boxminus$ to $Q(2)$.]

See [89, Scheme 5.1].

- [From $Q(\ell)$ to $Q(\ell + 1)$ for all $\ell = 2, \dots, \ell^\star - 1$.]

See [89, Scheme 5.2].

- [From $Q(D, D)$ to $Q(D + 1, D + 1)$ for all $D = \ell^\star, \dots, \sqrt{|\Lambda|} - 2$.]

This transition is based on [89, Scheme 5.5], and it goes for example as follows. A $(+1)$–spin is added somewhere at the interior of the E-edge of $Q(D, D)$. Afterwards, successively, adjacent (-1)–spins are flipped in this column until $Q(D + 1, D; \ell^\star + 1, \ell^\star, \ell^\star, \ell^\star + 1)$ is reached. Then a (-1)–spin is flipped at the upper end of the SE-edge. Now (-1)–spins are flipped until $Q(D + 1, D; \ell^\star + 1, \ell^\star, \ell^\star, \ell^\star)$ is reached. Next, the same is done at the NE-edge such that $Q(D + 1, D; \ell^\star, \ell^\star, \ell^\star, \ell^\star) = Q(D + 1, D)$ is reached. This procedure is repeated below $Q(D + 1, D)$, i.e. first (-1)–spins are flipped at the S-edge until $Q(D+1, D+1; \ell^\star, \ell^\star, \ell^\star + 1, \ell^\star + 1)$ is reached, then an oblique bar is added at the SW-edge to reach $Q(D+1, D+1; \ell^\star, \ell^\star, \ell^\star, \ell^\star + 1)$, and finally, (-1)–spins are flipped at the SE-edge, until we arrive at $Q(D + 1, D + 1)$.

• Lastly, flip all remaining (-1)–spins outside of $Q(\sqrt{|\Lambda|}-1, \sqrt{|\Lambda|}-1)$ until $\boxplus$ is reached.

Inequality (II.3.8) *holds.* The proof relies on the detailed computations made in [89, (3.4a)–(3.4e)]. Let $k^\star$ be such that $\gamma_{\mathrm{NN}}(k^\star) \in Q(D^\star - 1, D^\star)$. Then $\mathrm{H}_{\mathrm{NN}}(\gamma_{\mathrm{NN}}(k^\star)) = \mathrm{E}^\star_{\mathrm{NN}} - 2\tilde{J} + h < \mathrm{E}^\star_{\mathrm{NN}}$. If we go backwards in the path from that point on, then we will have to flip all $(+1)$–spins on the NE-edge of $Q(D^\star - 1, D^\star)$. This is an increase of the energy in each step until only one $(+1)$–spin remains on this edge (cf. [89, (3.4a)]). At this point the energy equals to

$$\mathrm{E}^\star_{\mathrm{NN}} - 2\tilde{J} + \ell^\star h < \mathrm{E}^\star_{\mathrm{NN}}, \tag{II.3.9}$$

where we have used Assumption II.14 b) and (II.3.4). Flipping the last $(+1)$–spin on this edge decreases the energy by $2K - h$ (cf. [89, (3.4c)], but with here we flip a $(+1)$–spin). Next, we do the same thing on the SE-edge, i.e. we flip all but one $(+1)$–spins on this edge and arrive at the energy

$$\mathrm{E}^\star_{\mathrm{NN}} - 2\tilde{J} - 2K + 2\ell^\star h < \mathrm{E}^\star_{\mathrm{NN}}. \tag{II.3.10}$$

Flipping the last $(+1)$–spin on this edge, we arrive at $\mathrm{E}^\star_{\mathrm{NN}} - 2J + (2\ell^\star + 1)h$. Finally, we need to flip all but one $(+1)$–spins on the E-edge, which leads to the energy level (cf. [89, (3.4a)])

$$\mathrm{E}^\star_{\mathrm{NN}} - 2J + (D^\star - 1)h < \mathrm{E}^\star_{\mathrm{NN}}, \tag{II.3.11}$$

and flipping the last $(+1)$–spin on this edge, we arrive at the energy $\mathrm{E}^\star_{\mathrm{NN}} - 4J + 4K + D^\star h$ (analogously to [89, (3.4e)]). With the same reasoning, if we keep on going backwards in the path of γ_{NN}, we will always stay below $\mathrm{E}^\star_{\mathrm{NN}}$, since the length of the edges of the circumscribing rectangles will be at most $D^\star - 1$. Hence, we get that

$$\max_{i=1,\dots,k^\star} \mathrm{H}_{\mathrm{NN}}(\gamma_{\mathrm{NN}}(i)) < \mathrm{E}^\star_{\mathrm{NN}}. \tag{II.3.12}$$

We now analyze the path of γ_{NN} after the step $k^\star + 2$. It holds that $\mathrm{H}_{\mathrm{NN}}(\gamma_{\mathrm{NN}}(k^\star + 2)) = \mathrm{E}^\star_{\mathrm{NN}} - h < \mathrm{E}^\star_{\mathrm{NN}}$. First, $L^\star - 4$ $(+1)$–spins are attached at the interior of the S-edge. The energy is decreased to $\mathrm{E}^\star_{\mathrm{NN}} - (L^\star - 3)h$. Afterwards, a $(+1)$–spin is added at the SW-edge, which leads to the energy (cf. [89, (3.4c)])

$$\mathrm{E}^\star_{\mathrm{NN}} + 2K - (L^\star - 2)h < \mathrm{E}^\star_{\mathrm{NN}}, \tag{II.3.13}$$

where we have used the inequality $L^\star \geq 2\ell^\star + 1$, which follows immediately from Assumption II.14 b). Filling the SW-edge decreases the energy by $(\ell^\star - 1)h$. Then we do the same things for the SE-edge by attaching first a $(+1)$–spin on this edge, which increases the energy to

$$\mathrm{E}^\star_{\mathrm{NN}} + 4K - (L^\star + \ell^\star - 2)h < \mathrm{E}^\star_{\mathrm{NN}}, \tag{II.3.14}$$

and then filling up this edge, which decreases the energy to $\mathrm{E}^\star_{\mathrm{NN}} + 4K - (D^\star - 1)h$. Next, a protuberance is added at the interior of the E-edge. We arrive at the energy level (cf. [89, (3.4e)])

$$\mathrm{E}^\star_{\mathrm{NN}} + 4K + 2\tilde{J} - D^\star h = \mathrm{E}^\star_{\mathrm{NN}} + 2J - D^\star h < \mathrm{E}^\star_{\mathrm{NN}}. \tag{II.3.15}$$

If we keep following the path of γ_{NN}, we will always stay below $\Gamma^\star_{\mathrm{NN}}$, since the length of the edges of the circumscribing rectangles will be at least $D^\star$. Combining this with (II.3.12) and the fact that $\mathrm{H}_{\mathrm{NN}}(\gamma_{\mathrm{NN}}(k^\star + 1)) = \mathrm{E}^\star_{\mathrm{NN}}$, we infer (II.3.8).

II.3.2 Proof of $\Phi(\boxminus, \boxplus) - \mathrm{H}_{\mathrm{NN}}(\boxminus) \geq \Gamma^{\star}_{\mathrm{NN}}$

We first list a few observations taken from [89].

Lemma II.17 *Let $\sigma \in S$ be a local minimum of H_{NN}. Then all clusters of σ have distance at least $\sqrt{2}$ from each other and each cluster is either a stable octagon or a rectangle that wraps around the torus.*

Proof. In [89, Lemma 2.1] the following fact was proven. Let $\sigma \in S$ be a local minimum and let σ_1 be a cluster of σ, then $\sigma_1 = Q(\sigma_1)$, where $Q(\sigma_1)$ is the *octagonal envelope* of σ_1, i.e.

- if σ_1 does not wind around the torus, then $Q(\sigma_1)$ is the smallest octagon containing σ_1, and
- if σ_1 winds around the torus, then $Q(\sigma_1) = R(\sigma_1)$.

See [89, p. 424 (before and after Scheme 3.2)] for the definition of the octagonal envelope. Moreover, it is shown, at the end of the proof of [89, Lemma 2.1], that all clusters of σ have distance at least $\sqrt{2}$ and that if a cluster of σ is a octagon, it must be a stable octagon. □

Lemma II.18 *Assume that $\sigma \in S$ consists of a unique cluster that does not wrap around the torus. Let $R(\sigma) \in R(D_n, D_w)$ with $D_n \geq D_w$.*

- *If $D_w \geq 2\ell^{\star} - 1$, then*

$$\mathrm{H}_{\mathrm{NN}}(\sigma) \geq \mathrm{H}_{\mathrm{NN}}(Q(D_n, D_w)), \tag{II.3.16}$$

  and equality holds if and only if $\sigma = Q(D_n, D_w)$.
- *If $D_w < 2\ell^{\star} - 1$ and D_w is odd, then*

$$\mathrm{H}_{\mathrm{NN}}(\sigma) \geq \mathrm{H}_{\mathrm{NN}}(Q(D_n, D_w; \tfrac{1}{2}(D_w + 1))), \tag{II.3.17}$$

  and equality holds if and only if $\sigma = Q(D_n, D_w; \frac{1}{2}(D_w + 1))$.
- *If $D_w < 2\ell^{\star} - 1$ and D_w is even, then*

$$\mathrm{H}_{\mathrm{NN}}(\sigma) \geq \mathrm{H}_{\mathrm{NN}}(Q(D_n, D_w; \tfrac{1}{2}D_w, \tfrac{1}{2}D_w, \tfrac{1}{2}D_w + 1, \tfrac{1}{2}D_w + 1)), \tag{II.3.18}$$

  and equality holds if and only if $\sigma = Q(D_n, D_w; \frac{1}{2}D_w, \frac{1}{2}D_w, \frac{1}{2}D_w + 1, \frac{1}{2}D_w + 1)$.

Proof. See [89, Lemma 3.2] and the proof of [89, Lemma 4.1A]. The main step is to show that the function $l \mapsto F(l)$ is minimized in $\ell^{\star}$. □

In the following lemma we show that every optimal path has to cross $Q(D^{\star} - 1, D^{\star})$.

Lemma II.19 *Let $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$. Then γ has to cross $Q(D^{\star} - 1, D^{\star})$.*

Proof. Assume the contrary, i.e. $\gamma \cap Q(D^{\star} - 1, D^{\star}) = \emptyset$. Using the same arguments as in the end of the proof of Lemma II.11, we can restrict to the case that throughout its whole path γ consists only of a unique cluster. On its way to $\boxplus$, γ has to cross a configuration, whose rectangular envelope has both horizontal and vertical length greater or equal to $D^{\star}$. Let

$$\bar{t} = \min\{l \geq 0 \,|\, P_H R(\gamma(l)), P_V R(\gamma(l)) \geq D^{\star}\}. \tag{II.3.19}$$

By the definition of $\bar{t}$, we have that $R(\gamma(\bar{t}-1)) \in R(D^\star+m, D^\star-1)$ for some $m \geq 0$.

Case 1 . $[m=0]$.
Obviously, $D^\star \geq 2\ell^\star - 1$. Hence, by Lemma II.18 and since γ does not cross $Q(D^\star-1, D^\star)$, we have that

$$\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) > \mathrm{H}_{\mathrm{NN}}(Q(D^\star-1, D^\star)) = \mathrm{E}^\star_{\mathrm{NN}} - 2\tilde{J} + h. \tag{II.3.20}$$

The minimal increase of energy to enlarge the rectangular envelope of a configuration is $2\tilde{J}-h$. Hence,

$$\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t})) \geq \mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) + 2\tilde{J} - h > \mathrm{E}^\star_{\mathrm{NN}}. \tag{II.3.21}$$

This contradicts the fact that $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$, since $\Phi(\boxminus, \boxplus) \leq \max_{\eta\in\gamma_{\mathrm{NN}}} \mathrm{H}_{\mathrm{NN}}(\eta) \leq \mathrm{E}^\star_{\mathrm{NN}}$, where γ_{NN} was constructed in Subsection II.3.1.

Case 2 . $[m \in [1, \sqrt{|\Lambda|} - D^\star)]$.
Again, by Lemma II.18 we have that

$$\begin{aligned}\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) &\geq \mathrm{H}_{\mathrm{NN}}(Q(D^\star+m, D^\star-1)) \\ &= \mathrm{H}_{\mathrm{NN}}(Q(D^\star, D^\star-1)) + m(2J - h(D^\star-1)) \\ &> \mathrm{E}^\star_{\mathrm{NN}} - 2\tilde{J} + h.\end{aligned} \tag{II.3.22}$$

As before, this leads to a contradiction, since

$$\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t})) \geq \mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) + 2\tilde{J} - h > \mathrm{E}^\star_{\mathrm{NN}}. \tag{II.3.23}$$

Case 3 . $[m = \sqrt{|\Lambda|} - D^\star]$.
In this case, $\gamma(\bar{t}-1)$ wraps around the torus. One can easily observe that $\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) \geq \mathrm{H}_{\mathrm{NN}}(R(\sqrt{|\Lambda|}, D^\star-1))$. We infer that

$$\begin{aligned}\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) &\geq \mathrm{H}_{\mathrm{NN}}(R(\sqrt{|\Lambda|}, D^\star-1)) \\ &= \mathrm{H}_{\mathrm{NN}}(Q(D^\star, D^\star-1)) + (\sqrt{|\Lambda|} - D^\star)(2J - h(D^\star-1)) - 2J(D^\star-1) - 4F(\ell^\star) \\ &> \mathrm{H}_{\mathrm{NN}}(Q(D^\star, D^\star-1)) = E^\star_{\mathrm{NN}} - 2\tilde{J} + h,\end{aligned} \tag{II.3.24}$$

where we have used that $F(\ell^\star) < 0$ and Assumption II.14 d). Finally,

$$\mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t})) \geq \mathrm{H}_{\mathrm{NN}}(\gamma(\bar{t}-1)) + 2\tilde{J} - h > E^\star_{\mathrm{NN}}. \tag{II.3.25}$$

This concludes the proof. □

Finally, the following lemma concludes the proof of $\Phi(\boxminus, \boxplus) - \mathrm{H}_{\mathrm{NN}}(\boxminus) \geq \Gamma^\star_{\mathrm{NN}}$.

Lemma II.20 *Let $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$. In order to cross a configuration whose rectangular envelope has both vertical and horizontal length greater or equal to $D^\star$, γ has to pass through $Q(D^\star - 1, D^\star)$ and $Q(D^\star-1, D^\star)^{\mathrm{1pr}}$. In particular, each optimal path between $\boxminus$ and $\boxplus$ has to cross $Q(D^\star-1, D^\star)^{\mathrm{1pr}}$.*

Proof. Consider the time step $\bar{t}$ defined in the proof of Lemma II.19. It was shown there that necessarily $\gamma(\bar{t}-1)$ needs to belong to $Q(D^\star-1, D^\star)$. Since $P_V R(\gamma(\bar{t})), P_H R(\gamma(\bar{t})) \geq D^\star$, $\gamma(\bar{t})$ must be obtained from $\gamma(\bar{t}-1)$ by flipping a (-1)–spin at a site that is attached at the coordinate edge of a longer side of the droplet. If it would not attach at the interior of the coordinate edge, then the energy level $\mathrm{E}^\star_{\mathrm{NN}} + 2K$ would be reached. Hence, the protuberance must be added at the interior of the coordinate edge, which implies that $\gamma(\bar{t})$ needs to belong to $Q(D^\star-1, D^\star)^{\mathrm{1pr}}$. □

II.3.3 Identification of $\mathcal{P}^\star$ and $\mathcal{C}^\star$

In Subsection II.3.1 we have seen that $Q(D^\star-1,D^\star)\subset\mathcal{P}^\star$. Now let $\sigma\in\mathcal{P}^\star$ and $x\in\Lambda$ be such that $\sigma^x\in\mathcal{C}^\star$. If follows from the definition of $\mathcal{P}^\star$ and $\mathcal{C}^\star$ that there exists $\bar\gamma\in(\boxminus,\boxplus)_{\text{opt}}$ and $\ell\in\mathbb{N}$ such that

(i) $\bar\gamma(\ell)=\sigma$ and $\bar\gamma(\ell+1)=\sigma^x$,

(ii) $\mathrm{H}_{\mathrm{NN}}(\bar\gamma(k))<\mathrm{E}^\star_{\mathrm{NN}}$ for all $k\in\{0,\dots,\ell\}$,

(iii) $\Phi(\boxminus,\bar\gamma(k))\geq\Phi(\bar\gamma(k),\boxplus)$ for all $k\geq\ell+1$.

By Lemma II.20, (ii) implies that $\min(P_HR(\sigma),P_VR(\sigma))\leq D^\star-1$, since otherwise the energy level $\mathrm{E}^\star_{\mathrm{NN}}$ would have been reached. There are two possible cases.

Case 1. $[P_HR(\sigma^x),P_VR(\sigma^x)\geq D^\star]$.
Lemma II.20 implies that necessarily $\sigma\in Q(D^\star-1,D^\star)$ and $\sigma^x\in Q(D^\star-1,D^\star)^{\mathrm{1pr}}$.

Case 2. $[\min(P_HR(\sigma^x),P_VR(\sigma^x))\leq D^\star-1]$.
Also by Lemma II.20, there must exist some $k^\star\geq\ell+2$ such that $\bar\gamma(k^\star)\in Q(D^\star-1,D^\star)$. But this contradicts (iii), since $\Phi(\boxminus,\gamma(k^\star))<\Phi(\bar\gamma(k^\star),\boxplus)=\mathrm{E}^\star_{\mathrm{NN}}$. Hence, only Case 1 can hold true.

We conclude that $\mathcal{P}^\star=Q(D^\star-1,D^\star)$ and $\mathcal{C}^\star=Q(D^\star-1,D^\star)^{\mathrm{1pr}}$.

II.3.4 Verification of (H1)

Obviously, $S_{\mathrm{stab}}=\{\boxplus\}$, since $\boxplus$ minimizes all three sums in (II.3.1). It remains to show that $S_{\mathrm{meta}}=\{\boxminus\}$.

Let $\sigma\in S\setminus\{\boxminus,\boxplus\}$. As in Subsection II.2.4, we have to show that there exists $\sigma'\in S$ such that $\mathrm{H}_{\mathrm{NN}}(\sigma')<\mathrm{H}_{\mathrm{NN}}(\sigma)$ and $\Phi(\sigma,\sigma')-\mathrm{H}_{\mathrm{NN}}(\sigma)<\Gamma^\star_{\mathrm{NN}}$.

Case 1. [σ contains a cluster, which is not a stable octagon and not a rectangle that wraps around the torus].
Lemma II.17 implies that σ is not a local minimum, i.e. there exists $x\in\Lambda$ such that $\mathrm{H}_{\mathrm{NN}}(\sigma^x)<\mathrm{H}_{\mathrm{NN}}(\sigma)$ and $\Phi(\sigma,\sigma^x)-\mathrm{H}_{\mathrm{NN}}(\sigma)=0<\Gamma^\star_{\mathrm{NN}}$.

Case 2. [σ contains a cluster Q, which is a stable octagon with $D^\star\leq P_VR(Q)\leq\sqrt{\Lambda}-1$].
Let σ' be obtained from σ by attaching at Q an oblique bar at its NE-edge and its SE-edge respectively, and a vertical bar at its E-edge in the same way that was described in the third step of the construction of γ_{NN} given in Subsection II.3.1. Then we obtain

$$\begin{aligned}&\mathrm{H}_{\mathrm{NN}}(\sigma')-\mathrm{H}_{\mathrm{NN}}(\sigma)\leq 2J-P_VR(Q)h\leq 2J-D^\star h<0,\quad\text{and}\\&\Phi(\sigma,\sigma')-\mathrm{H}_{\mathrm{NN}}(\sigma)\leq 2\tilde J-h<\Gamma^\star_{\mathrm{NN}}.\end{aligned}\tag{II.3.26}$$

Case 3. [σ contains a cluster Q, which is a stable octagon with $P_VR(Q)\leq D^\star-1$].
Let σ' be obtained from σ as follows. First the uppermost $(+1)$–spin at the NE-edge is flipped. Afterwards, successively, adjacent $(+1)$–spins are flipped until this oblique bar consist only of (-1)–spins. In the same way, the SE-edge and the E-edge of Q are detached by starting from the uppermost $(+1)$–spin and then successively flipping all adjacent $(+1)$–spins until the respective edge is detached from Q. Then

$$\begin{aligned}&\mathrm{H}_{\mathrm{NN}}(\sigma')-\mathrm{H}_{\mathrm{NN}}(\sigma)=-2J+P_VR(Q)h\leq-2J+(D^\star-1)h<0,\quad\text{and}\\&\Phi(\sigma,\sigma')-\mathrm{H}_{\mathrm{NN}}(\sigma)\leq(P_VR(Q)-1)h<\Gamma^\star_{\mathrm{NN}}.\end{aligned}\tag{II.3.27}$$

Case 4. [σ contains a cluster R that is a rectangle that wraps around the torus.]. Let σ' be obtained from σ by attaching at R a bar that also wraps around the torus. Then, by Assumption II.6 d), we have that

$$\begin{aligned} \mathrm{H}_{\mathrm{NN}}(\sigma') - \mathrm{H}_{\mathrm{NN}}(\sigma) &= 2\tilde{J} - \sqrt{|\Lambda|}h < 0, \quad \text{and} \\ \Phi(\sigma,\sigma') - \mathrm{H}_{\mathrm{NN}}(\sigma) &= 2\tilde{J} - h < \Gamma^{\star}_{\mathrm{NN}}. \end{aligned} \tag{II.3.28}$$

We conclude that $S_{\mathrm{meta}} = \{\boxminus\}$.

II.3.5 Verification of (H2)

Obviously, $|\{\sigma \in \mathcal{P}^{\star} \,|\, \sigma \sim \sigma'\}| = 1$ for all $\sigma' \in \mathcal{C}^{\star}$. Therefore, (H2) holds.

II.3.6 Computation of K

We proceed analogously to Subsection II.2.6.

Lower bound. Note that $\partial^{+}\mathcal{C}^{\star} \cap S^{\star} = Q(D^{\star}-1, D^{\star}) \cup Q(D^{\star}-1, D^{\star})^{\mathrm{2pr}}$, $Q(D^{\star}-1, D^{\star}) \subset S_{\boxminus}$ and $Q(D^{\star}-1, D^{\star})^{\mathrm{2pr}} \subset S_{\boxplus}$. Hence,

$$\begin{aligned} \frac{1}{K} &\geq \min_{C_1,\dots,C_I \in [0,1]} \min_{\substack{h:S^{\star}\to[0,1] \\ h|_{S_{\boxminus}}=1, h|_{S_{\boxplus}}=0, h|_{S_i}=C_i \,\forall i}} \frac{1}{2} \sum_{\eta,\eta' \in (\mathcal{C}^{\star})^{+}} \mathbb{1}_{\{\eta\sim\eta'\}} [h(\eta) - h(\eta')]^2 \\ &= \min_{h:\mathcal{C}^{\star}\to[0,1]} \sum_{\eta\in\mathcal{C}^{\star}} \left(\sum_{\eta'\in Q(D^{\star}-1,D^{\star}),\eta'\sim\eta} [1-h(\eta)]^2 + \sum_{\eta'\in Q(D^{\star}-1,D^{\star})^{\mathrm{2pr}},\eta'\sim\eta} h(\eta)^2 \right) \\ &= \sum_{\eta\in\mathcal{C}^{\star}} \min_{h\in[0,1]} \Big(|Q(D^{\star}-1,D^{\star}) \sim \eta| \; [1-h]^2 + |Q(D^{\star}-1,D^{\star})^{\mathrm{2pr}} \sim \eta| \; h^2 \Big) \\ &= \sum_{\eta\in\mathcal{C}^{\star}} \frac{|Q(D^{\star}-1,D^{\star}) \sim \eta| \cdot |Q(D^{\star}-1,D^{\star})^{\mathrm{2pr}} \sim \eta|}{|Q(D^{\star}-1,D^{\star}) \sim \eta| + |Q(D^{\star}-1,D^{\star})^{\mathrm{2pr}} \sim \eta|}. \end{aligned} \tag{II.3.29}$$

For all $\eta \in \mathcal{C}^{\star}$ we have that $|Q(D^{\star}-1, D^{\star}) \sim \eta| = 1$. Moreover, there are four sites at the longer coordinate edges of a critical droplet with $|Q(D^{\star}-1, D^{\star})^{\mathrm{2pr}} \sim \eta| = 1$, and $2(L^{\star}-4)$ sites with $|Q(D^{\star}-1, D^{\star})^{\mathrm{2pr}} \sim \eta| = 2$. Further, there are $|\Lambda|$ possible locations for a configuration in $\mathcal{C}^{\star}$, and there are two analogue rotations for each critical droplet. Therefore, we obtain that

$$\frac{1}{K} \geq \Big(2(L^{\star}-4)\frac{2}{3} + 4\frac{1}{2}\Big)2|\Lambda| = \frac{4(2L^{\star}-5)}{3}|\Lambda|. \tag{II.3.30}$$

Upper bound. The following proof uses the same arguments as in Subsection II.2.6. Hence, we shall only sketch the main arguments here. Define

$$\begin{aligned} \mathscr{S}^{-} &= \{\sigma \in S^{\star} \,|\, \min(P_H R(\eta), P_V R(\eta)) \leq D^{\star}-1 \text{ for all clusters } \eta \text{ of } \sigma\}, \text{ and} \\ \mathscr{S}^{+} &= \{\sigma \in S^{\star} \,|\, \text{ there exists a cluster } \eta \text{ of } \sigma \text{ such that } P_H R(\eta), P_V R(\eta) \geq D^{\star} \,\}. \end{aligned} \tag{II.3.31}$$

Note that $S^{\star} = \mathscr{S}^{-} \cup \mathscr{S}^{+}$, $\mathscr{S}^{-} \cap \mathscr{S}^{+} = \emptyset$, $\mathcal{P}^{\star} \subset \mathscr{S}^{-}$ and $\mathcal{C}^{\star} \subset \mathscr{S}^{+}$.

Lemma II.21 *Let $\sigma \in \mathscr{S}^-$ and $\sigma' \in \mathscr{S}^+$. Then $\sigma \sim \sigma'$ if and only if $\sigma \in \mathcal{P}^\star$ and $\sigma' \in \mathcal{C}^\star$.*

Proof. The proof is a straightforward adaptation of the proof of Lemma II.19. □

The same arguments as in Subsection II.2.6 yield that $S_{\boxminus} \subset \mathscr{S}^-$, $S_{\boxplus} \subset \mathscr{S}^+$ and for all $i = 0, \dots, I$ we have that either $S_i \subset \mathscr{S}^-$ or $S_i \subset \mathscr{S}^+$. Therefore, as in Subsection II.2.6, we can estimate the minimum in (II.1.16) from above by the minimum over all functions of the form (II.2.36) and use Lemma II.21 to infer that

$$\begin{aligned}
\frac{1}{K} &\leq \min_{\substack{h:S^\star\to[0,1]\\ h|_{\mathscr{S}^-}=1, h|_{\mathscr{S}^+\setminus\mathcal{C}^\star}=0}} \frac{1}{2}\sum_{\eta,\eta'\in S^\star} \mathbb{1}_{\{\eta\sim\eta'\}}[h(\eta)-h(\eta')]^2 \\
&= \min_{\substack{h:(\mathcal{C}^\star)^+\to[0,1]\\ h|_{\mathscr{S}^-\cap\partial^+\mathcal{C}^\star}=1, h|_{\mathscr{S}^+\cap\partial^+\mathcal{C}^\star}=0}} \frac{1}{2}\sum_{\eta,\eta'\in(\mathcal{C}^\star)^+} \mathbb{1}_{\{\eta\sim\eta'\}}[h(\eta)-h(\eta')]^2 \\
&= \min_{h:\mathcal{C}^\star\to[0,1]} \sum_{\eta\in\mathcal{C}^\star}\left(\sum_{\eta'\in Q(D^\star-1,D^\star),\eta'\sim\eta}[1-h(\eta)]^2 + \sum_{\eta'\in Q(D^\star-1,D^\star)^{2\mathrm{pr}},\eta'\sim\eta} h(\eta)^2\right) \\
&= \frac{4(2L^\star-5)}{3}|\Lambda|.
\end{aligned} \tag{II.3.32}$$

II.4 Ising model with alternating magnetic field

We adapt the same strategy as in the Sections II.2 and II.3 to a third modification of the Ising model (cf. Section I.4.4), where the Hamiltonian is given by

$$\mathrm{H}_\pm(\sigma) = -\frac{J}{2}\sum_{(x,y)\in\Lambda^\star}\sigma(x)\sigma(y) + \frac{h_2}{2}\sum_{x\in\Lambda_2}\sigma(x) - \frac{h_1}{2}\sum_{x\in\Lambda_1}\sigma(x), \tag{II.4.1}$$

where $\sigma \in S$, $J, h_2, h_1 > 0$, $\Lambda_2 = \{(x_1,x_2) \in \Lambda \,|\, x_2 \text{ is odd}\}$ are the *odd rows* in Λ, $\Lambda_1 = \Lambda \setminus \Lambda_2$ are the *even rows* and $\Lambda^\star$ is the set of *unordered nearest-neighbour bonds* in Λ. One can rewrite $\mathrm{H}_\pm(\sigma)$ geometrically as

$$\mathrm{H}_\pm(\sigma) = \mathrm{H}_\pm(\boxminus) + h_2|\sigma\cap\Lambda_2| - h_1|\sigma\cap\Lambda_1| + J|\partial(\sigma)|. \tag{II.4.2}$$

Under the assumptions below, the critical lengths in this model are given by

$$l_b^\star = \left\lceil \frac{\mu}{\varepsilon} \right\rceil \qquad \text{and} \qquad l_h^\star = 2l_b^\star - 1, \tag{II.4.3}$$

where

$$\begin{aligned}
&\varepsilon = h_1 - h_2, \quad \text{and} \\
&\mu = 2J - h_2.
\end{aligned} \tag{II.4.4}$$

$l_b^\star$ will be the length of the basis of the critical droplet, and $l_h^\star$ will be its height. We make the following assumptions in this section.

Assumption II.22 *a) $h_1 > h_2$,*

b) $J > h_1$,

c) $\frac{\mu}{\varepsilon} \notin \mathbb{N}$,

d) $|\Lambda| > \left(2\left\lceil \frac{2J(l_h^\star-1)+h_2}{4J-\varepsilon(l_b^\star-1)} \right\rceil + l_h^\star\right)^2$.

Assumption a) ensures that $\boxplus$ is the *stable configuration* in this system. Assumptions b), c) and d) are made due to similar reasons as in the Sections II.2 and II.3. Assumption b) can also be modified in various ways. E.g. one can take $J < h_1 < 2J$. We refer to [107, p. 10], where several other regimes are listed. In contrast to [107], in this text, we only consider the regime given in Assumption II.22, since all other regimes can be handled in a similar way without using new ideas. It immediately follows from Assumption II.22 c) that

$$(l_b^\star - 1)\varepsilon < \mu < l_b^\star \varepsilon. \tag{II.4.5}$$

In the following definition we define the protocritical and the critical configurations for this model. Figure II.4 below provides an example.

Definition II.23 *Let $\sigma \in S$ consist of a unique cluster. $l \in R(1\times 2)$ is called a* 2-protuberance attached at σ *if there exists $x \in l$ and $\bar{y} \in \sigma$ such that $|x - \bar{y}| = 1$ and $\sum_{y\in\Lambda:|y-x|=1} \sigma(y) = 0$ and $\sum_{y\in\Lambda:|y-x'|=1} \sigma(y) = -2$, where x' is the unique element in $l \setminus x$.*

We define the following subsets of S.

$\mathcal{P}_1$ denotes the set of all configurations consisting only of a rectangle from $R((l_b^\star - 1) \times l_h^\star)$ that starts and ends in Λ_1 (i.e. the bottom and the top row belong to Λ_1) and with an additional protuberance attached at one of its vertical sides on a row in Λ_1.

$\mathcal{C}_1$ denotes the set of all configurations that are obtained from a configuration in $\mathcal{P}_1$ by adding a second $(+1)$–spin in Λ_2 adjacent to the protuberance and attached at the rectangle.

$\mathcal{P}_2'$ denotes the set of all configurations consisting only of a rectangle from $R(l_b^\star \times (l_h^\star - 2))$ that starts and ends in Λ_1 and with an additional horizontal bar of length 2 attached at one of the horizontal sides of the droplet.

$\mathcal{P}_2''$ denotes the set of all configurations consisting only of a rectangle from $R(l_b^\star \times (l_h^\star - 2))$ that starts and ends in Λ_1 and with an additional 2-protuberance attached at one of the horizontal sides of the droplet.

Define $\mathcal{P}_2 = \mathcal{P}_2' \cup \mathcal{P}_2''$.

$\mathcal{C}_2'$ denotes the set of all configurations that are obtained from a configuration in $\mathcal{P}_2'$ by adding a $(+1)$–spin in Λ_1 attached to the horizontal bar of length 2.

$\mathcal{C}_2''$ denotes the set of all configurations that are obtained from a configuration in $\mathcal{P}_2''$ by adding a $(+1)$–spin, which is both attached to the 2-protuberance and to the rectangle.

We easily observe that $\mathcal{C}_2' = \mathcal{C}_2''$. Define $\mathcal{C}_2 = \mathcal{C}_2' = \mathcal{C}_2''$.

We now state the main result of this section.

Theorem II.24 *Under Assumption II.22, the pair $(\boxminus, \boxplus)$ satisfies* (H1) *so that Theorem II.3 a), Theorem II.4 and Theorem II.5 hold for the Ising model with alternating magnetic field. Moreover,*

- $\mathcal{P}^\star = \mathcal{P}_1 \cup \mathcal{P}_2$,
- $\mathcal{C}^\star = \mathcal{C}_1 \cup \mathcal{C}_2$,

- $\Phi(\boxminus,\boxplus) - \mathrm{H}_\pm(\boxminus) = 4J\, l_b^\star + \mu(l_b^\star - 1) - \varepsilon(l_b^\star(l_b^\star - 1) + 1) =: \Gamma_\pm^\star =: \mathrm{E}_\pm^\star - \mathrm{H}_\pm(\boxminus)$,
- $K^{-1} = \frac{14\,(l_b^\star - 1)}{3}|\Lambda|$.

Proof. The proof is divided into the Subsection II.4.1–II.4.5. □

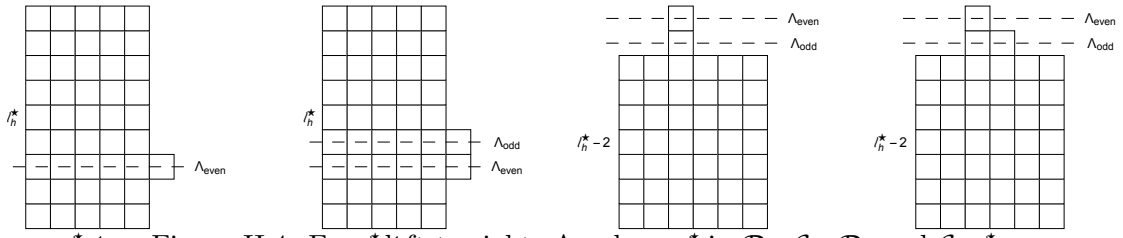


Figure II.4: From left to right: An element in $\mathcal{P}_1$, $\mathcal{C}_1$, $\mathcal{P}_2$ and $\mathcal{C}_2$.

II.4.1 Proof of $\Phi(\boxminus,\boxplus) - \mathrm{H}_\pm(\boxminus) \le \Gamma_\pm^\star$

As in Sections II.2 and II.3, we construct a reference path $\gamma_\pm : \boxminus \to \boxplus$ such that

$$\max_{\eta\in\gamma_\pm} \mathrm{H}_\pm(\eta) \le \mathrm{H}_\pm(\boxminus) + \Gamma_\pm^\star = \mathrm{E}_\pm^\star. \tag{II.4.6}$$

Construction of $\gamma_\pm$. $\gamma_\pm$ is given through the following scheme.

- Let $\gamma_\pm(0) = \boxminus$.
- In the first step an arbitrary (-1)–spin in Λ_1 is flipped.
- [From $R(l \times (2l-1))$ to $R((l+1)\times(2l+1))$ for $l \le l_b^\star - 1$.]

A protuberance is added to the right vertical side of the droplet at a row that belongs to Λ_1. Then successively adjacent (-1)–spins are flipped until the droplet belongs to the set $R\left((l+1)\times(2l-1)\right)$. Next, a protuberance is added to the above horizontal side of the droplet, which is an odd row. Afterwards, a second $(+1)$–spin is added above the protuberance on the even row. Hence, a 2-protuberance attached to the above horizontal side of the droplet was added. Then, analogously as for this 2-protuberance, one adds successively adjacent 1×2 rectangles at the above horizontal side of the droplet until $R((l+1)\times(2l+1))$ is reached.

- [From $R(l \times l_h^\star)$ to $R((l+1)\times l_h^\star)$ for $l \ge l_b^\star$.]

A protuberance is added on the right vertical side of the droplet at a row that belongs to Λ_1, and successively adjacent (-1)–spins are flipped until the droplet belongs to $R((l+1)\times l_h^\star)$.

- [From $R(\sqrt{|\Lambda|}\times l_h^\star)$ to $\boxplus$.]

As above, a 2-protuberance is added to the above horizontal side of the droplet, which is an odd row, and successively adjacent 1×2 rectangles are added at the above horizontal side of the droplet, until a configuration in $R(\sqrt{|\Lambda|}\times(l_h^\star+2))$ is reached. This procedure is repeated until the configuration $\boxplus$ appears.

Inequality (II.4.6) *holds.* Let $k^\star$ be such that $\gamma_\pm(k^\star) \in R(l_b^\star \times (l_h^\star - 2))$. Using (II.4.2) and Assumption II.22, we observe that

$$\begin{aligned}
\mathrm{H}_\pm(\gamma_\pm(k^\star)) &= \mathrm{H}_\pm(\boxminus) + 6J(l_b^\star - 1) - h_1 l_b^\star(l_b^\star - 1) + h_2 l_b^\star(l_b^\star - 2) \\
&= \mathrm{H}_\pm(\boxminus) + 4J\,(l_b^\star - 1) + \mu(l_b^\star - 1) - \varepsilon l_b^\star(l_b^\star - 1) - h_2 \\
&= \mathrm{E}_\pm^\star - 4J + \varepsilon - h_2 < \mathrm{E}_\pm^\star.
\end{aligned} \tag{II.4.7}$$

If we go backwards in the path from that point on, then we will have to cut the right vertical bar of the droplet. While cutting this vertical bar, the highest energy level is reached when only two adjacent (+1)–spins remain, one in Λ_1 and one in Λ_2. Indeed, at that point the energy in (II.4.7) is increased by $\varepsilon/2(l_b^\star - 3) + h_2$, so that it equals

$$\mathrm{E}_\pm^\star - 4J + \varepsilon(l_b^\star - 1) < \mathrm{E}_\pm^\star, \tag{II.4.8}$$

where we have used (II.4.5). Cutting the last (+1)–spins, we reach $R((l_b^\star - 1) \times (l_h^\star - 2))$ and the energy decreases to $\mathrm{E}_\pm^\star - 6J + \varepsilon l_b^\star$. Next, we have to cut the above two rows by successively cutting vertical bars of length 2 in these rows. Doing that, the highest energy point is the stage, where only one vertical bar of length 2 and a single (+1)–spin in Λ_2 next to it have remained. At this point the energy has increased by $\varepsilon(l_b^\star - 2) + h_1$ and it equals to

$$\mathrm{E}_\pm^\star - 6J + \varepsilon l_b^\star + \varepsilon(l_b^\star - 2) + h_1 = \mathrm{E}_\pm^\star - 6J + 2\varepsilon(l_b^\star - 1) + h_1 < \mathrm{E}_\pm^\star. \tag{II.4.9}$$

Using the same arguments, if we keep on going backwards in the path of $\gamma_\pm$, we will always stay below $\mathrm{E}_\pm^\star$, since the sizes of the cut columns and rows further decrease. Hence,

$$\max_{i=1,\dots,k^\star} \mathrm{H}_\pm(\gamma_\pm(i)) < \mathrm{E}_\pm^\star. \tag{II.4.10}$$

We now consider the path of $\gamma_\pm$ after the step $k^\star + 3$. We have that $\mathrm{H}_\pm(\gamma_\pm(k^\star + 3)) = \mathrm{E}_\pm^\star$. First, the two rows above the droplet are filled. This lowers the energy to $\mathrm{E}_\pm^\star - \varepsilon(l_b^\star - 1) - h_2$. Afterwards, a protuberance is attached on the right vertical side of the droplet in a row that belongs to Λ_1. The energy is increased by $2J - h_1$ and equals to

$$\mathrm{E}_\pm^\star + \mu - \varepsilon l_b^\star - h_2 < \mathrm{E}_\pm^\star. \tag{II.4.11}$$

Adding a second (+1)–spin adjacent to the protuberance further increases the energy by h_2. By (II.4.5), we still get

$$\mathrm{E}_\pm^\star + \mu - \varepsilon l_b^\star < \mathrm{E}_\pm^\star. \tag{II.4.12}$$

If we fill this column, we further decrease the energy so that the energy still remains below $\mathrm{E}_\pm^\star$. In the following, analogously, columns are added successively on the right vertical side of the droplet and each column decreases the energy by $\mu - \varepsilon l_b^\star$. This is repeated until the droplet wraps around the torus. It is easy to see that the remaining part of γ also stays below $\mathrm{E}_\pm^\star$. Hence,

$$\max_{i \geq k^\star + 3} \mathrm{H}_\pm(\gamma_\pm(i)) \leq \mathrm{E}_\pm^\star. \tag{II.4.13}$$

Finally, we have that $\mathrm{H}_\pm(\gamma_\pm(k^\star + 1)) = \mathrm{E}_\pm^\star - 2J + h_1 - h_2$ and $\mathrm{H}_\pm(\gamma_\pm(k^\star + 2)) = \mathrm{E}_\pm^\star - h_2$, which are clearly below $\mathrm{E}_\pm^\star$. Hence, together with (II.4.10) and (II.4.13), we conclude (II.4.6).

II.4.2 Proof of $\Phi(\boxminus, \boxplus) - \mathrm{H}_\pm(\boxminus) \geq \Gamma_\pm^\star$

Before we prove that $\Phi(\boxminus, \boxplus) - \mathrm{H}_\pm(\boxminus) \geq \Gamma_\pm^\star$, we need to collect some results that were established in [107].

Definition II.25 *Let $l_1, l_2 \in \mathbb{N}$. We say that $\sigma \in R(l_1 \times l_2)$ is a* stable rectangle *if σ starts and ends in Λ_1 (i.e. its bottom and top row belong to Λ_1), $l_1 \geq 2$, $l_2 \geq 3$ and l_2 is odd. Note that a stable rectangle can possibly wrap around the torus.*

Recall (II.1.11). Analogously to [107], we say that $\sigma \in S$ is h_2–stable if and only if $\sigma \in S_{h_2}$.

Lemma II.26 *$\sigma \in S$ is h_2–stable if and only if σ is a union of isolated stable rectangles.*

Proof. This is the content of [107, Proposition 3.1] and the comment after it. □

The following lemma is the analogue of Corollary II.10 for this model.

Lemma II.27 *Let $\sigma \in S$ be such that $R(\sigma)$ is a stable rectangle. Then*

$$\mathrm{H}_{\pm}(\sigma) \geq \mathrm{H}_{\pm}(R(\sigma)), \tag{II.4.14}$$

and equality holds, if and only if $\sigma = R(\sigma)$.

Proof. This is the content of [107, Lemma 3.3]. □

Let $l_1, l_2 \in \mathbb{N}$, and let $R, R' \in R(l_1 \times l_2)$. Note that if l_2 is an odd number, R starts in Λ_2 and R' starts in Λ_1, then $\mathrm{H}_{\pm}(R) > \mathrm{H}_{\pm}(R')$. And if l_2 is even, then $\mathrm{H}_{\pm}(R) = \mathrm{H}_{\pm}(R')$. Therefore, from now on, we set $\mathrm{H}_{\pm}(l_1 \times l_2) = \mathrm{H}_{\pm}(R')$, which is the energetically more profitable choice. We will use this fact tacitly several times in the remaining part of this section.

Lemma II.28 *Let $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$. Then γ has to cross $\mathcal{P}_1 \cup \mathcal{P}_2$.*

Proof. Assume the contrary, i.e. $\gamma \cap \{\mathcal{P}_1 \cup \mathcal{P}_2\} = \emptyset$. Suppose first that throughout its whole path γ consists of a unique cluster. At the end of this proof we treat the general case.

Since γ leads to $\boxplus$, there exists some time $\bar{t}$ such that $P_V R(\gamma(j)) \geq l_h^\star$ and $P_H R(\gamma(j)) \geq l_b^\star$ for all $j \geq \bar{t}$ and

$$\bar{t} - 1 = \max\left\{j \geq 0 \mid P_V R(\gamma(j)) < l_h^\star \text{ or } P_H R(\gamma(j)) < l_b^\star\right\}. \tag{II.4.15}$$

Note that $\gamma(\bar{t}-1)$ has to satisfy either

1.) $P_H R(\gamma(\bar{t}-1)) = l_b^\star - 1$ and $P_V R(\gamma(\bar{t}-1)) = l_h^\star + n$ for some $n \geq 0$, or

2.) $P_V R(\gamma(\bar{t}-1)) = l_h^\star - 1$ and $P_H R(\gamma(\bar{t}-1)) = l_b^\star + m$ for some $m \geq 0$.

Case 1. $[P_H R(\gamma(\bar{t}-1)) = l_b^\star - 1$ and $P_V R(\gamma(\bar{t}-1)) = l_h^\star + n$ for some $n \geq 0]$.

Case 1.1. $[n = 0]$.
Let τ be the first time that a second $(+1)$–spin is added outside of $R(\gamma(\bar{t}-1)) = R((l_b^\star - 1) \times l_h^\star)$, i.e.

$$\tau = \min\left\{j \geq \bar{t} + 1 \;\middle|\; |\gamma(j) \setminus R(\gamma(\bar{t}-1))| = 2\right\} \tag{II.4.16}$$

Note that $|\gamma(\tau-1) \setminus R(\gamma(\bar{t}-1))| = 1$ and that this protuberance is placed either at the right vertical side or at the left vertical side of $R(\gamma(\bar{t}-1))$, since $\gamma(\bar{t}-1)$ was the last configuration with the property $P_H R(\gamma(\bar{t}-1)) = l_b^\star - 1$. Analogously, $P_V R(\gamma(\tau-1)) = l_h^\star$, otherwise, this would also contradict the definition of $\bar{t}-1$. Now if $\gamma(\tau-1) \setminus R(\gamma(\bar{t}-1)) \in \Lambda_2$, we have that

$$\mathrm{H}_{\pm}(\gamma(\tau-1)) \geq \mathrm{H}_{\pm}((l_b^\star - 1) \times l_h^\star) + 2J + h_2 = \mathrm{E}_{\pm}^\star + h_1 > \mathrm{E}_{\pm}^\star. \tag{II.4.17}$$

This contradicts $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$, since we already know from Subsection II.4.1 that $\Phi(\boxminus, \boxplus) \leq E^\star_\pm$. But if $\gamma(\tau-1) \setminus R(\gamma(\bar{t}-1)) \in \Lambda_1$, then, since γ does not cross $\mathcal{P}_1$ and since the minimal increase of energy to enlarge the rectangular envelope is $2J - h_1$, we have by Lemma II.27 that

$$\mathrm{H}_\pm(\gamma(\tau-1)) > \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + 2J - h_1 = \mathrm{E}^\star_\pm - h_2. \tag{II.4.18}$$

$\gamma(\tau)$ is obtained from $\gamma(\tau-1)$ by flipping a (-1)–spin outside of $R(\gamma(\bar{t}-1))$. One can easily see that the most profitable way is to flip a (-1)–spin at a site that is adjacent to the protuberance of $\gamma(\tau-1)$, which consequently must belong to Λ_2. Hence,

$$\mathrm{H}_\pm(\gamma(\tau)) \geq \mathrm{H}_\pm(\gamma(\tau-1)) + h_2 > \mathrm{E}^\star_\pm, \tag{II.4.19}$$

which leads to a contradiction.

Case 1.2. $[n = 2k$ for some $k > 1]$.
According to Lemma II.27, we have that

$$\begin{aligned} \mathrm{H}_\pm(\gamma(\bar{t})) &\geq \mathrm{H}_\pm(\gamma(\bar{t}-1)) + 2J - h_1 \geq \mathrm{H}_\pm((l_b^\star - 1) \times (l_h^\star + 2k)) + 2J - h_1 \\ &= \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + k(4J - \varepsilon(l_b^\star - 1)) + 2J - h_1 > \mathrm{E}^\star_\pm. \end{aligned} \tag{II.4.20}$$

As before, this leads to a contradiction.

Case 1.3. $[n = 2k+1$ for some $k \geq 0]$.
It holds that either the top or bottom row of $\gamma(\bar{t})$ must belong to Λ_2. Similar to Case 1.2, we obtain a contradiction, since

$$\begin{aligned} \mathrm{H}_\pm(\gamma(\bar{t})) &\geq \mathrm{H}_\pm(\gamma(\bar{t}-1)) + 2J - h_1 \geq \mathrm{H}_\pm((l_b^\star - 1) \times (l_h^\star + 2k)) + 4J + h_2 - h_1 \\ &\geq \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + 4J + h_2 - h_1 > \mathrm{E}^\star_\pm. \end{aligned}$$

Case 1.4. $[P_V R(\gamma(\bar{t}-1)) = \sqrt{|\Lambda|}]$.
Using Assumption II.22 d), we observe that

$$\begin{aligned} \mathrm{H}_\pm(\gamma(\bar{t}-1)) &\geq \mathrm{H}_\pm((l_b^\star - 1) \times \sqrt{|\Lambda|}) \\ &\geq \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + \lfloor (\sqrt{|\Lambda|} - l_h^\star)/2 \rfloor (4J - \varepsilon(l_b^\star - 1)) + h_2(l_b^\star - 1) \\ &> \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + h_2. \end{aligned} \tag{II.4.21}$$

This leads to a contradiction, since

$$\mathrm{H}_\pm(\gamma(\bar{t})) \geq \mathrm{H}_\pm(\gamma(\bar{t}-1)) + 2J - h_1 > \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + h_2 + 2J - h_1 = \mathrm{E}^\star_\pm. \tag{II.4.22}$$

Case 2. $[P_V R(\gamma(\bar{t}-1)) = l_h^\star - 1$ and $P_H R(\gamma(\bar{t}-1)) = l_b^\star + m$ for some $m \geq 0]$.
Assume first that $\gamma(\bar{t})$ starts in Λ_2. Hence, the top and the bottom row of $\gamma(\bar{t})$ belong to Λ_2. Then, since $\gamma(\bar{t})$ is obtained from $\gamma(\bar{t}-1)$ by adding a protuberance at a horizontal side of $R(\gamma(\bar{t}-1))$, we have that

$$\mathrm{H}_\pm(\gamma(\bar{t})) \geq \mathrm{H}_\pm(\gamma(\bar{t}-1)) + 2J + h_2. \tag{II.4.23}$$

Note that either the top or the bottom row of $\gamma(\bar{t}-1)$ belongs to Λ_2. By cutting this row, we can estimate the right-hand side of (II.4.23) from below by

$$\mathrm{H}_\pm((l_b^\star + m) \times (l_h^\star - 2)) + 4J + 2h_2. \tag{II.4.24}$$

Moreover, (II.4.24) is bounded from below by

$$\mathrm{H}_{\pm}(l_b^\star \times (l_h^\star - 2)) + m(\mu - \varepsilon(l_b^\star - 1)) + 4J + 2h_2, \tag{II.4.25}$$

which is, obviously, strictly greater that $\mathrm{E}_{\pm}^\star$. This leads to a contradiction, and we can therefore, from now on, assume that $\gamma(\bar{t})$ starts in Λ_1.

Note that $\gamma(\bar{t})$ is obtained from $\gamma(\bar{t}-1)$ either by adding a protuberance at the above horizontal side of $R(\gamma(\bar{t}-1))$ or the below one. Without restriction, we suppose that a protuberance is added at the above horizontal side of $R(\gamma(\bar{t}-1))$. Moreover, let $P_H R(\gamma(\bar{t})) \times (l_h^\star - 2)$ denote the rectangle that is obtained from $R(\gamma(\bar{t}-1))$ by flipping all $(+1)$–spins from the top row of $R(\gamma(\bar{t}-1))$. Note that $P_H R(\gamma(\bar{t})) \times (l_h^\star - 2)$ starts from Λ_1, $P_H R(\gamma(\bar{t})) = l_b^\star + m$ and that $\mathrm{H}_{\pm}(\gamma(\bar{t})) \geq \mathrm{H}_{\pm}(\gamma(\bar{t}-1)) + 2J - h_1$.

Case 2.1. $[|\gamma(\bar{t}) \setminus \{P_H R(\gamma(\bar{t})) \times (l_h^\star - 2)\}| > 2]$.
In this case we necessarily have that $\gamma(\bar{t}-1)$ has at least two $(+1)$–spins in its uppermost row. If $m = 0$, then, since γ does not cross $\mathcal{P}_2'$, we have by Lemma II.27 that

$$\mathrm{H}_{\pm}(\gamma(\bar{t}-1)) > \mathrm{H}_{\pm}(l_b^\star \times (l_h^\star - 2)) + 2J + 2h_2 = \mathrm{E}_{\pm}^\star - 2J + h_1. \tag{II.4.26}$$

This leads to a contradiction, since

$$\mathrm{H}_{\pm}(\gamma(\bar{t})) \geq \mathrm{H}_{\pm}(\gamma(\bar{t}-1)) + 2J - h_1 > \mathrm{E}_{\pm}^\star. \tag{II.4.27}$$

If $m > 0$ and $P_H R(\gamma(\bar{t}-1)) < \sqrt{|\Lambda|}$, then similarly, we observe

$$\begin{aligned} \mathrm{H}_{\pm}(\gamma(\bar{t})) &\geq \mathrm{H}_{\pm}(\gamma(\bar{t}-1)) + 2J - h_1 \geq \mathrm{H}_{\pm}((l_b^\star + m) \times (l_h^\star - 2)) + 2J + 2h_2 + 2J - h_1 \\ &= \mathrm{H}_{\pm}(l_b^\star \times (l_h^\star - 2)) + m(\mu - \varepsilon(l_b^\star - 1)) + 4J + 2h_2 - h_1 > \mathrm{E}_{\pm}^\star, \end{aligned} \tag{II.4.28}$$

which is a contradiction. Finally, if $P_H R(\gamma(\bar{t}-1)) = \sqrt{|\Lambda|}$, we have that

$$\begin{aligned} &\mathrm{H}_{\pm}(\gamma(\bar{t})) \geq \mathrm{H}_{\pm}(\gamma(\bar{t}-1)) + 2J - h_1 \geq \mathrm{H}_{\pm}(\sqrt{|\Lambda|} \times (l_h^\star - 2)) + 4J + 2h_2 - h_1 \\ &= \mathrm{H}_{\pm}(l_b^\star \times (l_h^\star - 2)) + (\sqrt{|\Lambda|} - l_b^\star)(\mu - \varepsilon(l_b^\star - 1)) - 2J(l_h^\star - 1) + 4J + 2h_2 - h_1 > \mathrm{E}_{\pm}^\star. \end{aligned} \tag{II.4.29}$$

Case 2.2. $[|\gamma(\bar{t}) \setminus \{P_H R(\gamma(\bar{t})) \times (l_h^\star - 2)\}| = 2]$.
Define

$$T = \max\left\{ j \geq \bar{t} \,\middle|\, |\gamma(j) \setminus \{P_H R(\gamma(\bar{t})) \times (l_h^\star - 2)\}| \leq 2 \right\} \tag{II.4.30}$$

i.e. the last time that a configuration has only two $(+1)$–spins outside of $P_H R(\gamma(\bar{t})) \times (l_h^\star - 2)$. From the maximality property of $\bar{t}$, we have that $P_V R(\gamma(T)) = l_h^\star$ and $P_H R(\gamma(T)) = l_b^\star + m'$ for some $m' \geq 0$. Moreover, we easily observe that $\mathrm{H}_{\pm}(\gamma(T+1)) \geq \mathrm{H}_{\pm}(\gamma(T)) + h_2$. As in Case 2.1, we show that every possible value of m' leads to a contradiction. If $m' = 0$, then, since γ does not cross $\mathcal{P}_2'$, Lemma II.27 implies that

$$\mathrm{H}_{\pm}(\gamma(T)) > \mathrm{H}_{\pm}(l_b^\star \times (l_h^\star - 2)) + 4J - h_1 + h_2 = \mathrm{E}_{\pm}^\star - h_2, \tag{II.4.31}$$

and therefore

$$\mathrm{H}_{\pm}(\gamma(T+1) \geq \mathrm{H}_{\pm}(\gamma(T)) + h_2 > \mathrm{E}_{\pm}^\star. \tag{II.4.32}$$

If $m' > 0$ and $P_H R(\gamma(T)) < \sqrt{|\Lambda|}$, then

$$\begin{aligned} \mathrm{H}_\pm(\gamma(T+1)) &\geq \mathrm{H}_\pm(\gamma(T)) + h_2 \geq \mathrm{H}_\pm((l_b^\star + m) \times (l_h^\star - 2)) + 4J - h_1 + h_2 + h_2 \\ &> \mathrm{E}_\pm^\star. \end{aligned} \tag{II.4.33}$$

And if $P_H R(\gamma(T)) = \sqrt{|\Lambda|}$, we have that

$$\mathrm{H}_\pm(\gamma(T+1)) \geq \mathrm{H}_\pm(\gamma(T)) + h_2 \geq \mathrm{H}_\pm(\sqrt{|\Lambda|} \times (l_h^\star - 2)) + 4J - h_1 + 2h_2 > \mathrm{E}_\pm^\star. \tag{II.4.34}$$

Finally, we briefly sketch the proof for the case when γ can consist of several clusters. Recall the definitions of $(n_j)_j, ((\gamma^k(j))_{k\leq n_j})_j, ((\ell_V^k(j))_{k\leq n_j})_j, ((\ell_H^k(j))_{k\leq n_j})_j, \ell_V$ and ℓ_H from the proof of Lemma II.11. Similarly as in (II.2.24) and (II.2.25), we can show that for all $j \in \mathbb{N}$,

$$\begin{aligned} \mathrm{H}_\pm(\,\gamma(j)\,) &\geq \sum_{k=1}^{n_j} \mathrm{H}_\pm(\,R(\gamma^k(j))\,) - (n_j - 1)\,\mathrm{H}_\pm(\boxminus) \\ &= \mathrm{H}_\pm(\boxminus) + 2J\left(\sum_{k=1}^{n_j} \ell_V^k(j) + \sum_{k=1}^{n_j} \ell_H^k(j)\right) + h_2 \sum_{k=1}^{n_j} \ell_H^k(j)\lfloor \ell_V^k(j)/2 \rfloor - h_1 \sum_{k=1}^{n_j} \ell_H^k(j)\lceil \ell_V^k(j)/2 \rceil. \end{aligned} \tag{II.4.35}$$

Analogously to (II.2.26) and (II.4.15), define

$$\tilde{t} - 1 = \max\big\{ j \geq 0 \mid \ell_V(j) < l_h^\star \text{ or } \ell_H(j) < l_b^\star \big\}. \tag{II.4.36}$$

We have that either $\ell_H(\tilde{t}-1) = l_b^\star - 1$ or $\ell_V(\tilde{t}-1) = l_h^\star - 1$. Proceeding as in the first part of this proof and in the end of the proof of Lemma II.11, we can now show that, under the hypothesis that $\gamma \cap \{\mathcal{P}_1 \cup \mathcal{P}_2\} = \emptyset$, both cases lead to the fact that $\mathrm{H}_\pm(\,\gamma(\tilde{t})\,) > \mathrm{E}_\pm^\star$, which is a contradiction. We omit the details and conclude the proof of this lemma. □

The following observation concludes the proof of $\Phi(\boxminus, \boxplus) - \mathrm{H}_\pm(\boxminus) \geq \Gamma_\pm^\star$.

Lemma II.29 *Let $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$. In order to cross at a time $\bar{t}$ a configuration $\gamma(\bar{t})$ such that $P_V R(\gamma(j)) \geq l_h^\star$ and $P_H R(\gamma(j)) \geq l_b^\star$ for all $j \geq \bar{t}$, there must be some time $t' \geq \bar{t} - 1$ such that $\gamma(t') \in \mathcal{P}_1 \cup \mathcal{P}_2$ and $\gamma(t'+1) \in \mathcal{C}_1 \cup \mathcal{C}_2$. In particular, every optimal path between $\boxminus$ and $\boxplus$ has to cross $\mathcal{C}_1 \cup \mathcal{C}_2$.*

Proof. Consider the time step $\bar{t}$ defined in the proof of Lemma II.28. It was shown that there necessarily exists a time $t' \geq \bar{t} - 1$ such that $\gamma(t') \in \mathcal{P}_1 \cup \mathcal{P}_2$. Note that

$$P_V R(\gamma(j)) \geq l_h^\star \;\text{ and }\; P_H R(\gamma(j)) \geq l_b^\star \qquad \text{for all } j \geq t' + 1. \tag{II.4.37}$$

In the following we show that $\gamma(t'+1) \in \mathcal{C}_1 \cup \mathcal{C}_2$.

Case 1. [$\gamma(t') \in \mathcal{P}_1 \cup \mathcal{P}_2''$].
In this case, $\mathrm{H}_\pm(\gamma(t')) = \mathrm{E}_\pm^\star - h_2$. Then it is easy to see that $\gamma(t'+1)$ must belong to $\mathcal{C}_1 \cup \mathcal{C}_2''$. Indeed, for any other spin flip that fulfills the constraint (II.4.37), the energy level of $\gamma(t'+1)$ would exceed $\mathrm{E}_\pm^\star$, and this violates the fact that $\gamma \in (\boxminus, \boxplus)_{\mathrm{opt}}$. Note that we have tacitly used Assumption II.22 a).

Case 2. [$\gamma(t') \in \mathcal{P}_2'$].
By the definition of $\bar{t}$, we have that $t' = \bar{t} - 1$. And since $P_V R(\gamma(\bar{t})) = l_h^\star$ and $P_V R(\gamma(t')) = l_h^\star - 1$, we necessarily have that $\gamma(t'+1) \in \mathcal{C}_2'$. This concludes the proof. □

II.4.3 Identification of $\mathcal{P}^\star$ and $\mathcal{C}^\star$

Recall the definition of $\mathcal{P}^\star$ and $\mathcal{C}^\star$ from Definition II.2. Repeating similar computations as in Subsection II.4.1, it is clear that $\mathcal{P}_1 \cup \mathcal{P}_2 \subset \mathcal{P}^\star$. Now let $\sigma \in \mathcal{P}^\star$ and $x \in \Lambda$ be such that $\sigma^x \in \mathcal{C}^\star$. Then there exists $\gamma \in (\boxminus, \boxplus)_{\text{opt}}$ and $\ell \in \mathbb{N}$ such that

(i) $\gamma(\ell) = \sigma$ and $\gamma(\ell+1) = \sigma^x$,

(ii) $\mathrm{H}_\pm(\gamma(k)) < \mathrm{E}^\star_\pm$ for all $k \in \{0, \dots, \ell\}$,

(iii) $\Phi(\boxminus, \gamma(k)) \geq \Phi(\gamma(k), \boxplus)$ for all $k \geq \ell + 1$.

As in the proof of Lemma II.28 and in Lemma II.29, let

$$\bar{t} - 1 = \max\big\{j \geq 0 \mid P_V R(\gamma(j)) < l_h^\star \text{ or } P_H R(\gamma(j)) < l_b^\star\big\}. \tag{II.4.38}$$

We know from Lemma II.29 that there exists $t' \geq \bar{t}$ such that $\gamma(t') \in \mathcal{P}_1 \cup \mathcal{P}_2$ and $\gamma(t'+1) \in \mathcal{C}_1 \cup \mathcal{C}_2$. We get from fact (ii) that $\ell \leq t'$.

If $\ell = t'$, then we have that $\sigma \in \mathcal{P}_1 \cup \mathcal{P}_2$ and $\sigma^x \in \mathcal{C}_1 \cup \mathcal{C}_2$.

If $\ell < t'$, then fact (iii) is violated, since $\Phi(\boxminus, \gamma(t')) < \Phi(\gamma(t'), \boxplus) = \mathrm{E}^\star_\pm$. Hence, it must be the case that $\ell = t'$. We conclude that $\mathcal{P}^\star = \mathcal{P}_1 \cup \mathcal{P}_2$ and $\mathcal{C}^\star = \mathcal{C}_1 \cup \mathcal{C}_2$.

II.4.4 Verification of (H1)

Obviously, $S_{\text{stab}} = \{\boxplus\}$, since $h_1 > h_2$. It remains to show that $S_{\text{meta}} = \{\boxminus\}$. Let $\sigma \in S$. There are four cases.

Case 1. [σ contains a cluster, which is not a stable rectangle].
Lemma II.26 implies that σ is not h_2–stable, i.e. there exists $\sigma' \in S$ such that $\mathrm{H}_\pm(\sigma') < \mathrm{H}_\pm(\sigma)$ and $\Phi(\sigma, \sigma') - \mathrm{H}_\pm(\sigma) \leq h_2 < \Gamma^\star_\pm$.

Case 2. [σ contains a cluster R, which is a stable rectangle with $P_V R \geq l_h^\star$ and $P_H R < \sqrt{|\Lambda|}$].
Let σ' be obtained from σ by attaching at the right vertical side of R a column of length $P_V R$. We start to attach on an even row on the right vertical side of R and then successively flip adjacent spins until the column is filled. Then

$$\begin{aligned} &\mathrm{H}_\pm(\sigma') \leq \mathrm{H}_\pm(\sigma) + \mu - \frac{P_V R + 1}{2}\varepsilon \leq \mathrm{H}_\pm(\sigma) + \mu - l_b^\star \varepsilon < \mathrm{H}_\pm(\sigma), \text{ and} \\ &\Phi(\sigma, \sigma') - \mathrm{H}_\pm(\sigma) \leq 2J - h_1 < \Gamma^\star_\pm. \end{aligned} \tag{II.4.39}$$

Case 3. [σ contains a cluster R, which is a stable rectangle with $P_V R \leq l_h^\star - 2$ and $P_H R < \sqrt{|\Lambda|}$].
Let σ' be obtained from σ by cutting the right column of R. Then

$$\begin{aligned} &\mathrm{H}_\pm(\sigma') = \mathrm{H}_\pm(\sigma) - \mu + \frac{P_V R + 1}{2}\varepsilon \leq \mathrm{H}_\pm(\sigma) - \mu + (l_b^\star - 1)\varepsilon < \mathrm{H}_\pm(\sigma), \text{ and} \\ &\Phi(\sigma, \sigma') - \mathrm{H}_\pm(\sigma) \leq \frac{P_V R - 1}{2}\varepsilon + h_2 < \Gamma^\star_\pm. \end{aligned} \tag{II.4.40}$$

Case 4. [σ contains a cluster R, which is a stable rectangle with $P_H R = \sqrt{|\Lambda|}$].
Let σ' be obtained from σ by attaching above R successively vertical bars of length 2 until the two rows above R wrap around the torus. Then,

$$\begin{aligned} \mathrm{H}_\pm(\sigma') &= \mathrm{H}_\pm(\sigma) + 4J - l_1\varepsilon < \mathrm{H}_\pm(\sigma), \quad \text{and} \\ \Phi(\sigma,\sigma') - \mathrm{H}_\pm(\sigma) &\leq 4J - \varepsilon < \Gamma^\star_\pm. \end{aligned} \tag{II.4.41}$$

This proves that $S_{\mathrm{meta}} = \{\boxminus\}$.

II.4.5 Computation of K

Again, we proceed as in Subsection II.2.6 and in Subsection II.3.6. Before estimating K^{-1} from below and above, we define $\bar{\mathcal{C}} = \bar{\mathcal{C}}_1 \cup \bar{\mathcal{C}}_2$, where

- $\bar{\mathcal{C}}_1$ is the set of all configurations σ that are obtained from a configuration $\sigma' \in \mathcal{C}_1$ as follows. There is a column in σ' that has length 2. σ is obtained from σ' by adding a third (+1)–spin on the even row adjacent to this column, and
- $\bar{\mathcal{C}}_2$ is the set of all configurations σ that are obtained from a configuration $\sigma' \in \mathcal{C}_2$ as follows. There is a component of three (+1)–spins above or below the $l_b^\star \times (l_h^\star - 2)$-rectangle in σ'. σ is obtained from σ' by adding a (+1)–spin such that this component becomes a 2×2-square.

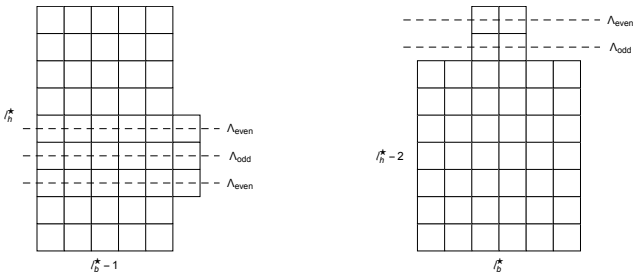


Figure II.5: The left object is an element in $\bar{\mathcal{C}}_1$ and the right object is an element in $\bar{\mathcal{C}}_2$.

It is easy to see that $\partial^+\mathcal{C}^\star \cap S^\star = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \bar{\mathcal{C}}_1 \cup \bar{\mathcal{C}}_2 = \mathcal{P}^\star \cup \bar{\mathcal{C}}$, $\mathcal{P}^\star \subset S_{\boxminus}$ and $\bar{\mathcal{C}} \subset S_{\boxplus}$.

Lower bound. Using these definitions and facts, we can estimate K^{-1} as follows.

$$\begin{aligned} \frac{1}{K} &\geq \min_{C_1,\dots,C_I \in [0,1]} \min_{\substack{h:S^\star \to [0,1] \\ h|_{S_{\boxminus}}=1, h|_{S_{\boxplus}}=0, h|_{S_i}=C_i\,\forall i}} \frac{1}{2} \sum_{\eta,\eta' \in (\mathcal{C}^\star)^+} \mathbb{1}_{\{\eta \sim \eta'\}} [h(\eta) - h(\eta')]^2 \\ &= \min_{h:\mathcal{C}^\star \to [0,1]} \sum_{\eta \in \mathcal{C}^\star} \left(\sum_{\eta' \in \mathcal{P}^\star, \eta' \sim \eta} [1 - h(\eta)]^2 + \sum_{\eta' \in \bar{\mathcal{C}}, \eta' \sim \eta} h(\eta)^2 \right) \\ &= \sum_{\eta \in \mathcal{C}_1} \frac{|\mathcal{P}^\star \sim \eta| \cdot |\bar{\mathcal{C}} \sim \eta|}{|\mathcal{P}^\star \sim \eta| + |\bar{\mathcal{C}} \sim \eta|} + \sum_{\eta \in \mathcal{C}_2} \frac{|\mathcal{P}^\star \sim \eta| \cdot |\bar{\mathcal{C}} \sim \eta|}{|\mathcal{P}^\star \sim \eta| + |\bar{\mathcal{C}} \sim \eta|}. \end{aligned} \tag{II.4.42}$$

For all $\eta \in \mathcal{C}_1$ we have that $|\mathcal{P}^\star \sim \eta| = 1$, whereas for all $\eta \in \mathcal{C}_2$ we have that $|\mathcal{P}^\star \sim \eta| = 2$. Moreover, $|\bar{\mathcal{C}} \sim \eta| = 1$ for all $\eta \in \mathcal{C}^\star$. Finally, it can be seen easily that $|\mathcal{C}_1| = |\mathcal{C}_2| = 4|\Lambda|(l_b^\star - 1)$. Hence,

$$\frac{1}{K} \geq |\mathcal{C}_1|\frac{1}{2} + |\mathcal{C}_2|\frac{2}{3} = \frac{14\,(l_b^\star - 1)}{3}|\Lambda|. \tag{II.4.43}$$

Upper bound. We say that a row or a column of a configuration is a *singleton* if it consists only of a single $(+1)$– spin. We define the following subsets of $S^\star$.

$$\begin{aligned}
\mathscr{S}^- = \{\sigma \in S^\star \mid\ & \text{for all clusters } \eta \text{ of } \sigma \text{ we have that either } (P_V R(\eta) < l_h^\star) \\
& \text{or } (P_H R(\eta) < l_b^\star) \\
& \text{or } (P_H R(\eta) \geq l_b^\star, P_V R(\eta) = l_h^\star \text{ and at least two rows of } \eta \text{ are singletons}) \\
& \text{or } (P_H R(\eta) = l_b^\star, P_V R(\eta) = l_h^\star \text{ and at least one column of } \eta \text{ is a singleton})\}, \\
\mathscr{S}_1^+ = \{\sigma \in S^\star \mid\ & \text{there exists a cluster } \eta \text{ of } \sigma \text{ such that } P_H R(\eta) = l_b^\star, P_V R(\eta) = l_h^\star \\
& \text{and no column of } \eta \text{ is a singleton} \\
& \text{and at most one row of } \eta \text{ is a singleton}\}, \\
\mathscr{S}_2^+ = \{\sigma \in S^\star \mid\ & \text{there exists a cluster } \eta \text{ of } \sigma \text{ such that } P_H R(\eta) > l_b^\star \text{ and } P_V R(\eta) = l_h^\star \\
& \text{and at most one row of } \eta \text{ is a singleton}\}, \\
\mathscr{S}_3^+ = \{\sigma \in S^\star \mid\ & \text{there exists a cluster } \eta \text{ of } \sigma \text{ such that } P_H R(\eta) = l_b^\star \text{ and } P_V R(\eta) > l_h^\star\}, \\
\mathscr{S}_4^+ = \{\sigma \in S^\star \mid\ & \text{there exists a cluster } \eta \text{ of } \sigma \text{ such that } P_H R(\eta) > l_b^\star \text{ and } P_V R(\eta) > l_h^\star\}.
\end{aligned} \tag{II.4.44}$$

Set $\mathscr{S}^+ = \mathscr{S}_1^+ \cup \mathscr{S}_2^+ \cup \mathscr{S}_3^+ \cup \mathscr{S}_4^+$. Then, $S^\star = \mathscr{S}^- \cup \mathscr{S}^+$, $\mathscr{S}^- \cap \mathscr{S}^+ = \emptyset$, $\mathcal{P}^\star \subset \mathscr{S}^-$ and $\mathcal{C}^\star \subset \mathscr{S}^+$.

Lemma II.30 *Let $\sigma \in \mathscr{S}^-$ and $\sigma' \in \mathscr{S}^+$. Then $\sigma \sim \sigma'$ if and only if $\sigma \in \mathcal{P}^\star$ and $\sigma' \in \mathcal{C}^\star$.*

Proof. In the following we show separately for all different cases that the assumption that either $\sigma \notin \mathcal{P}^\star$ or $\sigma' \notin \mathcal{C}^\star$ leads to $\sigma \notin S^\star$ or $\sigma' \notin S^\star$, which is a contradiction. Using the same arguments as in the proof of Lemma II.28, it is no restriction to assume that both σ' and σ consist of a unique cluster and that $R(\sigma)$ starts in Λ_1.

Case 1. [$\sigma' \in \mathscr{S}_1^+$].

Case 1.1. [$P_V R(\sigma) < l_h^\star$].
σ' is obtained from σ by adding a row to σ, which is a singleton. Therefore, since $\sigma' \in \mathscr{S}_1^+$, each row of σ needs to have at least two $(+1)$–spins. Now the same computations as in Case 2.1 from the proof of Lemma II.28 lead to a contradiction. Here $\gamma(\bar{t})$ is replaced by σ' and $\gamma(\bar{t}-1)$ is replaced by σ.

Case 1.2. [$P_H R(\sigma) < l_b^\star$].
σ' is obtained from σ by adding a column to σ, which is a singleton. Since no column of σ' is a singleton, $\sigma \sim \sigma'$ can not hold true, which implies that this case is not possible.

Case 1.3. [$P_H R(\sigma) \geq l_b^\star, P_V R(\sigma) = l_h^\star$ and at least two rows of σ are singletons].
σ' is obtained from σ by flipping a (-1)–spin in a row of σ that is a singleton. The same computations as in the Case 2.2 from the proof of Lemma II.28 lead to a contradiction. Here $\gamma(T+1)$ is replaced by σ' and $\gamma(T)$ is replaced by σ.

Case 1.4. [$P_H R(\sigma) = l_b^\star, P_V R(\sigma) = l_h^\star$ and at least one column of σ is a singleton].
σ' is obtained from σ by flipping a (-1)–spin in a column of σ that is a singleton. The same

computations as in the Case 1.1 from the proof of Lemma II.28 lead to a contradiction. Here $\gamma(\tau)$ is replaced by σ' and $\gamma(\tau-1)$ is replaced by σ.

Case 2. [$\sigma' \in \mathscr{S}_2^+$].

Case 2.1. [$P_V R(\sigma) < l_h^\star$].
We necessarily have that $P_H R(\sigma) = l_b^\star + m$ for some $m > 0$. σ' is obtained from σ by adding a row in Λ_1 to σ, which is a singleton. Since $\sigma' \in \mathscr{S}_2^+$, we have that the odd row below the added row contains at least two (+1)–spins. Now the same computations as in the equations (II.4.28)–(II.4.29) lead to a contradiction. Here $\gamma(\bar{t})$ is replaced by σ' and $\gamma(\bar{t}-1)$ by σ.

Case 2.2. [$P_H R(\sigma) < l_b^\star$].
It is easy to see that $\sigma \sim \sigma'$ can not hold true in this case.

Case 2.3. [$P_H R(\sigma) \geq l_b^\star, P_V R(\sigma) = l_h^\star$ and at least two rows of σ are singletons].
See Case 1.3.

Case 2.4. [$P_H R(\sigma) = l_b^\star, P_V R(\sigma) = l_h^\star$ and at least one column of σ is a singleton].
σ' is obtained from σ by adding a column to σ, which is a singleton. Hence, two columns of σ' are singletons. This implies that

$$\mathrm{H}_\pm(\sigma') \geq \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + 4J - 2h_1 = \mathrm{E}_\pm^\star + 2J - h_2 - h_1 > \mathrm{E}_\pm^\star. \tag{II.4.45}$$

Case 3. [$\sigma' \in \mathscr{S}_3^+$].

Case 3.1. [$P_V R(\sigma) < l_h^\star$].
It is easy to see that $\sigma \sim \sigma'$ can not hold true in this case.

Case 3.2. [$P_H R(\sigma) < l_b^\star$].
σ' is obtained from σ by adding a protuberance at the left vertical side or the right vertical side of $R(\sigma)$. The same computations as in the Cases 1.2, 1.3 and 1.4 from the proof of Lemma II.28 lead to a contradiction. Here $\gamma(\bar{t})$ is replaced by σ' and $\gamma(\bar{t}-1)$ by σ.

Case 3.3. [$P_H R(\sigma) \geq l_b^\star, P_V R(\sigma) = l_h^\star$ and at least two rows of σ are singletons].
σ' is obtained from σ by adding a protuberance at the top row or bottom row of $R(\sigma)$. Let $P_H R(\sigma) = l_b^\star + m$ for some $m \geq 0$. Obviously, $\mathrm{H}_\pm(\sigma) \geq \mathrm{H}_\pm((l_b^\star + m) \times (l_h^\star - 2)) + 4J - \varepsilon$. This implies that

$$\begin{aligned} \mathrm{H}_\pm(\sigma') &\geq \mathrm{H}_\pm(\sigma) + 2J - h_1 \geq \mathrm{H}_\pm((l_b^\star + m) \times (l_h^\star - 2)) + 6J - \varepsilon - h_1 \\ &= \mathrm{E}_\pm^\star + m(\mu - \varepsilon(l_b^\star - 1)) + 2J - h_2 - h_1 > \mathrm{E}_\pm^\star. \end{aligned} \tag{II.4.46}$$

Case 3.4. [$P_H R(\sigma) = l_b^\star, P_V R(\sigma) = l_h^\star$ and at least one column of σ is a singleton].
σ' is obtained from σ by adding a protuberance at the top row or bottom row of $R(\sigma)$. Hence, one column and one row of σ' are singletons. Then, as in Case 2.4 above,

$$\mathrm{H}_\pm(\sigma') \geq \mathrm{H}_\pm((l_b^\star - 1) \times l_h^\star) + 4J - 2h_1 > \mathrm{E}_\pm^\star. \tag{II.4.47}$$

Case 4. [$\sigma' \in \mathscr{S}_4^+$].

Case 4.1. [$P_V R(\sigma) < l_h^\star$].
It is easy to see that $\sigma \sim \sigma'$ can not hold true in this case.

Case 4.2. [$P_H R(\sigma) < l_b^\star$].
$\sigma \sim \sigma'$ can not hold true in this case.

Case 4.3. [$P_H R(\sigma) \geq l_b^\star, P_V R(\sigma) = l_h^\star$ and at least two rows of σ are singletons].
See Case 3.3.

Case 4.4. [$P_H R(\sigma) = l_b^\star, P_V R(\sigma) = l_h^\star$ and at least one column of σ is a singleton]. $\sigma \sim \sigma'$ can not hold true in this case. □

As in Subsection II.2.6 and in Subsection II.3.6, we have that $S_{\boxminus} \subset \mathscr{S}^-$, $S_{\boxplus} \subset \mathscr{S}^+$. Moreover, for all $i = 0, \dots, I$ either $S_i \subset \mathscr{S}^-$ or $S_i \subset \mathscr{S}^+$ holds true. Therefore, again as in Subsection II.2.6 and in Subsection II.3.6, we estimate the minimum in (II.1.16) from above by the minimum over all functions of the form (II.2.36). Using Lemma II.30 we infer that

$$\begin{aligned}
\frac{1}{K} &\leq \min_{\substack{h:S^\star\to[0,1]\\ h|_{\mathscr{S}^-}=1, h|_{\mathscr{S}^+\setminus\mathcal{C}^\star}=0}} \frac{1}{2}\sum_{\eta,\eta'\in S^\star} \mathbb{1}_{\{\eta\sim\eta'\}}[h(\eta)-h(\eta')]^2 \\
&= \min_{\substack{h:(\mathcal{C}^\star)^+\to[0,1]\\ h|_{\mathscr{S}^-\cap\partial^+\mathcal{C}^\star}=1, h|_{\mathscr{S}^+\cap\partial^+\mathcal{C}^\star}=0}} \frac{1}{2}\sum_{\eta,\eta'\in(\mathcal{C}^\star)^+} \mathbb{1}_{\{\eta\sim\eta'\}}[h(\eta)-h(\eta')]^2 \\
&= \min_{h:\mathcal{C}^\star\to[0,1]} \sum_{\eta\in\mathcal{C}^\star} \left(\sum_{\eta'\in\mathcal{P}^\star,\eta'\sim\eta} [1-h(\eta)]^2 + \sum_{\eta'\in\bar{\mathcal{C}},\eta'\sim\eta} h(\eta)^2 \right) \\
&= \frac{14\,(l_b^\star-1)}{3}|\Lambda|.
\end{aligned} \tag{II.4.48}$$

Chapter III

Gradient flow approach to local mean-field spin systems

The results of the present chapter have already been published online as the paper [13] in joint work with Anton Bovier. Some typos and minor mistakes of [13] are corrected.

Recall Section I.5, where we provide a motivation and a first formulation of the main results of this chapter. This chapter is organized as follows. First we introduce some notation that is used in this chapter. Then, in Section III.1 we introduce a *modified Wasserstein distance* and establish a *gradient flow formalism* for the partial differential equation (I.5.10) with respect to the resulting metric. In Section III.2 we use the *Fathi-Sandier-Serfaty approach* and the results of Section III.1 to prove a *large deviation principle* for the local mean-field interacting spin system, which we introduced in Subsection I.5.1. Finally, in Section III.3 we use the *Sandier-Serfaty approach* to prove a *law of large numbers* for the system in Subsection I.5.1.

Notation

In the following let $n \in \mathbb{N}$ and $(Y, \mathrm{d}), (\bar{Y}, \mathrm{e}), (Y_1, \mathrm{d}_1), \dots, (\bar{Y}_n, \mathrm{d}_n)$ be Polish spaces.

Measure theoretic notations.

- If $Y \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$, we denote by Leb_Y the Lebesgue measure restricted to Y.
- We often denote elements in $\mathbb{R}$ by θ or $\bar{\theta}$ and write $d\theta$ instead of $\mathrm{Leb}_{\mathbb{R}}$. In the same manner, for $N \in \mathbb{N}$, we often denote elements in $\mathbb{R}^N$ by $\Theta = (\theta^k)_{k=0}^{N-1}$ and write $d\Theta$ instead of $\mathrm{Leb}_{\mathbb{R}^N}$.
- Let $\mathbb{T}^d$ denote the d-dimensional unit torus. We usually denote elements in $\mathbb{T}^d$ by x or $\bar{x}$ and write dx instead of $\mathrm{Leb}_{\mathbb{T}^d}$.
- Define
$$\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times Y) := \big\{\mu \in \mathcal{M}_1(\mathbb{T}^d \times Y) \,\big|\, \mathbf{p}^1_{\#}\mu = \mathrm{Leb}_{\mathbb{T}^d}\big\}. \tag{III.0.1}$$
By the disintegration theorem (see e.g. [3, 5.3.1]), for each $\mu \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times Y)$, there exists a family $(\mu^x)_{x\in\mathbb{T}^d}$ of probability measures on Y such that $x \mapsto \mu^x$ is Borel-measurable and $\mu = \mu^x\, dx$, i.e.
$$\int_{\mathbb{T}^d \times Y} f(x,y)\, d\mu(x,y) = \int_{\mathbb{T}^d} \int_Y f(x,y)\, d\mu^x(y) dx \tag{III.0.2}$$

for all measurable and bounded $f : \mathbb{T}^d \times Y \to \mathbb{R}$.

- Let μ and ν be two measures on Y. Define the *relative entropy* between μ and ν by

$$\mathcal{H}(\mu \,|\, \nu) := \begin{cases} \int_{\mathbb{T}^d \times \mathbb{R}} \log\left(\frac{d\mu}{d\nu}\right) d\mu & : \mu \ll \nu, \\ \infty & : \text{else.} \end{cases} \tag{III.0.3}$$

  By abuse of notation we use the same letter $\mathcal{H}$ for all Polish spaces.

Wasserstein spaces.

- By abuse of notation, for all Polish spaces (Y, d), W_2 denotes the L^2-*Wasserstein distance induced by* d on $\mathcal{M}_1(Y)$, i.e.

$$W_2(\mu,\nu)^2 := \inf_{\gamma \in \mathrm{Cpl}(\mu,\nu)} \int_{Y^2} \mathrm{d}(y,y')^2 \, d\gamma(y,y'), \tag{III.0.4}$$

  where $\mu, \nu \in \mathcal{M}_1(Y)$ and $\mathrm{Cpl}(\mu,\nu)$ denotes the space of all probability measures on Y^2 that have μ and ν as marginals. We denote by $\mathrm{Opt}(\mu,\nu) \subset \mathrm{Cpl}(\mu,\nu)$ the set of all measures that realize the infimum in (III.0.4) (cf. [127, 4.1]).

- Set

$$\mathcal{P}_2(Y) := \{\mu \in \mathcal{M}_1(Y) \,|\, \exists y_0 \in Y \,:\, \int_Y d(y,y_0)^2 d\mu(y) < \infty\}. \tag{III.0.5}$$

  Then $(\mathcal{P}_2(Y), W_2)$ is a Polish space (cf. [127, 6.18]). If $Y \subset \mathbb{R}^d$, then we denote by $\mathcal{P}_2^a(Y)$ the subset of $\mathcal{P}_2(Y)$ that consists of those measures that are absolutely continuous with respect to Leb_Y.

- $\widetilde{W}$ denotes the L^2-Wasserstein distance on $\mathcal{M}_1(Y)$ induced by the distance $\tilde{d} = d/(d+1)$. Then it is known that $\widetilde{W}$ metrizes the weak topology on $\mathcal{M}_1(Y)$ (cf. [127, 6.13]).

Some maps.

- For $i \leq n$, let $\mathbf{p}^i : Y_1 \times \cdots \times Y_n \to Y_i$ denote the projection on the i-th component, i.e. $\mathbf{p}^i(y_1,\ldots,y_n) = y_i$. Whenever it is necessary, we write $\mathbf{p}^i_{Y_1\times\cdots\times Y_n}$ instead of $\mathbf{p}^i$ in order to be able to distinguish different projection maps.
- For $t > 0$, we denote by e_t the evaluation map at t, i.e. $e_t(f) = f(t)$ for all $f : (0,\infty) \to Y$.
- $\mathrm{Id}_Y : Y \to Y$ denotes the identity map on Y.

Abbreviations.

- A function is d-l.s.c. if it is lower semi-continuous with respect to d.
- For $\varphi \in C^{1,0,1}((0,T)\times\mathbb{T}^d\times\mathbb{R})$ we often write ∂_t and ∂_θ to denote the partial derivative with respect to the parameter in $(0,T)$ and $\mathbb{R}$, respectively.
- For $a \in [-\infty,\infty]$, let $a^+ := \max\{0,a\}$ and $a^- := \max\{0,-a\}$.
- We sometimes write $(y_t)_t := (y_t)_{t\in[0,T]}$ for curves $(y_t)_{t\in[0,T]} \subset Y$.

III.1 Gradient flow representation

The outline of this section is given after Theorem I.18 in Section I.5.

III.1.1 Preliminaries

In this subsection we introduce a modification of the Wasserstein space and list some of its metric properties. This space will provide the framework to derive a gradient flow representation for the system in Subsection I.5.2.

The underlying space for this representation is given by

$$\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) := \left\{ \mu \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \,\middle|\, \int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\mu(x,\theta) < \infty \right\}. \tag{III.1.1}$$

We equip $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ with the distance

$$\mathrm{W}^{\mathrm{L}}(\mu,\nu)^2 := \int_{\mathbb{T}^d} W_2(\mu^x,\nu^x)^2 \, dx, \quad \mu,\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}). \tag{III.1.2}$$

Here we have used that the map $x \mapsto W_2(\mu^x,\nu^x)$ is measurable. This is true, since, by the measurable selection lemma ([127, 5.22]), for all $\mu,\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ there exists a family $(\pi^x)_{x\in\mathbb{T}^d}$ of probability measures on $\mathbb{R}^2$ such that $x \mapsto \pi^x$ is Borel-measurable and $\pi^x \in \mathrm{Opt}(\mu^x,\nu^x)$ for almost every $x \in \mathbb{T}^d$. Defining $\pi \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R})$ by $\pi = \pi^x \, dx$, we observe that the set

$$\mathrm{Opt}^{\mathrm{L}}(\mu,\nu) := \Big\{ \pi \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}) \,\Big|\, \pi = \pi^x \, dx, \text{ where } x \mapsto \pi^x \text{ is Borel-measurable and } \pi^x \in \mathrm{Opt}(\mu^x,\nu^x) \text{ for almost every } x \in \mathbb{T}^d \Big\} \tag{III.1.3}$$

is non-empty. Note that

$$\mathrm{W}^{\mathrm{L}}(\mu,\nu)^2 = \int_{\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}} |\theta - \theta'|^2 d\pi(x,\theta,\theta') \quad \text{for all } \pi \in \mathrm{Opt}^{\mathrm{L}}(\mu,\nu). \tag{III.1.4}$$

Moreover, W^{L} can be connected more directly to an optimal transportation problem, since [3, 12.4.6] shows that for all $\mu,\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$

$$\mathrm{W}^{\mathrm{L}}(\mu,\nu)^2 = \inf_{\gamma \in \mathrm{Cpl}^{\mathrm{L}}(\mu,\nu)} \int_{\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}} |\theta - \theta'|^2 d\gamma(x,\theta,\theta'), \tag{III.1.5}$$

where

$$\mathrm{Cpl}^{\mathrm{L}}(\mu,\nu) := \Big\{ \gamma \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}) \,\Big|\, \mathbf{p}^{1,2}_{\#}\gamma = \mu, \, \mathbf{p}^{1,3}_{\#}\gamma = \nu \Big\}. \tag{III.1.6}$$

Using (III.1.5), it is easy to extend the definition of W^{L} to the whole space $\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Further, [3, 5.3.2] yields that

$$\mathrm{Cpl}^{\mathrm{L}}(\mu,\nu) = \Big\{ \gamma \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}) \,\Big|\, \gamma = \gamma^x \, dx, \text{ where } x \mapsto \gamma^x \text{ is Borel-measurable and } \gamma^x \in \mathrm{Cpl}(\mu^x,\nu^x) \text{ for almost every } x \in \mathbb{T}^d \Big\}. \tag{III.1.7}$$

This implies that $\mathrm{Opt}^{\mathrm{L}}(\mu,\nu) \subset \mathrm{Cpl}^{\mathrm{L}}(\mu,\nu)$. Therefore, it is easy to see that $\mathrm{Opt}^{\mathrm{L}}(\mu,\nu)$ is the set of minimizers in (III.1.5). From now on, we call the elements of $\mathrm{Opt}^{\mathrm{L}}(\mu,\nu)$ *L-optimal plans between μ and ν*, and the elements of $\mathrm{Cpl}^{\mathrm{L}}(\mu,\nu)$ *L-couplings of μ and ν*.

Comparison between W^{L} and W_2. Let W_2 denote the Wasserstein distance on $\mathcal{P}_2(\mathbb{T}^d \times \mathbb{R})$. Then we have

$$\mathrm{W}^{\mathrm{L}}(\mu,\nu) \geq W_2(\mu,\nu) \quad \text{for all } \mu,\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}). \tag{III.1.8}$$

Indeed, this can be shown by estimating the Wasserstein distance by the L^2-norm with respect to $(\mathbf{p}^1,\mathbf{p}^2,\mathbf{p}^1,\mathbf{p}^3)_{\#}\pi \in \mathrm{Cpl}(\mu,\nu)$, where $\pi \in \mathrm{Opt}^{\mathrm{L}}(\mu,\nu)$. However, there is no equality in general as it can be seen from the following example. Let $A := \{x \in \mathbb{T}^d \,|\, x_1 \leq \frac{1}{2}\}$ and define $\mu,\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ by

$$\begin{aligned} \mu(dx,d\theta) &:= \mathbb{1}_A(x)\delta_0(d\theta)dx + \mathbb{1}_{A^c}(x)\delta_1(d\theta)dx, \\ \nu(dx,d\theta) &:= \mathbb{1}_A(x)\delta_1(d\theta)dx + \mathbb{1}_{A^c}(x)\delta_0(d\theta)dx. \end{aligned} \tag{III.1.9}$$

Then it is easy to see that $\mathrm{W}^{\mathrm{L}}(\mu,\nu) = 1$ and $W_2(\mu,\nu) \leq \frac{1}{4}$.

The absolutely continuous case. Let us consider the special case, when the measures are absolutely continuous with respect to $\mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}$. Set

$$\mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R}) = \big\{\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \,\big|\, \mu \ll \mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}\big\}. \tag{III.1.10}$$

It is clear that, if $\mu \in \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R})$, then $\mu^x \in \mathcal{P}_2^a(\mathbb{R})$ for almost every $x \in \mathbb{T}^d$. Consequently, if $\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, then $\mathrm{Opt}(\mu^x,\nu^x) = \{(\mathrm{Id}_{\mathbb{R}}, \mathrm{T}_{\mu^x}^{\nu^x})_{\#}\mu^x\}$ for some $\mathrm{T}_{\mu^x}^{\nu^x} \in \mathrm{L}^2(\mu^x)$ for almost every $x \in \mathbb{T}^d$ (cf. [127, 10.42]). Hence, $\mathrm{Opt}^{\mathrm{L}}(\mu,\nu) = \{(\mathrm{Id}_{\mathbb{R}}, \mathrm{T}_{\mu^x}^{\nu^x})_{\#}\mu^x\, dx\}$.

Lemma III.1 *Let $\mu \in \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R})$ and $\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then there exists a unique map $\mathrm{T}_\mu^\nu \in \mathrm{L}^2(\mu)$ such that*

- *$\mathrm{T}_\mu^\nu(x,\theta) = \mathrm{T}_{\mu^x}^{\nu^x}(\theta)$ for almost every $x \in \mathbb{T}^d$,*
- *$\mathrm{W}^{\mathrm{L}}(\mu,\nu) = \|\boldsymbol{p}^2 - \mathrm{T}_\mu^\nu\|_{\mathrm{L}^2(\mu)}$.*

In the following we call T_μ^ν the L-optimal map *between μ and ν.*

Proof. Let $\pi \in \mathrm{Opt}^{\mathrm{L}}(\mu,\nu)$. Define a linear map $L : \mathrm{L}^2(\mu) \to \mathbb{R}$ by

$$L(g) := \int_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} g(x,\theta)(\theta-\theta')d\pi(x,\theta,\theta'). \tag{III.1.11}$$

Due to the monotone-class theorem and the fact that $x \mapsto \pi^x$ is Borel-measurable, the integrand is measurable. Next we apply the Cauchy-Schwartz inequality to obtain

$$|L(g)| \leq \|g\|_{\mathrm{L}^2(\pi)}\mathrm{W}^{\mathrm{L}}(\mu,\nu) = \|g\|_{\mathrm{L}^2(\mu)}\mathrm{W}^{\mathrm{L}}(\mu,\nu). \tag{III.1.12}$$

Hence, the Riesz representation theorem yields the existence of a unique element $f \in \mathrm{L}^2(\mu)$ such that $L(g) = \int fg\, d\mu$ for all $g \in \mathrm{L}^2(\mu)$. Thus

$$\begin{aligned} \int_{\mathbb{T}^d\times\mathbb{R}} fg\, d\mu = L(g) &= \int_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} g(x,\theta)(\theta-\theta')d\pi^x dx \\ &= \int_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} g(x,\theta)(\theta - \mathrm{T}_{\mu^x}^{\nu^x}(\theta))d\mu. \end{aligned} \tag{III.1.13}$$

Hence, $f(x,\theta) = \theta - \mathrm{T}_{\mu^x}^{\nu^x}(\theta)$ μ-a.e. Defining $\mathrm{T}_\mu^\nu := \mathbf{p}^2 - f$ yields the desired results. □

Stability of L-couplings and L-optimal plans. First we want to show that a sequence of L-couplings converges weakly if the corresponding sequences of marginals converge. For $\mathcal{K}, \mathcal{L} \subset \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, define

$$\mathrm{Cpl}^{\mathrm{L}}(\mathcal{K}, \mathcal{L}) := \big\{\gamma \in \mathcal{M}_1(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}) \,\big|\, \exists \mu \in \mathcal{K}, \nu \in \mathcal{L} \,:\, \gamma \in \mathrm{Cpl}^{\mathrm{L}}(\mu, \nu)\big\}. \tag{III.1.14}$$

Lemma III.2 *(i) If $\mathcal{K}$ and $\mathcal{L}$ are both tight subsets of $\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, then $\mathrm{Cpl}^{\mathrm{L}}(\mathcal{K}, \mathcal{L})$ is a tight subset of $\mathcal{M}_1(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R})$.*

(ii) If $\mathcal{K}$ and $\mathcal{L}$ are both compact with respect to the weak topology in $\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, then $\mathrm{Cpl}^{\mathrm{L}}(\mathcal{K}, \mathcal{L})$ is compact with respect to the weak topology in $\mathcal{M}_1(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R})$.

Proof. We skip this proof as it is a straightforward modification of the analogous result in the setting of the Kantorovich problem; see e.g. [127, 4.4]. □

We prove the analogous result for L-*optimal* plans only in the following special case.

Lemma III.3 *Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ and $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ be such that for all subsequences $(\mu_k)_k$, there exists a subsequence $(\mu_{k_l})_l$ and a $\mathrm{Leb}_{\mathbb{T}^d}$-null-set $\mathcal{N}_k$ such that*

$$\mu_{k_l}^x \rightharpoonup \mu^x \quad \text{for all } x \in \mathbb{T}^d \setminus \mathcal{N}_k. \tag{III.1.15}$$

Let $\pi_n \in \mathrm{Opt}^{\mathrm{L}}(\mu_n, \mu)$ for all n. Then,

$$\pi_n \rightharpoonup (\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x dx. \tag{III.1.16}$$

Proof. Let $(\mu_k)_k$ be a subsequence. From the assumptions and from the stability of optimal plans in $(\mathcal{P}_2(\mathbb{R}), W_2)$ ([127, 5.21]) and since $\mathrm{Opt}(\mu^x, \mu^x) = \{(\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x\}$, we have that

$$\pi_{k_l}^x \rightharpoonup (\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x \quad \text{for all } x \in \mathbb{T}^d \setminus \mathcal{N}_k. \tag{III.1.17}$$

Let $f \in C_b(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R})$. Then the dominated convergence theorem yields

$$\lim_{l\to\infty} \int_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} f \, d\pi_{k_l} = \int_{\mathbb{T}^d} \lim_{l\to\infty} \int_{\mathbb{R}\times\mathbb{R}} f d\pi_{k_l}^x dx = \int_{\mathbb{T}^d} \int_{\mathbb{R}\times\mathbb{R}} f \, d(\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x dx \tag{III.1.18}$$

Hence, $\pi_{k_l} \rightharpoonup (\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x dx$. And since the weak topology in $\mathcal{M}_1(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R})$ is metrizable, we infer the weak convergence of the whole sequence $(\pi_n)_n$ towards $(\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x dx$. □

Weak lower semi-continuity of W^{L}. In the following lemma we show that W^{L} is lower semi-continuous with respect to weak convergence. Recall that we have extended the definition of W^{L} to the whole space $\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$.

Lemma III.4 *Let $(\mu_n)_n, (\nu_n)_n \subset \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ and $\mu, \nu \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ be such that $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Then,*

$$\liminf_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n, \nu_n) \geq \mathrm{W}^{\mathrm{L}}(\mu, \nu). \tag{III.1.19}$$

Proof. Consider a subsequence such that $\lim_{k\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_k,\nu_k) = \liminf_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n,\nu_n)$. Let $\pi_k \in \mathrm{Opt}^{\mathrm{L}}(\mu_k,\nu_k)$ for all k. Lemma III.2 yields the existence of a subsequence $(\pi_{k_l})_l$ such that $\pi_{k_l} \rightharpoonup \pi$ for some $\pi \in \mathrm{Cpl}^{\mathrm{L}}(\mu,\nu)$. Then

$$\begin{aligned}\liminf_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n,\nu_n)^2 &= \lim_{l\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_{k_l},\nu_{k_l})^2 = \lim_{l\to\infty} \int_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} |\theta-\theta'|^2\, d\pi_{k_l} \\ &\geq \int_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} |\theta-\theta'|^2\, d\pi \;\geq\; \mathrm{W}^{\mathrm{L}}(\mu,\nu),\end{aligned} \tag{III.1.20}$$

where the first inequality is due to a standard lower semi-continuity result for integrals (see e.g. [3, 5.1.7]) and the second inequality is due to (III.1.5). □

Characterization of convergence in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$. Convergence with respect to the Wasserstein distance can be characterized by weak convergence plus convergence of the moments; see (I.2.16). A similar fact is true for convergence in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$.

Proposition III.5 *Let $(\mu_n)_n \subset \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ and $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$. Then $\lim_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n,\mu) = 0$ if and only if*

(i) $\lim_{n\to\infty} \int_{\mathbb{T}^d\times\mathbb{R}} |\theta|^2 d\mu_n = \int_{\mathbb{T}^d\times\mathbb{R}} |\theta|^2 d\mu$, *and*

(ii) for all subsequences $(\mu_k)_k$, there exists a subsequence $(\mu_{k_l})_l$ and a $\mathrm{Leb}_{\mathbb{T}^d}$-null-set $\mathcal{N}_k$ such that

$$\mu^x_{k_l} \rightharpoonup \mu^x \quad \text{for all } x \in \mathbb{T}^d \setminus \mathcal{N}_k. \tag{III.1.21}$$

Proof. Assume that $\lim_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n,\mu) = 0$. (i) is a simple consequence of the triangle inequality for W^{L}, which we prove below in Lemma III.6. Indeed,

$$\begin{aligned}\left| \left(\int_{\mathbb{T}^d\times\mathbb{R}} |\theta|^2 d\mu_n \right)^{\frac{1}{2}} - \left(\int_{\mathbb{T}^d\times\mathbb{R}} |\theta|^2 d\mu \right)^{\frac{1}{2}} \right| &= \left| \mathrm{W}^{\mathrm{L}}(\mu_n, \delta_0\otimes\mathrm{Leb}_{\mathbb{T}^d}) - \mathrm{W}^{\mathrm{L}}(\mu, \delta_0\otimes\mathrm{Leb}_{\mathbb{T}^d}) \right| \\ &\leq \mathrm{W}^{\mathrm{L}}(\mu_n,\mu) \;\longrightarrow 0.\end{aligned} \tag{III.1.22}$$

To show (ii), let $(\mu_k)_k$ be a subsequence. Note that the function $x \mapsto W_2(\mu^x_k,\mu^x)$ converges to 0 in $\mathrm{L}^2(\mathbb{T}^d)$. Hence, there exists a further subsequence $(\mu_{k_l})_l$ and a $\mathrm{Leb}_{\mathbb{T}^d}$-null-set $\mathcal{N}_k$ such that

$$\lim_{l\to\infty} W_2(\mu^x_{k_l},\mu^x) = 0 \quad \text{for all } x \in \mathbb{T}^d \setminus \mathcal{N}_k. \tag{III.1.23}$$

This yields (III.1.21), since Wasserstein convergence implies weak convergence.

Conversely, assume (i) and (ii). Let $\pi_n \in \mathrm{Opt}^{\mathrm{L}}(\mu_n,\mu)$ for all n. Lemma III.3 shows that (ii) implies

$$\pi_n \rightharpoonup (\mathrm{Id}_{\mathbb{R}}, \mathrm{Id}_{\mathbb{R}})_{\#}\mu^x dx. \tag{III.1.24}$$

It is a simple consequence of (ii), the dominated convergence theorem and the metrizability of weak convergence that

$$\mu_n \rightharpoonup \mu. \tag{III.1.25}$$

Proceeding exactly as in the Wasserstein case (see e.g. the last part of the proof of [127, 6.9]), we can show that (i), (III.1.24) and (III.1.25) imply $\lim_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n,\mu) = 0$. Again, we skip the details as there will be no new insights. □

$(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ is a Polish space. Here we show that $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ is a complete and separable metric space.

Lemma III.6 *$(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ is a metric space.*

Proof. W^{L} is well-defined on $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, since for all $\mu, \nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$

$$\begin{aligned}\mathrm{W}^{\mathrm{L}}(\mu,\nu)^2 &\leq \int_{\mathbb{T}^d} (W_2(\mu^x, \delta_0) + W_2(\delta_0, \nu^x))^2 dx \leq 4 \int_{\mathbb{T}^d} (W_2(\mu^x, \delta_0)^2 + W_2(\delta_0, \nu^x)^2) dx \\ &= 4 \int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\mu + 4 \int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\nu < \infty. \end{aligned} \tag{III.1.26}$$

W^{L} is symmetric, since the Wasserstein distance on $\mathbb{R}$ is symmetric. Let $\mu, \nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. If $\mu = \nu$, then $\mu^x = \nu^x$ for a.e. $x \in \mathbb{T}^d$ by the uniqueness claim in the disintegration theorem, and therefore $\mathrm{W}^{\mathrm{L}}(\mu, \nu) = 0$. And if $\mathrm{W}^{\mathrm{L}}(\mu, \nu) = 0$, then necessarily $W_2(\mu^x, \nu^x) = 0$ for a.e. $x \in \mathbb{T}^d$. This implies that $\mu^x = \nu^x$ for a.e. $x \in \mathbb{T}^d$, and hence $\mu = \nu$. It remains to show the triangle inequality. Let $\mu, \nu, \sigma \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then

$$\begin{aligned}\mathrm{W}^{\mathrm{L}}(\mu,\nu) &= \left(\int_{\mathbb{T}^d} W_2(\mu^x, \nu^x)^2 dx \right)^{\frac{1}{2}} \leq \left(\int_{\mathbb{T}^d} (W_2(\sigma^x, \mu^x) + W_2(\sigma^x, \nu^x))^2 dx \right)^{\frac{1}{2}} \\ &\leq \left(\int_{\mathbb{T}^d} W_2(\sigma^x, \mu^x)^2 dx \right)^{\frac{1}{2}} + \left(\int_{\mathbb{T}^d} W_2(\sigma^x, \nu^x)^2 dx \right)^{\frac{1}{2}} = \mathrm{W}^{\mathrm{L}}(\sigma,\mu) + \mathrm{W}^{\mathrm{L}}(\sigma,\nu), \end{aligned} \tag{III.1.27}$$

where we have used the triangle inequality for the Wasserstein distance and Minkowski's inequality. □

Lemma III.7 *$(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ is complete.*

Proof. Let $(\mu_n)_n$ be a Cauchy sequence in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$. Let $\varepsilon > 0$. There exists $N_\varepsilon > 0$ such that $\mathrm{W}^{\mathrm{L}}(\mu_n, \mu_m) < \varepsilon$ for all $n, m \geq N_\varepsilon$. Then if $n \geq N_e$

$$\begin{aligned}\left(\int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\mu_n \right)^{\frac{1}{2}} &\leq \mathrm{W}^{\mathrm{L}}(\mu_n, \mu_{N_\varepsilon}) + \mathrm{W}^{\mathrm{L}}(\mu_{N_\varepsilon}, \delta_0 \otimes \mathrm{Leb}_{\mathbb{T}^d}) \\ &\leq \varepsilon + \max_{i \leq N_\varepsilon} \left(\int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\mu_i \right)^{\frac{1}{2}}. \end{aligned} \tag{III.1.28}$$

Therefore,

$$\sup_{n \in \mathbb{N}} \left(\int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\mu_n \right)^{\frac{1}{2}} \leq \varepsilon + \max_{i \leq N_\varepsilon} \left(\int_{\mathbb{T}^d \times \mathbb{R}} |\theta|^2 d\mu_i \right)^{\frac{1}{2}} < \infty, \tag{III.1.29}$$

and we infer the existence of a weakly converging subsequence $(\mu_k)_k$ with limit point $\hat{\mu} \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. The weak lower semi-continuity of $\nu \mapsto \int |\theta|^2 d\nu$ and (III.1.29) imply that even $\hat{\mu} \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Finally, the weak lower semi-continuity of W^{L} yields

$$\lim_{n \to \infty} \mathrm{W}^{\mathrm{L}}(\mu_n, \hat{\mu}) \leq \lim_{n \to \infty} \liminf_{k \to \infty} \mathrm{W}^{\mathrm{L}}(\mu_n, \mu_k) = 0, \tag{III.1.30}$$

since $(\mu_n)_n$ is Cauchy. Thus $(\mu_n)_n$ is a converging sequence in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$. □

Lemma III.8 *$(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ is separable.*

Proof. To simplify the notation, we only give the proof for the case $d = 1$. Let $D \subset \mathcal{P}_2(\mathbb{R})$ be countable and dense with respect to W_2. Let for all $n \in \mathbb{N}$ and $k \leq 2^n - 1$, $A_{k,n} = [k2^{-n}, (k+1)2^{-n})$. Define

$$\mathcal{D} := \bigcup_{n\in\mathbb{N}} \bigcup_{\{\nu_k^n\}_{k=0,\dots,2^n-1}\subset D} \left\{ \sum_{k=0}^{2^n-1} \mathbb{1}_{A_{k,n}}(x)\, \nu_k^n \, dx \right\} \tag{III.1.31}$$

Then $\mathcal{D}$ is countable and $\mathcal{D} \subset \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R})$. In the following we show that $\mathcal{D}$ is dense in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$.

Define for all n, the operator $\mathrm{S}_n : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R}) \to \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R})$ by

$$\mathrm{S}_n(\mu) := \sum_{k=0}^{2^n-1} \mathbb{1}_{A_{k,n}}(x)\, S_{k,n}(\mu)\, dx, \quad \mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R}), \tag{III.1.32}$$

where for all $k \leq 2^n - 1$, $S_{k,n} : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ is the operator that sends $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R})$ to the averaged measure $S_{k,n}(\mu) = 2^n \int_{A_{k,n}} d\mu^x dx$ defined by

$$\int_{\mathbb{R}} f \, dS_{k,n}(\mu) = 2^n \int_{A_{k,n}} \int_{\mathbb{R}} f \, d\mu^x dx, \quad \text{for all measurable, bounded } f : \mathbb{R} \longrightarrow \mathbb{R}. \tag{III.1.33}$$

Let $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R})$. The proof of this lemma consists of showing the following two facts.

(i) For all $\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $\nu^n \in \mathcal{D}$ such that $\mathrm{W}^{\mathrm{L}}(\mathrm{S}_n(\mu), \nu^n) < \varepsilon$.

(ii) $\lim_{n\uparrow\infty} \mathrm{W}^{\mathrm{L}}(\mathrm{S}_n(\mu), \mu) = 0$.

Indeed, statements (i) and (ii) imply that for any μ, there exists a sequence $(\nu^n)_n \subset \mathcal{D}$ such that $\mathrm{W}^{\mathrm{L}}(\nu^n, \mu) \to 0$, that is, $\mathcal{D}$ is dense in $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T} \times \mathbb{R})$.

We now show statement (i). Since D is dense in $\mathcal{P}_2(\mathbb{R})$, there exists $\nu_{k,n} \in D$ such that $W_2(\nu_{k,n}, S_{k,n}(\mu)) < \varepsilon$ for all $k \leq 2^n - 1$. Set $\nu^n = \sum_{k=0}^{2^n-1} \mathbb{1}_{A_{k,n}}(x)\nu_{k,n}\, dx$. We immediately observe that $\mathrm{W}^{\mathrm{L}}(\mathrm{S}_n(\mu), \nu^n) < \varepsilon$.

Next we prove (ii). In view of Proposition III.5, it is enough to show that

(A) $\int |\theta|^2 d\mathrm{S}_n(\mu) = \int |\theta|^2 d\mu$ for all n, and

(B) $\mathrm{S}_n(\mu)^x \rightharpoonup \mu^x$ for almost every $x \in \mathbb{T}$.

(A) is a simple consequence of (III.1.33). It remains to show (B), which is done in six steps. The main problem is to avoid the non-separability of the space $C_b(\mathbb{R})$. We do this in a standard way, which was done e.g. in the proof of [54, 11.4.1]. This means, we push the measures down from $\mathbb{T} \times \mathbb{R}$ to a bounded set. Consider $h(\theta) = \arctan(\theta)$ and abbreviate $O := (\pi/2, \pi/2)$. Set $\sigma = (\mathbf{p}^1, h)_{\#}\mu$. Consequently, σ is supported in $\mathbb{T} \times O$. Let $\mathrm{BL}(O)$ be the set of real-valued bounded Lipschitz functions on O.

Step 1. $\big[\forall f \in \mathrm{BL}(O)\, \exists \,\text{null-set}\, \mathcal{N}^f : \int f \, d\mathrm{S}_n(\sigma)^x \to \int f \, d\sigma^x \quad \forall x \in \mathbb{T} \setminus \mathcal{N}^f.\big]$
Let $\mathbb{T} \setminus \mathcal{N}^f$ be the set of Lebesgue-points of $x \mapsto \int_{\mathbb{R}} f d\sigma^x \in \mathrm{L}^1(\mathbb{T})$. For each $x \in \mathbb{T}$, let $k_x(n) = \lfloor x2^n \rfloor$. Hence, $x \in A_{k_x(n),n}$ for each n. Denote by $B(x, 2^{-n})$ the ball of radius 2^{-n} around $x \in \mathbb{T}$. Then we observe that for each $x \in \mathbb{T} \setminus \mathcal{N}^f$

$$\begin{aligned}\left|\int_O f d\mathrm{S}_n(\sigma)^x - \int_O f d\sigma^x\right| &= \left|\int_O f dS_{k_x(n),n}(\sigma) - \int_O f d\sigma^x\right| \\ &\leq 2^n \int_{A_{k_x(n),n}} \left|\int_O f d\sigma^y - \int_O f d\sigma^x\right| dy \\ &\leq \frac{2}{\mathrm{Leb}_{\mathbb{T}}(B(x,2^{-n}))} \int_{B(x,2^{-n})} \left|\int_O f d\sigma^y - \int_O f d\sigma^x\right| dy \\ &\longrightarrow 0 \quad \text{as} \quad n \to \infty,\end{aligned} \tag{III.1.34}$$

since x is a Lebesgue point.

Step 2. $\big[$*Let $\iota : O \to \bar{O}$ be the canonical inclusion, then*
$\forall \bar{f} \in \mathrm{BL}(\bar{O}) \, \exists$ null-set $\mathcal{N}^{\bar{f}} : \int \bar{f} \, d\iota_\# \mathrm{S}_n(\sigma)^x \to \int f \, d\iota_\# \sigma^x \quad \forall x \in \mathbb{T} \setminus \mathcal{N}^{\bar{f}}.\big]$
$\bar{f}$ has the representation

$$\bar{f}(\theta) = \inf_{\vartheta \in \bar{O}} \bar{f}(\vartheta) + \mathrm{Lip}(\bar{f}) \, |\theta - \vartheta| = \inf_{\vartheta \in O} \bar{f}(\vartheta) + \mathrm{Lip}(\bar{f}) \, |\theta - \vartheta|, \tag{III.1.35}$$

where $\mathrm{Lip}(\bar{f})$ is the Lipschitz-constant of $\bar{f}$. Define $f \in \mathrm{BL}(O)$ by $f(\theta) := \inf_{\vartheta \in O} \bar{f}(\vartheta) + \mathrm{Lip}(\bar{f}) \, |\theta - \vartheta|$. Then $\bar{f} = f$ on O. Set $\mathcal{N}^{\bar{f}} := \mathcal{N}^f$, where $\mathcal{N}^f$ is the null-set from Step 1. Then, since $\iota_\# \mathrm{S}_n(\sigma)$ and $\iota_\# \sigma$ are supported on O, we obtain that for all $x \in \mathbb{T} \setminus \mathcal{N}^{\bar{f}}$

$$\begin{aligned}\lim_{n\to\infty} \int_{\bar{O}} \bar{f} \, d\iota_\# \mathrm{S}_n(\sigma)^x &= \lim_{n\to\infty} \int_O f \, d\iota_\# \mathrm{S}_n(\sigma)^x = \lim_{n\to\infty} \int_O f \, d\mathrm{S}_n(\sigma)^x \\ &= \int_O f \, d\sigma^x = \int_{\bar{O}} \bar{f} \, d\iota_\# \sigma^x.\end{aligned} \tag{III.1.36}$$

Step 3. $\big[\exists$ null-set $\mathcal{N} : \int \bar{f} \, d\iota_\# \mathrm{S}_n(\sigma)^x \to \int f \, d\iota_\# \sigma^x \quad \forall x \in \mathbb{T} \setminus \mathcal{N} \quad \forall \bar{f} \in \mathrm{BL}(\bar{O}).\big]$
$\mathrm{BL}(\bar{O})$ is separable, i.e. there exists a countable set $E \subset \mathrm{BL}(\bar{O})$, which is dense with respect to $\|\cdot\|_\infty$. Set $\mathcal{N} := \cup_{\bar{f} \in E} \mathcal{N}_k^{\bar{f}}$. Since E is dense in $\mathrm{BL}(\bar{O})$, this concludes the claim.

Step 4. $\big[\exists$ null-set $\mathcal{N} : \int f \, d\mathrm{S}_n(\sigma)^x \to \int f \, d\sigma^x \quad \forall x \in \mathbb{T} \setminus \mathcal{N} \quad \forall f \in \mathrm{BL}(O).\big]$
Using [54, 6.1.1] we know that there exists $\bar{f} \in \mathrm{BL}(\bar{O})$ such that $\bar{f} = f$ on O. Now the claim follows immediately from Step 3.

Step 5. $\big[\exists$ null-set $\mathcal{N} : \int f \, d\mathrm{S}_n(\sigma)^x \to \int f \, d\sigma^x \quad \forall x \in \mathbb{T} \setminus \mathcal{N} \quad \forall f \in C_b(O).\big]$
The claim follows from Step 4 and [54, 11.3.3].

Step 6. $\big[\exists$ null-set $\mathcal{N} : \int f \, d\mathrm{S}_n(\mu)^x \to \int f \, d\mu^x \quad \forall x \in \mathbb{T} \setminus \mathcal{N} \quad \forall f \in C_b(\mathbb{R}).\big]$
Note that $\mathrm{S}_n(\sigma)^x = (h^{-1})_\# \mathrm{S}_n(\mu)^x$ and $\sigma^x = (h^{-1})_\# \mu^x$ for all $x \in \mathbb{T} \setminus \mathcal{N}$. Hence, the claim follows from the continuous mapping theorem (see e.g. [3, 5.2.1]). This concludes the proof. □

III.1.2 Curves in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$

In this subsection we analyse *geodesics* and *absolutely continuous curves* in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$. For the latter we show that these curves are characterized by weak solutions of some type of *continuity equation* and we introduce a notion of *tangent velocity* at these curves (cf. Lemma I.3). This fact is the key ingredient later to represent weak solutions of (I.5.10) as gradient flows in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ (see Theorem III.41).

Geodesics. Let $T \in (0,\infty)$. A curve $(\mu_t)_{t\in[0,T]}$ in a metric space (X,d) is called *geodesic* (between μ_0 and μ_T) if $\mathrm{d}(\mu_s,\mu_t) = (t-s)/T$ for all $0 \le s \le t \le T$. In the following we show that $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$ is a geodesic space, i.e. between each pair of measures there exists a geodesic.

Proposition III.9 *Let $\mu_0,\mu_T \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$. Let $\pi \in \mathrm{Opt}^{\mathrm{L}}(\mu_0,\mu_T)$. Define the curve $(\mu_t)_{t\in[0,T]}$ by*

$$\mu_t := (\boldsymbol{p}^1_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}}\,,\,(1-t)\boldsymbol{p}^2_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}} + t\boldsymbol{p}^3_{\mathbb{T}^d\times\mathbb{R}\times\mathbb{R}})_{\#}\pi, \quad t\in[0,T]. \tag{III.1.37}$$

Then $(\mu_t)_t$ is a geodesic. Moreover, if in addition $\mu_0 \ll \mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}$, then $\mu_t = (\boldsymbol{p}^1_{\mathbb{T}^d\times\mathbb{R}}, (1-t)\boldsymbol{p}^2_{\mathbb{T}^d\times\mathbb{R}} + t\,\mathrm{T}_{\mu_0}^{\mu_T})_{\#}\mu_0$ and we also have that $\mu_t \ll \mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}$ for all $t\in(0,T)$.

Proof. Note that for each t, the disintegration of μ_t with respect to $\mathrm{Leb}_{\mathbb{T}^d}$ is given by $\mu_t^x = ((1-t)\mathbf{p}^1_{\mathbb{R}\times\mathbb{R}} + t\mathbf{p}^2_{\mathbb{R}\times\mathbb{R}})_{\#}\pi^x$ for almost every $x\in\mathbb{T}^d$. Hence, we know that $(\mu_t^x)_t$ is a geodesic in $(\mathcal{P}_2(\mathbb{R}), W_2)$ for almost every x (see e.g. [3, 7.2.2]). We infer

$$\mathrm{W}^{\mathrm{L}}(\mu_s,\mu_t)^2 = \frac{(t-s)^2}{T^2}\int_{\mathbb{T}^d} W_2(\mu_0^x,\mu_T^x)^2 dx = \frac{(t-s)^2}{T^2}\mathrm{W}^{\mathrm{L}}(\mu_0,\mu_T)^2. \tag{III.1.38}$$

The second claim follows from the observation that $\pi = (\mathbf{p}^1_{\mathbb{T}^d\times\mathbb{R}}, \mathbf{p}^2_{\mathbb{T}^d\times\mathbb{R}}, \mathrm{T}_{\mu_0}^{\mu_T})_{\#}\mu_0$. The third claim follows from the analogue statement in the Wasserstein space (see e.g. [4, 2.4]). □

Absolutely continuous curves. Let I be an open and bounded (or unbounded) interval. A curve $(\mu_t)_{t\in\mathrm{I}}$ in a metric space (X,d) is called *absolutely continuous* and we write $(\mu_t)_t \in \mathcal{AC}(\mathrm{I};X)$ if there exists $m\in\mathrm{L}^2(\mathrm{I})$ (or $m\in\mathrm{L}^2_{\mathrm{loc}}(\mathrm{I})$ if I is unbounded) such that

$$\mathrm{d}(\mu_s,\mu_t) \le \int_s^t m(r)dr \qquad \forall\, s,t\in\mathrm{I}, s\le t. \tag{III.1.39}$$

If (X,d) is a Polish space, [3, 1.1.2] yields the existence of the *metric derivative* $|\mu'|\in\mathrm{L}^2(\mathrm{I})$ (or $|\mu'|\in\mathrm{L}^2_{\mathrm{loc}}(\mathrm{I})$ if I is unbounded) defined by

$$|\mu'|(t) = \lim_{s\to t}\frac{\mathrm{d}(\mu_s,\mu_t)}{|s-t|} \qquad \text{for almost every } t\in\mathrm{I}. \tag{III.1.40}$$

In the following we analyse absolutely continuous curves in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$ and show that some analogous results as in Wasserstein spaces (cf. Lemma I.3) hold true.

Proposition III.10 ***(A)*** *Let $T\in(0,\infty)$ (or $T=\infty$) and $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$. Then there exists $v : (0,T)\times\mathbb{T}^d\times\mathbb{R}\to\mathbb{R}$ jointly measurable such that*

(i) $\partial_t\mu_t + \partial_\theta(\mu_t\, v) = 0$ in $(0,T)\times\mathbb{T}^d\times\mathbb{R}$ in the sense of distributions, i.e. for all $\varphi\in C_c^\infty((0,T)\times\mathbb{T}^d\times\mathbb{R})$,

$$\int_{(0,T)\times\mathbb{T}^d\times\mathbb{R}} \Big(\partial_t\varphi_t(x,\theta) + \partial_\theta\varphi_t(x,\theta)\, v_t^x(\theta)\Big) d\mu_t(x,\theta)dt = 0, \tag{III.1.41}$$

(ii) $\|v_t\|_{\mathrm{L}^2(\mu_t)} \le |\mu'|(t)$ for almost every t,

(iii) $v_t \in \overline{\{\partial_\theta\varphi \,|\, \varphi \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})\}}^{\mathrm{L}^2(\mu_t)}$ *for almost every* t,

(iv) $v_t^x \in \overline{\{\varphi' \,|\, \varphi \in C_c^\infty(\mathbb{R})\}}^{\mathrm{L}^2(\mu_t^x)}$ *for almost every* t *and* x,

(B) *Conversely, let* $(\mu_t)_{t\in(0,T)} \subset \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ *and let* $v \in \mathrm{L}^2((0,T) \times \mathbb{T}^d \times \mathbb{R}\,;\, \mu_t dt)$ *(or* $t \mapsto \|v_t\|_{\mathrm{L}^2(\mu_t)} \in \mathrm{L}^2_{\mathrm{loc}}((0,T))$ *if* $T = \infty$*). Suppose that* (III.1.41) *holds. Then* $(\mu_t)_t \in \mathcal{AC}((0,T); \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ *and* $\|v_t\|_{\mathrm{L}^2(\mu_t)} \geq |\mu'|(t)$ *for almost every* t.

Proof. Without restriction we can assume that $T < \infty$, since otherwise we can exhaust $(0,\infty)$ with bounded intervals.

We now show ***(A)***. We proceed analogously to the proof of [3, 8.3.1]. Let $\mathcal{T} = \{\partial_\theta\varphi \,|\, \varphi \in C_c^\infty((0,T) \times \mathbb{T}^d \times \mathbb{R})\}$. Define a linear map $L : \mathcal{T} \to \mathbb{R}$ by

$$L(\partial_\theta\varphi) := \int_{(0,T)\times\mathbb{T}^d\times\mathbb{R}} \partial_t\varphi \, d\mu_t \, dt. \tag{III.1.42}$$

Performing the very same steps as in the proof of [3, 8.3.1], we obtain

$$|L(\partial_\theta\varphi)| \leq \| \, |\mu'| \, \|_{\mathrm{L}^2((0,T))} \, \|\partial_\theta\varphi\|_{\mathrm{L}^2((0,T)\times\mathbb{T}^d\times\mathbb{R}\,;\,\mu_t dt)}, \tag{III.1.43}$$

which resembles equation (8.3.10) in [3]. Note that we have tacitly used Lemma III.3. Let $\overline{\mathcal{T}}$ denote the closure of $\mathcal{T}$ with respect to $\|\cdot\|_{\mathrm{L}^2((0,T)\times\mathbb{T}^d\times\mathbb{R}\,;\,\mu_t dt)}$. Then, using the Riesz representation theorem, (III.1.43) implies that there exists a unique $v \in \overline{\mathcal{T}}$ such that

$$L(w) = \int_{(0,T)\times\mathbb{T}^d\times\mathbb{R}} v\, w \, d\mu_t \, dt \quad \forall\, w \in \overline{\mathcal{T}}. \tag{III.1.44}$$

In particular, since we can take $w = \partial_\theta\varphi$ for $\varphi \in C_c^\infty((0,T) \times \mathbb{T}^d \times \mathbb{R})$, (III.1.44) yields *(i)*. Again, using the same arguments as in [3, 8.3.1], we obtain that for all intervals $J \subset (0,T)$

$$\int_J \|v_t\|^2_{\mathrm{L}^2(\mu_t)} dt \leq \int_J |\mu'|^2(t) dt, \tag{III.1.45}$$

which is equation (8.3.13) in [3]. As J was arbitrary, this implies *(ii)*. To show *(iii)*, take $(\varphi_n)_n \subset C_c^\infty((0,T) \times \mathbb{T}^d \times \mathbb{R})$ such that $\partial_\theta\varphi_n \to v$ in $\mathrm{L}^2((0,T) \times \mathbb{T}^d \times \mathbb{R}\,;\, \mu_t dt)$. Hence, the function $t \mapsto \|\partial_\theta\varphi_n(t,\cdot) - v_t\|_{\mathrm{L}^2(\mathbb{T}^d\times\mathbb{R}\,;\,\mu_t)}$ converges to 0 in $\mathrm{L}^2((0,T)\,;\,dt)$. This yields that, up to subsequences, $t \mapsto \|\partial_\theta\varphi_n(t,\cdot) - v_t\|_{\mathrm{L}^2(\mathbb{T}^d\times\mathbb{R}\,;\,\mu_t)}$ converges to 0 point-wise almost everywhere. Since $\varphi_n(t,\cdot) \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})$ for all t, we conclude the proof of *(iii)*. In the same way, one proves the claim *(iv)*.

Next we prove ***(B)***. Let $D \subset C_c^\infty((0,T)\times\mathbb{R})$ be countable and dense with respect to $\|\cdot\|_\infty$. Let $\varphi \in D$. Then (III.1.41) implies that

$$\int_{\mathbb{T}^d} \zeta(x) \int_{(0,T)\times\mathbb{R}} (\partial_t\varphi + \partial_\theta\varphi \, v^x) d\mu_t^x \, dt \, dx = 0 \quad \forall\, \zeta \in C_c^\infty(\mathbb{T}^d). \tag{III.1.46}$$

Hence, there exists a $\mathrm{Leb}_{\mathbb{T}^d}$-null-set $\mathcal{N}^\varphi$ such that

$$\int_{(0,T)\times\mathbb{R}} (\partial_t\varphi + \partial_\theta\varphi \, v^x) d\mu_t^x \, dt = 0 \quad \forall\, x \in \mathbb{T}^d \setminus \mathcal{N}^\varphi. \tag{III.1.47}$$

Set $\mathcal{N}' = \cup_{\varphi\in D}\mathcal{N}^\varphi$. Moreover, the assumption that $t \mapsto \|v_t\|_{\mathrm{L}^2(\mu_t)} \in \mathrm{L}^2((0,T))$ assures that there exists a further null-set $\mathcal{N}''$ such that

$$\int_{(0,T)\times\mathbb{R}} |v^x|^2 d\mu_t^x \, dt < \infty \quad \forall\, x \in \mathbb{T}^d \setminus \mathcal{N}''. \tag{III.1.48}$$

Using that D is dense, the dominated convergence theorem yields that

$$\int_{(0,T)\times\mathbb{R}} (\partial_t\varphi + \partial_\theta\varphi\, v^x) d\mu_t^x \, dt = 0 \quad \forall\, \varphi \in C_c^\infty((0,T)\times\mathbb{R})\ \forall\, x \in \mathbb{T}^d \setminus (\mathcal{N}' \cup \mathcal{N}''). \tag{III.1.49}$$

Therefore, for each $x \in \mathbb{T}^d \setminus (\mathcal{N}' \cup \mathcal{N}'')$, the pair $\big((\mu_t^x)_t, (v_t^x)_t\big)$ fulfils the assumptions of the converse implication of [59, 2.5]. In particular, we obtain

$$W_2(\mu_s^x, \mu_t^x)^2 \le (t-s) \int_s^t \|v_r^x\|^2_{\mathrm{L}^2(\mu_r^x)} \, dr \quad \forall\, 0 < s \le t < T \quad \forall\, x \in \mathbb{T}^d \setminus (\mathcal{N}' \cup \mathcal{N}''). \tag{III.1.50}$$

This inequality was shown at the end of the proof of [59, 2.5]. (III.1.50) easily implies that for all $0 < s \le t < T$

$$\mathrm{W}^{\mathrm{L}}(\mu_s, \mu_t)^2 \le (t-s) \int_s^t \|v_r\|^2_{\mathrm{L}^2(\mu_r)} \, dr \le \left(\int_s^t \max\{1, \|v_r\|^2_{\mathrm{L}^2(\mu_r)}\} \, dr \right)^2. \tag{III.1.51}$$

We infer that $(\mu_t)_{t\in(0,T)}$ is an absolutely continuous curve in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}), \mathrm{W}^{\mathrm{L}})$. Finally, the first inequality in (III.1.51) shows that $\|v_t\|_{\mathrm{L}^2(\mu_t)} \ge |\mu'|(t)$ almost everywhere. □

The previous result introduced a few important objects that have to emphasized.

Definition III.11 *Let $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$, $T \in (0,\infty)$ (or $T = \infty$) and suppose that $(\mu_t)_t \in \mathcal{AC}((0,T); \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$. Define*

(i) $\mathrm{Tan}_\mu \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}) := \overline{\{\partial_\theta\varphi \,|\, \varphi \in C_c^\infty(\mathbb{T}^d\times\mathbb{R})\}}^{\mathrm{L}^2(\mu)}$, *the* tangent space *at μ,*

(ii) $\mathrm{Tan}_{\mu^x} \mathcal{P}_2(\mathbb{R}) := \overline{\{\varphi' \,|\, \varphi \in C_c^\infty(\mathbb{R})\}}^{\mathrm{L}^2(\mu^x)}$ *for $x \in \mathbb{T}^d$,*

(iii) $v : (0,T)\times\mathbb{T}^d\times\mathbb{R} \to \mathbb{R}$ *is called* tangent velocity *for $(\mu_t)_t$ if*

- $v \in \mathrm{L}^2((0,T)\times\mathbb{T}^d\times\mathbb{R}\,;\, \mu_t dt)$ *(or $t \mapsto \|v_t\|_{\mathrm{L}^2(\mu_t)} \in \mathrm{L}^2_{\mathrm{loc}}((0,T))$ if $T = \infty$),*
- $\partial_t\mu_t + \partial_\theta(\mu_t\, v) = 0$ *in $(0,T)\times\mathbb{T}^d\times\mathbb{R}$ in the sense of distributions,*
- $v_t \in \mathrm{Tan}_{\mu_t}\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ *for almost every t.*

The following lemma is an easy consequence of the above definition and can be proven exactly as in [3, Chapter 8.4].

Lemma III.12 *(i)* $\mathrm{Tan}_\mu \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}) = \{w \in \mathrm{L}^2(\mu_t) \,|\, \partial_\theta(w\mu) = 0\}^\perp$ *for $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$, where ∂_θ is meant in the sense of distributions.*

(ii) $v \in \mathrm{Tan}_\mu \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ *if and only if* $\|v\|_{L^2(\mu)} = \inf\{\|v+w\|_{L^2(\mu)} \,|\, w \in L^2(\mu), \partial_\theta(w\mu) = 0\}$.

(iii) Let $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$, $v \in \mathrm{Tan}_\mu \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ and $w \in \mathrm{L}^2(\mu)$ be such that $\partial_\theta(w\mu) = 0$. Then $\|v\|_{L^2(\mu)} = \|v+w\|_{L^2(\mu)}$ if and only if $\|w\|_{L^2(\mu)} = 0$.

We can summarize the previous results in the following statement.

Corollary III.13 *Let $T \in (0,\infty]$. $(\mu_t)_{t\in(0,T)}$ is absolutely continuous in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}),\mathrm{W}^{\mathrm{L}})$ if and only if there exists a tangent velocity v for $(\mu_t)_t$. Moreover, $\|v_t\|_{\mathrm{L}^2(\mu_t)} = |\mu'|(t)$ for almost every t and v is uniquely determined* $\mathrm{Leb}_{(0,T)}$*-a.e.*

Proof. Obviously, Proposition III.10 shows each claim except of the uniqueness result. Let w be an other tangent velocity for $(\mu_t)_t$. Note that $\partial_\theta((w_t - v_t)\mu_t) = 0$ for almost every t. Therefore, Lemma III.12 (ii) implies that $\|w_t\|_{L^2(\mu_t)} \le \|w_t + (v_t - w_t)\|_{L^2(\mu_t)} = \|v_t\|_{L^2(\mu_t)}$ for almost every t. Analogously, applying Lemma III.12 (ii) for v shows that $\|v_t\|_{L^2(\mu_t)} = \|w_t\|_{L^2(\mu_t)} = \|v_t + (w_t - v_t)\|_{L^2(\mu_t)}$ for almost every t. Using Lemma III.12 (iii), this yields that $\|w_t - v_t\|_{L^2(\mu_t)} = 0$ for almost every t. □

L-optimal maps vs. $\mathrm{Tan}_\mu\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$. In the following we show that, if $\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ and $\mu \in \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d\times\mathbb{R})$, then $\mathrm{T}_\mu^\nu - \mathbf{p}^2 \in \mathrm{Tan}_\mu\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$. This will be a consequence of the following observation.

Lemma III.14 *Let $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ and $w \in L^2(\mu)$. Then $w \in \mathrm{Tan}_\mu\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})^\perp$ if and only if $w(x,\cdot) \in \mathrm{Tan}_{\mu^x}\mathcal{P}_2(\mathbb{R})^\perp$ for almost every $x \in \mathbb{T}^d$.*

Proof. The proof relies on Lemma III.12 (i). Note that the same statements as in Lemma III.12 also hold for $\mathrm{Tan}_{\mu^x}\mathcal{P}_2(\mathbb{R})$ ([3, Chapter 8.4]). Therefore, the "if"-part is trivial. To show the "only if"-part, we apply the same arguments as in the proof of Proposition III.10 *(B)* to obtain a $\mathrm{Leb}_{\mathbb{T}^d}$−null-set $\mathcal{N}$ such that for all $x \in \mathbb{T}^d \setminus \mathcal{N}$

$$\int_{\mathbb{R}} \varphi'\, w(x,\cdot)\, d\mu^x = 0 \quad \forall\, \varphi \in C_c^\infty(\mathbb{R}). \tag{III.1.52}$$

We conclude that $w(x,\cdot) \in \{w \in \mathrm{L}^2(\mu_t) \,|\, \partial_\theta(w\mu) = 0\} = \mathrm{Tan}_{\mu^x}\mathcal{P}_2(\mathbb{R})^\perp$ for almost every $x \in \mathbb{T}^d$. □

Corollary III.15 *Let $\mu \in \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d\times\mathbb{R})$ and $\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$. Then $\mathrm{T}_\mu^\nu - \boldsymbol{p}^2 \in \mathrm{Tan}_\mu\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$.*

Proof. It is enough to show that for all $w \in \mathrm{Tan}_\mu\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})^\perp$

$$\int_{\mathbb{T}^d\times\mathbb{R}} (\mathrm{T}_\mu^\nu - \mathbf{p}^2)\, w\, d\mu = 0. \tag{III.1.53}$$

[3, 8.5.2] states that $\mathrm{T}_\mu^\nu(x,\cdot) - \mathbf{p}^2 = \mathrm{T}_{\mu^x}^{\nu^x} - \mathrm{Id}_{\mathbb{R}} \in \mathrm{Tan}_{\mu^x}\mathcal{P}_2(\mathbb{R})$ for almost every $x \in \mathbb{T}^d$. Therefore, Lemma III.14 implies that $\int_{\mathbb{R}}(\mathrm{T}_\mu^\nu - \mathbf{p}^2)(x,\cdot)w(x,\cdot)d\mu^x = 0$ for almost every x, which immediately implies (III.1.53). □

$\mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$ vs. $\mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}))$. Here we show that a curve $(\mu_t)_t$ is absolutely continuous in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}),\mathrm{W}^{\mathrm{L}})$ if and only if $(\mu_t^x)_t$ is absolutely continuous in $(\mathcal{P}_2(\mathbb{R}),W_2)$ for almost every x.

Lemma III.16 *Let $T \in (0,\infty)$ or $T = \infty$, $(\mu_t)_{t\in(0,T)} \subset \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ and $v \in \mathrm{L}^2((0,T) \times \mathbb{T}^d \times \mathbb{R}\,;\, \mu_t dt)$ (or $t \mapsto \|v_t\|_{\mathrm{L}^2(\mu_t)} \in \mathrm{L}^2_{\mathrm{loc}}((0,T))$ if $T = \infty$). Then $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ and v is the tangent velocity for $(\mu_t)_t$ if and only if for almost every $x \in \mathbb{T}^d$, $(\mu_t^x)_t \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}))$ and v^x is the tangent velocity for $(\mu_t^x)_t$ in the Wasserstein sense, i.e. there exists a $\mathrm{Leb}_{(0,T)}$-null-set $\mathcal{N}_x$ such that*

(i) $v^x \in \mathrm{L}^2((0,T) \times \mathbb{R}\,;\, \mu_t^x dt)$ *(or $t \mapsto \|v_t^x\|_{\mathrm{L}^2(\mu_t^x)} \in \mathrm{L}^2_{\mathrm{loc}}((0,T))$ if $T = \infty$),*

(ii) $\partial_t \mu_t^x + \partial_\theta(\mu_t^x\, v^x) = 0$ *in $(0,T) \times \mathbb{R}$ in the sense of distributions,*

(iii) $v_t^x \in \mathrm{Tan}_{\mu_t^x}\mathcal{P}_2(\mathbb{R})$ *for all $t \in (0,T) \setminus \mathcal{N}_x$.*

In particular, $|\mu'|^2(t) = \|v_t\|^2_{\mathrm{L}^2(\mu_t)} = \int_{\mathbb{T}^d} \|v_t^x\|^2_{\mathrm{L}^2(\mu_t^x)} dx = \int_{\mathbb{T}^d} |(\mu^x)'|^2(t)\, dx$ for almost every t.

Proof. Assume $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ with tangent velocity v. *(i)* follows from the corresponding integrability condition on v being the tangent velocity of $(\mu_t)_t$. *(ii)* was shown in the proof of Proposition III.10 *(B)*. *(iii)* follows from Proposition III.10 *(A)*. By [59, 2.5], these facts imply that

$$(\mu_t^x)_t \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R})) \quad \text{and} \quad \|v_t^x\|_{\mathrm{L}^2(\mu_t^x)} = |(\mu^x)'|(t) \ \text{ for almost every } t. \tag{III.1.54}$$

Conversely, it is an easy observation that (ii) implies that $\partial_t \mu_t + \partial_\theta(\mu_t\, v) = 0$ in the sense of distributions. Hence, Proposition III.10 *(B)* yields that $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ and $\|v_t\|_{\mathrm{L}^2(\mu_t)} \geq |\mu'|(t)$ for almost every t. It remains to show that v is the tangent velocity for $(\mu_t)_t$. An easy application of Fubini's theorem shows that *(iii)* can be reformulated as follows: For almost every t, there exists a $\mathrm{Leb}_{\mathbb{T}^d}$-null-set $\mathcal{N}_t$ such that $v_t^x \in \mathrm{Tan}_{\mu_t^x}\mathcal{P}_2(\mathbb{R})$ for all $x \in \mathbb{T}^d \setminus \mathcal{N}_t$. Using this formulation, we can argue in the same way as in Corollary III.15 to conclude that $v_t \in \mathrm{Tan}_{\mu_t}\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ for a.e. t, which shows that v is the tangent velocity for $(\mu_t)_t$. □

Infinitesimal behaviour. The goal of this paragraph is to show differentiability of W^{L} along absolutely continuous curves. We start with the following observation, which, again, is also true in the analogue setting of the Wasserstein distance.

Lemma III.17 *Let $T \in (0,\infty]$ and $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ with tangent velocity v. Suppose that $(\mu_t)_t \subset \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R})$. Then*

$$\lim_{h\to 0} \left\| \frac{1}{h}(\mathrm{T}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{p}^2) - v_t \right\|_{\mathrm{L}^2(\mu_t)} = 0 \quad \textit{for almost every } t \in (0,T). \tag{III.1.55}$$

Proof. Let $s_h^t := \frac{1}{h}(\mathrm{T}_{\mu_t}^{\mu_{t+h}} - \mathbf{p}^2)$. By Lemma III.16, we have that $(\mu_t^x)_t \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}))$ with Wasserstein tangent velocity v^x for almost every x. Therefore, we can apply [3, 8.4.6] to see that for almost every x there exists a null-set $\mathcal{N}_x$ such that

$$\lim_{h\to 0} \left\| s_h^t(x,\cdot) - v_t^x \right\|_{\mathrm{L}^2(\mu_t^x)} = 0 \quad \text{for all } t \in (0,T) \setminus \mathcal{N}_x. \tag{III.1.56}$$

As above, using Fubini's theorem, we can reformulate (III.1.56) in such a way that for almost every t, there exists a null-set $\mathcal{N}_t$ such that

$$\lim_{h\to 0} \left\| s_h^t(x,\cdot) - v_t^x \right\|_{\mathrm{L}^2(\mu_t^x)} = 0 \quad \text{for all } x \in \mathbb{T}^d \setminus \mathcal{N}_t. \tag{III.1.57}$$

In particular, this shows that for almost every t, $x \mapsto \|s_h^t(x,\cdot)\|^2_{\mathrm{L}^2(\mu_t^x)}$ converges to $x \mapsto \|v_t^x\|^2_{\mathrm{L}^2(\mu_t^x)}$ point-wise almost everywhere. However, since for almost every t

$$\int_{\mathbb{T}^d} \left\|s_h^t(x,\cdot)\right\|^2_{\mathrm{L}^2(\mu_t^x)} dx = \frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\mu_t,\mu_{t+h}) \longrightarrow |\mu'|^2(t) = \int_{\mathbb{T}^d} \|v_t^x\|^2_{\mathrm{L}^2(\mu_t^x)} dx, \tag{III.1.58}$$

we even have that $x \mapsto \|s_h^t(x,\cdot)\|^2_{\mathrm{L}^2(\mu_t^x)}$ converges to $x \mapsto \|v_t^x\|^2_{\mathrm{L}^2(\mu_t^x)}$ in $\mathrm{L}^1(\mathbb{T}^d)$ for almost every t. Hence, for each h, the function $x \mapsto \|s_h^t(x,\cdot) - v_t^x\|^2_{\mathrm{L}^2(\mu_t^x)}$ is majorized by the function $x \mapsto 4(\|s_h^t(x,\cdot)\|^2_{\mathrm{L}^2(\mu_t^x)} + \|v_t^x\|^2_{\mathrm{L}^2(\mu_t^x)})$, which is a converging sequence in $\mathrm{L}^1(\mathbb{T}^d)$. Therefore, we can apply the (generalized) dominated convergence theorem to obtain that for almost every t

$$\lim_{h\to 0} \left\|s_h^t - v_t\right\|^2_{\mathrm{L}^2(\mu_t)} = \lim_{h\to 0} \int_{\mathbb{T}^d} \left\|s_h^t(x,\cdot) - v_t^x\right\|^2_{\mathrm{L}^2(\mu_t^x)} dx = \int_{\mathbb{T}^d} \lim_{h\to 0} \left\|s_h^t(x,\cdot) - v_t^x\right\|^2_{\mathrm{L}^2(\mu_t^x)} dx = 0, \tag{III.1.59}$$

which concludes the proof of this lemma. □

Proposition III.18 *Let $T \in (0,\infty]$ and $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ with tangent velocity v. Let $(\mu_t)_t \subset \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R})$ and $\sigma \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then*

$$\frac{d}{dt}\mathrm{W}^{\mathrm{L}}(\mu_t,\sigma)^2 = 2\int_{\mathbb{T}^d\times\mathbb{R}} (\boldsymbol{p}^2 - \mathrm{T}^{\sigma}_{\mu_t})\, v_t \, d\mu_t \quad \text{for almost every } t \in (0,T). \tag{III.1.60}$$

Proof. As above, the proof relies on the analogous result for $(\mu_t^x)_t$ and the dominated convergence theorem. Let for all $t \in (0,T)$ and $h > 0$

$$f_h^t : x \mapsto \frac{1}{h^2} W_2(\mu_t^x,\mu_{t+h}^x)^2 + 4(W_2(\mu_t^x,\sigma^x)^2 + W_2(\mu_{t+h}^x,\sigma^x)^2). \tag{III.1.61}$$

It will turn out that f_h^t is the majorizing sequence that we need. Thus, we need to show that $(f_h^t)_h$ converges in $\mathrm{L}^1(\mathbb{T}^d)$ for a.e. t. Indeed, we observe that as $h \to 0$ (again, after an application of Fubini's theorem) for almost every t

$$f_h^t(x) \longrightarrow |(\mu^x)'|^2(t) + 8W_2(\mu_t^x,\sigma^x)^2) =: f^t(x) \quad \text{for almost every } x \in \mathbb{T}^d. \tag{III.1.62}$$

Moreover, for almost every t

$$\begin{aligned} \|f_h^t\|_{\mathrm{L}^1(\mathbb{T}^d)} &= \frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\mu_t,\mu_{t+h})^2 + 4(\mathrm{W}^{\mathrm{L}}(\mu_t,\sigma)^2 + \mathrm{W}^{\mathrm{L}}(\mu_{t+h},\sigma)^2) \\ &\longrightarrow |\mu'|^2(t) + 8\mathrm{W}^{\mathrm{L}}(\mu_t,\sigma)^2 = \|f^t\|_{\mathrm{L}^1(\mathbb{T}^d)}. \end{aligned} \tag{III.1.63}$$

(III.1.62) and (III.1.63) show that $\lim_{h\to 0} f_h^t = f^t$ in $\mathrm{L}^1(\mathbb{T}^d)$ for a.e. t.

Note that from [3, 8.4.7] we get that for almost every t and x

$$\frac{d}{dt} W_2(\mu_t^x,\sigma^x)^2 = 2\int_{\mathbb{R}} (\mathrm{Id}_{\mathbb{R}} - \mathrm{T}^{\sigma^x}_{\mu_t^x})\, v_t^x \, d\mu_t^x. \tag{III.1.64}$$

Further, as a consequence of the triangle inequality and Young's inequality, we observe that for all t, h and x

$$\frac{1}{h^2}(W_2(\mu_{t+h}^x,\sigma^x)^2 - W_2(\mu_t^x,\sigma^x)^2) \le \frac{1}{2} f_h^t(x). \tag{III.1.65}$$

Therefore, the dominated convergence theorem yields (III.1.60). □

III.1.3 Gradient flows in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ for λ-convex functionals

In this subsection we introduce the notion of a *subdifferential* for a certain class of functionals in $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then we define *gradient flows* in $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ for such functionals and prove in Theorem III.27 their existence, uniqueness and some properties.

In this chapter, we only consider functionals that satisfy the following convexity property (cf. [3, 4.0.1]).

Definition III.19 *Let (X, d) be a Polish space. Then $\phi : X \to (-\infty, \infty]$ is called* strongly λ-convex *if $\lambda \in \mathbb{R}$ and for all $\sigma, \mu_0, \mu_1 \in D(\phi)$ there exists a curve $(\gamma_t)_{t\in[0,1]}$ with $\gamma_0 = \mu_0$, $\gamma_1 = \mu_1$ such that for all $0 < \tau < \frac{1}{\lambda^-}$ (with the convention that $1/0 = \infty$), the functional*

$$\Phi(\tau, \sigma; \cdot) := \frac{1}{2\tau}\mathrm{d}(\cdot, \sigma)^2 + \phi(\cdot) \tag{III.1.66}$$

is $(\frac{1}{\tau} + \lambda)$-convex along $(\gamma_t)_t$, i.e. for all $t \in [0, 1]$

$$\Phi(\tau, \sigma; \gamma_t) \le (1-t)\Phi(\tau, \sigma; \gamma_0) + t\,\Phi(\tau, \sigma; \gamma_1) - \frac{1}{2}\left(\frac{1}{\tau} + \lambda\right) t(1-t)\,\mathrm{d}(\mu_0, \mu_1)^2. \tag{III.1.67}$$

However, in most of the cases, it suffices to demand the following weaker form of convexity.

Definition III.20 *$\phi : X \to (-\infty, \infty]$ is called* λ-convex (along geodesics) *if $\lambda \in \mathbb{R}$ and for all $\mu_0, \mu_1 \in D(\phi)$ there exists a geodesic $(\mu_t)_{t\in[0,1]}$ such that ϕ is λ-convex along $(\mu_t)_t$.*

In our case, i.e. if $X = \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, $(\mu_t)_{t\in[0,1]}$ will always be the geodesic induced by some $\pi \in \mathrm{Opt}^{\mathrm{L}}(\mu_0, \mu_1)$ as in (III.1.37).

Subdifferential calculus. As we already motivated in Section I.2, instead of working with gradients in $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, we prefer to work with (strong) subdifferentials.

Definition III.21 *Let $\phi : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ be proper*[1]*, λ-convex and W^{L}-l.s.c. Let $\mu \in D(\phi) \cap \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R})$ and $\xi \in \mathrm{L}^2(\mu)$. Then we say that ξ* belongs to the subdifferential of *ϕ at μ and we write $\xi \in \partial\phi(\mu)$ if*

$$\phi(\nu) - \phi(\mu) \ge \int_{\mathbb{T}^d\times\mathbb{R}} \xi\,(\mathrm{T}_\mu^\nu - \boldsymbol{p}^2)\; d\mu + \frac{\lambda}{2}\mathrm{W}^{\mathrm{L}}(\mu, \nu)^2 \quad \forall\, \nu \in D(\phi). \tag{III.1.68}$$

Further, we say that $\xi \in \partial\phi(\mu)$ is a strong subdifferential of *ϕ at μ if*

$$\phi((\boldsymbol{p}^1, T)_\#\mu) - \phi(\mu) \ge \int_{\mathbb{T}^d\times\mathbb{R}} \xi\,(\mathrm{T} - \boldsymbol{p}^2)\; d\mu + o(\|\mathrm{T} - \boldsymbol{p}^2\|_{\mathrm{L}^2(\mu)}) \quad \text{as } \|\mathrm{T} - \boldsymbol{p}^2\|_{\mathrm{L}^2(\mu)} \to 0. \tag{III.1.69}$$

Lemma III.22 *Let $\phi : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ be proper, λ-convex and W^{L}-l.s.c. Let $\mu \in D(\phi) \cap \mathcal{P}_2^{\mathrm{L},a}(\mathbb{T}^d \times \mathbb{R})$ and $\xi \in \partial\phi(\mu) \cap \mathrm{Tan}_\mu\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then ξ is a strong subdifferential of ϕ at μ.*

[1] This means that $\phi(\mu) > -\infty$ for all $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ and there exists $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ such that $\phi(\mu) < \infty$.

Proof. The proof follows the same lines as in the Wasserstein case (see [59, 3.2]). Therefore, we omit the details. □

Definition III.23 *Let (X, d) be a Polish space. Let $\phi : X \to (-\infty, \infty]$ be proper and* d*-l.s.c. Then the* metric slope *$|\partial\phi| : D(\phi) \to [0, \infty]$ is defined by*

$$|\partial\phi|(\mu) = \limsup_{\nu \to \mu} \frac{(\phi(\mu) - \phi(\nu))^+}{\mathrm{d}(\mu, \nu)}. \tag{III.1.70}$$

Next we show that λ-convex functionals are differentiable almost everywhere along curves in $\mathcal{AC}((0,T); \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ and compute the derivative.

Lemma III.24 *Let $\phi : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ be proper, λ-convex and W^{L}-l.s.c. Let $T \in (0, \infty]$ and $(\mu_t)_t \in \mathcal{AC}((0,T); \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ with tangent velocity v. Suppose that*

$$\int_s^t |\partial\phi|(\mu_r)|\mu'|(r)\, dr < \infty \qquad \forall 0 < s < t < T. \tag{III.1.71}$$

Then

(i) $t \mapsto \phi(\mu_t)$ is absolutely continuous,

(ii) there exists a $\mathrm{Leb}_{(0,T)}$-null-set $\mathcal{N}$ such that for all $\xi \in \partial\phi(\mu_t)$

$$\frac{d}{dt}\phi(\mu_t) = \int_{\mathbb{T}^d \times \mathbb{R}} \xi\, v_t\, d\mu_t \quad \text{for all } t \in (0,T) \setminus \mathcal{N}. \tag{III.1.72}$$

Proof. (i) is the content of [3, 2.4.10]. To show *(ii)*, let $\mathcal{N}$ be such that for all $t \in (0,T) \setminus \mathcal{N}$, (III.1.55) holds, $\frac{d}{dt}\phi(\mu_t)$ exists and $|\partial\phi|(\mu_t) < \infty$. From here we proceed as in [3, 10.3.18]. □

Gradient flows in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$. We are now able to define the notion of gradient flows in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$.

Definition III.25 *Let $\phi : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ be proper, λ-convex and W^{L}-l.s.c. Let $T \in (0, \infty]$ and $(\mu_t)_t \in \mathcal{AC}((0,T); \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$ with tangent velocity v. Then $(\mu_t)_t$ is called* gradient flow for *ϕ, if*

$$-v_t \in \partial\phi(\mu_t) \qquad \text{for almost every } t \in (0,T). \tag{III.1.73}$$

Further, $\mu_0 \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ is called initial value *of $(\mu_t)_t$ if $\lim_{t\to 0} \mathrm{W}^{\mathrm{L}}(\mu_t, \mu_0) = 0$.*

Let us first note that, as in the Wasserstein case, gradient flows in $(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}), \mathrm{W}^{\mathrm{L}})$ are equivalent to the solutions of a system of *evolution variational inequalities (E.V.I.)*.

Lemma III.26 *Let $\phi : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ be proper, λ-convex and W^{L}-l.s.c. Let $T \in (0, \infty]$ and $(\mu_t)_t \in \mathcal{AC}((0,T); \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}))$. Then $(\mu_t)_t$ is a gradient flow for ϕ if and only if for all $\nu \in D(\phi)$ there exists a $\mathrm{Leb}_{(0,T)}$-null-set $\mathcal{N}_\nu$ such that for all $t \in (0,T) \setminus \mathcal{N}_\nu$*

$$\frac{1}{2}\frac{d}{dt}\mathrm{W}^{\mathrm{L}}(\mu_t, \nu)^2 \le \phi(\nu) - \phi(\mu_t) - \frac{\lambda}{2}\mathrm{W}^{\mathrm{L}}(\mu_t, \nu)^2. \tag{III.1.74}$$

Proof. Again, the proof consists of adapting the analogous proof in the Wasserstein case (see [3, 11.1.4]), which is based on Proposition III.18. We omit the details. □

In the following theorem we obtain the existence of gradient flows and further properties such as uniqueness, an energy identity and a regularisation estimate. The result is limited to the case, when the functional ϕ is proper, strongly λ-convex, W^{L}-l.s.c and *coercive*, where we say that a functional $\phi : X \to (-\infty, \infty]$ on a Polish space (X, d) is coercive if there exists $\mu^* \in X$ and $r^* > 0$ such that

$$\inf\{\phi(\nu) \,|\, \nu \in X\,,\, \mathrm{d}(\nu, \mu^*) \leq r^*\} > -\infty \quad \text{(cf. [3, (2.4.10)])}. \tag{III.1.75}$$

Theorem III.27 *Let $T \in (0, \infty]$ and $\phi : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ be proper, strongly λ-convex, W^{L}-l.s.c and coercive. Then:*

(i) (Existence) *For each $\mu_0 \in \overline{D(\phi)}$, there exists a gradient flow for ϕ with initial value μ_0.*

(ii) (λ-contraction and uniqueness) *Let $(\mu_t)_t$ and $(\nu_t)_t$ be gradient flows for ϕ with initial value $\mu_0 \in \overline{D(\phi)}$ and $\nu_0 \in \overline{D(\phi)}$, respectively. Then, for all $t \in (0, T)$*

$$\mathrm{W}^{\mathrm{L}}(\mu_t, \nu_t) \leq \mathrm{e}^{-\lambda t} \mathrm{W}^{\mathrm{L}}(\mu_0, \nu_0). \tag{III.1.76}$$

In particular, for each $\mu_0 \in \overline{D(\phi)}$, the gradient flow for ϕ with initial value μ_0 is unique.

(iii) (Energy identity) *Let $(\mu_t)_t$ be the gradient flow for ϕ with initial value $\mu_0 \in D(\phi)$, then for all $t \in (0, T)$*

$$\phi(\mu_t) - \phi(\mu_0) + \frac{1}{2} \int_0^t \left(|\partial\phi|^2(\mu_s) + |\mu'|^2(s) \right) ds = 0. \tag{III.1.77}$$

(iv) (Monotonicity along gradient flows) *Let $(\mu_t)_t$ be the gradient flow for ϕ with initial value $\mu_0 \in \overline{D(\phi)}$, then for almost every $t \in (0, T)$*

$$\frac{d}{dt} \phi(\mu_t) = -\|v_t\|^2_{\mathrm{L}^2(\mu_t)}. \tag{III.1.78}$$

(v) (Regularization estimate) *Let $(\mu_t)_t$ be the gradient flow for ϕ with initial value $\mu_0 \in \overline{D(\phi)}$, then for all $t \in (0, T)$ and all $\nu \in D(\phi)$*

$$\phi(\mu_t) \leq \begin{cases} \phi(\nu) + \frac{\lambda}{2(\mathrm{e}^{\lambda t} - 1)} \mathrm{W}^{\mathrm{L}}(\mu_0, \nu)^2 & : \lambda \neq 0, \\ \phi(\nu) + \frac{1}{2t} \mathrm{W}^{\mathrm{L}}(\mu_0, \nu)^2 & : \lambda = 0. \end{cases} \tag{III.1.79}$$

Proof. Again, we benefit from the work that was done in [3].

For $\mu_0 \in \overline{D(\phi)}$, we introduce the following implicit Euler scheme. Let $\tau > 0$. Define recursively:

$$\begin{cases} \mu_0^\tau := \mu_0, \\ \mu_n^\tau \in \underset{\nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})}{\operatorname{argmin}} \left(\phi(\nu) + \frac{1}{2\tau} \mathrm{W}^{\mathrm{L}}(\mu_{n-1}^\tau, \nu)^2 \right) \text{ for } n \in \mathbb{N}. \end{cases} \tag{III.1.80}$$

[3, 2.2.2] shows that this scheme is well-defined. Define the piecewise constant interpolating trajectory $(\bar{\mu}_t^\tau)_{t\in[0,T]}$ by

$$\begin{cases} \bar{\mu}_0^\tau := \mu_0, \\ \bar{\mu}_t^\tau := \mu_n^\tau \quad \text{for } t \in ((n-1)\tau, n\tau] \text{ for all } n \in \mathbb{N} \text{ such that } n\tau \leq T. \end{cases} \tag{III.1.81}$$

Then [3, 4.0.4] yields the convergence of this scheme with respect to W^{L} towards a curve $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$ with initial value μ_0, which solves (III.1.74) and satisfies *(ii)*. In addition, Lemma III.26 yields *(i)*.

[3, 4.0.4] shows that the gradient flow $(\mu_t)_t$ is a so-called minimizing movement (see [3, 2.0.6] for the definition). Hence, [3, 2.3.3] implies *(iii)*. (Note that in our case the object $|\partial^-\phi|$ from this theorem is just $|\partial\phi|$ and that the assumption that $|\partial\phi|$ is a *strong upper gradient* is also fulfilled by [3, 2.4.10].)

(iv) follows from the chain rule given in Lemma III.24.

(v) follows from [3, 4.3.2][2] and [3, (3.1.1)]. □

III.1.4 Local McKean-Vlasov equation

In this subsection we apply Theorem III.27 to a functional $\mathcal{F}$ that is of the form

$$\mathcal{F}(\mu) := \mathcal{S}(\mu) + \mathcal{W}(\mu) + \mathcal{V}(\mu), \tag{III.1.82}$$

where $\mathcal{S}, \mathcal{W}$ and $\mathcal{V}$ are called *entropy*, *interaction energy* and *potential energy*, respectively. In order to apply Theorem III.27 for $\mathcal{F}$, we show separately that each of its summands $\mathcal{S}, \mathcal{W}$ and $\mathcal{V}$ are well-defined, proper, strongly λ-convex, W^{L}-l.s.c and coercive in the Lemmas III.28, III.30 and III.32, respectively. (It will turn out that $\mathcal{F}$ is trivially proper.) Moreover, we compute a directional derivative of $\mathcal{F}$ (Proposition III.36), analyse the subdifferential of $\mathcal{F}$ (Proposition III.38) and derive a variational characterisation of gradient flows for $\mathcal{F}$ (Theorem III.40), which will be a key fact in the forthcoming sections (cf. Lemma I.8 and Lemma I.11). Finally, we show in Theorem III.41 the equivalence of the gradient flow for $\mathcal{F}$ and the weak solution to the partial differential equation (I.5.10).

Entropy. Define the *entropy* $\mathcal{S} : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}) \to (-\infty,\infty]$ by

$$\mathcal{S}(\mu) := \begin{cases} \int_{\mathbb{T}^d\times\mathbb{R}} \log(\rho) d\mu & : \mu \ll \mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}},\ \mu = \rho\,\mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}, \\ \infty & : \text{else.} \end{cases} \tag{III.1.83}$$

A very useful observation is that for each $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$

$$\mathcal{S}(\mu) = \int_{\mathbb{T}^d\times\mathbb{R}} S_1(\mu^x) dx, \tag{III.1.84}$$

where $S_1 : \mathcal{P}_2(\mathbb{R}) \to (-\infty,\infty]$ is the entropy functional on $\mathcal{P}_2(\mathbb{R})$, i.e.

$$S_1(\mu^x) := \begin{cases} \int_{\mathbb{R}} \log(\rho^x) d\mu^x & : \mu^x \ll \mathrm{Leb}_{\mathbb{R}},\ \mu^x = \rho^x\,\mathrm{Leb}_{\mathbb{R}}, \\ \infty & : \text{else.} \end{cases} \tag{III.1.85}$$

This fact simplifies our analysis, since we benefit from the already known results for S_1; see e.g. in [3]. In the following lemma we show that Theorem III.27 is applicable for $\mathcal{S}$.

[2] Note that there is a typo in [3, (4.3.2)]: It must be $\frac{\mathrm{e}^{\lambda T}-1}{\lambda}$ instead of $\frac{\mathrm{e}^{\lambda T}-1}{T}$.

Lemma III.28 *(i)* (Well-defined) *Let $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ and $\varepsilon > 0$. Then there exists $C_\varepsilon > 0$ such that $\mathcal{S}(\mu) \geq -C_\varepsilon - \varepsilon \int |\theta|^2 d\mu$ $(> -\infty)$.*

(ii) (Coercivity) *For all $r > 0$ we have*

$$\inf \left\{ \mathcal{S}(\nu) \,\middle|\, \nu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \,, \int |\theta|^2 d\nu \leq r \right\} > -\infty. \tag{III.1.86}$$

In particular, $\mathcal{S}$ is coercive.

(iii) (W^{L}-l.s.c) *Let $(\mu_n)_{n\in\mathbb{N}}$ be such that $\sup_n \int |\theta|^2 d\mu_n < \infty$ and $\mu_n \rightharpoonup \mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then*

$$\liminf_{n\to\infty} \mathcal{S}(\mu_n) \geq \mathcal{S}(\mu). \tag{III.1.87}$$

In particular, $\mathcal{S}$ is W^{L}-l.s.c.

(iv) (Strong 0-convexity) *$\mathcal{S}$ is strongly 0-convex.*

Proof. The corresponding statement for S_1 (see [83, (29)]) and (III.1.84) imply *(i)*.

(ii) is an immediate consequence of *(i)*.

To show *(iii)*, set $\nu := \mathrm{e}^{-|\theta|-\beta} d\theta dx \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, where $\beta > 0$ is a normalization constant. Recall the definition of the relative entropy given in (III.0.3). Then for $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$

$$\mathcal{S}(\mu) = \mathcal{H}(\mu \,|\, \nu) - \tilde{\mathcal{V}}(\mu), \tag{III.1.88}$$

where $\tilde{\mathcal{V}}(\mu) := \int (|\theta| + \beta)\, d\mu$. Since $\sup_n \int |\theta|^2 d\mu_n < \infty$, [3, 5.1.7] implies that

$$\lim_{n\to\infty} \tilde{\mathcal{V}}(\mu_n) = \tilde{\mathcal{V}}(\mu). \tag{III.1.89}$$

And by the dual representation of $\mathcal{H}$ (see [3, 9.4.4]), we have that $\mathcal{H}(\cdot \,|\, \nu)$ is the supremum of functionals that are continuous with respect to weak convergence. Hence, $\mathcal{H}(\cdot \,|\, \nu)$ is lower semi-continuous with respect to weak convergence. This fact together with (III.1.89) yields *(iii)*.

It remains to prove *(iv)*. Let $\sigma, \mu_0, \mu_1 \in D(\mathcal{S})$ and Φ be as in (III.1.66) for the functional $\mathcal{S}$. Analogously, define $\Phi_1(\tau, \sigma^x; \cdot) = \frac{1}{2\tau} W_2(\sigma^x, \cdot) + S_1(\cdot)$. Then we observe that for all $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$

$$\Phi(\tau, \sigma; \mu) = \int_{\mathbb{T}^d} \Phi_1(\tau, \sigma^x; \mu^x)\, dx. \tag{III.1.90}$$

Moreover, we show at the end of this proof that there exits a measure $\omega \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}^3)$ such that for almost every $x \in \mathbb{T}^d$

$$(\mathbf{p}^{1,2}_{\mathbb{R}^3})_{\#}\omega^x \in \mathrm{Opt}(\sigma^x, \mu_0^x) \quad \text{and} \quad (\mathbf{p}^{1,3}_{\mathbb{R}^3})_{\#}\omega^x \in \mathrm{Opt}(\sigma^x, \mu_1^x). \tag{III.1.91}$$

Set for all $t \in [0,1]$

$$\gamma_t = \left((1-t)\, \mathbf{p}^2_{\mathbb{R}^3} + t\, \mathbf{p}^3_{\mathbb{R}^3}\right)_{\#} \omega^x\, dx \in \mathcal{M}_1(\mathbb{T}^d \times \mathbb{R} \times \mathbb{R}). \tag{III.1.92}$$

Then, [3, 9.3.9] and [3, 9.2.7] show that for almost every $x \in \mathbb{T}^d$ and for all $t \in [0,1]$

$$\Phi_1(\tau, \sigma^x; \gamma_t^x) \leq (1-t)\Phi_1(\tau, \sigma^x; \gamma_0^x) + t\, \Phi_1(\tau, \sigma^x; \gamma_1^x) - \frac{1}{2\tau} t(1-t) W_2(\mu_0^x, \mu_1^x)^2. \tag{III.1.93}$$

Using (III.1.90), this implies that $\mathcal{S}$ is strongly 0-convex. It remains to show the existence of the measure ω. Let $\pi_0 \in \mathrm{Opt}^{\mathrm{L}}(\sigma, \mu_0)$ and $\pi_1 \in \mathrm{Opt}^{\mathrm{L}}(\sigma, \mu_1)$. Using the disintegration theorem, we obtain the existence of Borel measurable families $(\pi_0^{x,m})_{x\in\mathbb{T}^d, m\in\mathbb{R}}, (\pi_1^{x,m})_{x\in\mathbb{T}^d,m\in\mathbb{R}} \subset \mathcal{M}_1(\mathbb{R})$ such that

$$\pi_0 = \pi_0^{x,m}\, d\sigma^x(m)\, dx \quad \text{and} \quad \pi_1 = \pi_1^{x,m}\, d\sigma^x(m)\, dx. \tag{III.1.94}$$

Using the measurable selection lemma ([127, 5.22]), we know that there exists a family $(\omega^{x,m})_{x\in\mathbb{T}^d,m\in\mathbb{R}} \subset \mathcal{M}_1(\mathbb{R}^2)$ such that

$$\omega^{x,m} \in \mathrm{Opt}(\pi_0^{x,m}, \pi_1^{x,m}) \quad \text{and} \quad (x,m) \mapsto \omega^{x,m} \text{ is measurable.} \tag{III.1.95}$$

Define $\omega := \omega^{x,m}\, d\sigma^x(m)\, dx \in \mathcal{M}_1(\mathbb{T}^d \times \mathbb{R}^3)$. It is easy to see that ω fulfils (III.1.91). Indeed, for all Borel-measurable $M, A \subset \mathbb{R}$

$$\omega^x(M \times A \times \mathbb{R}) = \int_M \omega^{x,m}(A \times \mathbb{R})\, d\sigma^x(m) = \int_M \pi_0^{x,m}(A)\, d\sigma^x(m) = \pi_0^x(M \times A). \tag{III.1.96}$$

Therefore, $(\mathbf{p}^{1,2}_{\mathbb{R}^3})_{\#}\omega^x \in \mathrm{Opt}(\sigma^x, \mu_0^x)$, since we chose $\pi_0 \in \mathrm{Opt}^{\mathrm{L}}(\sigma, \mu_0)$. Analogously, one can show that $(\mathbf{p}^{1,3}_{\mathbb{R}^3})_{\#}\omega^x \in \mathrm{Opt}(\sigma^x, \mu_1^x)$. □

Interaction energy. Define the *interaction energy* $\mathcal{W} : \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}) \to (-\infty, \infty]$ by

$$\mathcal{W}(\mu) := \frac{1}{2} \int_{\mathbb{T}^d\times\mathbb{R}} \int_{\mathbb{T}^d\times\mathbb{R}} W(x, \bar{x}, \theta, \bar{\theta})\, d\mu(x,\theta) d\mu(\bar{x}, \bar{\theta}), \tag{III.1.97}$$

where $W \in C^{0,0,1,1}(\mathbb{T}^d \times \mathbb{T}^d \times \mathbb{R} \times \mathbb{R})$ satisfies the following assumptions.

Assumption III.29 *(1) $W(x, \bar{x}, \theta, \bar{\theta}) \geq -\alpha(|(\theta, \bar{\theta})|^2 + 1)$ for some $\alpha > 0$.*

(2) There exists $\bar{\lambda} \in \mathbb{R}$ such that for all $(x, \bar{x}) \in \mathbb{T}^d \times \mathbb{T}^d$, $(\theta, \bar{\theta}) \mapsto W(\bar{x}, x, \theta, \bar{\theta})$ is $\bar{\lambda}$-convex, i.e. for all $(\theta_1, \bar{\theta}_1), (\theta_2, \bar{\theta}_2) \in \mathbb{R}^2$

$$\begin{aligned} W(x, \bar{x}, (1-t)\theta_1 + t\theta_2, (1-t)\bar{\theta}_1 + t\bar{\theta}_2) \leq{}& (1-t)W(x, \bar{x}, \theta_1, \bar{\theta}_1) + t\, W(x, \bar{x}, \theta_2, \bar{\theta}_2) \\ & - \tfrac{\bar{\lambda}}{2} t(1-t) \left|(\theta_1, \bar{\theta}_1) - (\theta_2, \bar{\theta}_2)\right|^2. \end{aligned} \tag{III.1.98}$$

Lemma III.30 *Suppose that Assumption III.29 is satisfied. Then $\mathcal{W}$ is well-defined, coercive, strongly $\bar{\lambda}$-convex and W^{L}-l.s.c.*

Proof. Assumption III.29 (1) implies that $\mathcal{W}(\mu) \geq -\alpha \int |\theta|^2 d\mu - \alpha$ for all $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. This shows that $\mathcal{W}$ is well-defined and coercive.

Let $(\mu_n)_n$ and $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ be such that $\lim_{n\to\infty} \mathrm{W}^{\mathrm{L}}(\mu_n, \mu) = 0$. From [26, Theorem 2.8] and Lemma III.5, we obtain that $\mu_n \otimes \mu_n \rightharpoonup \mu \otimes \mu$ and $\lim_{n\to\infty} \int |\theta|^2 d(\mu_n \otimes \mu_n) = \int |\theta|^2 d(\mu \otimes \mu)$. Therefore, by Assumption III.29 (1), it is straightforward to see that W^- is uniformly integrable with respect to $(\mu_n)_n$. Hence, [3, 5.1.7] implies that $\liminf_{n\to\infty} \mathcal{W}(\mu_n) \geq \mathcal{W}(\mu)$.

It remains to show the strong $\bar{\lambda}$-convexity. Let $\sigma, \mu_0, \mu_1 \in D(\mathcal{W})$ and Φ be as in (III.1.66) for the functional $\mathcal{W}$. Let $\omega \in \mathcal{M}_1^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R}^3)$ and $(\gamma_t)_{t\in[0,1]}$ be as in the proof of Proposition

III.28 (iv). Since $\left(\mathbf{p}^1_{\mathbb{R}^3}, (1-t)\,\mathbf{p}^2_{\mathbb{R}^3} + t\,\mathbf{p}^3_{\mathbb{R}^3}\right)_{\#}\omega^x$ is a coupling of σ^x and γ^x_t for almost every $x \in \mathbb{T}^d$ and using (III.1.91), we obtain that for all $t \in [0,1]$

$$\begin{aligned}\int_{\mathbb{T}^d} W_2(\sigma^x,\gamma^x_t)^2\,dx &\le \int_{\mathbb{T}^d}\int_{\mathbb{R}^3} |(1-t)\theta_2 + t\,\theta_3 - \theta_1|^2 d\omega^x(\theta_1,\theta_2,\theta_3)\,dx \\ &= (1-t)\mathrm{W}^{\mathrm{L}}(\sigma,\mu_0)^2 + t\,\mathrm{W}^{\mathrm{L}}(\sigma,\mu_1)^2 - t(1-t)\int_{\mathbb{T}^d}\int_{\mathbb{R}^3} |\theta_2-\theta_3|^2 d\omega^x\,dx.\end{aligned} \tag{III.1.99}$$

Moreover, Assumption III.29 (2) implies that

$$\begin{aligned}\mathcal{W}(\gamma_t) &= \frac{1}{2}\int_{(\mathbb{T}^d\times\mathbb{R}^3)^2} W(x,\bar{x},(1-t)\theta_2+t\theta_3,(1-t)\bar{\theta}_2+t\bar{\theta}_3)d\omega(x,\theta_1,\theta_2,\theta_3)d\omega(\bar{x},\bar{\theta}_1,\bar{\theta}_2,\bar{\theta}_3) \\ &\le (1-t)\mathcal{W}(\mu_0) + t\,\mathcal{W}(\mu_1) - \frac{\bar{\lambda}}{2}t(1-t)\int_{\mathbb{T}^d\times\mathbb{R}^3} |\theta_2-\theta_3|^2\,d\omega(x,\theta_1,\theta_2,\theta_3).\end{aligned} \tag{III.1.100}$$

(III.1.99) and (III.1.100) yield that for all $\tau \in (0, \frac{1}{\bar{\lambda}^-})$ and for all $t \in [0,1]$

$$\begin{aligned}\Phi(\tau,\sigma;\gamma_t) &\le (1-t)\Phi(\tau,\sigma;\gamma_0) + t\,\Phi(\tau,\sigma;\gamma_1) - \left(\frac{1}{2\tau}+\frac{\bar{\lambda}}{2}\right)t(1-t)\int_{\mathbb{T}^d\times\mathbb{R}^3}|\theta_2-\theta_3|^2\,d\omega \\ &\le (1-t)\Phi(\tau,\sigma;\gamma_0) + t\,\Phi(\tau,\sigma;\gamma_1) - \left(\frac{1}{2\tau}+\frac{\bar{\lambda}}{2}\right)t(1-t)\mathrm{W}^{\mathrm{L}}(\mu_0,\mu_1),\end{aligned} \tag{III.1.101}$$

which is also a consequence of (III.1.91). This concludes the proof. □

Potential energy. Define the *potential energy* $\mathcal{V} : \mathcal{P}^{\mathrm{L}}_2(\mathbb{T}^d\times\mathbb{R}) \to (-\infty,\infty]$ by

$$\mathcal{V}(\mu) := \int_{\mathbb{T}^d\times\mathbb{R}} V\,d\mu, \tag{III.1.102}$$

where $V \in C^{0,1}(\mathbb{T}^d\times\mathbb{R})$ satisfies the following assumptions.

Assumption III.31 *(1) $V(x,\theta) \ge -\alpha(|\theta|^2+1)$ for some $\alpha > 0$.*

(2) There exists $\hat{\lambda} \in \mathbb{R}$ such that for all $x \in \mathbb{T}^d$, $\theta \mapsto V(x,\theta)$ is $\hat{\lambda}$-convex.

It turns out that under these assumptions, the potential energy is just the special case of the interaction energy, when $W(x,\bar{x},\theta,\bar{\theta}) = V(x,\theta) + V(\bar{x},\bar{\theta})$. Therefore, all the results for the interaction energy carry over to the potential energy and we have nothing to prove here.

Lemma III.32 *Suppose that Assumption III.31 is satisfied. Then, $\mathcal{V}$ is well-defined, coercive, strongly $\hat{\lambda}$-convex and W^{L}-l.s.c.*

The McKean-Vlasov-functional $\mathcal{F}$. From now on, we specify the functionals $\mathcal{W}$ and $\mathcal{V}$ as follows.

Assumption III.33 *(1) $W(x,\bar{x},\theta,\bar{\theta}) = -J(x-\bar{x})\,\theta\,\bar{\theta}$, where $J : \mathbb{T}^d \to \mathbb{R}$ is continuous and symmetric. It is easy to see that Assumption III.29 is satisfied. Indeed, as an immediate consequence of Young's inequality, Assumption III.29 (2) is satisfied for $\bar{\lambda} := -\|J\|_\infty$.*

(2) $V(x,\theta) = \Psi(\theta)$, *where* $\Psi : \mathbb{R} \to \mathbb{R}$ *is assumed to be a polynomial of degree* 2ℓ *for some* $\ell \in \mathbb{N}$, *and it is such that Assumption III.31 (2) is satisfied for some* $\hat{\lambda} \in \mathbb{R}$, *and*

$$\Psi(\theta) \geq C_\Psi \theta^{2\ell} + C'_\Psi \theta^2 - C''_\Psi \quad \text{for all } \theta \in \mathbb{R}, \tag{III.1.103}$$

for some $C_\Psi, C''_\Psi \geq 0$ *and* $C'_\Psi > \|J\|_\infty$.

For example, if Ψ is a polynomial of degree 2ℓ, then Ψ satisfies Assumption III.33 if the coefficient of degree 2ℓ is positive.

Assumption III.33 implies that $\mathcal{F}$ has the form

$$\mathcal{F}(\mu) = \int_{\mathbb{T}^d\times\mathbb{R}} \log(\rho) d\mu + \int_{\mathbb{T}^d\times\mathbb{R}} \Psi \, d\mu - \frac{1}{2}\int_{\mathbb{T}^d\times\mathbb{R}}\int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\theta\bar{\theta} d\mu(x,\theta) d\mu(\bar{x},\bar{\theta}) \tag{III.1.104}$$

if μ has a density ρ with respect to $\mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}$, and $\mathcal{F}(\mu) = \infty$ otherwise. Note that $\mathcal{F}$ is proper (e.g. $\mathcal{F}(\exp(-\theta^2/2)(2\pi)^{-1/2} d\theta dx) < \infty$). Furthermore, the definition of $\mathcal{F}$ can be naturally extended to $\mathcal{M}_1(\mathbb{T}^d \times \mathbb{R})$. We observe the following lower bound on $\mathcal{F}$.

Lemma III.34 *We have, for some constant* $C'' > 0$,

$$\mathcal{F}(\mu) \geq \int_{\mathbb{T}^d\times\mathbb{R}} \Big(C_\Psi |\theta|^{2\ell} + (C'_\Psi - \|J\|_\infty)|\theta|^2 \Big) d\mu - C'' \qquad \text{for all } \mu \in \mathcal{M}_1(\mathbb{T}^d \times \mathbb{R}). \tag{III.1.105}$$

In particular, there exists $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ *such that* $\inf_{\sigma\in\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})} \mathcal{F}(\sigma) = \mathcal{F}(\mu)$ *and* $D(\mathcal{F}) \subset \{\mu \in \mathcal{M}_1(\mathbb{T}^d \times \mathbb{R}) \,|\, \int |\theta|^2 \, d\mu < \infty\}$.

Proof. Let $\mu \in \mathcal{M}_1(\mathbb{T}^d \times \mathbb{R})$ and assume that μ has a density ρ, since otherwise the claim is trivial. Notice that we can rewrite $\mathcal{F}$ as

$$\mathcal{F}(\mu) = \mathcal{H}\left(\mu \,\middle|\, \mathrm{e}^{-\frac{1}{2}\Psi(\theta)} d\theta dx\right) + \frac{1}{2}\int_{(\mathbb{T}^d\times\mathbb{R})^2} \left(\frac{1}{2}\big(\Psi(\theta) + \Psi(\bar{\theta})\big) - J(x-\bar{x})\theta\bar{\theta} \right) d\mu d\mu. \tag{III.1.106}$$

Then, since

$$\mathcal{H}\Big(\mu \,\Big|\, \mathrm{e}^{-\frac{1}{2}\Psi(\theta)} d\theta dx\Big) \geq -\log \int_{\mathbb{T}\times\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\Psi(\theta)} d\theta dx, \tag{III.1.107}$$

and by Young's inequality and (III.1.103),

$$\frac{1}{2}\big(\Psi(\theta) + \Psi(\bar{\theta})\big) - J(x-\bar{x})\theta\bar{\theta} \geq \frac{1}{2} C_\Psi(\theta^{2\ell} + \bar{\theta}^{2\ell}) + \frac{1}{2}(C'_\Psi - \|J\|_\infty)(\theta^2 + \bar{\theta}^2) - C''_\Psi, \tag{III.1.108}$$

we infer (III.1.105).

For the second claim, note that (III.1.105) implies the weak compactness of the level sets of $\mathcal{F}$ and that Theorem III.35 below shows the weak lower semi-continuity of $\mathcal{F}$. Therefore, the direct method of the calculus of variation is applicable and we infer the existence of a minimizer. □

As a consequence of the observations on $\mathcal{S}, \mathcal{V}$ and $\mathcal{W}$, we obtain the following result for $\mathcal{F}$.

Theorem III.35 *$\mathcal{F}$ is well-defined, proper, coercive, $(\bar\lambda+\hat\lambda)$-convex, strongly λ-convex for some $\lambda\in\mathbb{R}$ and lower semi-continuous with respect to weak convergence. In particular, $\mathcal{F}$ is W^{L}-l.s.c. Therefore, Theorem III.27 is applicable for $\mathcal{F}$.*

Proof. It remains to show that $\mathcal{F}$ is weakly lower semi-continuous. Let $(\mu_n)_{n\in\mathbb{N}}\subset\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})$ and $\mu\in\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})$ be such that $\mu_n\rightharpoonup\mu$. Without restriction suppose that $\mathcal{F}(\mu)<\infty$. We show the lower semi-continuity for both summands on the right-hand side of (III.1.106) separately. In the proof of Lemma III.28 we have already seen that the functional $\mathcal{H}\big(\,\cdot\,\big|\,\frac{1}{\alpha}\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta dx\big)$ is weakly lower semi-continuous, where $\alpha=\int\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta dx$. Therefore,

$$\begin{aligned}\liminf_{n\to\infty}\mathcal{H}\Big(\mu^n\,\Big|\,\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta dx\Big)&=\liminf_{n\to\infty}\mathcal{H}\Big(\mu^n\,\Big|\,\tfrac{1}{\alpha}\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta dx\Big)-\log(\alpha)\\&\geq\mathcal{H}\Big(\mu\,\Big|\,\tfrac{1}{\alpha}\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta dx\Big)-\log(\alpha)=\mathcal{H}\Big(\mu\,\Big|\,\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta dx\Big).\end{aligned}\tag{III.1.109}$$

Moreover, the integrand of the second summand in (III.1.106) is lower semi-continuous and bounded from below due to (III.1.108). Therefore, [3, 5.1.7] yields

$$\begin{aligned}\liminf_{n\to\infty}\int_{(\mathbb{T}^d\times\mathbb{R})^2}&\Big(\frac{1}{2}\big(\Psi(\theta)+\Psi(\bar\theta)\big)-J(x-\bar x)\theta\bar\theta\Big)d(\mu^n\otimes\mu^n)\\&\geq\int_{(\mathbb{T}^d\times\mathbb{R})^2}\left(\frac{1}{2}\big(\Psi(\theta)+\Psi(\bar\theta)\big)-J(x-\bar x)\theta\bar\theta\right)d(\mu\otimes\mu),\end{aligned}\tag{III.1.110}$$

which concludes the proof. □

Directional derivative. In order to find a characterisation of the (strong) subdifferential of $\mathcal{F}$, it is useful to study the infinitesimal behaviour of $\mathcal{F}$ along curves that are pushed along smooth functions. This is the content of the following proposition.

Proposition III.36 *Let $\mu\in D(\mathcal{F})$ and $\beta\in C^2_c(\mathbb{T}^d\times\mathbb{R};\mathbb{R})$. For all $t\in\mathbb{R}$ define*

$$\mu_{t,\beta}=(\boldsymbol{p}^1,\boldsymbol{p}^2+t\,\beta)_{\#}\mu.\tag{III.1.111}$$

Then

$$\frac{d}{dt}\bigg|_{t=0}\mathcal{F}(\mu_{t,\beta})=\int_{\mathbb{T}^d\times\mathbb{R}}\left(\beta(x,\theta)\left[\Psi'(\theta)-\int_{\mathbb{T}^d\times\mathbb{R}}J(x-\bar x)\bar\theta\,d\mu(\bar x,\bar\theta)\right]-\partial_\theta\beta(x,\theta)\right)d\mu(x,\theta).\tag{III.1.112}$$

Proof. We compute the derivative for each summand separately. We begin with $\mathcal{S}$. Again, the proof is similar to the Wasserstein case. Indeed, if we consider the function $\hat\beta=(0,\beta)\in C^2_c(\mathbb{T}^d\times\mathbb{R};\mathbb{T}^d\times\mathbb{R})$, then $\mu_{t,\beta}=(\mathrm{Id}_{\mathbb{T}^d\times\mathbb{R}}+t\,\hat\beta)_{\#}\mu$ for all $t\in\mathbb{R}$. Then, [83, (38)] implies that

$$\frac{d}{dt}\bigg|_{t=0}\mathcal{S}(\mu_{t,\beta})=-\int_{\mathbb{T}^d\times\mathbb{R}}\mathrm{div}\hat\beta\,d\mu=-\int_{\mathbb{T}^d\times\mathbb{R}}\partial_\theta\beta\,d\mu.\tag{III.1.113}$$

To compute the directional derivative of $\mathcal{W}$, observe that

$$\begin{aligned}\frac{d}{dt}\bigg|_{t=0}\mathcal{W}(\mu_{t,\beta})&=-\frac{1}{2}\frac{d}{dt}\bigg|_{t=0}\int_{\mathbb{T}^d\times\mathbb{R}}\int_{\mathbb{T}^d\times\mathbb{R}}J(x-\bar x)(\theta+t\beta(x,\theta))(\bar\theta+t\beta(\bar x,\bar\theta))d\mu(x,\theta)d\mu(\bar x,\bar\theta)\\&=-\frac{1}{2}\int_{\mathbb{T}^d\times\mathbb{R}}\int_{\mathbb{T}^d\times\mathbb{R}}J(x-\bar x)\frac{d}{dt}\bigg|_{t=0}(\theta+t\beta(x,\theta))(\bar\theta+t\beta(\bar x,\bar\theta))d\mu(x,\theta)d\mu(\bar x,\bar\theta)\\&=-\int_{\mathbb{T}^d\times\mathbb{R}}\int_{\mathbb{T}^d\times\mathbb{R}}J(x-\bar x)\,\bar\theta\,\beta(x,\theta)\,d\mu(x,\theta)\;d\mu(\bar x,\bar\theta),\end{aligned}\tag{III.1.114}$$

where we have used the symmetry of the integrand. To exchange differentiation and integration, we have used the Leibniz-integral-rule, which is applicable, since all functions are continuous and β has compact support. In the same way, one computes that

$$\frac{d}{dt}\Big|_{t=0} \mathcal{V}(\mu_{t,\beta}) = \int_{\mathbb{T}^d\times\mathbb{R}} \beta\, \Psi'\, d\mu. \tag{III.1.115}$$

□

Subdifferential of $\mathcal{F}$. Note that for all $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$, $(x,\theta) \mapsto \int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\bar{\theta}\, d\mu(\bar{x},\bar{\theta}) \in \mathrm{L}^2(\mu)$. This observation is important in order to compute an element of the subdifferential of $\mathcal{F}$ in the following proposition.

Lemma III.37 *Let $\mu \in D(\mathcal{F})$. Therefore, μ has a density ρ with respect to $\mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}$. Suppose that $\partial_\theta\rho$ exists weakly in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{T}^d\times\mathbb{R})$ and $\frac{\partial_\theta\rho}{\rho} + \Psi' \in \mathrm{L}^2(\mu)$. Then*

$$\left((x,\theta)\mapsto \frac{\partial_\theta\rho(x,\theta)}{\rho(x,\theta)} + \Psi'(\theta) - \int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\bar{\theta}\, d\mu(\bar{x},\bar{\theta})\right) \in \partial\mathcal{F}(\mu) \tag{III.1.116}$$

Proof. We first show that $\Big((x,\theta)\mapsto -\int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\bar{\theta}\, d\mu(\bar{x},\bar{\theta})\Big) \in \partial\mathcal{W}(\mu)$. Note that for all $(\theta_1,\bar{\theta}_1),(\theta_2,\bar{\theta}_2)\in\mathbb{R}^2$

$$\begin{aligned}
-J(x-\bar{x})\big(\theta_1\bar{\theta}_1 - \theta_2\bar{\theta}_2\big) &= -J(x-\bar{x})\Big(\bar{\theta}_2(\theta_1-\theta_2) + \theta_2(\bar{\theta}_1-\bar{\theta}_2) + (\theta_1-\theta_2)(\bar{\theta}_1-\bar{\theta}_2)\Big)\\
&\geq -J(x-\bar{x})\Big(\bar{\theta}_2(\theta_1-\theta_2) + \theta_2(\bar{\theta}_1-\bar{\theta}_2)\Big) + \frac{\bar{\lambda}}{2}\left|(\theta_1,\bar{\theta}_1)-(\theta_2,\bar{\theta}_2)\right|^2
\end{aligned} \tag{III.1.117}$$

This yields for all $\nu\in D(\mathcal{F})\subset D(\mathcal{W})$

$$\begin{aligned}
\mathcal{W}(\nu)-\mathcal{W}(\mu) =& \frac{1}{2}\int_{(\mathbb{T}^d\times\mathbb{R})^2} -J(x-\bar{x})\Big(\mathrm{T}_\mu^\nu(x,\theta)\mathrm{T}_\mu^\nu(\bar{x},\bar{\theta}) - \theta\bar{\theta}\Big)d(\mu\otimes\mu)\\
\geq& \frac{1}{2}\int_{(\mathbb{T}^d\times\mathbb{R})^2} -J(x-\bar{x})\bar{\theta}(\mathrm{T}_\mu^\nu(x,\theta)-\theta)d(\mu\otimes\mu)\\
&+\frac{1}{2}\int_{(\mathbb{T}^d\times\mathbb{R})^2} -J(x-\bar{x})\theta(\mathrm{T}_\mu^\nu(\bar{x},\bar{\theta})-\bar{\theta})d(\mu\otimes\mu)\\
&+\frac{\bar{\lambda}}{4}\int_{(\mathbb{T}^d\times\mathbb{R})^2}\Big|(\mathrm{T}_\mu^\nu(x,\theta),\mathrm{T}_\mu^\nu(\bar{x},\bar{\theta})) - (\theta,\bar{\theta})\Big|^2 d(\mu\otimes\mu)\\
=& \int_{\mathbb{T}^d\times\mathbb{R}}\left(-\int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\bar{\theta}d\mu(\bar{x},\bar{\theta})\right)(\mathrm{T}_\mu^\nu(x,\theta)-\theta)\, d\mu(x,\theta) + \frac{\bar{\lambda}}{2}\mathrm{W}^{\mathrm{L}}(\mu,\nu)^2.
\end{aligned} \tag{III.1.118}$$

It remains to show that $\partial_\theta\rho/\rho + \Psi' \in \partial(\mathcal{S}+\mathcal{V})$. Notice that $\theta\mapsto\Psi(\theta)-\frac{1}{2}\hat{\lambda}|\theta|^2$ is convex. Set $\tilde{V}(\theta) := \Psi(\theta) - \frac{1}{2}\hat{\lambda}|\theta|^2 + \beta$, where $\beta\in\mathbb{R}$ is such that $\exp(-\tilde{V}(\theta))d\theta$ is a probability measure. Define $\tilde{\mathcal{V}}(\mu) := \int \tilde{V}d\mu$. Then, similarly as in (III.1.88) and (III.1.84), we have that

$$\mathcal{S}(\mu)+\tilde{\mathcal{V}}(\mu) = \mathcal{H}\Big(\mu\,\Big|\,\mathrm{e}^{-\tilde{V}(\theta)}d\theta dx\Big) = \int_{\mathbb{T}^d\times\mathbb{R}}\mathcal{H}\Big(\mu^x\,\Big|\,\mathrm{e}^{-\tilde{V}(\theta)}d\theta\Big)dx. \tag{III.1.119}$$

This fact allows us to use the results from the Wasserstein case. By taking a compact exhaustion of $\mathbb{R}$, one can see immediately that there exists a null-set $\mathcal{N} \subset \mathbb{T}^d$ such that for all $x \in \mathbb{T}^d \setminus \mathcal{N}$

$$\rho(x,\cdot) \in \mathrm{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}), \qquad \frac{\partial_\theta \rho(x,\cdot)}{\rho(x,\cdot)} + \Psi' \in \mathrm{L}^2(\mu^x) \qquad \text{and} \qquad \int_{\mathbb{R}} |\theta|^2 d\mu^x < \infty. \tag{III.1.120}$$

Moreover, if we set $\sigma^x(\theta) = \rho(x,\theta)\exp(\tilde{V}(\theta))$, then (III.1.120) implies that

$$\sigma^x \in \mathrm{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}) \qquad \text{and} \qquad \frac{\partial_\theta \sigma^x}{\sigma^x} \in \mathrm{L}^2(\mu^x) \qquad \text{for almost every } x. \tag{III.1.121}$$

Therefore, [3, 10.4.9] is applicable and we obtain that for all $\nu \in D(\mathcal{F})$

$$\mathcal{H}\Big(\nu^x \,\Big|\, \mathrm{e}^{-\tilde{V}(\theta)} d\theta\Big) - \mathcal{H}\Big(\mu^x \,\Big|\, \mathrm{e}^{-\tilde{V}(\theta)} d\theta\Big) \geq \int_{\mathbb{R}} \frac{\partial_\theta \sigma^x}{\sigma^x} (\mathrm{T}^{\nu^x}_{\mu^x} - \mathrm{Id}_{\mathbb{R}})\, d\mu^x \quad \text{for a.e. } x. \tag{III.1.122}$$

Using (III.1.119), this implies that

$$(\mathcal{S} + \tilde{\mathcal{V}})(\nu) - (\mathcal{S} + \tilde{\mathcal{V}})(\mu) \geq \int_{\mathbb{T}^d \times \mathbb{R}} \Big(\frac{\partial_\theta \rho}{\rho} + \Psi' - \hat{\lambda}\mathbf{p}^2\Big)(\mathrm{T}^\nu_\mu - \mathbf{p}^2)\, d\mu. \tag{III.1.123}$$

Since $\mathrm{W}^{\mathrm{L}}(\mu,\nu) = \|\mathrm{T}^\nu_\mu - \mathbf{p}^2\|_{\mathrm{L}^2(\mu)}$, we infer

$$(\mathcal{S} + \mathcal{V})(\nu) - (\mathcal{S} + \mathcal{V})(\mu) \geq \int_{\mathbb{T}^d \times \mathbb{R}} \Big(\frac{\partial_\theta \rho}{\rho} + \Psi'\Big)(\mathrm{T}^\nu_\mu - \mathbf{p}^2)\, d\mu + \frac{\hat{\lambda}}{2} \mathrm{W}^{\mathrm{L}}(\mu,\nu), \tag{III.1.124}$$

which concludes the proof. □

Proposition III.38 *Let $\mu = \rho\,\mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}} \in D(\mathcal{F})$. Then the following statements are equivalent.*

(i) $|\partial\mathcal{F}|(\mu) < \infty$,

(ii) $\partial_\theta\rho$ *exists weakly in* $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{T}^d \times \mathbb{R})$ *and there exists* $w \in \mathrm{L}^2(\mu)$ *such that* $\partial_\theta\rho(x,\theta) = \rho(x,\theta)(w(x,\theta) - \Psi'(\theta) + \int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\bar{\theta}\, d\mu(\bar{x},\bar{\theta}))$.

Moreover, in this case, $w \in \mathrm{Tan}_\mu \mathcal{P}^{\mathrm{L}}_2(\mathbb{T}^d \times \mathbb{R}) \cap \partial\mathcal{F}(\mu)$, $|\partial\mathcal{F}|(\mu) = \|w\|_{\mathrm{L}^2(\mu)}$ *and* w *is the* μ*-a.e. unique strong subdifferential at* μ.

Proof. Again, the proof is very similar to the Wasserstein case (cf. [59, 4.3]). However, here we include the details, since the statement and the proof will become very crucial for the remainder of this chapter.

(ii) $\Rightarrow$ *(i)*. Lemma III.37 shows that under the conditions of *(ii)*, $w \in \partial\mathcal{F}(\mu)$. Hence, $|\partial\mathcal{F}|(\mu) \leq \|w\|_{\mathrm{L}^2(\mu)} < \infty$, which is an immediate consequence of the definition of the metric slope (cf. [3, 10.3.10]).

(i) $\Rightarrow$ *(ii)*. Define a linear operator $L : C^\infty_c(\mathbb{T}^d \times \mathbb{R}) \to \mathbb{R}$ by

$$L(\beta) := \int_{\mathbb{T}^d\times\mathbb{R}} \left(\beta(x,\theta)\left[\Psi'(\theta) - \int_{\mathbb{T}^d\times\mathbb{R}} J(x-\bar{x})\bar{\theta}\, d\mu(\bar{x},\bar{\theta})\right] - \partial_\theta\beta(x,\theta)\right) d\mu(x,\theta). \tag{III.1.125}$$

Let $\beta \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})$ and $(\mu_{t,\beta})_t$ be as in Proposition III.36. Using the representation (III.1.5) and that $(\mathbf{p}^1, \mathbf{p}^2 + t\beta, \mathbf{p}^2)_\# \mu \in \mathrm{Cpl}^{\mathrm{L}}(\mu_{t,\beta}, \mu)$, it is easy to see that $\mathrm{W}^{\mathrm{L}}(\mu_{t,\beta}, \mu) \le |t| \cdot \|\beta\|_{\mathrm{L}^2(\mu)}$. Then, as in [59, p. 13, l. 12], via Proposition III.36, we observe that if $L(\beta) > 0$,

$$\begin{aligned} L(\beta) &= \lim_{t \downarrow 0} \frac{(\mathcal{F}(\mu) - \mathcal{F}(\mu_{-t,\beta}))^+}{t} \le \limsup_{t \downarrow 0} \frac{(\mathcal{F}(\mu) - \mathcal{F}(\mu_{t,-\beta}))^+}{\mathrm{W}^{\mathrm{L}}(\mu_{t,-\beta}, \mu)} \|\beta\|_{\mathrm{L}^2(\mu)} \\ &\le |\partial \mathcal{F}|(\mu)\ \|\beta\|_{\mathrm{L}^2(\mu)}, \end{aligned} \tag{III.1.126}$$

and if $L(\beta) < 0$,

$$\begin{aligned} L(\beta) &= \lim_{t \downarrow 0} \frac{(\mathcal{F}(\mu) - \mathcal{F}(\mu_{t,\beta}))^+}{-t} \ge -\liminf_{t \downarrow 0} \frac{(\mathcal{F}(\mu) - \mathcal{F}(\mu_{t,\beta}))^+}{\mathrm{W}^{\mathrm{L}}(\mu_{t,\beta}, \mu)} \|\beta\|_{\mathrm{L}^2(\mu)} \\ &\ge -|\partial \mathcal{F}|(\mu)\ \|\beta\|_{\mathrm{L}^2(\mu)}. \end{aligned} \tag{III.1.127}$$

Thus, $|L(\beta)| \le |\partial \mathcal{F}|(\mu)\ \|\beta\|_{\mathrm{L}^2(\mu)}$. Extending L to the $\mathrm{L}^2(\mu)$-closure of $C_c^\infty(\mathbb{T}^d \times \mathbb{R})$, the Riesz representation theorem yields the existence of a unique $w \in \mathrm{L}^2(\mu)$ such that

- $\int w\beta\rho\, d\theta dx = \int \left(\beta \left[\Psi' - \int J(\cdot - \bar{x})\bar{\theta}\, d\mu\right] - \partial_\theta \beta\right) \rho\, d\theta dx$ for all $\beta \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})$, and
- $|\partial \mathcal{F}|(\mu) \ge \|w\|_{\mathrm{L}^2(\mu)}$.

This shows that the weak derivative $\partial_\theta \rho$ exists and equals $\rho(w - \Psi' + \int_{\mathbb{T}^d \times \mathbb{R}} J(x - \bar{x})\bar{\theta}\, d\mu)$, which clearly belongs to $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{T}^d \times \mathbb{R})$. We infer *(ii)*.

It remains to show the other claims. Let $\mathbf{p}^{\mathrm{Tan}}$ denote the orthogonal projection onto $\mathrm{Tan}_\mu \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Then $\mathbf{p}^{\mathrm{Tan}}(w) \in \partial \mathcal{F}(\mu)$, since $w \in \partial \mathcal{F}(\mu)$. Indeed, this follows immediately from the definition of the subdifferential and Corollary III.15. Hence, by Lemma III.12

$$|\partial \mathcal{F}|(\mu) \le \|\mathbf{p}^{\mathrm{Tan}}(w)\|_{\mathrm{L}^2(\mu)} \le \|\mathbf{p}^{\mathrm{Tan}}(w) + w - \mathbf{p}^{\mathrm{Tan}}(w)\|_{\mathrm{L}^2(\mu)} = \|w\|_{\mathrm{L}^2(\mu)} \le |\partial \mathcal{F}|(\mu), \tag{III.1.128}$$

which, again by Lemma III.12, shows that $w \in \mathrm{Tan}_\mu \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$ and $|\partial \mathcal{F}|(\mu) = \|w\|_{\mathrm{L}^2(\mu)}$.

Finally, let z be another strong subdifferential of $\mathcal{F}$ at μ. Then, for all $\beta \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})$

$$\int w\beta\, d\mu = L(\beta) = \frac{d}{dt}\bigg|_{t=0} \mathcal{F}(\mu_{t,\beta}) = \lim_{t \downarrow 0} \frac{\mathcal{F}(\mu_{t,\beta}) - \mathcal{F}(\mu)}{t} \ge \int z\beta\, d\mu, \tag{III.1.129}$$

since z is a strong subdifferential. Considering $\lim_{t \uparrow 0}$, we obtain the other inequality. Therefore, $\int w\beta\, d\mu = \int z\beta\, d\mu$, for all $\beta \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})$, which implies that $z = w$ μ-a.e. □

Corollary III.39 *Let $\mu \in D(\mathcal{F})$. Then*

$$|\partial \mathcal{F}|(\mu) = \sup_{\beta \in C_c^\infty(\mathbb{T}^d \times \mathbb{R}),\, \|\beta\|_{\mathrm{L}^2(\mu)} > 0} \frac{\left| \int_{\mathbb{T}^d \times \mathbb{R}} \left(\beta \left[\Psi' - \int_{\mathbb{T}^d \times \mathbb{R}} J(\cdot - \bar{x})\bar{\theta}\, d\mu(\bar{x}, \bar{\theta})\right] - \partial_\theta \beta\right) d\mu \right|}{\|\beta\|_{\mathrm{L}^2(\mu)}}. \tag{III.1.130}$$

Moreover, if $(\mu_n)_{n \in \mathbb{N}}$ is such that $\sup_n \int |\theta|^2 d\mu_n < \infty$ and $\mu_n \rightharpoonup \mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$, then

$$\liminf_{n \to \infty} |\partial \mathcal{F}|(\mu_n) \ge |\partial \mathcal{F}|(\mu). \tag{III.1.131}$$

Proof. If $|\partial\mathcal{F}|(\mu) < \infty$, then (III.1.130) follows from the proof of Proposition III.38, since the right-hand side of (III.1.130) equals $\|w\|_{\mathrm{L}^2(\mu)}$. Here we have used that the extension of the operator L from (III.1.125) has the same operator norm as L. And if the right-hand side of (III.1.130) is finite, then L is bounded. Therefore, repeating the above arguments, we infer part *(ii)* of Proposition III.38, which leads to $|\partial\mathcal{F}|(\mu) < \infty$ and finally to (III.1.130).

The proof of (III.1.131) is a straightforward consequence of (III.1.130), the fact that $\beta \in C_c^\infty(\mathbb{T}^d \times \mathbb{R})$, and [3, 5.1.7]. □

Characterisation as curves of maximal slope. The following characterisation of gradient flows for $\mathcal{F}$ is a key fact in order to apply the Fathi-Sandier-Serfaty approach, which we introduced in Section I.3. We motivated this statement already in Lemma I.8 by the analogous Wasserstein setting.

Theorem III.40 *Let $T \in (0,\infty)$. Define $\mathscr{J} : C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R})) \to [0,\infty]$ by*

$$\mathscr{J}\,[(\nu_t)_t] := \mathcal{F}(\nu_T) - \mathcal{F}(\nu_0) + \frac{1}{2}\int_0^T \big(|\partial\mathcal{F}|^2(\nu_t) + |\nu'|^2(t)\big)\,dt, \tag{III.1.132}$$

if $(\nu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$ and $\mathscr{J}\,[(\nu_t)_t] = \infty$ else. Let $\mu_0 \in D(\mathcal{F})$. For any curve $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$ such that $\lim_{t\to 0}\mathrm{W}^{\mathrm{L}}(\mu_t,\mu_0) = 0$ we have that $\mathscr{J}\,[(\mu_t)_t] \geq 0$. Equality holds if and only if $(\mu_t)_t$ is the gradient flow for $\mathcal{F}$ with initial value μ_0.

Proof. Since $\mathcal{F}$ is $(\bar\lambda+\hat\lambda)$-convex, we can apply [3, 2.4.10] to see that

$$\mathcal{F}(\mu_\varepsilon) - \mathcal{F}(\mu_T) \leq \int_\varepsilon^T |\partial\mathcal{F}|(\mu_t)\,|\mu'|(t)\,dt \qquad \text{for all } \varepsilon \in (0,T). \tag{III.1.133}$$

Thus, Young's inequality and the W^{L}-l.s.c. of $\mathcal{F}$ yield the first claim. The "if"-part of the second claim is the content of Theorem III.27 *(iii)*. To show the "only if"-part, assume that $\mathscr{J}\,[(\mu_t)_t] = 0$. Hence, $|\partial\mathcal{F}|(\mu_t) < \infty$ for almost every t and Proposition III.38 is applicable. Let $(v_t)_t$ be the tangent velocity of $(\mu_t)_t$ and $(\rho_t)_t$ be the curve of the probability densities of $(\mu_t)_t$. Recall that $\|v_t\|^2_{\mathrm{L}^2(\mu_t)} = |\mu'|(t)$ for a.e. t. Then, using the chain rule from Lemma III.24 and the characterisation of the metric slope from Proposition III.38, we obtain that

$$\frac{1}{2}\int_0^T \left\| v_t + \frac{\partial_\theta\rho_t}{\rho_t} + \Psi' - \int_{\mathbb{T}^d\times\mathbb{R}} J(\cdot-\bar x)\bar\theta\,d\mu_t \right\|^2_{\mathrm{L}^2(\mu_t)} dt = \mathscr{J}\,[(\mu_t)_t] = 0, \tag{III.1.134}$$

which, again by Proposition III.38, implies that $-v_t \in \partial\mathcal{F}(\mu_t)$ for a.e. t. Therefore, $(\mu_t)_t$ is the gradient flow for $\mathcal{F}$. □

Local McKean-Vlasov equation. Now we are able to build the bridge to (I.5.10) in the following theorem. See also Lemma I.9 for the analogous Wasserstein setting.

Theorem III.41 *Let $\mu_0 \in D(\mathcal{F})$. Let $T \in (0,\infty)$ and $(\mu_t)_{t\in[0,T]} \subset \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R})$ be such that $\lim_{t\to 0}\mathrm{W}^{\mathrm{L}}(\mu_t,\mu_0) = 0$. Then $(\mu_t)_t$ is the gradient flow for $\mathcal{F}$ if and only if*

(i) $\mu_t = \rho_t\,\mathrm{Leb}_{\mathbb{T}^d\times\mathbb{R}}$ *for all* $t \in [0,T]$,

(ii) the curve of densities $(\rho_t)_t$ is a weak solution to

$$\partial_t \rho_t(x,\theta) = \partial^2_{\theta\theta}\rho_t(x,\theta) + \partial_\theta\left(\rho_t(x,\theta)\left(\Psi'(\theta) - \int J(x-\bar{x})\bar{\theta}\rho_t(\bar{x},\bar{\theta})\,d\bar{\theta}d\bar{x}\right)\right), \tag{III.1.135}$$

(iii) $\int_0^T |\partial\mathcal{F}|^2(\mu_t)\,dt < \infty$.

Proof. If $(\mu_t)_t$ is the gradient flow for $\mathcal{F}$, then Theorem III.27 *(v)* and *(iii)* imply the claims *(i)* and *(iii)*, respectively. Claim *(ii)* follows immediately from Proposition III.38 and the fact that $(\mu_t)_t$ satisfies the continuity equation.

Conversely, assume *(i)–(iii)*. *(iii)* implies that $|\partial\mathcal{F}|(\mu_t) < \infty$ for almost every t. Therefore, Proposition III.38 is applicable and we obtain that for almost every t, $\partial_\theta\rho_t$ exists weakly and

$$w_t := \frac{\partial_\theta\rho_t}{\rho_t} + \Psi' - \int J(\cdot-\bar{x})\bar{\theta}\,d\mu_t(\bar{x},\bar{\theta}) \in \mathrm{Tan}_{\mu_t}\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}) \cap \partial\mathcal{F}(\mu_t). \tag{III.1.136}$$

Moreover, *(ii)* shows that $(\mu_t)_t$ solves the continuity equation with respect to $(-w_t)_t$. And since *(iii)* also shows that $w \in \mathrm{L}^2((0,T)\times\mathbb{T}^d\times\mathbb{R};\mu_t dt)$, we infer via Proposition III.10 *(B)* that $(\mu_t)_t \in \mathcal{AC}((0,T);\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d\times\mathbb{R}))$ with tangent velocity $-w$. And since $w_t \in \partial\mathcal{F}(\mu_t)$ for almost every t, we conclude that $(\mu_t)_t$ must be the gradient flow for $\mathcal{F}$ with initial value μ_0. □

III.2 Large deviation principle

In this section we derive the large deviation principle for the system introduced in Subsection I.5.1. First, in Subsection III.2.1, we rigorously introduce the model and state some properties. Then we define in Subsection III.2.2 the empirical measure map and in Subsection III.2.3 we state the main result and its proof. The proof of the lower bound and the recovery sequence are moved to Subsection III.2.4 and Subsection III.2.5, respectively. For convenience purposes, from now on we restrict to the case $d = 1$. Throughout the remaining part of this chapter suppose Assumption III.33 and let $T \in (0,\infty)$.

III.2.1 The microscopic system

Let $N \in \mathbb{N}$. Recall that the *microscopic Hamiltonian* $\mathrm{H}^N : \mathbb{R}^N \to \mathbb{R}$ is given by

$$\mathrm{H}^N(\Theta) = \sum_{i=0}^{N-1}\left(\Psi\left(\theta^i\right) - \frac{1}{2N}\sum_{j=0}^{N-1} J\left(\frac{i-j}{N}\right)\theta^i\theta^j\right). \tag{III.2.1}$$

Define $\mathcal{H}^N : \mathcal{P}_2(\mathbb{R}^N) \to (-\infty,\infty]$ by

$$\mathcal{H}^N(\cdot) := \mathcal{H}(\cdot \mid \exp(\mathrm{H}^N)\mathrm{Leb}_{\mathbb{R}^N}). \tag{III.2.2}$$

Analogously to (III.1.132), define $\mathscr{J}^N : C([0,T];\mathcal{M}_1(\mathbb{R}^N)) \to [0,\infty]$ by

$$\mathscr{J}^N[(\nu_t^N)_t] := \mathcal{H}^N(\nu_T^N) - \mathcal{H}^N(\nu_0^N) + \frac{1}{2}\int_0^T \left(|\partial\mathcal{H}^N|^2(\nu_t^N) + |(\nu^N)'|^2(t)\right)dt \tag{III.2.3}$$

if $(\nu_t^N)_t \in \mathcal{AC}((0,T)\,;\,\mathcal{P}_2(\mathbb{R}^N))$ and $\mathscr{J}^N[(\nu_t^N)_{t\in[0,T]}] = \infty$ else.

Lemma III.42 *Recall the parameters from Assumption III.33, Lemma III.34 and Theorem III.35. Then,*

(i) $\mathcal{H}^N$ *is proper,* $(\bar{\lambda}+\hat{\lambda})$*-convex, strongly* λ*-convex,* W_2*-l.s.c. and coercive,*

(ii) for all $\nu^N \in \mathcal{M}_1(\mathbb{R}^N)$*, for some constant* $C'' > 0$*,*

$$\frac{1}{N}\mathcal{H}^N(\nu^N) \geq \frac{1}{N}\int_{\mathbb{R}^N}\Big(C_\Psi \sum_{i=0}^{N-1}|\theta^i|^{2\ell} + (C'_\Psi - \|J\|_\infty)\,|\Theta|^2\Big)d\nu^N(\Theta) - C'', \tag{III.2.4}$$

(iii) for all $\mu_0^N \in D(\mathcal{H}^N)$*, there exists a unique curve* $(\mu_t^N)_t \in \mathcal{AC}((0,T)\,;\,\mathcal{P}_2(\mathbb{R}^N))$ *such that* $\lim_{t\to 0} W_2(\mu_t^N,\mu_0^N) = 0$ *and* $\mathscr{J}^N[(\mu_t^N)_t] = 0$*. We call* $(\mu_t^N)_t$ the Wasserstein gradient flow for $\mathcal{H}^N$ *with initial value* μ_0^N*,*

(iv) there exists $Q^N \in \mathcal{M}_1(C([0,T];\mathbb{R}^N))$ *such that* $(e_t)_\# Q^N = \mu_t^N$ *for all* $t \in [0,T]$ *and* Q^N *is the law of the trajectories on* $[0,T]$ *of a solution to*

$$d\Theta_t^N = -\nabla \mathrm{H}^N(\Theta_t^N)\,dt + \sqrt{2}\,dB_t^N \quad \text{and} \quad \Theta_0^N \sim \mu_0^N. \tag{III.2.5}$$

Proof. (i) follows from [3, 9.3.9], [3, 9.3.2] and [3, 9.2.7].

To show *(ii)*, let, without restriction, $\nu^N \in \mathcal{M}_1(\mathbb{R}^N)$ be such that $\mathcal{H}^N(\nu^N) < \infty$. Then,

$$\begin{aligned}\frac{1}{N}\mathcal{H}^N(\nu^N) = \frac{1}{N}\mathcal{H}\left(\mu^N \,\middle|\, \exp\left(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\right)d\Theta\right)\\ + \frac{1}{2N^2}\sum_{k,j=0}^{N-1}\int_{\mathbb{R}^N}\left(\frac{1}{2}\Psi(\theta^k)+\frac{1}{2}\Psi(\theta^j) - J\left(\frac{k-j}{N}\right)\theta^k\theta^j\right)d\nu^N(\Theta).\end{aligned} \tag{III.2.6}$$

Proceeding as in the proof of Lemma III.34 yields part *(ii)*.

(iii) is a consequence of [3, 11.2.1] and part *(i)*.

(iv) follows from [105, 3.3]. □

For technical reasons we have to restrict the choice of the sequence $(\mu_0^N)_N$ of initial values in the following way.

Assumption III.43 *For all* $N \in \mathbb{N}$*,* $\mu_0^N \in \mathcal{M}_1(\mathbb{R}^N)$ *is given by* $d\mu_0^N(\Theta) = \rho_0^N(\Theta)\,d\Theta$*, where*

$$\rho_0^N(\Theta) := \prod_{k=0}^{N-1}\kappa\left(\tfrac{k}{N},\theta^k\right)\mathrm{e}^{-\Psi(\theta^k)}, \tag{III.2.7}$$

where $\kappa : \mathbb{T}\times\mathbb{R} \to [0,\infty)$ *is upper semi-continuous and such that*

- $\int_{\mathbb{R}} \kappa(x,\theta)\,\mathrm{e}^{-\Psi(\theta)}\,d\theta = 1$ *for each* $x \in \mathbb{T}$*,*
- *the restriction* $\kappa : \{\kappa > 0\} \to (0,\infty)$ *is a continuous map, where* $\{\kappa>0\} := \{(x,\theta) \in \mathbb{T}\times\mathbb{R} \,|\, \kappa(x,\theta) > 0\}$*,*
- $\kappa(x,\theta) \leq C_\kappa \exp(\frac{1}{8}C_\Psi\,\theta^{2\ell} + \frac{1}{8}(C'_\Psi - \|J\|_\infty)\theta^2)$ *for some* $C_\kappa > 0$*, and*
- *either* $\kappa(x,\theta) \geq c'_\kappa \exp(-c_\kappa\Psi(\theta))$ *on* $\{\kappa>0\}$ *for some* $c_\kappa, c'_\kappa > 0$*, or* $x \mapsto \kappa(x,\theta)$ *is constant for all* $\theta \in \mathbb{R}$*.*

III.2.2 The empirical measure map.

For all $N \in \mathbb{N}$, define the *empirical measure map* $K^N : \mathbb{R}^N \to \mathcal{M}_1(\mathbb{T} \times \mathbb{R})$ by

$$K^N(\Theta) = \frac{1}{N} \sum_{k=0}^{N-1} \delta_{(\frac{k}{N}, \theta^k)}. \tag{III.2.8}$$

Moreover, let $K_T^N : C([0,T]\,; \mathbb{R}^N) \to C([0,T]\,; \mathcal{M}_1(\mathbb{T} \times \mathbb{R}))$ be defined by

$$K_T^N((\Theta_t)_{t\in[0,T]}) = (K^N(\Theta_t))_{t\in[0,T]}. \tag{III.2.9}$$

For technical reasons it is sometimes useful to consider a modification of K^N defined by

$$\begin{aligned} L^N : \mathbb{R}^N &\to \mathcal{M}_1^{\mathrm{L}}(\mathbb{T} \times \mathbb{R}) \\ \Theta &\mapsto \sum_{k=0}^{N-1} \mathrm{Leb}_{A_{k,N}} \otimes \delta_{\theta^k}, \end{aligned} \tag{III.2.10}$$

where $(A_{k,N})_{k=0}^{N-1}$ is a partition of $\mathbb{T}$ given by

$$A_{k,N} = [kN, (k+1)N), \qquad k = 0, \dots, N-1. \tag{III.2.11}$$

In the following lemma we show that L^N is indeed just a small modification of K^N.

Lemma III.44 *(i) Let $N \in \mathbb{N}$ and $\Theta \in \mathbb{R}^N$. Then $W_2(K^N(\Theta), L^N(\Theta)) \le \frac{1}{N}$.*

(ii) Let $\widetilde{\mathbb{W}}$ denote the Wasserstein distance on $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T} \times \mathbb{R}))$ induced by the distance $\widetilde{W}$ on $\mathcal{M}_1(\mathbb{T} \times \mathbb{R})$. Then

$$\widetilde{\mathbb{W}}\left((L^N)_{\#}\mu^N, (K^N)_{\#}\mu^N\right) \le \frac{1}{N} \qquad \forall\, \mu^N \in \mathcal{M}_1(\mathbb{R}^N). \tag{III.2.12}$$

Proof. Define $G : \mathbb{T} \times \mathbb{R} \to \mathbb{T} \times \mathbb{R}$ by

$$G(x,\theta) = \sum_{k=0}^{N-1} \mathbb{1}_{A_{k,n}}(x) \left(\frac{k}{N}, \theta\right). \tag{III.2.13}$$

Then $(G, \mathrm{Id}_{\mathbb{T}\times\mathbb{R}})_{\#} L^N(\Theta) \in \mathrm{Cpl}(K^N(\Theta), L^N(\Theta))$. Estimating $W_2(K^N(\Theta), L^N(\Theta))$ by the cost with respect to this coupling yields *(i)*. Finally, *(ii)* follows immediately from part *(i)*. □

III.2.3 The large deviation principle.

Definition III.45 *Let (X, d) be a Polish space. Let $(\Pi_n)_{n\in\mathbb{N}}$ be a family of probability measures on X and let $I : X \to [0,\infty]$ be* d*-l.s.c. Then $(\Pi_n)_n$ is said to satisfy a* large deviation principle (LDP) *on X with* rate function *I if*

(i) for any closed set $C \subset X$, $\limsup_{n\to\infty} \frac{1}{n} \log \Pi_n(C) \le -\inf_{x\in C} I(x)$, and

(ii) for any open set $O \subset X$, $\liminf_{n\to\infty} \frac{1}{n} \log \Pi_n(O) \ge -\inf_{x\in O} I(x)$.

Recall that $\mathcal{M}_1(\mathbb{T}\times\mathbb{R})$ is equipped with the metric $\widetilde{W}$. Let $C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ be equipped with the supremum norm induced by $\widetilde{W}$. Theorem III.47 below states the LDP result for the sequence $\{(K_T^N)_\# Q^N\}_N$ on $C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$, where, for all $N\in\mathbb{N}$, Q^N is the measure from Lemma III.42 *(iv)*. The rate function will be given by

$$I[(\nu_t)_t] = \begin{cases} \frac{1}{2}\mathscr{J}[(\nu_t)_t] + \mathcal{H}(\nu_0\,|\,\mu_0) & \text{if } (\nu_t)_t \in \mathcal{AC}([0,T]\,;\,\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})),\\ \infty & \text{else,}\end{cases} \tag{III.2.14}$$

where $\mu_0 = \rho_0\,\mathrm{Leb}_{\mathbb{T}\times\mathbb{R}} \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})$ with $\rho_0(x,\theta) = \kappa(x,\theta)\,\mathrm{e}^{-\Psi(\theta)}$. Before we state and prove the LDP result, we need to show the lower semi-continuity of I.

Lemma III.46 *$(\nu_t)_t \mapsto I[(\nu_t)_t]$ is lower semi-continuous in $C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$.*

Proof. Let $\lim_{m\to\infty}(\nu_t^m)_t = (\nu_t)_t$ in $C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$. In particular, $\nu_t^m \rightharpoonup \nu_t$ for all $t\geq 0$. Without restriction assume that $\liminf_{m\to\infty} I[(\nu_t^m)_t] < \infty$, since otherwise, the claim is trivial. Moreover, by considering appropriate subsequences, we can even suppose that $\sup_{m\in\mathbb{N}} I[(\nu_t^m)_t] < \infty$. In particular, $\sup_{m\in\mathbb{N}}\mathscr{J}[(\nu_t^m)_t]$, $\sup_{m\in\mathbb{N}}\mathcal{H}(\nu_0^m\,|\,\mu_0) < \infty$, since both terms are non-negative. The proof is divided into seven steps.

Step 1. [$\inf_{m\in\mathbb{N}}\mathcal{H}(\nu_0^m\,|\,\mu_0) - \frac{1}{2}\mathcal{F}(\nu_0^m) > -\infty$.]
Note that, since $\sup_{m\in\mathbb{N}}\mathcal{H}(\nu_0^m\,|\,\mu_0) < \infty$, κ is strictly positive inside the support of ν_0^m. Then, similarly as in the proof of Theorem III.35

$$\begin{aligned}\mathcal{H}(\nu_0^m\,|\,\mu_0) - \tfrac{1}{2}\mathcal{F}(\nu_0^m) &= \frac{1}{2}\mathcal{H}\Big(\nu_0^m\,\Big|\,\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\Big) \\ &\quad + \frac{1}{4}\int_{(\mathbb{T}\times\mathbb{R})^2}\Big(\tfrac{1}{2}\Psi(\theta) + \tfrac{1}{2}\Psi(\bar\theta) + J(x-\bar x)\theta\bar\theta - 2\log\kappa(x,\theta) - 2\log\kappa(\bar x,\bar\theta)\Big)\,d(\nu_0^m\otimes\nu_0^m).\end{aligned} \tag{III.2.15}$$

By using Assumption III.33 and Assumption III.43, we have that

$$\mathcal{H}\Big(\nu_0^m\,\Big|\,\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\Big) \geq -\log\int \mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta, \quad \text{and} \tag{III.2.16}$$

$$\tfrac{1}{2}\Psi(\theta) + \tfrac{1}{2}\Psi(\bar\theta) + J(x-\bar x)\theta\bar\theta - 2\log\kappa(x,\theta) - 2\log\kappa(\bar x,\bar\theta) \geq -C_\Psi'' - 4\log(C_\kappa). \tag{III.2.17}$$

Combining (III.2.15), (III.2.16) and (III.2.17) concludes the claim of Step 1.

Step 2. [$\sup_{m\in\mathbb{N}}\int_0^T |(\nu^m)'|^2(r)\,dr < \infty$ and $\sup_{m\in\mathbb{N}}\int_0^T |\partial\mathcal{F}|^2(\nu_r^m)\,dr < \infty$.]
Using Step 1, the fact that $\sup_{m\in\mathbb{N}} I[(\nu_t^m)_t] < \infty$, and Lemma III.34, we infer the claim.

Step 3. [$\sup_{m\in\mathbb{N}}\sup_{t\in[0,T]}\mathcal{F}(\nu_t^m) < \infty$ and $\sup_{m\in\mathbb{N}}\sup_{t\in[0,T]}\int_{\mathbb{T}\times\mathbb{R}}|\theta|^2\,d\nu^m < \infty$.]
Since $|\partial\mathcal{F}|$ is a so-called strong upper gradient ([3, 1.2.1 and 2.4.10]), we infer that

$$\sup_{m\in\mathbb{N}}\sup_{t\in[0,T]}\mathcal{F}(\nu_t^m) \leq \sup_{m\in\mathbb{N}}\sup_{t\in[0,T]}\int_t^T |\partial\mathcal{F}|(\nu_r^m)\,|(\nu^m)'|(r)\,dr + \mathcal{F}(\nu_T^m) < \infty, \tag{III.2.18}$$

where we used Step 2 in the last step. The second claim is shown by combining (III.2.18) with Lemma III.34.

Step 4. [$\liminf_{m\to\infty}\Big(\mathcal{F}(\nu_T^m) + \frac{1}{2}\int_0^T |\partial\mathcal{F}|^2(\nu_t^m)\,dt\Big) \geq \mathcal{F}(\nu_T) + \frac{1}{2}\int_0^T |\partial\mathcal{F}|^2(\nu_t)\,dt$.]
The claim follows from a combination of Theorem III.35, Fatou's lemma, Step 3 and Corollary III.39.

Step 5. [$\liminf_{m\to\infty} \mathcal{H}(\nu_0^m \,|\, \mu_0) - \frac{1}{2}\mathcal{F}(\nu_0^m) \geq \mathcal{H}(\nu_0^m \,|\, \mu_0) - \frac{1}{2}\mathcal{F}(\nu_0^m)$.]
Recall (III.2.15), and recall that we have already seen in the proof of Theorem III.35 that

$$\liminf_{m\to\infty} \frac{1}{2}\mathcal{H}\Big(\nu_0^m \,\Big|\, \mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\Big) \geq \frac{1}{2}\mathcal{H}\Big(\nu_0 \,\Big|\, \mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\Big). \tag{III.2.19}$$

The integrand in the second term on the right-hand side of (III.2.15) is lower semi-continuous and bounded from below by Assumption III.33 and Assumption III.43. Therefore, analogously to (III.1.110), [3, 5.1.7] yields the lower semi-continuity of this term.

Step 6. [$(\nu_t)_t \in \mathcal{AC}([0,T]\,;\,\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$.]
According to [96, Lemma 1], it suffices to show that

$$\sup_{0<h<T}\int_0^{T-h}\frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\nu_t,\nu_{t+h})^2\,dt<\infty \quad\text{and}\quad \int_0^T \mathrm{W}^{\mathrm{L}}(\nu_t,\delta_0\otimes\mathrm{Leb}_{\mathbb{T}})^2\,dt<\infty. \tag{III.2.20}$$

Since $\mathrm{W}^{\mathrm{L}}(\nu_t,\delta_0\otimes\mathrm{Leb}_{\mathbb{T}})^2=\int|\theta|^2\,d\nu_t$, Step 3 and [3, 5.1.7] imply the second claim in (III.2.20). In order to show the first claim in (III.2.20), note that $(\nu_t^m)_t\in\mathcal{AC}([0,T]\,;\,\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$ for all m. Then, using Fatou's lemma and Lemma III.4, we obtain that

$$\begin{aligned}\sup_{0<h<T}\int_0^{T-h}\frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\nu_t,\nu_{t+h})^2\,dt &\leq \sup_{0<h<T}\liminf_{m\to\infty}\int_0^{T-h}\frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\nu_t^m,\nu_{t+h}^m)^2\,dt\\ &\leq \sup_{0<h<T}\liminf_{m\to\infty}\int_0^{T-h}\frac{1}{h}\int_t^{t+h}|(\nu^m)'|^2(r)\,dr\,dt\\ &\leq \liminf_{m\to\infty}\int_0^T|(\nu^m)'|^2(r)\,dr<\infty,\end{aligned} \tag{III.2.21}$$

where we have used Fubini's theorem in the last step.

Step 7. [$\int_0^T|\nu'|^2(t)\,dt\leq\liminf_{m\to\infty}\int_0^T|(\nu^m)'|^2(r)\,dr$.]
Let $\varepsilon\in(0,T/2)$. Then, repeating the arguments from (III.2.21),

$$\int_0^{T-\varepsilon}|\nu'|^2(t)\,dt \leq \liminf_{h\downarrow 0,h<\varepsilon}\int_0^{T-\varepsilon}\frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\nu_t,\nu_{t+h})^2\,dt\leq\liminf_{m\to\infty}\int_0^T|(\nu^m)'|^2(r)\,dr. \tag{III.2.22}$$

Letting $\varepsilon\downarrow 0$ concludes the proof. □

Theorem III.47 *Let $(\mu_0^N)_N$ satisfy Assumption III.43. For all $N\in\mathbb{N}$, let $(\mu_t^N)_{t\in[0,T]}$ be the Wasserstein gradient flow for $\mathcal{H}^N$ with initial value μ_0^N and $(Q^N)_N$ be the corresponding representation measures from Lemma III.42 (iv). Then the sequence $\{(K_T^N)_\# Q^N\}_N$ satisfies a large deviation principle on $C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ with rate function I.*

Proof. In [100, Theorem 3.4 and Theorem 3.5], it is shown that the above LDP result for $\{(K_T^N)_\# Q^N\}_N$ is true if and only if the following three conditions are satisfied.

(i) The family $\{(K_T^N)_\# Q^N\}_N$ is exponentially tight, i.e. for all $s>0$ there exists a compact set $\mathcal{K}_s\subset C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ such that

$$\limsup_{N\to\infty}\frac{1}{N}\log\Big((K_T^N)_\# Q^N(\mathcal{K}_s^c)\Big)\leq -s. \tag{III.2.23}$$

(ii) For all $(\nu_t)_t \in C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ and for all $(\Gamma^N)_N \subset \mathcal{M}_1(C([0,T];\mathcal{M}_1(\mathbb{T}\times\mathbb{R})))$ that converge to $\delta_{(\nu_t)_t}$ weakly in $\mathcal{M}_1(\,C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))\,)$, it holds

$$\liminf_{N\to\infty}\frac{1}{N}\mathcal{H}\left(\Gamma^N\,\middle|\,(K_T^N)_\#Q^N\right)\geq I[(\nu_t)_t]. \tag{III.2.24}$$

(iii) For all $(\nu_t)_t \in C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ there exists $(\Gamma^N)_N \subset \mathcal{M}_1(\,C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))\,)$ such that $(\Gamma^N)_N$ converges to $\delta_{(\nu_t)_t}$ weakly in $\mathcal{M}_1(\,C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))\,)$ and

$$\limsup_{N\to\infty}\frac{1}{N}\mathcal{H}\left(\Gamma^N\,\middle|\,(K_T^N)_\#Q^N\right)\leq I[(\nu_t)_t]. \tag{III.2.25}$$

Fact *(i)* was proven in [105, 3.29].

To prove *(ii)*, note that if the left-hand side of (III.2.24) is infinite, the claim is trivial. Therefore, we assume without restriction that

$$\mathcal{H}\left(\Gamma^N\,\middle|\,(K_T^N)_\#Q^N\right)<\infty\quad\text{ for all } N\in\mathbb{N}. \tag{III.2.26}$$

This implies in particular that Γ^N is absolutely continuous with respect to $(K_T^N)_\#Q^N$ for all N. Since the map K_T^N is injective, we infer that for all N there is a $P^N\in\mathcal{M}_1(\,C([0,T]\,;\,\mathbb{R}^N)\,)$ such that $\Gamma^N=(K_T^N)_\#P^N$. Moreover,

$$\mathcal{H}\left((K_T^N)_\#P^N\,\middle|\,(K_T^N)_\#Q^N\right)=\mathcal{H}\left(P^N\,\middle|\,Q^N\right)\quad\text{ for all } N\in\mathbb{N}, \tag{III.2.27}$$

which is again a consequence of the injectivity of K_T^N. Now we can use [70, 4.1.(i)] to observe that

$$\mathcal{H}\left(P^N\,\middle|\,Q^N\right)\geq\frac{1}{2}\mathscr{I}^N[(\nu_t^N)_t]+\mathcal{H}\left(\nu_0^N\,\middle|\,\mu_0^N\right)\quad\text{ for all } N\in\mathbb{N}, \tag{III.2.28}$$

where $\nu_t^N:=(e_t)_\#P^N$ for all t. In particular, the right-hand side in (III.2.28) is finite, which implies that $(\nu_t^N)_t\in\mathcal{AC}([0,T];\mathcal{P}_2(\mathbb{R}^N))$. Hence, in order to prove *(ii)*, it is enough to show that

$$\liminf_{N\to\infty}\frac{1}{N}\left(\frac{1}{2}\mathscr{I}^N[(\nu_t^N)_t]+\mathcal{H}\left(\nu_0^N\,\middle|\,\mu_0^N\right)\right)\geq I[(\nu_t)_t], \tag{III.2.29}$$

whenever $((K^N)_\#\nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all t, where $(\nu_t^N)_t\in\mathcal{AC}([0,T];\mathcal{P}_2(\mathbb{R}^N))$ for all N. This is the content of Proposition III.48 below.

It remains to prove *(iii)*. If $I[(\nu_t)_t]=\infty$, we take $\Gamma^N=\delta_{(\nu_t)_t}$ for all N and (III.2.25) is trivially satisfied. So assume that $I[(\nu_t)_t]<\infty$. In particular, $(\nu_t)_t\in\mathcal{AC}([0,T];\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$. Proposition III.56 below shows that there exists $(\nu_t^N)_t\in C([0,T];\mathcal{M}_1(\mathbb{R}^N))$ such that $((K^N)_\#\nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all t and

$$\limsup_{N\to\infty}\frac{1}{N}\left(\frac{1}{2}\mathscr{I}^N[(\nu_t^N)_t]+\mathcal{H}(\nu_0^N\,|\,\mu_0^N)\right)\leq I[(\nu_t)_t]. \tag{III.2.30}$$

Further, for all N, [70, 4.1.(ii)] yields the existence of $\tilde{P}^N\subset\mathcal{M}_1(\,C([0,T]\,;\,\mathbb{R}^N)\,)$ such that

$$\frac{1}{2}\mathscr{I}^N[(\nu_t^N)_t]+\mathcal{H}(\nu_0^N\,|\,\mu_0^N)=\mathcal{H}\left(\tilde{P}^N\,\middle|\,Q^N\right) \tag{III.2.31}$$

and $\nu_t^N=(e_t)_\#\tilde{P}^N$ for all t. Hence, in order to prove *(iii)*, it only remains to show that $(\,(K_T^N)_\#\tilde{P}^N\,)_N$ converges to $\delta_{(\nu_t)_t}$ weakly in $\mathcal{M}_1(\,C([0,T]\,;\,\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))\,)$.

Since $((K^N)_\# \nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all t, it suffices to show that $((K_T^N)_\# \tilde{P}^N)_N$ is tight. Let $\varepsilon > 0$. Let $\mathcal{K}$ be the compact set from part *(i)* according to the choice $s = I[(\nu_t)_t]/\varepsilon$. Then, via the entropy inequality (see e.g. [100, (3.7)]), (III.2.23), (III.2.27), (III.2.30) and (III.2.31), we obtain

$$\begin{aligned} \limsup_{N\to\infty} (K_T^N)_\# \tilde{P}^N(\mathcal{K}^c) &\le \limsup_{N\to\infty} \frac{\log 2 + \mathcal{H}\left((K_T^N)_\# \tilde{P}^N \,\middle|\, (K_T^N)_\# Q^N\right)}{\log\left(1 + 1/(K_T^N)_\# Q^N(\mathcal{K}^c)\right)} \\ &\le \limsup_{N\to\infty} \frac{\frac{1}{N}\mathcal{H}\left((K_T^N)_\# \tilde{P}^N \,\middle|\, (K_T^N)_\# Q^N\right)}{-\frac{1}{N}\log\left((K_T^N)_\# Q^N(\mathcal{K}^c)\right)} \quad \le \ \varepsilon, \end{aligned} \tag{III.2.32}$$

which implies the tightness of $((K_T^N)_\# \tilde{P}^N)_N$. □

III.2.4 Lower Bound

Proposition III.48 *Let $(\nu_t)_t \in C([0,T];\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ and $(\nu_t^N)_t \in \mathcal{AC}([0,T];\mathcal{P}_2(\mathbb{R}^N))$ for all $N \in \mathbb{N}$. Suppose that $((K^N)_\# \nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all $t \in [0,T]$. Then*

$$\liminf_{N\to\infty} \frac{1}{N}\left(\frac{1}{2}\mathscr{J}^N[(\nu_t^N)_t] + \mathcal{H}(\nu_0^N \,|\, \mu_0^N)\right) \ge \frac{1}{2}\mathscr{J}[(\nu_t)_t] + \mathcal{H}(\nu_0 \,|\, \mu_0). \tag{III.2.33}$$

Proof. Assume that the left-hand side of (III.2.33) is finite. Otherwise, the claim is trivial. In particular, since both summands are non-negative, we have

$$\liminf_{N\to\infty} \frac{1}{N}\mathscr{J}^N[(\nu_t^N)_t] < \infty \qquad \text{and} \qquad \liminf_{N\to\infty} \frac{1}{N}\mathcal{H}(\nu_0^N \,|\, \mu_0^N) < \infty. \tag{III.2.34}$$

Under this assumption, we show (III.2.33) for each part separately in the forthcoming paragraphs. Hence, the claim follows from the Lemmas III.52–III.55. □

Preliminaries. We first list some consequences of (III.2.34) in the following lemma.

Lemma III.49 *Under the same assumptions as in Proposition III.48 and under* (III.2.34)*, we have*

$$\liminf_{N\to\infty} \frac{1}{N}\int_0^T |(\nu^N)'|^2(t)\,dt < \infty. \tag{III.2.35}$$

Moreover, $(\nu_t)_t$ is an absolutely continuous curve in $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})$.

Proof. **Step 1.** $\left[\ \inf_{N\in\mathbb{N}} \frac{1}{N}\left(\mathcal{H}(\nu_0^N \,|\, \mu_0^N) - \frac{1}{2}\mathcal{H}^N(\nu_0^N)\right) > -\infty.\ \right]$

Analogously to (III.2.15) and in view of (III.2.6) and (III.2.17) we observe that for all $N \in \mathbb{N}$

$$\begin{aligned}\frac{1}{N}\big(\mathcal{H}(\nu_0^N\,|\,\mu_0^N)-\tfrac{1}{2}\mathcal{H}^N(\nu_0^N)\big)&=\frac{1}{2N}\mathcal{H}\Big(\nu_0^N\,\Big|\,\exp\Big(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\Big)d\Theta\Big)\\&+\frac{1}{4}\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int_{(\mathbb{T}\times\mathbb{R})^2}\Big(\tfrac{1}{2}\Psi(\theta)+\tfrac{1}{2}\Psi(\bar\theta)+J(x-\bar x)\theta\bar\theta\\&\qquad-2\log\kappa(x,\theta)-2\log\kappa(\bar x,\bar\theta)\Big)\,d\gamma\,d\gamma\,d(K^N)_\#\nu_0^N(\gamma)\qquad\text{(III.2.36)}\\&\geq\frac{1}{2N}\mathcal{H}\Big(\nu_0^N\,\Big|\,\exp\Big(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\Big)d\Theta\Big)-\frac{1}{4}C''_\Psi-\log(C_\kappa)\\&\geq-\frac{1}{2}\log\int\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta-\frac{1}{4}C''_\Psi-\log(C_\kappa)>-\infty.\end{aligned}$$

Step 2. $\big[\ \liminf_{N\to\infty}\frac{1}{N}\int_0^T|(\nu^N)'|^2(t)\,dt<\infty.\ \big]$

Step 1, Lemma III.42 *(ii)* and the finiteness of the left-hand side of (III.2.33) yield the claim.

Step 3. $\big[\ \liminf_{N\to\infty}\frac{1}{N}\int_{\mathbb{R}^N}(|\Theta|^2+\sum_{i=0}^{N-1}|\theta^i|^{2\ell})\,d\nu_0^N<\infty.\ \big]$

By a similar computation as in Step 1, (III.2.34) yields that

$$\begin{aligned}\infty>\liminf_{N\to\infty}\frac{1}{N}\mathcal{H}(\nu_0^N\,|\,\mu_0^N)&=\liminf_{N\to\infty}\frac{1}{N}\mathcal{H}\Big(\nu_0^N\,\Big|\,\exp\Big(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\Big)d\Theta\Big)\\&+\liminf_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\left(-\int_{\mathbb{R}^N}\log\kappa(\tfrac{k}{N},\theta^k)d\nu_0^N+\frac{1}{2}\int_{\mathbb{R}^N}\Psi(\theta^k)d\nu_0^N\right)\qquad\text{(III.2.37)}\\&\geq C\ +\ \frac{1}{4}\ \liminf_{N\to\infty}\frac{1}{N}\int_{\mathbb{R}^N}\sum_{k=0}^{N-1}\Psi(\theta^k)\,d\nu_0^N\end{aligned}$$

for some $C\in\mathbb{R}$. Finally, (III.1.103) implies Step 3.

Step 4. $\big[\ \liminf_{N\to\infty}\frac{1}{N}\sup_{t\in[0,T]}\int_{\mathbb{R}^N}|\Theta|^2\,d\nu_t^N<\infty.\ \big]$

Step 2 and 3 imply that

$$\begin{aligned}\liminf_{N\to\infty}\frac{1}{N}\sup_{t\in[0,T]}\int_{\mathbb{R}^N}|\Theta|^2\,d\nu_t^N&\leq\liminf_{N\to\infty}4\frac{1}{N}\left(\sup_{t\in[0,T]}W_2(\nu_t^N,\nu_0^N)^2dt+\int_{\mathbb{R}^N}|\Theta|^2\,d\nu_0^N\right)\\&\leq\liminf_{N\to\infty}4\frac{1}{N}\left(T\int_0^T|(\nu^N)'|^2(t)dt+\int_{\mathbb{R}^N}|\Theta|^2\,d\nu_0^N\right)<\infty.\qquad\text{(III.2.38)}\end{aligned}$$

Step 5. $\big[\ \sup_{t\in[0,T]}\int_{\mathbb{T}\times\mathbb{R}}|\theta|^2\ d\nu_t(x,\theta)\ <\infty.\ \big]$

Using that $((K^N)_\#\nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$, and using [3, 5.1.7], Fatou's lemma and Step 4 we obtain that

$$\sup_{t\in[0,T]}\int_{\mathbb{T}\times\mathbb{R}}|\theta|^2\ d\nu_t\ \leq\liminf_{N\to\infty}\sup_{t\in[0,T]}\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int_{\mathbb{T}\times\mathbb{R}}|\theta|^2\,d\gamma\ d(K^N)_\#\nu_t^N(\gamma)\ <\infty.\qquad\text{(III.2.39)}$$

Step 6. $\big[\ \nu_t\in\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})\text{ for all }t\in[0,T].\ \big]$

Since $((K^N)_\#\nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$, we have that for all $f\in C_b(\mathbb{T})$

$$\begin{aligned}\int_{\mathbb{T}\times\mathbb{R}} f(x)\,d\nu_t(x,\theta) &= \lim_{N\to\infty}\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int_{\mathbb{T}\times\mathbb{R}} f(x)\,d\gamma(x,\theta)\,d(K^N)_\#\nu_t^N(\gamma)\\ &= \lim_{N\to\infty}\int_{\mathbb{T}\times\mathbb{R}} f\left(\frac{\lfloor xN\rfloor}{N}\right)\,dx = \int_{\mathbb{T}\times\mathbb{R}} f\,(x)\;dx.\end{aligned} \tag{III.2.40}$$

Step 7. [$t\mapsto\nu_t$ is absolutely continuous.]
Analogously to the proof of Lemma III.46, it suffices to show that

$$\sup_{0<h<T}\int_0^{T-h}\frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\nu_t,\nu_{t+h})^2\,dt<\infty \qquad\text{and}\qquad \int_0^T\int_{\mathbb{T}\times\mathbb{R}}|\theta|^2\,d\nu_t(x,\theta)\,dt<\infty. \tag{III.2.41}$$

The second claim was shown in Step 5. The first claim in (III.2.41) follows from similar arguments as in (III.2.21). Indeed, using Lemma III.44 and the Lemmas III.50 and III.51 below, we observe that

$$\begin{aligned}\sup_{0<h<T}\int_0^{T-h}\frac{1}{h^2}\mathrm{W}^{\mathrm{L}}(\nu_t,\nu_{t+h})^2dt &\le \sup_{0<h<T}\int_0^{T-h}\liminf_{N\to\infty}\frac{1}{h^2}\mathbb{W}^{\mathrm{L}}\Big((L^N)_\#\nu_t^N\,,\,(L^N)_\#\nu_{t+h}^N\Big)^2dt\\ &\le \sup_{0<h<T}\int_0^{T-h}\liminf_{N\to\infty}\frac{1}{h^2N}W_2(\nu_t^N,\nu_{t+h}^N)^2\,dt\\ &\le \liminf_{N\to\infty}\frac{1}{N}\int_0^T|(\nu^N)'|^2(r)\,dr<\infty,\end{aligned} \tag{III.2.42}$$

where we have used Fatou's lemma, Fubini's theorem and Step 2. We conclude the proof. □

Lemma III.50 *Let $\mathbb{W}^{\mathrm{L}}$ denote the Wasserstein distance on $\mathcal{M}_1(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$ induced by W^{L}. Let $(A^N)_N,(B^N)_N\subset\mathcal{M}_1(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$ and $A,B\in\mathcal{M}_1(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$ be such that A^N converges to A and B^N converges to B weakly in $\mathcal{M}_1(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$. Then*

$$\liminf_{N\to\infty}\mathbb{W}^{\mathrm{L}}(A^N,B^N)\ge\mathbb{W}^{\mathrm{L}}(A,B). \tag{III.2.43}$$

Proof. In view of Lemma III.4, the claim is an application of [127, 4.3]. □

Lemma III.51 *Let $\mu^N,\nu^N\in\mathcal{P}_2(\mathbb{R}^N)$. Then*

$$\mathbb{W}^{\mathrm{L}}\Big((L^N)_\#\mu^N\,,\,(L^N)_\#\nu^N\Big)\le\frac{1}{\sqrt{N}}W_2(\mu^N,\nu^N). \tag{III.2.44}$$

Proof. Let $\pi^N\in\mathrm{Opt}(\mu^N,\nu^N)$. Define

$$\begin{aligned}G^N:\mathbb{R}^N\times\mathbb{R}^N&\to\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})\times\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})\\ (\Theta,\bar\Theta)&\mapsto\Big(L^N(\Theta),L^N(\bar\Theta)\Big).\end{aligned} \tag{III.2.45}$$

Set $\gamma^N=(G^N)_\#\pi^N\in\mathcal{M}_1(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})\times\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$. Then γ^N has $(L^N)_\#\mu^N$ and $(L^N)_\#\nu^N$ as marginals. Therefore,

$$\begin{aligned}\mathbb{W}^{\mathrm{L}}\Big((L^N)_\#\mu^N\,,\,(L^N)_\#\nu^N\Big)^2 &\le \int_{(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))^2}\mathrm{W}^{\mathrm{L}}(\sigma,\bar\sigma)^2\;d\gamma^N(\sigma,\bar\sigma)\\ &= \int_{(\mathbb{R}^N)^2}\frac{1}{N}\sum_{k=0}^{N-1}W_2(\delta_{\theta^k},\delta_{\bar\theta^k})^2\;d\pi^N(\Theta,\bar\Theta)=\frac{1}{N}W_2(\mu^N,\nu^N)^2,\end{aligned} \tag{III.2.46}$$

which concludes the proof. □

McKean-Vlasov-functional. Here we can even show a more general statement which will be useful in the next section.

Lemma III.52 *Let $\mu^N \in \mathcal{P}_2(\mathbb{R}^N)$ for all $N \in \mathbb{N}$ and let $A \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$. Assume that $((K^N)_\# \mu^N)_N$ converges weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ to A. Then*

$$\liminf_{N\to\infty} \frac{1}{N}\mathcal{H}^N(\mu^N) \geq \int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} \mathcal{F}(\gamma)\, dA(\gamma). \tag{III.2.47}$$

Proof. Recall (III.2.6). Then we observe that

$$\begin{aligned}\frac{1}{N}\mathcal{H}^N(\mu^N) = \frac{1}{N}\mathcal{H}\left(\mu^N \,\middle|\, \exp\left(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\right) d\Theta\right) \\ + \frac{1}{2}\int_{\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})}\int_{(\mathbb{T}^d\times\mathbb{R})^2}\left(\frac{1}{2}\big(\Psi(\theta)+\Psi(\bar\theta)\big) - J(x-\bar x)\theta\bar\theta\right) d\gamma d\gamma\, d(K^N)_\#\mu^N(\gamma).\end{aligned} \tag{III.2.48}$$

Similar arguments as in the proof of Theorem III.35 show that

$$\begin{aligned}\liminf_{N\to\infty} \int_{\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})}\int_{(\mathbb{T}^d\times\mathbb{R})^2}\left(\frac{1}{2}\big(\Psi(\theta)+\Psi(\bar\theta)\big) - J(x-\bar x)\theta\bar\theta\right) d\gamma d\gamma\, d(K^N)_\#\mu^N(\gamma) \\ \geq \int_{\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})}\int_{(\mathbb{T}^d\times\mathbb{R})^2}\left(\frac{1}{2}\big(\Psi(\theta)+\Psi(\bar\theta)\big) - J(x-\bar x)\theta\bar\theta\right) d\gamma d\gamma\, dA(\gamma).\end{aligned} \tag{III.2.49}$$

It remains to show that

$$\liminf_{N\to\infty} \frac{1}{N}\mathcal{H}\left(\mu^N \,\middle|\, \exp\left(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\right) d\Theta\right) \geq \int_{\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})} \mathcal{H}\left(\gamma \,\middle|\, \mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\right) dA(\gamma). \tag{III.2.50}$$

Let $\alpha := \int \mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta$ and for all $n \in \mathbb{N}$, set

$$B^N := (K^N)_\#\left(\frac{1}{\alpha^N}\exp\left(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\right) d\Theta\right) \quad \text{and} \quad A^N := (K^N)_\#\mu^N. \tag{III.2.51}$$

Since the map K^N is injective, we have that

$$\mathcal{H}\Big(A^N \,\Big|\, B^N\Big) = \mathcal{H}\left(\mu^N \,\middle|\, \frac{1}{\alpha^N}\exp\left(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\right) d\Theta\right). \tag{III.2.52}$$

It is an easy adaptation of Sanov's theorem that $(B^N)_N$ satisfies a large deviation principle with rate function $\mathcal{H}\left(\cdot \,\middle|\, \alpha^{-1}\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\right)$; see e.g. [113, Theorem 17] for the details. Therefore, [100, 3.5] implies that

$$\liminf_{N\to\infty} \frac{1}{N}\mathcal{H}\left(A^N \,\middle|\, B^N\right) \geq \int_{\mathcal{M}_1(\mathbb{T}^d\times\mathbb{R})} \mathcal{H}\left(\gamma \,\middle|\, \alpha^{-1}\mathrm{e}^{-\frac{1}{2}\Psi(\theta)}d\theta\right) dA(\gamma). \tag{III.2.53}$$

(III.2.53) and (III.2.52) yield (III.2.50). This concludes the proof. □

Initialization.

Lemma III.53 *Under the same assumptions as in Proposition III.48 and under* (III.2.34)*, we have*

$$\liminf_{N\to\infty} \frac{1}{N}\left(\mathcal{H}(\nu_0^N \,|\, \mu_0^N) - \frac{1}{2}\mathcal{H}^N(\nu_0^N)\right) \geq \mathcal{H}(\nu_0 \,|\, \mu_0) - \frac{1}{2}\mathcal{F}(\nu_0). \tag{III.2.54}$$

Proof. Similarly as in the proof of Lemma III.46 and in Step 1 of the proof of Lemma III.49, we observe that

$$\begin{aligned}
&\liminf_{N\to\infty} \frac{1}{N}\big(\mathcal{H}(\nu_0^N \,|\, \mu_0^N) - \tfrac{1}{2}\mathcal{H}^N(\nu_0^N)\big) \\
&\geq \liminf_{N\to\infty} \frac{1}{N}\mathcal{H}\Big(\nu_0^N \,\Big|\, \exp\Big(-\frac{1}{2}\sum_{k=0}^{N-1}\Psi(\theta^k)\Big)d\Theta\Big) \\
&\quad + \frac{1}{2}\int_{(\mathbb{T}\times\mathbb{R})^2}\Big(\tfrac{1}{2}\Psi(\theta) + \tfrac{1}{2}\Psi(\bar\theta) + J(x-\bar x)\theta\bar\theta - 2\log\kappa(x,\theta) - 2\log\kappa(\bar x,\bar\theta)\Big)\,d\nu_0 d\nu_0,
\end{aligned} \tag{III.2.55}$$

where we have used (III.2.17) and [3, 5.1.7]. Combining (III.2.55) with (III.2.50) yields (III.2.54). □

Metric derivative. Also here we can show directly a more general statement.

Lemma III.54 *Let $(c_t)_{t\in[0,T]} \subset \mathcal{M}_1(\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$ be absolutely continuous with respect to the metric $\mathbb{W}^{\mathrm{L}}$ from Lemma III.50. Let $(\nu_t^N)_t \in \mathcal{AC}([0,T];\mathcal{P}_2(\mathbb{R}^N))$ for all $N\in\mathbb{N}$. Suppose that $((K^N)_\#\nu_t^N)_N$ converges to c_t weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all $t\in[0,T]$. Then,*

$$\liminf_{N\to\infty} \frac{1}{N}\int_0^T |(\nu^N)'|^2(t)\,dt \geq \int_0^T |c'|^2(t)\,dt. \tag{III.2.56}$$

Proof. Similarly as in (III.2.22) and in (III.2.42), we obtain that for all $\varepsilon\in(0,T/2)$

$$\begin{aligned}
\int_0^{T-\varepsilon} |c'|^2(t)\,dt &\leq \liminf_{h\downarrow 0, h<\varepsilon}\int_0^{T-\varepsilon}\frac{1}{h^2}\mathbb{W}^{\mathrm{L}}(c_t,c_{t+h})^2\,dt \\
&\leq \liminf_{h\downarrow 0, h<\varepsilon}\int_0^{T-\varepsilon}\liminf_{N\to\infty}\frac{1}{h^2}\mathbb{W}^{\mathrm{L}}\Big((L^N)_\#\nu_t^N\,,\,(L^N)_\#\nu_{t+h}^N\Big)^2 dt \\
&\leq \liminf_{N\to\infty}\frac{1}{N}\int_0^T |(\nu^N)'|^2(r)\,dr.
\end{aligned} \tag{III.2.57}$$

Letting $\varepsilon\downarrow 0$ concludes the proof. □

Metric slope. Here we postpone the more general statement to Section III.3.

Lemma III.55 *Under the same assumptions as in Proposition III.48 and under* (III.2.34)*, we have*

$$\liminf_{N\to\infty} \frac{1}{N}\int_0^T |\partial\mathcal{H}^N|^2(\nu_t^N)\,dt \geq \int_0^T |\partial\mathcal{F}|^2(\nu_t)\,dt. \tag{III.2.58}$$

Proof. Similarly as in Corollary III.39 one can show that (cf. [59, 4.3] or [3, 10.4.9])

$$|\partial\mathcal{H}^N|(\nu_t^N) = \sup_{\varphi\in C_c^\infty(\mathbb{R}^N;\mathbb{R}^N),\, \|\varphi\|_{\mathrm{L}^2(\nu_t^N)}>0} \frac{\left|\int_{\mathbb{R}^N}\left(\varphi\nabla \mathrm{H}^N - \mathrm{div}\varphi\right) d\nu_t^N\right|}{\|\varphi\|_{\mathrm{L}^2(\nu_t^N)}} \tag{III.2.59}$$

for almost every $t\in[0,T]$. Let $\varphi(\Theta) = (\beta(\frac{k}{N}\,,\,\theta^k))_{k=0}^{N-1}$ for some arbitrary $\beta\in C_c^\infty(\mathbb{T}\times\mathbb{R})$ such that $\|\beta\|_{\mathrm{L}^2(\nu_t)}>0$. This is admissible, since

$$\|\varphi\|^2_{\mathrm{L}^2(\nu_t^N)} = N\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} \|\beta\|^2_{\mathrm{L}^2(\gamma)}\, d(K^N)_\#\nu_t^N(\gamma) \tag{III.2.60}$$

and the right-hand side is greater than zero for N large enough, since $((K^N)_\#\nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$. We obtain

$$\begin{aligned}
&\liminf_{N\to\infty}\frac{1}{N}|\partial\mathcal{H}^N|^2(\nu_t^N)\\
&\qquad\geq \liminf_{N\to\infty}\frac{\left(\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int\left(\beta\left[\Psi'-\int J(\cdot-\bar{x})\bar{\theta}\,d\gamma\right]-\partial_\theta\beta\right)d\gamma\; d(K^N)_\#\nu_t^N(\gamma)\right)^2}{\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int\beta^2 d\gamma\; d(K^N)_\#\nu_t^N(\gamma)}\\
&\qquad= \frac{1}{\|\beta\|^2_{\mathrm{L}^2(\nu_t)}}\left(\int\left(\beta\left[\Psi'-\int J(\cdot-\bar{x})\bar{\theta}\,d\nu_t\right]-\partial_\theta\beta\right)d\nu_t\right)^2.
\end{aligned} \tag{III.2.61}$$

Here we used [3, 5.1.7] in the last step, and that, in view of Step 4 of the proof of Lemma III.49,

$$\begin{aligned}
&\limsup_{k\to\infty}\liminf_{N\to\infty}\left|\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} - \int\int\beta J(\cdot-\bar{x})\left((\bar{\theta}\vee(-k))\wedge k-\bar{\theta}\right)d\gamma d\gamma\; d(K^N)_\#\nu_t^N(\gamma)\right|\\
&\leq \limsup_{k\to\infty}\liminf_{N\to\infty}\|\beta\cdot J\|_\infty\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int_{|\theta|\geq k}(|\bar{\theta}|-k)d\gamma\; d(K^N)_\#\nu_t^N(\gamma)\\
&\leq \limsup_{k\to\infty}\liminf_{N\to\infty}\frac{1}{k}\|\beta\cdot J\|_\infty\int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})}\int|\bar{\theta}|^2 d\gamma\; d(K^N)_\#\nu_t^N(\gamma) \;=\; 0.
\end{aligned} \tag{III.2.62}$$

Then, taking the supremum over β in (III.2.61), we get via Corollary III.39

$$\liminf_{N\to\infty}\frac{1}{N}|\partial\mathcal{H}^N|^2(\nu_t^N)\geq|\partial\mathcal{F}|^2(\nu_t). \tag{III.2.63}$$

Finally, Fatou's lemma yields (III.2.58). □

III.2.5 Recovery sequence

Proposition III.56 *Let $(\nu_t)_{t\in[0,T]}\in\mathcal{AC}([0,T];\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))$ be such that $I[(\nu_t)_t]<\infty$. Then for all $N\in\mathbb{N}$ there exists $(\nu_t^N)_{t\in[0,T]}\in C([0,T];\mathcal{M}_1(\mathbb{R}^N))$ such that $((K^N)_\#\nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all $t\in[0,T]$ and*

$$\limsup_{N\to\infty}\frac{1}{N}\left(\frac{1}{2}\mathscr{J}^N[(\nu_t^N)_t]+\mathcal{H}(\nu_0^N\,|\,\mu_0^N)\right)\leq\frac{1}{2}\mathscr{J}[(\nu_t)_t]+\mathcal{H}(\nu_0\,|\,\mu_0). \tag{III.2.64}$$

Proof. First, we observe that, since $I[(\nu_t)_t] < \infty$, we also have that

$$\mathscr{J}[(\nu_t)_t] < \infty \qquad \text{and} \qquad \mathcal{H}(\nu_0 \,|\, \mu_0) < \infty. \tag{III.2.65}$$

The recovery sequence is given as follows. Recall the partition $(A_{k,N})_{k=0}^{N-1}$ of $\mathbb{T}$ introduced in (III.2.11). Then define, for all $N \in \mathbb{N}$ and for all $t \in [0,T]$, $\nu_t^N \in \mathcal{M}_1(\mathbb{R}^N)$ by

$$d\nu_t^N(\Theta) = \prod_{k=0}^{N-1} N\nu_t(A_{k,N} \times d\theta^k). \tag{III.2.66}$$

Lemma III.57 below shows that $((K^N)_\# \nu_t^N)_N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ for all t. We show (III.2.64) for each part separately. Hence, the claim follows from the Lemmas III.58–III.61 and Lemma III.52. □

Preliminaries. First we note that (III.2.65) implies that ν_0 has a density f_0 with respect to $\mathrm{Leb}_{\mathbb{T}\times\mathbb{R}}$. Moreover, a similar computation as in (III.2.15) shows that

$$\mathcal{H}(\nu_0 \,|\, \mu_0) - \frac{1}{2}\mathcal{F}(\nu_0) > -\infty. \tag{III.2.67}$$

Together with Lemma III.34 and (III.2.65), this yields that

$$\int_0^T \Big(|\nu'|^2(t) + |\partial\mathcal{F}|^2(\nu_t) \Big)\, dt < \infty \qquad \text{and} \qquad \mathcal{F}(\nu_T) < \infty. \tag{III.2.68}$$

Since $|\partial\mathcal{F}|$ is a strong upper gradient ([3, 1.2.1 and 2.4.10]), from (III.2.68) we infer that

$$\sup_{t\in[0,T]} \mathcal{F}(\nu_t) \le \sup_{t\in[0,T]} \int_t^T |\partial\mathcal{F}|(\nu_r)\, |\nu'|(r)\, dr + \mathcal{F}(\nu_T) < \infty. \tag{III.2.69}$$

Therefore, for all t, ν_t has a density f_t with respect to $\mathrm{Leb}_{\mathbb{T}\times\mathbb{R}}$. And combining (III.2.69) with the lower bound on $\mathcal{F}$ (Lemma III.34), we infer that

$$\sup_{t\in[0,T]} \int_{\mathbb{T}\times\mathbb{R}} (|\theta|^2 + |\theta|^{2\ell})\, d\nu_t \, < \infty. \tag{III.2.70}$$

Finally, Lemma III.34 and (III.2.67) yield that $\int (|\theta|^2 + |\theta|^{2\ell}) d\nu_0 < \infty$.

Convergence.

Lemma III.57 *Under the same setting as in the proof of Proposition III.56, we have that for all $t \in [0,T]$*

$$(K^N)_\# \nu_t^N \text{ converges to } \delta_{\nu_t} \text{ weakly in } \mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R})). \tag{III.2.71}$$

Proof. For all $N \in \mathbb{N}$ and $t \in [0,T]$ let $\Upsilon_t^N = (\vartheta_t^{k,N})_{k=0,\dots,N-1}$ be a random variable with law ν_t^N on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Step 1. [Let $f \in C_b(\mathbb{T}\times\mathbb{R})$, then $\lim_{N\to\infty} \int f\, d(K^N(\Upsilon_t^N)) = \int f d\nu_t$ a.s.]
The proof is a standard application of Kolmogorov's maximum inequality [54, 9.7.4]. For the sake of completeness, we provide the details. Let $\varepsilon > 0$ and set for all $N \in \mathbb{N}$

$$\frac{1}{N}S_N := \frac{1}{N}\sum_{k=0}^{N-1}\left(f(\tfrac{k}{N},\vartheta_t^{k,N}) - \mathbb{E}\left[f(\tfrac{k}{N},\vartheta_t^{k,N})\right]\right). \tag{III.2.72}$$

Then

$$\begin{aligned}\sum_{p\in\mathbb{N}}\mathbb{P}\left[\max_{2^{p-1}+1\le n\le 2^p}|S_n| > n\varepsilon\right] &\le \sum_{p\in\mathbb{N}}\mathbb{P}\left[\max_{n\le 2^p}|S_n| > 2^p\frac{\varepsilon}{2}\right] \le \sum_{p\in\mathbb{N}}\frac{4}{\varepsilon^2 2^{2p}}\mathrm{Var}[S_{2^p}]\\ &\le \sum_{p\in\mathbb{N}}\frac{4}{\varepsilon^2 2^p}\|f\|_\infty^2 < \infty.\end{aligned} \tag{III.2.73}$$

Hence, the Borel-Cantelli Lemma yields that $\lim_{N\to\infty} S_N/N = 0$ a.s. Finally,

$$\begin{aligned}\int f\,dK^N(\Upsilon_t^N) &= \frac{1}{N}S_N + \frac{1}{N}\sum_{k=0}^{N-1}\mathbb{E}\left[f(\tfrac{k}{N},\vartheta_t^{k,N})\right] = \frac{1}{N}S_N + \sum_{k=0}^{N-1}\int_{A_{k,N}}\int_{\mathbb{R}} f(\tfrac{k}{N},\vartheta)d\nu_t\\ &= \frac{1}{N}S_N + \int_{\mathbb{T}\times\mathbb{R}} f\left(\tfrac{\lfloor xN\rfloor}{N},\vartheta\right)d\nu_t(x,\vartheta) \;\longrightarrow\; \int_{\mathbb{T}\times\mathbb{R}} f\,d\nu_t \quad \text{a.s.}\end{aligned} \tag{III.2.74}$$

Step 2. $\left[\ \mathbb{P}\left[\lim_{N\to\infty}\widetilde{W}(K^N(\Upsilon_t^N),\nu_t) = 0\right] = 1.\ \right]$

The claim follows from Step 1 once we apply the same arguments as in the proof of [54, 11.4.1]. Recall that we have used those arguments already to prove Lemma III.8.

Step 3. $\left[\ (K^N)_\#\nu_t^N$ converges to δ_{ν_t} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))\ \right]$

Step 2 and [54, 9.2.1] yield that $\lim_{N\to\infty}\mathbb{P}\left[\widetilde{W}(K^N(\Upsilon_t^N),\nu_t) > \varepsilon\right] = 0$ for all $\varepsilon > 0$. Hence, [54, 9.3.5] implies the claim. This concludes the proof. □

McKean-Vlasov-functional.

Lemma III.58 *Under the same setting as in the proof of Proposition III.56 and under* (III.2.65)*, we have*

$$\limsup_{N\to\infty}\frac{1}{N}\,\mathcal{H}^N(\nu_t^N) \;\le\; \mathcal{F}(\nu_t) \quad \textit{for all } t\in[0,T]. \tag{III.2.75}$$

In particular, (III.2.75) *holds true for $t = 0$ and $t = T$.*

Proof. Let $t\in[0,T]$. Recall that ν_t has a density f_t and $\int|\theta|^2 d\nu_t < \infty$. We observe that

$$\begin{aligned}\frac{1}{N}\,\mathcal{H}^N(\nu_t^N) &= \frac{1}{2}\sum_{k,j=0}^{N-1}\int_{A_{k,N}}\int_{A_{j,N}}\int_{\mathbb{R}}\int_{\mathbb{R}} J\left(\tfrac{k-j}{N}\right)\theta\bar\theta\,d\nu_t(\bar x,\bar\theta)\,d\nu_t(x,\theta) + O\left(\frac{1}{N}\right)\\ &\quad + \frac{1}{N}\sum_{k=0}^{N-1}\int_{\mathbb{R}}\log\left(\int_{A_{k,N}} f_t(x,\theta)\mathrm{e}^{\Psi(\theta)}N dx\right)\int_{A_{k,N}} f_t(x,\theta)N dx\,d\theta\\ &= \frac{1}{2}\int_{(\mathbb{T}\times\mathbb{R})^2} J\left(\tfrac{\lfloor xN\rfloor-\lfloor\bar xN\rfloor}{N}\right)\theta\bar\theta\,d\nu_t(\bar x,\bar\theta)\,d\nu_t(x,\theta) + O\left(\frac{1}{N}\right)\\ &\quad + \frac{1}{N}\sum_{k=0}^{N-1}\int_{\mathbb{R}}\log\left(\int_{A_{k,N}} f_t(x,\theta)\mathrm{e}^{\Psi(\theta)}N dx\right)\int_{A_{k,N}} f_t(x,\theta)\mathrm{e}^{\Psi(\theta)}N dx\,\mathrm{e}^{-\Psi(\theta)}d\theta.\end{aligned} \tag{III.2.76}$$

Since $N \cdot \mathrm{Leb}_{A_{k,N}}$ is a probability measure and $s \mapsto s \log s$ is convex on $(0,\infty)$, Jensen's inequality yields

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}\int_{\mathbb{R}}\log\left(\int_{A_{k,N}} f_t(x,\theta)\mathrm{e}^{\Psi(\theta)}N dx\right)\int_{A_{k,N}} f_t(x,\theta)\mathrm{e}^{\Psi(\theta)}N dx\,\mathrm{e}^{-\Psi(\theta)}d\theta \\ \leq \sum_{k=0}^{N-1}\int_{\mathbb{R}}\int_{A_{k,N}}\log\left(f_t(x,\theta)\mathrm{e}^{\Psi(\theta)}\right)\,f_t(x,\theta)\,dx\,d\theta = \int_{\mathbb{T}\times\mathbb{R}}\log\left(f_t\,\mathrm{e}^{\Psi}\right)d\nu_t.\end{aligned} \tag{III.2.77}$$

Moreover, the continuity of J, the fact that $\int|\theta|^2 d\nu_t < \infty$ and the dominated convergence theorem yield

$$\lim_{N\to\infty}\frac{1}{2}\int_{(\mathbb{T}\times\mathbb{R})^2} J\left(\tfrac{\lfloor xN\rfloor - \lfloor \bar{x}N\rfloor}{N}\right)\theta\bar{\theta}\,d\nu_t\,d\nu_t = \frac{1}{2}\int_{(\mathbb{T}\times\mathbb{R})^2} J\left(x-\bar{x}\right)\theta\bar{\theta}\,d\nu_t\,d\nu_t. \tag{III.2.78}$$

Lastly, (III.2.76), (III.2.77) and (III.2.78) yield (III.2.75). □

Initialization.

Lemma III.59 *Under the same setting as in the proof of Proposition III.56 and under* (III.2.65)*, we have*

$$\limsup_{N\to\infty}\frac{1}{N}\,\mathcal{H}(\nu_0^N\,|\,\mu_0^N) \leq\ \mathcal{H}(\nu_0\,|\,\mu_0)\quad \textit{for all } t\in[0,T]. \tag{III.2.79}$$

Proof. Set $\rho_0(x,\theta) := \mathrm{e}^{-\Psi(\theta)}\kappa(x,\theta)$. Then, as in the proof of Lemma III.58, we observe that

$$\begin{aligned}\frac{1}{N}\mathcal{H}(\nu_0^N|\mu_0^N) = \frac{1}{N}\sum_{k=0}^{N-1}\int_{\mathbb{R}}\log\left(\int_{A_{k,N}} f_0(x,\theta)\rho_0(\tfrac{k}{N},\theta)^{-1}N dx\right)\int_{A_{k,N}} f_0(x,\theta)N dx\,d\theta \\ \leq \int_{\mathbb{T}\times\mathbb{R}}\log\left(f_0(x,\theta)\rho_0(\tfrac{\lfloor xN\rfloor}{N},\theta)^{-1}\right)\,f_0(x,\theta)\,dx\,d\theta.\end{aligned} \tag{III.2.80}$$

Under Assumption III.43, we either have that $\rho_0(\frac{\lfloor xN\rfloor}{N},\theta) \geq c'_\kappa\,\mathrm{e}^{-(c_\kappa+1)\Psi(\theta)}$ on the set $\{\rho_0>0\}$ or that $\rho_0(\frac{\lfloor xN\rfloor}{N},\theta) = \rho_0(x,\theta)$ for all $(x,\theta)\in\mathbb{T}\times\mathbb{R}$. In the latter case, we trivially obtain (III.2.79). In the former case, the integrand on the right-hand side of (III.2.80) is bounded from above by $g := \log(f_0\ \exp((c_\kappa+1)\Psi)/c'_\kappa)f_0$, which is integrable. Indeed, from (III.2.67) and (III.2.70) we infer that $\mathcal{H}(\nu_0\,|\,\mathrm{e}^{-\Psi(\theta)}dxd\theta)$ is finite. Combined with (III.2.70) and the fact that Ψ is a polynomial of degree 2ℓ, this immediately implies the integrability of g. Hence, we can apply the dominated convergence theorem to interchange the integral and the limit. Finally, the regularity assumptions on ρ_0 from Assumption III.43 lead to (III.2.79). □

Metric derivative.

Lemma III.60 *Under the same setting as in the proof of Proposition III.56 and under* (III.2.65)*, we have that for all* $N\in\mathbb{N}$

$$\frac{1}{\sqrt{N}}\,|(\nu^N)'|(t)\ \leq\ |\nu'|(t)\quad \textit{for almost every } t\in[0,T]. \tag{III.2.81}$$

Proof. Let $s < t$. Let $\pi \in \mathrm{Opt}^{\mathrm{L}}(\nu_s, \nu_t)$ and define $\gamma \in \mathcal{P}_2(\mathbb{R}^N \times \mathbb{R}^N)$ by

$$d\gamma(\Theta, \bar{\Theta}) = \prod_{k=0}^{N-1} N\pi(A_{k,N} \times d\theta^k \times d\bar{\theta}^k). \tag{III.2.82}$$

It is readily checked that $\gamma \in \mathrm{Cpl}(\nu_s^N, \nu_t^N)$. Therefore,

$$W_2(\nu_s^N, \nu_t^N)^2 \le \int_{\mathbb{R}^N \times \mathbb{R}^N} |\Theta - \bar{\Theta}|^2 d\gamma(\Theta, \bar{\Theta}) = N\mathrm{W}^{\mathrm{L}}(\nu_s, \nu_t)^2, \tag{III.2.83}$$

which immediately implies (III.2.81). □

Metric slope.

Lemma III.61 *Under the same setting as in the proof of Proposition III.56 and under* (III.2.65)*, we have*

$$\limsup_{N\to\infty} \frac{1}{N} \int_0^T |\partial \mathcal{H}^N|^2(\nu_t^N)\, dt \;\le\; \int_0^T |\partial \mathcal{F}|^2(\nu_t)\, dt. \tag{III.2.84}$$

Proof. Recalling the definition of the weak derivative, one can easily show that for all $k \le N-1$

$$\partial_\theta \int_{A_{k,N}} f_t(x,\theta)\, dx = \int_{A_{k,N}} \partial_\theta f_t(x,\theta)\, dx \quad \text{for almost every } t \text{ and } \theta. \tag{III.2.85}$$

Let f_t^N be the density of ν_t^N with respect to $\mathrm{Leb}_{\mathbb{R}^N}$. In view of (III.2.85), we observe that

$$\begin{aligned}
&\frac{1}{N} \int_0^T \int_{\mathbb{R}^N} \left| \frac{\nabla f_t^N(\Theta)}{f_t^N(\Theta)} + \nabla \mathrm{H}^N(\Theta) \right|^2 d\nu_t^N(\Theta)\, dt \\
&= \frac{1}{N} \sum_{k=0}^{N-1} \int_0^T \int_{\mathbb{R}^N} \left(\frac{N \int_{A_{k,N}} \partial_\theta f_t(x,\theta^k)\, dx}{N \int_{A_{k,N}} f_t(x,\theta^k)\, dx} + \Psi'(\theta^k) - \frac{1}{2N} \sum_{j=0}^{N-1} J\left(\tfrac{k-j}{N}\right) \theta^j \right)^2 d\nu_t^N\, dt \\
&= \frac{1}{N} \sum_{k=0}^{N-1} \int_0^T \int_{\mathbb{R}^N} \left(\frac{N \int_{A_{k,N}} \partial_\theta f_t(x,\theta^k)\, dx}{N \int_{A_{k,N}} f_t(x,\theta^k)\, dx} + \Psi'(\theta^k) \right)^2 d\nu_t^N\, dt && \text{(III.2.86)} \\
&\quad - \frac{1}{N} \sum_{k=0}^{N-1} \int_0^T \int_{\mathbb{R}^N} \left(\frac{N \int_{A_{k,N}} \partial_\theta f_t(x,\theta^k)\, dx}{N \int_{A_{k,N}} f_t(x,\theta^k)\, dx} + \Psi'(\theta^k) \right) \frac{1}{N} \sum_{j=0}^{N-1} J\left(\tfrac{k-j}{N}\right) \theta^j\, d\nu_t^N\, dt && \text{(III.2.87)} \\
&\quad + \frac{1}{N} \sum_{k=0}^{N-1} \int_0^T \int_{\mathbb{R}^N} \left(\frac{1}{2N} \sum_{j=0}^{N-1} J\left(\tfrac{k-j}{N}\right) \theta^j \right)^2 d\nu_t^N\, dt. && \text{(III.2.88)}
\end{aligned}$$

We treat each term (III.2.86)–(III.2.88) separately. First, we compute

$$\text{(III.2.86)} = \frac{1}{N} \sum_{k=0}^{N-1} \int_0^T \int_{\mathbb{R}} \frac{\left(N \int_{A_{k,N}} \left(\partial_\theta f_t(x,\theta) + \Psi'(\theta) f_t(x,\theta) \right) dx \right)^2}{N \int_{A_{k,N}} f_t(x,\theta)\, dx}\, d\theta\, dt. \tag{III.2.89}$$

In the same way as in the proof of [3, 8.1.10], we are allowed to apply Jensen's inequality for the integrand, since the function $(x,z)\mapsto x^2/z$ is convex on $\mathbb{R}\times(0,\infty)$. Hence,

$$\begin{aligned}\text{(III.2.86)} &\le \sum_{k=0}^{N-1}\int_0^T\int_{\mathbb{R}}\int_{A_{k,N}} \frac{\Big(\partial_\theta f_t(x,\theta)+\Psi'(\theta)f_t(x,\theta)\Big)^2}{f_t(x,\theta)}\,dx\,d\theta\,dt \\ &= \int_0^T\int_{\mathbb{T}\times\mathbb{R}}\left(\frac{\partial_\theta f_t(x,\theta)}{f_t(x,\theta)}+\Psi'(\theta)\right)^2 d\nu_t\,dt.\end{aligned} \tag{III.2.90}$$

Next, we observe that (III.2.87) is equal to

$$\begin{aligned}&-\sum_{k,j=0}^{N-1}\int_0^T\int_{\mathbb{R}^2}\int_{A_{k,N}}\int_{A_{j,N}}\Big(\partial_\theta f_t(x,\theta)+\Psi'(\theta)f_t(x,\theta)\Big)J\left(\tfrac{k-j}{N}\right)\bar\theta f_t(\bar x,\bar\theta)\,d\bar x\,dx\,d\bar\theta\,d\theta dt\\ &=-\int_0^T\int_{\mathbb{T}\times\mathbb{R}}\left(\frac{\partial_\theta f_t(x,\theta)}{f_t(x,\theta)}+\Psi'(\theta)\right)\int_{\mathbb{T}\times\mathbb{R}} J\left(\tfrac{\lfloor xN\rfloor-\lfloor\bar xN\rfloor}{N}\right)\bar\theta\,d\nu_t(\bar x,\bar\theta)\,d\nu_t(x,\theta)dt \qquad \text{(III.2.91)}\\ &\longrightarrow \int_0^T\int_{\mathbb{T}\times\mathbb{R}}\left(\frac{\partial_\theta f_t(x,\theta)}{f_t(x,\theta)}+\Psi'(\theta)\right)\int_{\mathbb{T}\times\mathbb{R}} J\,(x-\bar x)\,\bar\theta\,d\nu_t(\bar x,\bar\theta)\,d\nu_t(x,\theta)dt,\end{aligned}$$

where we have used the continuity of J and the dominated convergence theorem, which is applicable, since by Young's inequality, (III.2.68) and (III.2.70)

$$\begin{aligned}&\left(\frac{\partial_\theta f_t(x,\theta)}{f_t(x,\theta)}+\Psi'(\theta)\right)\int_{\mathbb{T}\times\mathbb{R}} J\,(x-\bar x)\,\bar\theta\,d\nu_t(\bar x,\bar\theta)\\ &\quad\le \frac12\left(\frac{\partial_\theta f_t(x,\theta)}{f_t(x,\theta)}+\Psi'(\theta)\right)^2+\frac{\|J\|_\infty}{2}\int_{\mathbb{T}\times\mathbb{R}}\bar\theta^2\,d\nu_t \;\in\; \mathrm{L}^1([0,T]\times\mathbb{T}\times\mathbb{R}\,;\,\nu_t dt).\end{aligned} \tag{III.2.92}$$

For the term (III.2.88), we apply similar arguments to obtain that

$$\begin{aligned}\text{(III.2.88)} &= \frac14\sum_{k,j,l=0}^{N-1}\int_0^T\int_{\mathbb{R}^3}\int_{A_{k,N}}\int_{A_{j,N}}\int_{A_{l,N}} J\Big(\tfrac{k-j}{N}\Big)\bar\theta\,J\Big(\tfrac{k-l}{N}\Big)\hat\theta\,d\nu_t d\nu_t d\nu_t dt+O\left(\frac1N\right)\\ &= \int_0^T\int_{\mathbb{T}\times\mathbb{R}}\left(\frac12\int_{\mathbb{T}\times\mathbb{R}} J\left(\tfrac{\lfloor xN\rfloor-\lfloor\bar xN\rfloor}{N}\right)\bar\theta\,d\nu_t(\bar x,\bar\theta)\right)^2 d\nu_t(x,\theta)dt+O\left(\frac1N\right) \qquad \text{(III.2.93)}\\ &\longrightarrow \int_0^T\int_{\mathbb{T}\times\mathbb{R}}\left(\frac12\int_{\mathbb{T}\times\mathbb{R}} J\,(x-\bar x)\,\bar\theta\,d\nu_t(\bar x,\bar\theta)\right)^2 d\nu_t(x,\theta)dt.\end{aligned}$$

Hence, (III.2.90), (III.2.91), (III.2.93) and Proposition III.38 show that

$$\limsup_{N\to\infty}\frac1N\int_0^T\int_{\mathbb{R}^N}\left|\frac{\nabla f_t^N}{f_t^N}+\nabla \mathrm{H}^N\right|^2 d\nu_t^N\,dt\le\int_0^T|\partial\mathcal{F}|^2(\nu_t)\,dt. \tag{III.2.94}$$

Finally, it is known that (see for instance, [3, 10.4.9]) since the right-hand side (and hence also the left-hand side) of (III.2.94) is finite, we have

$$\limsup_{N\to\infty}\frac1N\int_0^T\int_{\mathbb{R}^N}\left|\frac{\nabla f_t^N}{f_t^N}+\nabla \mathrm{H}^N\right|^2 d\nu_t^N\,dt=\limsup_{N\to\infty}\frac1N\int_0^T|\partial\mathcal{H}^N|^2(\nu_t^N)\,dt. \tag{III.2.95}$$

(III.2.94) and (III.2.95) conclude the proof. □

III.3 Law of large numbers

In this section we derive a law of large numbers for the system introduced in Subsection III.2.1.

Theorem III.62 *For all $N \in \mathbb{N}$, let $(\mu_t^N)_{t\in[0,T]}$ be the Wasserstein gradient flow for $\mathcal{H}^N$ with initial value μ_0^N. Assume either*

a) Assumption III.43 on the sequence of initial data $\{\mu_0^N\}_N$ and let $\mu_0 = \rho_0 \operatorname{Leb}_{\mathbb{T}\times\mathbb{R}}$ with $\rho_0(x,\theta) = \kappa(x,\theta)\,\mathrm{e}^{-\Psi(\theta)}$, or

b) $\mu_0 \in D(\mathcal{F})$ and $\{\mu_0^N\}_N$ is such that $((K^N)_\# \mu_0^N)_N$ converges to δ_{μ_0} weakly in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ and $\lim_{N\to\infty} \frac{1}{N}\mathcal{H}^N(\mu_0^N) = \mathcal{F}(\mu_0)$.

Let $(\mu_t)_t$ be the gradient flow for $\mathcal{F}$ with initial value μ_0. Then,

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \widetilde{\mathbb{W}}((K^N)_\# \nu_t^N, \delta_{\nu_t}) \;=\; 0 \quad \text{and} \tag{III.3.1}$$

$$\lim_{N\to\infty} \frac{1}{N}\mathcal{H}^N(\mu_t^N) = \mathcal{F}(\mu_t) \quad \text{for all } t \in [0,T], \tag{III.3.2}$$

where $\widetilde{\mathbb{W}}$ was defined in Lemma III.44. (Recall that $\widetilde{\mathbb{W}}$ metrizes the weak convergence in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$.). Moreover, in the situation of b) and if $C_\Psi > 0$ and $\ell \geq 2$ in Assumption III.33, then we even have that

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \mathbb{W}_2((K^N)_\# \nu_t^N, \delta_{\nu_t}) \;=\; 0, \tag{III.3.3}$$

where $\mathbb{W}_2$ denotes the Wasserstein distance on $\mathcal{P}_2\big((\mathcal{P}_2(\mathbb{T}\times\mathbb{R}), W_2)\big)$ induced by W_2.

Proof. In the situation of *a)*, the proof follows immediately from Theorem III.47, since the corresponding rate function in the LDP result has a unique minimum at $(\mu_t)_t$ by Theorem III.40. So assume *b)*. First notice that

$$\sup_{N\in\mathbb{N}} \frac{1}{N}\int_0^T |(\mu^N)'|^2(t)\,dt, \quad \sup_{N\in\mathbb{N}} \frac{1}{N}\int_0^T |\partial\mathcal{H}^N|^2(\mu_t^N)\,dt, \quad \sup_{N\in\mathbb{N}} \frac{1}{N}\mathcal{H}^N(\mu_T^N) \;< \infty, \tag{III.3.4}$$

since $\mathscr{J}^N[(\mu_t^N)_t] = 0$ for all N, $\lim_{N\to\infty} \frac{1}{N}\mathcal{H}^N(\mu_0^N) = \mathcal{F}(\mu_0)$ and by Lemma III.42 *(ii)*. Moreover, arguing as in (III.2.69), we infer that

$$\sup_{N\in\mathbb{N}} \frac{1}{N} \sup_{t\in[0,T]} \mathcal{H}^N(\mu_t^N) \leq \sup_{N\in\mathbb{N}} \frac{1}{N}\left(\sup_{t\in[0,T]} \int_t^T |\partial\mathcal{H}^N|(\mu_r^N)\cdot|(\mu^N)'|(r)\,dr + \mathcal{H}^N(\mu_T^N)\right) < \infty. \tag{III.3.5}$$

By Lemma III.42 *(ii)* this implies that

$$\sup_{N\in\mathbb{N}} \frac{1}{N} \sup_{t\in[0,T]} \int_{\mathbb{R}^N} \sum_{i=0}^{N-1} |\theta^i|^{2\ell}\,d\mu_t^N(\Theta) \;< \infty. \tag{III.3.6}$$

Step 1. [Compactness.]
Lemma III.63 yields the existence of a subsequence $\{(\mu_t^n)_t\}_n$ and a continuous curve $(c_t)_t \in C([0,T]\,;\,\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R})))$ such that

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \widetilde{\mathbb{W}}((K^n)_\# \nu_t^n, c_t) \;=\; 0, \tag{III.3.7}$$

and if $\ell \geq 2$ in Assumption III.33,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \mathbb{W}_2((K^n)_\# \nu_t^n, c_t) \;=\; 0. \tag{III.3.8}$$

Step 2. [Superposition.]
Lemma III.64 below shows that there exists a measure $\Upsilon \in \mathcal{M}_1(\mathcal{AC}([0,T];\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})))$ such that $(e_t)_\#\Upsilon = c_t$ for all $t \in [0,T]$.

Step 3. [Lower semi-continuity.]
Assumption b) and the Lemmas III.52, III.54 and III.65 show that

$$\int_{\mathcal{AC}([0,T];\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))} \mathbb{1}_{\mu_0}(\eta_0)\cdot \mathscr{J}[\,(\eta_t)_t\,]\, d\Upsilon((\eta_t)_t) \leq \liminf_{n\to\infty} \frac{1}{n}\mathscr{J}^n[\,(\mu_t^n)_t\,]. \tag{III.3.9}$$

Step 4. [Convergence towards δ_{μ_t} for all $t \in [0,T]$.]
Step 3 shows that

$$\int_{\mathcal{AC}([0,T];\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R}))} \mathbb{1}_{\mu_0}(\eta_0)\cdot \mathscr{J}[\,(\eta_t)_t\,]\, d\Upsilon((\eta_t)_t) \leq 0. \tag{III.3.10}$$

Since the integrand on the left-hand side is non-negative (see Theorem III.40), this implies that $\mathbb{1}_{\mu_0}(\eta_0)\cdot \mathscr{J}[\,(\eta_t)_t\,] = 0$ for Υ-a.e. $(\eta_t)_t$. Thus, by the uniqueness claim in Theorem III.40, we infer that Υ must be concentrated on $(\mu_t)_t$. Together with Step 2, this shows that $c_t = \delta_{\mu_t}$ for all $t \in [0,T]$. Therefore, $(\delta_{\mu_t})_t$ is the unique limit point of the sequence $\{((K^N)_\#\nu_t^N)_T\}_N$. Combining this fact with Step 1 yields (III.3.1) and (III.3.3), respectively.

Step 5. [Proof of (III.3.2).]
From the previous steps, we infer that $\lim_{N\to\infty}\frac{1}{N}\mathscr{J}^N[\,(\mu_t^N)_t\,] = \mathscr{J}[\,(\mu_t)_t\,]$. Using again the Lemmas III.52, III.54 and III.65 and Assumption *b)*, this easily implies that

$$\lim_{N\to\infty}\frac{1}{N}\mathcal{H}^N(\mu_T^N) = \mathcal{F}(\mu_T). \tag{III.3.11}$$

We can now replace T by some arbitrary $t \in (0,T)$ and repeat the above proof to obtain (III.3.2). □

Lemma III.63 (Compactness) *Let $(\mu_t^N)_t \in \mathcal{AC}([0.T]\,;\mathcal{P}_2(\mathbb{R}^N))$ for all $N \in \mathbb{N}$. Assume that*

$$\begin{aligned} &\sup_{N\in\mathbb{N}}\frac{1}{N}\int_0^T |(\mu^N)'|^2(t)\,dt < \infty \quad \text{and} \\ &\sup_{N\in\mathbb{N}}\frac{1}{N}\int_{\mathbb{R}^N}\sum_{i=0}^{N-1}|\theta^i|^{2\ell}\,d\mu_t^N(\Theta) < \infty \;\; \forall\, t\in[0,T]. \end{aligned} \tag{III.3.12}$$

Then, there exists a subsequence $\{(\mu_t^n)_t\}_n$ and a curve $(c_t)_t \in C([0,T];\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R})))$ such that

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\widetilde{\mathbb{W}}((K^n)_\#\mu_t^n, c_t) \;=\; 0. \tag{III.3.13}$$

Moreover, if $\ell \geq 2$ in Assumption III.33, we even have that

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\mathbb{W}_2((K^n)_\#\mu_t^n, c_t) \;=\; 0. \tag{III.3.14}$$

Proof. In view of Lemma III.44, it is equivalent to show the claim with L^N replacing K^N. The proof uses the Arzelá-Ascoli theorem ([86, Chapter 7, Theorem 17]). Hence, we have to show that

(i) the sequence $\{((L^N)_\# \mu_t^N)_t\}_N$ is equi-continuous with respect to $\mathbb{W}_2$ and $\widetilde{\mathbb{W}}$,

(ii) for fixed $t \in [0,T]$, the sequence $\{(L^N)_\# \mu_t^N\}_N$ is compact with respect to $\widetilde{\mathbb{W}}$, and

(iii) if $\ell \geq 2$ in Assumption III.33, for fixed $t \in [0,T]$, the sequence $\{(L^N)_\# \mu_t^N\}_N$ is compact with respect to $\mathbb{W}_2$.

We first show claim (i). Recall the definition of $\mathbb{W}^{\mathrm{L}}$ from Lemma III.50. In Lemma III.51 we have seen that

$$\mathbb{W}^{\mathrm{L}}\Big((L^N)_\# \mu_s^N \,,\, (L^N)_\# \mu_t^N\Big) \leq \frac{1}{\sqrt{N}} W_2(\mu_s^N, \mu_t^N) \quad \text{for all } 0 \leq s < t \leq T. \tag{III.3.15}$$

Using (III.3.12) and that both $\mathbb{W}_2$ and $\widetilde{\mathbb{W}}$ are dominated by $\mathbb{W}^{\mathrm{L}}$, this implies the equi-continuity with respect to both $\mathbb{W}_2$ and $\widetilde{\mathbb{W}}$.

Next we show part (ii). Note that by (III.3.12),

$$\sup_{N\in\mathbb{N}} \int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} \int_{\mathbb{T}\times\mathbb{R}} |\theta|^{2\ell} \, d\gamma \; d(L^N)_\# \mu_t^N(\gamma) = \sup_{N\in\mathbb{N}} \frac{1}{N} \int_{\mathbb{R}^N} \sum_{i=0}^{N-1} |\theta^i|^{2\ell} \, d\mu_t^N(\Theta) < \infty. \tag{III.3.16}$$

Since the map $\gamma \mapsto \int_{\mathbb{T}\times\mathbb{R}} |\theta|^{2\ell} \, d\gamma$ has compact sublevels in $\mathcal{M}_1(\mathbb{T}\times\mathbb{R})$, this implies part (ii) (cf. [3, 5.1.5]).

It remains to show part (iii). From part (ii) we know that there exists a converging subsequence $\{(L^n)_\# \mu_t^n\}_n$. Let $c_t \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$ denote the limit. By using [127, 6.8 (iii)], it suffices to show that

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int_{W_2(\delta_0,\gamma)\geq R} W_2(\delta_0,\gamma)^2 \; d(L^n)_\# \mu_t^n(\gamma) = 0. \tag{III.3.17}$$

But if $\ell \geq 2$, we have that by Jensen's inequality and the fact that $|x| \leq 1$ for all $x \in \mathbb{T}$,

$$\begin{aligned}
&\lim_{R\to\infty} \limsup_{n\to\infty} \int_{W_2(\delta_0,\gamma)\geq R} W_2(\delta_0,\gamma)^2 \; d(L^n)_\# \mu_t^n(\gamma) \\
&\quad \leq \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{R^{2\ell-2}} \int_{W_2(\delta_0,\gamma)\geq R} W_2(\delta_0,\gamma)^{2\ell} \; d(L^n)_\# \mu_t^n(\gamma) \\
&\quad = \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{R^{2\ell-2}} \int_{W_2(\delta_0,\gamma)\geq R} \left(\int_{\mathbb{T}\times\mathbb{R}} (|x|^2 + |\theta|^2) \, d\gamma \right)^{2\ell-2} \; d(L^n)_\# \mu_t^n(\gamma) \\
&\quad \leq \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{R^{2\ell-2}} \int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} \int_{\mathbb{T}\times\mathbb{R}} (1+|\theta|^2)^{2\ell-2} d\gamma \; d(L^n)_\# \mu_t^n(\gamma) \\
&\quad \leq \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{R^{2\ell-2}} \int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} \int_{\mathbb{T}\times\mathbb{R}} 2^{2\ell-2}(1+|\theta|^{2\ell}) d\gamma \; d(L^n)_\# \mu_t^n(\gamma) \\
&\quad = 0,
\end{aligned} \tag{III.3.18}$$

where we used the uniform boundedness (III.3.16) in the last step. This shows part (iii). □

Lemma III.64 (Superposition) *Consider the same setting as in Lemma III.63 and assume in addition that*

$$\sup_{N\in\mathbb{N}} \frac{1}{N} \int_0^T \int_{\mathbb{R}^N} |\Theta|^2 \, d\mu_t^N(\Theta)\, dt < \infty. \tag{III.3.19}$$

Then $(c_t)_t$ is absolutely continuous with respect to $\mathbb{W}^{\mathrm{L}}$, and there exists a measure $\Upsilon \in \mathcal{M}_1(\mathcal{AC}([0,T];\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})))$ such that

$$(e_t)_\#\Upsilon = c_t \text{ for all } t\in[0,T] \quad \text{and} \quad \int |\eta'|^2(t)\, d\Upsilon((\eta_t)_t) = |c'|^2(t) \text{ for a.e. } t\in[0,T]. \tag{III.3.20}$$

Proof. Note that $\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})$ is a closed subspace of $\mathcal{M}_1(\mathbb{T}\times\mathbb{R})$. Therefore, the Portmanteau theorem ([54, 11.1.1]) yields that for almost every $t\in[0,T]$

$$c_t(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})) \geq \limsup_{n\to\infty} (L^n)_\#\mu_t^n(\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})) = 1. \tag{III.3.21}$$

Hence, c_t is supported in $\mathcal{M}_1^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})$ for almost every t. Moreover, (III.3.12) and [3, 5.1.7] show that c_t is supported in $\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})$ for all t. To show the absolute continuity of $(c_t)_t$ with respect to $\mathbb{W}^{\mathrm{L}}$ we proceed as in the proofs of the Lemmas III.46 and III.49. We have that

$$\begin{aligned}\sup_{0<h<T} \int_0^{T-h} \frac{1}{h^2}\mathbb{W}^{\mathrm{L}}(c_t,c_{t+h})^2 dt &\leq \sup_{0<h<T} \int_0^{T-h} \liminf_{n\to\infty} \frac{1}{h^2}\mathbb{W}^{\mathrm{L}}\Big((L^n)_\#\mu_t^n\,,\,(L^n)_\#\mu_{t+h}^n\Big)^2 dt \\ &\leq \liminf_{n\to\infty} \frac{1}{n}\int_0^T |(\mu^n)'|^2(r)\, dr < \infty. \end{aligned} \tag{III.3.22}$$

Moreover, by (III.3.19),

$$\begin{aligned}\int_0^T \mathbb{W}^{\mathrm{L}}(c_t,\delta_{\delta_0\otimes\mathrm{Leb}_{\mathbb{T}}})^2 dt &= \int_0^T \int_{\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})} \mathrm{W}^{\mathrm{L}}(\gamma,\delta_0\otimes\mathrm{Leb}_{\mathbb{T}})^2\, dc_t(\gamma)dt \\ &= \int_0^T \int\int |\theta|^2 d\gamma\, dc_t(\gamma)dt \\ &\leq \liminf_{n\to\infty} \int_0^T \int_{\mathcal{P}_2^{\mathrm{L}}(\mathbb{T}\times\mathbb{R})} \int_{\mathbb{T}\times\mathbb{R}} |\theta|^2 d\gamma\, d(L^n)_\#\mu_t^n(\gamma)\, dt \\ &= \liminf_{n\to\infty} \frac{1}{n}\int_0^T \int_{\mathbb{R}^n} |\Theta|^2\, d\mu_t^n(\Theta)\, dt < \infty. \end{aligned} \tag{III.3.23}$$

By [96, Lemma 1], (III.3.22) and (III.3.23) yield the absolute continuity of $(c_t)_t$. Finally, [96, Theorem 5] shows that this already implies the second claim. □

Lemma III.65 (Lower semi-continuity, metric slope) *Let $\mu^n\in\mathcal{P}_2(\mathbb{R}^n)\cap D(\mathcal{H}^n)$ for all $n\in\mathbb{N}$. Assume that $(K^n)_\#\mu^n \rightharpoonup c$ for some $c\in\mathcal{M}_1(\mathcal{M}_1(\mathbb{T}\times\mathbb{R}))$, and that*

$$\liminf_{n\to\infty} \frac{1}{n}\int_{\mathbb{R}^n} |\Theta|^2\, d\mu^n(\Theta)\, dt < \infty. \tag{III.3.24}$$

Then

$$\liminf_{n\to\infty} \frac{1}{n}|\partial\mathcal{H}^n|^2(\mu^n) \geq \int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} |\partial\mathcal{F}|^2(\sigma)\, dc(\sigma). \tag{III.3.25}$$

Proof. We use the same strategy as in the proof of [100, 3.5]. From [100, 3.9], we know that there exists a sequence $\{(E^i_{\delta,l})_{i=0}^{N_{\delta,l}}\}_{\delta>0,l\in\mathbb{N}}$ of subsets of $\mathcal{M}_1(\mathbb{T}\times\mathbb{R})$ such that

1) $\lim_{l\to\infty} c\left(\cup_{i=1}^{N_{\delta,l}} E^i_{\delta,l}\right) = 1$ and $\cup_{i=0}^{N_{\delta,l}} E^i_{\delta,l} = \mathcal{M}_1(\mathbb{T}\times\mathbb{R})$,
2) $E^i_{\delta,l} \cap E^j_{\delta,l} = \emptyset$ if $j \neq i$,
3) $\widetilde{W}(\sigma,\eta) < \delta$ for all $\sigma,\eta \in E^i_{\delta,l}$ and $i = 1,\dots,N_{\delta,l}$,
4) $c(\partial E^i_{\delta,l}) = 0$ for all $i = 1,\dots,N_{\delta,l}$,
5) each $E^i_{\delta,l}$ has non-empty interior,
6) $(E^i_{\delta,l})_{i=0}^{N_{\delta,l}}$ is finer than $(E^i_{\delta',l'})_{i=0}^{N_{\delta',l'}}$ if $\delta \le \delta'$ and $l \ge l'$.

Assume that the left-hand side of (III.3.25) is finite, since the claim would be trivial otherwise. Let $(\mu^m)_m$ be a subsequence such that

$$\lim_{m\to\infty} \frac{1}{m}|\partial\mathcal{H}^m|^2(\mu^m) = \liminf_{n\to\infty} \frac{1}{n}|\partial\mathcal{H}^n|^2(\mu^n) \quad \text{and} \quad \sup_{m\in\mathbb{N}} \frac{1}{m}|\partial\mathcal{H}^m|^2(\mu^m) < \infty. \tag{III.3.26}$$

In particular, by [3, 10.4.9], this implies that

$$|\partial\mathcal{H}^m|^2(\mu^m) = \int_{\mathbb{R}^m} \left|\frac{\nabla\rho^m}{\rho^m} + \nabla H^m\right|^2 d\mu^m, \tag{III.3.27}$$

where for all m, ρ^m denotes the density of μ^m with respect to $\mathrm{Leb}_{\mathbb{R}^m}$. For each m,δ,l,i, define the measure $\mu^{m,\delta,l,i} \in \mathcal{P}_2(\mathbb{R}^n)$ by

$$\int_{\mathbb{R}^m} f\, d\mu^{m,\delta,l,i} = \frac{1}{(K^m)_\#\mu^m(E^i_{\delta,l})} \int_{(K^m)^{-1}(E^i_{\delta,l})} f\, d\mu^m \tag{III.3.28}$$

for all measurable and bounded $f : \mathbb{R}^N \to \mathbb{R}$. Then,

$$\begin{aligned}
&\lim_{m\to\infty} \frac{1}{m}|\partial\mathcal{H}^m|^2(\mu^m) \\
&\quad= \lim_{m\to\infty} \sum_{i=0}^{N_{\delta,l}} \frac{1}{m} \int_{(K^m)^{-1}(E^i_{\delta,l})} \left|\frac{\nabla\rho^{m,\delta,l,i}}{\rho^{m,\delta,l,i}} + \nabla H^m\right|^2 d\mu^{m,\delta,l,i} \cdot (K^m)_\#\mu^m(E^i_{\delta,l}) \\
&\quad= \sum_{i=0}^{N_{\delta,l}} \lim_{m\to\infty} \frac{1}{m}|\partial\mathcal{H}^m|^2(\mu^{m,\delta,l,i}) \cdot c(E^i_{\delta,l}),
\end{aligned} \tag{III.3.29}$$

where we have used property 4). If we define a piecewise constant function $I_{\delta,l}$ by

$$I_{\delta,l}(\gamma) = \lim_{m\to\infty} \frac{1}{m}|\partial\mathcal{H}^m|^2(\mu^{m,\delta,l,i}), \quad \text{if } \gamma \in E^i_{\delta,l} \tag{III.3.30}$$

and use Fatou's Lemma, we obtain that

$$\lim_{m\to\infty} \frac{1}{m}|\partial\mathcal{H}^m|^2(\mu^m) \ge \int_{\mathcal{M}_1(\mathbb{T}\times\mathbb{R})} \liminf_{l\to\infty} \liminf_{\delta\to 0} I_{\delta,l}(\gamma)\, dc(\gamma). \tag{III.3.31}$$

By a straightforward modification of the proof of Lemma III.55, we can show that for c-a.e. γ

$$\liminf_{l\to\infty} \liminf_{\delta\to 0} I_{\delta,l}(\gamma) \ge |\partial\mathcal{F}|^2(\gamma). \tag{III.3.32}$$

This concludes the proof. □

On the convergence with respect to $\mathbb{W}^{\mathbf{L}}$. We conclude this chapter with the following remark. Theorem III.62 leads to the question, whether the convergence can also hold with respect to the distance $\mathbb{W}^{\mathrm{L}}$ that we used in Lemma III.63. The answer to this question is negative. The reason is the following lemma.

Lemma III.66 *Let $\mu^N \in \mathcal{P}_2(\mathbb{R}^N)$ for all $N \in \mathbb{N}$. Let $\mu \in \mathcal{P}_2^{\mathrm{L}}(\mathbb{T}^d \times \mathbb{R})$. Suppose that*

$$(L^N)_{\#}\mu^N \text{ converges to } \delta_\mu \text{ weakly in } \mathcal{M}_1(\mathcal{P}_2^L(\mathbb{T}^d \times \mathbb{R})). \tag{III.3.33}$$

Then, $\mu = \delta_{\theta_x}\, dx$ for some family $(\theta_x)_x \subset \mathbb{R}$.

In particular, the convergence result from Theorem III.62 can not hold true in $\mathcal{M}_1(\mathcal{P}_2^L(\mathbb{T}^d \times \mathbb{R}))$.

Proof. Let $(\mathrm{Z}^N)_N$ be $\mathcal{P}_2^L(\mathbb{T}^d \times \Theta)$-valued random variables on a common probability space $(\mathbb{P}, \Omega, \mathbb{F})$ such that for all N, $(L^N)_{\#}\mu^N$ is the law of Z^N. Then, combining (III.3.33), [54, p. 297, Problem 6] and [54, 9.2.1] implies that $(\mathrm{Z}^N)_N$ converges almost sure along a subsequence. That is, there exists a subsequence $(\mathrm{Z}^m)_m$ such that

$$\lim_{m\to\infty} \mathrm{W}^{\mathrm{L}}(\mathrm{Z}^m(\omega), \mu) = 0 \qquad \text{for almost all } \omega \in \Omega. \tag{III.3.34}$$

In view of Proposition III.5, this yields that for almost all $\omega \in \Omega$, there exists a further subsequence $(\mathrm{Z}^{m_k}(\omega))_k$ and a null-set $\mathcal{N}_{m,\omega}$ such that

$$\mathrm{Z}^{m_k}(\omega)^x \rightharpoonup \mu^x \quad \text{for all } x \in \mathbb{T}^d \setminus \mathcal{N}_{m,\omega}. \tag{III.3.35}$$

However, since $(L^{m_k})_{\#}\mu^{m_k}$ is the law of Z^{m_k}, we have that for all k and for all $x \in \mathbb{T}^d \setminus \mathcal{N}_m$,

$$\mathrm{Z}^{m_k}(\omega)^x = \delta_{\theta_{m_k,x,\omega}} \qquad \text{for some } \theta_{m_k,x,\omega} \in \Theta. \tag{III.3.36}$$

Therefore, for all $x \in \mathbb{T}^d \setminus \mathcal{N}_m$, μ^x is also concentrated on a single point, since it is the weak limit of a sequence of Dirac measures; see for instance [3, 5.1.8]. This concludes the proof. □

Chapter IV

Metastability in a continuous mean-field model

The results of the present chapter were established in joint work with Georg Menz (UCLA), and are contained in the preprint [14].

Recall Section I.6, where we provide a motivation and a first formulation of the main results of this chapter. This chapter is organized as follows. First we introduce some notation. Then, in Chapter IV.1, we show Kramers' law for the system defined in Subsection I.6.1 in the low-temperature regime and under the assumption that $J > 1$. In Chapter IV.2 we compute estimates on the average transition time in the high-temperature regime. Finally, in the appendix, we state some general properties of Legendre transforms, compute certain asymptotic integrals by using Laplace's method, and provide the proofs of the local Cramér theorem and the equivalence of ensembles, which are the key ingredients in this chapter.

Notation

- In this chapter, x is always an element of $\mathbb{R}^N$ for $N \in \mathbb{N}$, and its components are denoted by x_i. z and m are always elements of $\mathbb{R}$.
- Let $K, A, B \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. Let $f : A \times B \to [0, \infty)$. In this chapter, $O_K(f(\varepsilon, N))$ always stands for a function, whose absolute value is bounded by f uniformly in K. That is, $O_K(f(\varepsilon, N)) = R_K(m, \varepsilon, N)$ for some function $R_K : K \times A \times B \to [0, \infty)$ such that $|R_K(m, \varepsilon, N)| \leq C_K f(\varepsilon, N)$ for all $(m, \varepsilon, N) \in K \times A \times B$ for some constant $C_K > 0$.
  If, in addition, we have that $c_K f(\varepsilon, N) \leq |R_K(m, \varepsilon, N)|$ for some $c_K > 0$, we write $\Omega_K(f(\varepsilon, N))$ instead of $O_K(f(\varepsilon, N))$.
- Similarly, $O(f(\varepsilon, N))$ stands for a function $R : A \times B \to [0, \infty)$ such that $O(f(\varepsilon, N)) = R(\varepsilon, N)$ and there exists a constant $C' > 0$ such that $|R(\varepsilon, N)| \leq C f(\varepsilon, N)$. Finally, we define $\Omega(f(\varepsilon, N))$ analogously as $\Omega_K(f(\varepsilon, N))$.
- Let (S, d) be a metric space, $\rho > 0$ and $s \in S$. Then, define $B_\rho(s) = \{r \in S \,|\, d(s, r) < \rho\}$.
- Let Y be an Euclidean space. Then we say that $\mu \in \mathcal{M}_1(Y)$ satisfies the *Poincaré inequality* with constant ϱ if for all $f \in H^1(\mu)$,

$$\mathrm{Var}_\mu(f) := \int \left| f - \int f d\mu \right|^2 d\mu \leq \frac{1}{\varrho} \int |\nabla f|^2 \, d\mu, \tag{IV.0.1}$$

  where ∇ denotes the gradient determined by the Euclidean structure of Y.

IV.1 The Eyring-Kramers formula at low temperature

In order to simplify the notation, we omit in this section the superscripts N and ε. For example, we abbreviate $x = x^{N,\varepsilon}$, $\mu = \mu^{N,\varepsilon}$ and $\mathrm{H} = \mathrm{H}^{N,\varepsilon}$ (cf. (I.6.2), (I.6.4) and (I.6.5)). Moreover, we rewrite the microscopic Hamiltonian H as

$$\mathrm{H}(x) \;=\; \frac{1}{\varepsilon}\sum_{i=0}^{N-1}\psi_J(x_i) - \frac{1}{\varepsilon}\frac{J}{2N}\sum_{i,j=0}^{N-1}x_i x_j, \tag{IV.1.1}$$

where the *(effective) single-site potential* $\psi_J : \mathbb{R} \to \mathbb{R}$ is defined by

$$\psi_J(z) \;=\; \psi(z) \;+\; \frac{J}{2}z^2 \;=\; \frac{1}{4}z^4 \;+\; \frac{J-1}{2}z^2. \tag{IV.1.2}$$

Recall that, for the strength of the interaction part in this model, we assume that

$$J > 1. \tag{IV.1.3}$$

The outline of this section is given after Theorem I.22 in Subsection I.6.3.

IV.1.1 Preliminaries

Local Cramér theorem. In this paragraph we extend the results from [80, Proposition 31] or [101, Section 3]. The goal is to find an asymptotic representation for the measure $\bar{\mu} = P_{\#}\mu$.

The first observation is that we can disintegrate μ with respect to $\bar{\mu}$ explicitly using the *co-area formula* ([68, Section 3.4.2]). Indeed, as in [80, p. 306], we obtain that

$$\int_{\mathbb{R}^N} f(x)\,d\mu(x) \;=\; \int_{\mathbb{R}}\int_{P^{-1}(m)} f(x)\,d\mu_m(x)\,d\bar{\mu}(m) \tag{IV.1.4}$$

for all bounded and measurable $f : \mathbb{R}^N \to \mathbb{R}$, where the conditional measures (or fluctuation measures) μ_m are given by

$$d\mu_m(x) \;=\; \mathbb{1}_{P^{-1}(m)}(x)\,\mathrm{e}^{-\frac{1}{\varepsilon}\sum_{i=0}^{N-1}\psi_J(x_i)}\,d\mathcal{H}^{N-1}(x)\,\mathrm{e}^{N\varphi_{N,\varepsilon}(m)}, \tag{IV.1.5}$$

and $\varphi_{N,\varepsilon} : \mathbb{R} \mapsto \mathbb{R}$ is defined by

$$\varphi_{N,\varepsilon}(m) \;=\; -\frac{1}{N}\log\int_{P^{-1}(m)} \mathrm{e}^{-\frac{1}{\varepsilon}\sum_{i=0}^{N-1}\psi_J(x_i)}\,d\mathcal{H}^{N-1}(x). \tag{IV.1.6}$$

Moreover, for $\bar{\mu}$, we obtain the representation

$$d\bar{\mu}(m) \;=\; \frac{1}{Z_{\bar{\mu}}}\,\mathrm{e}^{-N\varphi_{N,\varepsilon}(m)+\frac{1}{\varepsilon}N\frac{J}{2}m^2}\,dm \tag{IV.1.7}$$

for some normalization constant $Z_{\bar{\mu}}$.

It turns out that the asymptotic behaviour of $\bar{\mu}$ will be determined by the *Cramér transform* φ_ε of the measure $\mathrm{e}^{-\frac{1}{\varepsilon}\psi_J(z)}dz$, which is defined as the *Legendre transform* of the function

$$\mathbb{R} \ni \sigma \;\mapsto\; \varphi_\varepsilon^*(\sigma) \;=\; \log\int_{\mathbb{R}} \mathrm{e}^{\sigma z-\frac{1}{\varepsilon}\psi_J(z)}\,dz \;\in\; \mathbb{R}. \tag{IV.1.8}$$

That is,

$$\varphi_\varepsilon(m) \;=\; \sup_{\sigma\in\mathbb{R}} \left(\sigma m - \varphi_\varepsilon^*(\sigma)\right). \tag{IV.1.9}$$

Moreover, for $\sigma \in \mathbb{R}$, we define the probability measure $\mu^{\varepsilon,\sigma} \in \mathcal{M}_1(\mathbb{R})$ by

$$d\mu^{\varepsilon,\sigma}(z) \;=\; \mathrm{e}^{-\varphi_\varepsilon^*(\sigma)+\sigma z-\frac{1}{\varepsilon}\psi_J(z)}\,dz \;=\; \frac{\mathrm{e}^{\sigma z-\frac{1}{\varepsilon}\psi_J(z)}}{\int_{\mathbb{R}} \mathrm{e}^{\sigma \bar z-\frac{1}{\varepsilon}\psi_J(\bar z)}\,d\bar z}\,dz. \tag{IV.1.10}$$

$\mu^{\varepsilon,\sigma}$ is closely related to φ_ε^* and φ_ε. This can be seen in Section IV.A.1 in the appendix, where we list several properties of the Cramér transform that are used in this chapter. In particular, we have that φ_ε^* and φ_ε are strictly convex and smooth, and hence, $\varphi_\varepsilon''(m)^{\frac{1}{2}}$ is well defined for all $m \in \mathbb{R}$.

In the following proposition we state the *local Cramér theorem*. (Recall that in Subsection I.6.2 we explain why this result is called like that.) Very similar versions of this result are already known in the literature; see for instance [80, Proposition 31] or [101, Section 3]. The main novelty here is that the result is uniform in $\varepsilon \ll 1$.

Proposition IV.1 (Local Cramér theorem) *Suppose* (IV.1.3). *Let $K \subset \mathbb{R}$ be compact. Then, there exist $N_K \in \mathbb{N}$ and $\varepsilon_K > 0$ such that, for all $\varepsilon < \varepsilon_K$, $N \geq N_K$ and $m \in K$,*

$$\mathrm{e}^{-N\varphi_{N,\varepsilon}(m)} \;=\; \mathrm{e}^{-N\varphi_\varepsilon(m)}\;\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}\;\left(1+O_K\left(\frac{1}{\sqrt{N}}\right)\right). \tag{IV.1.11}$$

In particular, this implies that

$$d\bar\mu(m) \;=\; \frac{1}{Z_{\bar\mu}}\,\mathrm{e}^{-N\bar H_\varepsilon(m)}\;\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}\;\left(1+O_K\left(\frac{1}{\sqrt{N}}\right)\right)\,dm, \tag{IV.1.12}$$

where

$$\bar H_\varepsilon(z) \;=\; \varphi_\varepsilon(z) \;-\; \frac{1}{\varepsilon}\frac{J}{2}z^2. \tag{IV.1.13}$$

Proof. The proof is postponed to Section IV.A.4 in the appendix. □

Analysis of the energy landscape. Proposition IV.1 indicates that the graph of $\bar H_\varepsilon$ determines the macroscopic energy landscape of our system under the order parameter P (see also Subsection I.6.2 for more comments). This suggests to study the analytic properties of $\bar H_\varepsilon$, which is the content of the following lemma.

Lemma IV.2 *Suppose that $J > 0$. Then,*

(i) $\lim_{|t|\mapsto\infty} \frac{1}{t^2}\varphi_\varepsilon(t) \;=\; \infty$, $\lim_{|t|\mapsto\infty} \frac{1}{t^2}\bar H_\varepsilon(t) \;=\; \infty$, *and*

(ii) for all $\varepsilon > 0$ small enough, $\bar H_\varepsilon$ has exactly three critical points located at $-m_\varepsilon^\star$, 0 and $m_\varepsilon^\star$ for some $m_\varepsilon^\star = 1 + \Omega(\varepsilon)$. Moreover, $\bar H_\varepsilon''(0) < 0$, $\bar H_\varepsilon''(m_\varepsilon^\star) > 0$ and $\bar H_\varepsilon''(-m_\varepsilon^\star) > 0$. That is, $\bar H_\varepsilon$ has a local maximum at 0, and the two global minima of $\bar H_\varepsilon$ are located at $\pm m_\varepsilon^\star$.

Proof. Part *(i)* follows from a simple argument, which is based on the fact that ψ_J is superquadratic at infinity and on Hölder's inequality. For instance, a proof can be found in [106, III.2.6] for a slightly more general setting.

To show part *(ii)*, first note that by Lemma IV.A.1, the condition $\bar{H}'_\varepsilon(m) = 0$ is equivalent to

$$m \;=\; (\varphi^*_\varepsilon)'\left(\frac{1}{\varepsilon}Jm\right) \;=\; \int_{\mathbb{R}} z\,\mathrm{e}^{-\varphi^*_\varepsilon\left(\frac{1}{\varepsilon}Jm\right)+\frac{1}{\varepsilon}Jzm-\frac{1}{\varepsilon}\psi_J(z)}\,dz. \tag{IV.1.14}$$

We know from [82, 3.1 and 3.2] that, for ε small enough, there exist exactly three solutions $\pm m^\star_\varepsilon$ and 0 for (IV.1.14), where $m^\star_\varepsilon = 1 + \Omega(\varepsilon)$.

We now show that $\bar{H}''_\varepsilon(0) < 0$ in the case $J > 1$. Using that $\varphi'_\varepsilon(0) = 0$, Lemma IV.A.1 and Corollary IV.A.3 yield that

$$\bar{H}''_\varepsilon(0) \;=\; \left(\int_{\mathbb{R}} z^2\,d\mu^{\varepsilon,0}(z)\right)^{-1} \;-\; \frac{1}{\varepsilon}\,J \;=\; \frac{J-1}{\varepsilon}\,(1+\Omega(\varepsilon)) \;-\; \frac{1}{\varepsilon}\,J \;<\; 0 \tag{IV.1.15}$$

for ε small enough. In the case $J < 1$, we have by standard Laplace asymptotics that for ε small enough, $\bar{H}''_\varepsilon(0) = \Omega(1) - \frac{1}{\varepsilon}J < 0$. The same result holds also for the case $J = 1$, since $\bar{H}''_\varepsilon(0)$ depends continuously on J (cf. Step 5.3 in the proof of [82, 3.2]).

By the symmetry of $\bar{H}_\varepsilon$, it only remains to show that $\bar{H}''_\varepsilon(m^\star_\varepsilon) > 0$. First note that, since $m^\star_\varepsilon = 1+\Omega(\varepsilon)$, for all $J > 0$, the function $z \mapsto \psi_J(z) - Jm^\star z$ admits a unique global minimum at some point $z_\varepsilon = 1 + \Omega(\varepsilon)$. Indeed, in the case $J > 1$, this follows by simply observing that ψ'_J is invertible, and in the case $J \le 1$, we have to apply Cardano's formula (see [23, Chapter 1 and 2]). (We omit the details in the latter case, since we do not use the claim of this lemma for the case $J \le 1$ in the remaining part of this chapter.) Then, as above, using Lemma IV.A.1, Corollary IV.A.3 and that $\varphi'_\varepsilon(m^\star_\varepsilon) = Jm^\star_\varepsilon$ implies that for ε small enough,

$$\begin{aligned}
\bar{H}''_\varepsilon(m^\star_\varepsilon) \;&=\; \left(\int_{\mathbb{R}}\left(z - \int_{\mathbb{R}} \bar{z}\,d\mu^{\varepsilon,Jm^\star_\varepsilon}(\bar{z})\right)^2 d\mu^{\varepsilon,Jm^\star_\varepsilon}(z)\right)^{-1} \;-\; \frac{1}{\varepsilon}\,J \\
&=\; \frac{1}{\varepsilon}\psi''_J\left(z_\varepsilon\right)\left(1 + O\left(\varepsilon\sqrt{\log(\varepsilon^{-1})^3}\right)\right) \;-\; \frac{1}{\varepsilon}\,J \\
&=\; \frac{1}{\varepsilon}\left(3z_\varepsilon^2 - 1\right)\left(1 + O\left(\varepsilon\sqrt{\log(\varepsilon^{-1})^3}\right)\right) \;>\; 0,
\end{aligned} \tag{IV.1.16}$$

which concludes the proof. □

Remark IV.3 *In the remaining part of this section, we suppose that ε is small enough such that $[-m^\star_\varepsilon, m^\star_\varepsilon] \subset [-2, 2]$.*

Potential-theoretic approach to metastability. In this paragraph, we quickly review the key ingredients form the potential-theoretic approach to metastability that we need in our setting. We follow [32, Chapter 2], where all the omitted details can be found.

The generator of the stochastic process $(x(t))_{t\in(0,\infty)}$ introduced in Subsection I.6.1 is given by

$$\mathcal{L} \;=\; \varepsilon\,\mathrm{e}^{\mathrm{H}}\left(\nabla\,\mathrm{e}^{-\mathrm{H}}\,\nabla\right), \tag{IV.1.17}$$

where H is the microscopic Hamiltonian (recall (IV.1.1)). We need the following definitions.

Definition IV.4 *Let $A, D \subset \mathbb{R}^N$ be open and regular and such that $A \cap D = \emptyset$ and $(A \cup D)^c$ is connected. For any $B \subset \mathbb{R}^N$, we write $\mathcal{T}_B = \inf\{t > 0 \,|\, x(t) \in B\}$.*

(i) The equilibrium potential between A and D, *$f^*_{A,D}$, is defined as the unique solution to the Dirichlet problem*

$$\begin{aligned} (-\mathcal{L}f)(x) &= 0, \qquad \text{for } x \in (A \cup D)^c, \\ f(x) &= 1, \qquad \text{for } x \in A, \\ f(x) &= 0, \qquad \text{for } x \in D. \end{aligned} \tag{IV.1.18}$$

For $x \in (A \cup D)^c$, we have the probabilistic interpretation that $f^*_{A,D} = \mathbb{P}_x[\mathcal{T}_A < \mathcal{T}_D]$.

(ii) The equilibrium measure, *$e_{A,D}$, is defined as the unique measure on ∂A such that*

$$f^*_{A,D}(x) = \int_{\partial A} G_{D^c}(x,y)\, e_{A,D}(dy) \quad \text{for } x \in (A \cup D)^c, \tag{IV.1.19}$$

where G_{D^c} is the Green function *corresponding to $\mathcal{L}$ on D^c (cf. [32, (2.2)]).*

(iii) The capacity, $\mathrm{Cap}(A, D)$, *of the* capacitor (A, D) *is defined by*

$$\mathrm{Cap}(A, D) = \int_{\partial A} \mathrm{e}^{-\mathrm{H}(y)}\, e_{A,D}(dy). \tag{IV.1.20}$$

(iv) The last-exit biased distribution on A, *$\nu_{A,D}$, is the probability measure on ∂A defined by*

$$\nu_{A,D}(dy) = \frac{\mathrm{e}^{-\mathrm{H}(y)}\, e_{A,D}(dy)}{\mathrm{Cap}(A, D)}. \tag{IV.1.21}$$

Using these notions, one can rewrite the average hitting time of B in the case that the initial condition is randomly chosen according to the last-exit distribution. This is the content of the following lemma.

Lemma IV.5 *Consider the same setting as in Definition IV.4. Then,*

$$\mathbb{E}_{\nu_{A,D}}[\mathcal{T}_D] := \int_{\partial A} \mathbb{E}_y[\mathcal{T}_D]\, \nu_{A,D}(dy) = \frac{\int_{D^c} f^*_{A,D}(y)\, e^{-\mathrm{H}(y)}\, dy}{\mathrm{Cap}(A, D)}. \tag{IV.1.22}$$

Proof. The proof can be found in [29, 7.30]. See also [32, (2.27)]. □

As we already mentioned, the main advantage to use Lemma IV.5 is the availability of variational principles for the capacity. In this chapter, we use the so-called *Dirichlet principle*, which is stated in the following lemma.

Lemma IV.6 (Dirichlet principle) *Consider the same setting as in Definition IV.4. Let*

$$\mathcal{H}_{A,D} = \left\{ f \in H^1(\mathbb{R}^N;\, \mathrm{e}^{-\mathrm{H}(x)}\, dx) \;\middle|\; f|_A = 1,\ f|_D = 0, \forall x \in \mathbb{R}^N : f(x) \in [0,1], \right\}, \tag{IV.1.23}$$

and define the Dirichlet form on $(A \cup D)^c$, $\mathcal{E}_{(A\cup D)^c} : \mathcal{H}_{A,D} \to [0,\infty]$, *by*

$$\mathcal{E}_{(A\cup D)^c}(f) \;=\; \varepsilon \int_{(A\cup D)^c} |\nabla f(x)|^2 \, \mathrm{e}^{-\mathrm{H}(x)} \, dx \quad \text{for } f \in \mathcal{H}_{A,D}. \tag{IV.1.24}$$

Then,

$$\mathrm{Cap}(A,D) \;=\; \inf_{f\in\mathcal{H}_{A,D}} \mathcal{E}_{(A\cup D)^c}(f) \;=\; \mathcal{E}_{(A\cup D)^c}(f^*_{A,D}). \tag{IV.1.25}$$

Proof. The proof can be found in [29, 7.33]. See also [32, (2.15)]. □

IV.1.2 The Eyring-Kramers formula

We have now collected all the notions that we need to formulate the main result in this chapter. Recall that, under (IV.1.3) and for ε small enough, the macroscopic Hamiltonian admits exactly two global minima $\pm m^\star_\varepsilon$. We therefore consider the hyperplanes $P^{-1}(-m^\star_\varepsilon)$ and $P^{-1}(m^\star_\varepsilon)$ as the *metastable sets* in our system.

The goal in this chapter is to use the potential-theoretic setting to compute the average transition time from $P^{-1}(-m^\star_\varepsilon)$ to $P^{-1}(m^\star_\varepsilon)$ for the stochastic process $(x(t))_{t\in(0,\infty)}$ introduced in Subsection I.6.1. However, due to technical reasons, we have to modify this goal in two ways.

First, instead of considering $P^{-1}(-m^\star_\varepsilon)$ and $P^{-1}(m^\star_\varepsilon)$ as the metastable sets, we rather consider $P^{-1}(-m^\star_\varepsilon + \eta)$ and $P^{-1}(m^\star_\varepsilon - \eta)$, where

$$\eta = \frac{\sqrt{2}}{\sqrt{N\bar{H}''_\varepsilon(-m^\star_\varepsilon)}} \sqrt{\log(N\varepsilon^{-1})}. \tag{IV.1.26}$$

(By using (IV.1.16), we have that $\eta = \Omega(\sqrt{\log(N)/N}\,\sqrt{\varepsilon\log(\varepsilon^{-1})})$.) Heuristically, the reason for this shift is the following. In the proof of our main result, we have to compute the integral in the numerator on the right-hand side of (IV.1.22). Using the disintegration (IV.1.4), Proposition IV.1 and the fact that $\bar{H}_\varepsilon$ has its global minima at $-m^\star_\varepsilon$ and $m^\star_\varepsilon$, we see that this integral is concentrated on the sets $\{\, x \,|\, Px \in [\pm m^\star_\varepsilon - \eta, \pm m^\star_\varepsilon + \eta] \,\}$. Hence, in order to apply Laplace's method, we need that the equilibrium potential is equal to 1 or equal to 0 on these sets, respectively.

Second, instead of running the system from some specific point in $P^{-1}(-m^\star_\varepsilon+\eta)$, we rather have to initialise our system randomly according to the last-exit biased distribution ν_{B^-,B^+}, where $B^-, B^+ \subset \mathbb{R}^N$ are defined by

$$\begin{aligned} B^- \;&=\; \{\, x\in\mathbb{R}^N \mid Px \;\le\; -m^\star_\varepsilon + \eta \,\} \quad \text{and} \\ B^+ \;&=\; \{\, x\in\mathbb{R}^N \mid Px \;\ge\; m^\star_\varepsilon - \eta \,\}. \end{aligned} \tag{IV.1.27}$$

Note that ν_{B^-,B^+} is a probability measure supported on $\partial B^- = P^{-1}(-m^\star_\varepsilon + \eta)$. The main reason for the choice of this initial distribution is that we can exploit the formula (IV.1.22). However, in a finite-dimensional setting, such as in [32], we could also obtain an asymptotic expression for $\mathbb{E}_y[\mathcal{T}_{B^+}]$ for $y \in \partial B^-$. This is done by using Harnack inequalities. But since these inequalities depend on the dimension of the base space, we are not able to transfer the strategy used in [32] to our high-dimensional setting.

In the following theorem we formulate the first main result of this chapter.

Theorem IV.7 *Suppose* (IV.1.3), *and recall the definition of* B^- *and* B^+ *in* (IV.1.27). *Then, for* N *large enough and* ε *small enough,*

$$\mathbb{E}_{\nu_{B^-,B^+}}[\mathcal{T}_{B^+}] \;=\; \frac{2\pi\,\sqrt{\varphi_\varepsilon''(-m_\varepsilon^\star)}\mathrm{e}^{N(\bar H_\varepsilon(0)-\bar H_\varepsilon(-m_\varepsilon^\star))}}{\varepsilon\,\sqrt{\bar H_\varepsilon''(-m_\varepsilon^\star)\,|\bar H_\varepsilon''(0)|\,\varphi_\varepsilon''(0)}}\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)+O\left(\varepsilon\right)\right). \tag{IV.1.28}$$

Proof. Combining (IV.1.22) with the Propositions IV.8, IV.9 and IV.12 concludes the proof. □

IV.1.3 Upper bound on the capacity

Proposition IV.8 *Consider the same setting as in Theorem IV.7. Then, for* N *large enough and* ε *small enough,*

$$\mathrm{Cap}(B^-,B^+) \;\le\; \varepsilon\,\frac{1}{2\pi}\,\mathrm{e}^{-N\bar H_\varepsilon(0)}\,\sqrt{|\bar H_\varepsilon''(0)|}\,\sqrt{\varphi_\varepsilon''(0)}\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right). \tag{IV.1.29}$$

Proof. We will obtain the upper bound by using the Dirichlet principle (Lemma IV.6). That is, we introduce a suitable test function and show that the corresponding Dirichlet form is asymptotically given by the right-hand side of (IV.1.29).

Step 1. [Choice of the test function f.]
Let

$$\rho=\frac{1}{\sqrt{N|\bar H_\varepsilon''(0)|}}\sqrt{\log(N)} \quad\text{and}\quad h^*(m) \;=\; \frac{\int_m^\rho \varphi_\varepsilon''(z)^{-\frac12}\;\mathrm{e}^{N\bar H_\varepsilon(z)}\,dz}{\int_{-\rho}^\rho \varphi_\varepsilon''(z)^{-\frac12}\;\mathrm{e}^{N\bar H_\varepsilon(z)}\,dz}, \tag{IV.1.30}$$

which is well-defined, since φ_ε is strictly convex. Then, h^* is the equilibrium potential corresponding to the invariant measure $\mathbb{1}_{(-\rho,\rho)}(z)\,\varphi_\varepsilon''(z)^{\frac12}\;\mathrm{e}^{-N\bar H_\varepsilon(z)}\,dz$; see [29, Section 7.2.5]. The test function that we use in this proof is given by

$$f(x)=\begin{cases}1 & \text{if } Px \;\le\; -\rho,\\ 0 & \text{if } Px \;\ge\; \rho,\\ h^*(Px) & \text{if } Px \;\in\; (-\rho,\rho).\end{cases} \tag{IV.1.31}$$

Step 2. [Estimation of the Dirichlet form of f.]
Using Lemma IV.6 and (IV.1.4), we have the following upper bound for the capacity.

$$\begin{aligned}\frac{1}{Z_\mu}\mathrm{Cap}(B^-,B^+) &\le \varepsilon\int_{\{x\in\mathbb{R}^N\,|\,Px\in(-\rho,\rho)\}}\sum_{i=0}^{N-1}\left|\partial_i h^*\left(\frac1N\sum_{i=0}^{N-1}x_i\right)\right|^2\;d\mu\\ &= \varepsilon\,\frac1N\int_{\{x\in\mathbb{R}^N\,|\,Px\in(-\rho,\rho)\}}|(h^*)'(Px)|^2\;d\mu\\ &= \varepsilon\,\frac1N\int_{-\rho}^{\rho}\int_{P^{-1}(m)}|(h^*)'(m)|^2\;d\mu_m d\bar\mu(m).\end{aligned} \tag{IV.1.32}$$

Applying Proposition IV.1 for $K = [-2, 2]$ and the definition of h^* yields that

$$\begin{aligned}\frac{1}{Z_\mu}\mathrm{Cap}(B^-, B^+) &\leq \frac{\varepsilon}{\sqrt{2\pi}}\frac{1}{N}\frac{1}{Z_{\bar\mu}}\int_{-\rho}^{\rho} |(h^*)'(m)|^2 \sqrt{\varphi_\varepsilon''(m)}\, \mathrm{e}^{-N\bar H_\varepsilon(m)}\, dm \left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right) \\ &= \frac{\varepsilon}{\sqrt{2\pi}}\frac{1}{N}\frac{1}{Z_{\bar\mu}}\left(\int_{-\rho}^{\rho} \frac{1}{\sqrt{\varphi_\varepsilon''(m)}}\, \mathrm{e}^{N\bar H_\varepsilon(m)}\, dm\right)^{-1} \left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right). \qquad \text{(IV.1.33)}\end{aligned}$$

In Step 4 and 5 of this proof we show that for $m \in [-\rho, \rho]$,

$$\sqrt{\varphi_\varepsilon''(m)} = \sqrt{\varphi_\varepsilon''(0)}\left(1 + O_K\left(\sqrt{\frac{\varepsilon\, \log(N)}{N}}\right)\right), \text{ and} \qquad \text{(IV.1.34)}$$

$$\bar H_\varepsilon(m) = \bar H_\varepsilon(0) + \frac{1}{2}m^2 \bar H_\varepsilon''(0) + O_K\left(\sqrt{\varepsilon}\sqrt{\frac{\log(N)}{N}}^3\right). \qquad \text{(IV.1.35)}$$

And since, by the co-area formula, $Z_{\bar\mu}\sqrt{N} = Z_\mu$, (IV.1.34) and (IV.1.35) imply that

$$\mathrm{Cap}(B^-, B^+) \leq \frac{\varepsilon\sqrt{\varphi_\varepsilon''(0)}\, \mathrm{e}^{-N\bar H_\varepsilon(0)}}{\sqrt{2\pi}\sqrt{N}}\left(\int_{-\rho}^{\rho} \mathrm{e}^{\frac{1}{2}Nm^2\bar H_\varepsilon''(0)}\, dm\right)^{-1}\left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right). \qquad \text{(IV.1.36)}$$

Combining this with the fact that

$$\begin{aligned}\int_{-\rho}^{\rho} \mathrm{e}^{\frac{1}{2}Nm^2\bar H_\varepsilon''(0)}\, dm &\geq \sqrt{\frac{2\pi}{N|\bar H_\varepsilon''(0)|}}\left(1 - \mathrm{e}^{\frac{1}{2}N\rho^2\bar H_\varepsilon''(0)}\right)^{\frac{1}{2}} \\ &= \sqrt{\frac{2\pi}{N|\bar H_\varepsilon''(0)|}}\left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right), \end{aligned} \qquad \text{(IV.1.37)}$$

concludes the proof of (IV.1.29).

Step 3. [Some a priori estimates.]
Before we show (IV.1.34) and (IV.1.35), we collect some a priori estimates. First, we use (IV.A.5) and Lemma IV.A.4 *(iii)* to see that there exists $c > 0$ such that for all $m \in K$ and ε small enough,

$$|\varphi_\varepsilon''(m)| = \varphi_\varepsilon''(m) = \frac{1}{(\varphi_\varepsilon^*)''(\varphi_\varepsilon'(m))} \in \left[\frac{c^{-1}}{\varepsilon}, \frac{c}{\varepsilon}\right]. \qquad \text{(IV.1.38)}$$

Moreover, recall that in the proof of Lemma IV.2, we have seen that $|\bar H_\varepsilon''(0)| = 1/\varepsilon(1+O(\varepsilon))$. Therefore, for ε small enough, $|\bar H_\varepsilon''(0)| \geq 1/(4\varepsilon)$. Next, we recall the definition of $\mu^{\varepsilon,\sigma}$ in (IV.1.10) and use Lemma IV.A.1, (IV.1.38) and Corollary IV.A.3 to see that there exists $c' > 0$ such that for all $m \in [-\rho, \rho] \subset K$,

$$\begin{aligned}|\varphi_\varepsilon'''(m)| &= \left|\frac{(\varphi_\varepsilon^*)'''(\varphi_\varepsilon'(m))}{(\varphi_\varepsilon^*)''(\varphi_\varepsilon'(m))^3}\right| = \left|\int_{\mathbb{R}} (z-m)^3\ d\mu^{\varepsilon,\varphi_\varepsilon'(\theta m)}(z)\right| \varphi_\varepsilon''(m)^3 \\ &\in \left[\frac{(c')^{-1}}{\varepsilon}, \frac{c'}{\varepsilon}\right]. \end{aligned} \qquad \text{(IV.1.39)}$$

Step 4. [Proof of (IV.1.34).]
By Taylor's formula, we have for some $\theta \in [0,1]$,

$$\sqrt{\varphi_\varepsilon''(m)} \;=\; \sqrt{\varphi_\varepsilon''(0)}\,\left(1 \;+\; m\,\frac{\varphi_\varepsilon'''(\theta m)}{2\sqrt{\varphi_\varepsilon''(0)}\sqrt{\varphi_\varepsilon''(\theta m)}}\right). \tag{IV.1.40}$$

Then, by the estimates from Step 3,

$$\begin{aligned}\left|m\frac{\varphi_\varepsilon'''(\theta m)}{2\sqrt{\varphi_\varepsilon''(0)}\sqrt{\varphi_\varepsilon''(\theta m)}}\right| \;&\le\; \frac{1}{2}\,\rho\,c\,c' \;=\; \frac{1}{2}\sqrt{\frac{\log(N)}{N}}\,\frac{c\,c'}{\sqrt{|\bar H_\varepsilon''(0)|}}\\ &\le\; c\,c'\sqrt{\frac{\varepsilon\,\log(N)}{N}}.\end{aligned} \tag{IV.1.41}$$

In combination with (IV.1.40), this yields (IV.1.34).

Step 5. [Proof of (IV.1.35).]
Again by Taylor's formula, for some $\theta' \in [0,1]$,

$$\bar H_\varepsilon(m) \;=\; \bar H_\varepsilon(0) \;+\; \frac{1}{2}m^2\bar H_\varepsilon''(0) \;+\; \frac{1}{6}m^3\bar H_\varepsilon'''(\theta' m). \tag{IV.1.42}$$

Similarly as in Step 4, we have that

$$\begin{aligned}\left|m^3\bar H_\varepsilon'''(\theta' m)\right| \;&\le\; \rho^3\,\left|\varphi_\varepsilon'''(\theta' m)\right| \;\le\; \sqrt{\frac{\log(N)}{N}}^{\,3}\,\frac{1}{\sqrt{|\bar H_\varepsilon''(0)|}^{\,3}}\,\frac{c'}{\varepsilon}\\ &\le\; 8\,c'\,\sqrt{\varepsilon}\,\sqrt{\frac{\log(N)}{N}}^{\,3},\end{aligned} \tag{IV.1.43}$$

which concludes the proof of (IV.1.35). □

IV.1.4 Lower bound on the capacity

In this section, we prove the lower bound on the capacity. The proof is inspired by the two-scale approach, which was initiated in [80]. Moreover, we use that by the *Bakry-Émery theorem* (see for instance [101, A.3] or [80, p. 305] combined with [102, Remark 1.2]), μ_m satisfies the *Poincaré inequality* (recall (IV.0.1) from the introduction) with constant $(J-1)/\varepsilon$. That is, for all $N \in \mathbb{N}$, $m \in \mathbb{R}$ and $f \in H^1(\mu_m)$,

$$\mathrm{Var}_{\mu_m}(f) \;:=\; \int\left|f - \int f d\mu_m\right|^2 d\mu_m \;\le\; \frac{\varepsilon}{J-1}\int\left|(\mathrm{id} - NP^tP)\nabla f\right|^2 d\mu_m, \tag{IV.1.44}$$

where $P^t m = (1/N)(m,\dots,m) \in \mathbb{R}^N$ for $m \in \mathbb{R}$.

Proposition IV.9 *Consider the same setting as in Theorem IV.7. Then, for N large enough and ε small enough,*

$$\mathrm{Cap}(B^-,B^+) \;\ge\; \frac{\varepsilon}{2\pi}\,\mathrm{e}^{-N\bar H_\varepsilon(0)}\,\sqrt{|\bar H_\varepsilon''(0)|}\,\sqrt{\varphi_\varepsilon''(0)}\,\left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right) + O\left(\varepsilon\right)\right). \tag{IV.1.45}$$

Proof. Let $f = f^*_{B^-,B^+}(x)$ (recall Definition IV.4 and Lemma IV.6) and, for $m \in K := [-2,2]$, let

$$\bar{f}(m) \;=\; \int_{P^{-1}(m)} f \, d\mu_m. \tag{IV.1.46}$$

As in [80, Section 2.1], we split the gradient ∇f into its fluctuation part $(\mathrm{id} - NP^tP)\nabla f$ and its macroscopic part $NP^tP\nabla f$. Note that

$$\left|(\mathrm{id} - NP^tP)\nabla f\right|^2 \;+\; \left|NP^tP\nabla f\right|^2 \;=\; |\nabla f|^2 . \tag{IV.1.47}$$

Using (IV.1.4), the fact that $|NP^tPx|^2 = N|Px|^2$ for all $x \in \mathbb{R}^N$, Jensen's inequality and [80, Lemma 21], we obtain that

$$\begin{aligned}
\int \left|NP^tP\nabla f\right|^2 d\mu &\geq N \int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} \int_{P^{-1}(m)} |P\nabla f|^2 \, d\mu_m \, d\bar{\mu}(m) \\
&\geq N \int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} \left| \int_{P^{-1}(m)} P\nabla f \, d\mu_m \right|^2 d\bar{\mu}(m) \\
&= N \int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} \left| \frac{\bar{f}'(m)}{N} + P\mathrm{Cov}_{\mu_m}(f, \nabla \mathrm{H}) \right|^2 d\bar{\mu}(m),
\end{aligned} \tag{IV.1.48}$$

where H is the microscopic Hamiltonian defined in (IV.1.1), and, for two functions $g, h \in L^1(\mu_m)$,

$$\mathrm{Cov}_{\mu_m}(g,h) \;=\; \int g \left(h - \int h d\mu_m \right) d\mu_m. \tag{IV.1.49}$$

Then, using Young's inequality, we have that for all $\tau \in [0,1]$,

$$\begin{aligned}
\int \left|NP^tP\nabla f\right|^2 d\mu \;\geq\; & (1-\tau)\frac{1}{N} \int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} \left|\bar{f}'(m)\right|^2 d\bar{\mu}(m) \\
& + \left(1 - \frac{1}{\tau}\right) N \int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} |P\mathrm{Cov}_{\mu_m}(f, \nabla \mathrm{H})|^2 \, d\bar{\mu}(m).
\end{aligned} \tag{IV.1.50}$$

Later in this proof we show that

$$\frac{1}{N} \int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} \left|\bar{f}'(m)\right|^2 d\bar{\mu}(m) \geq \frac{\mathrm{e}^{-N\bar{H}_\varepsilon(0)}}{2\pi \, Z_\mu} \sqrt{|\bar{H}''_\varepsilon(0)|} \sqrt{\varphi''_\varepsilon(0)} \left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}} \right) \right), \tag{IV.1.51}$$

and that for some constant $c > 0$, which is independent of ε and N,

$$\begin{aligned}
\int_{-m^\star_\varepsilon+\eta_\varepsilon}^{m^\star_\varepsilon-\eta_\varepsilon} |P\mathrm{Cov}_{\mu_m}(f, \nabla \mathrm{H})|^2 \, d\bar{\mu}(m) \;\leq\; & \frac{c}{N} \left(\varepsilon + \frac{1}{\sqrt{N}} \right) \int \left|(\mathrm{id} - NP^tP)\nabla f\right|^2 d\mu \\
& \times \left(1 + O(\varepsilon) + O\left(\frac{1}{\sqrt{N}} \right) \right).
\end{aligned} \tag{IV.1.52}$$

Combining Lemma IV.6 with (IV.1.50), (IV.1.51), (IV.1.52) and (IV.1.47), and choosing

$$\tau \;=\; \frac{c\left(\varepsilon+\frac{1}{\sqrt{N}}\right)}{1+c\left(\varepsilon+\frac{1}{\sqrt{N}}\right)}, \tag{IV.1.53}$$

yields (IV.1.45). It only remains to show (IV.1.51) and (IV.1.52).

Proof of (IV.1.51). Note that by Proposition IV.1,

$$\begin{aligned}&\frac{1}{N}\int_{-m_\varepsilon^\star+\eta_\varepsilon}^{m_\varepsilon^\star-\eta_\varepsilon}\left|\bar f'(m)\right|^2\,d\bar\mu(m)\\&\quad=\frac{1}{NZ_{\bar\mu}}\int_{-m_\varepsilon^\star+\eta_\varepsilon}^{m_\varepsilon^\star-\eta_\varepsilon}|\bar f'|^2\,\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}\,\mathrm{e}^{-N\bar H_\varepsilon(m)}\,dm\,\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right).\end{aligned} \tag{IV.1.54}$$

Then, by the fact that $1=\bar f(-m_\varepsilon^\star+\eta_\varepsilon)=1-\bar f(m_\varepsilon^\star-\eta_\varepsilon)$ and by our knowledge on one-dimensional capacities (see for instance [29, Section 7.2.5]),

$$\begin{aligned}&\int_{-m_\varepsilon^\star+\eta_\varepsilon}^{m_\varepsilon^\star-\eta_\varepsilon}|\bar f'|^2\,\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}\,\mathrm{e}^{-N\bar H_\varepsilon(m)}\,dm\\&\qquad\geq\inf_{\substack{h:\\ h(-m_\varepsilon^\star+\eta_\varepsilon)=1,\\ h(m_\varepsilon^\star-\eta_\varepsilon)=0}}\int_{-m_\varepsilon^\star+\eta_\varepsilon}^{m_\varepsilon^\star-\eta_\varepsilon}|h'|^2\,\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}\,\mathrm{e}^{-N\bar H_\varepsilon(m)}\,dm\\&\qquad=\frac{1}{\sqrt{2\pi}}\left(\int_{-m_\varepsilon^\star+\eta_\varepsilon}^{m_\varepsilon^\star-\eta_\varepsilon}\sqrt{\varphi_\varepsilon''(r)}^{-1}\,\mathrm{e}^{N\bar H_\varepsilon(r)}\,dr\right)^{-1}.\end{aligned} \tag{IV.1.55}$$

Recalling that $\max_{m\in[-m_\varepsilon^\star+\eta_\varepsilon,m_\varepsilon^\star-\eta_\varepsilon]}\bar H_\varepsilon(m)=\bar H_\varepsilon(0)$ and $\sqrt{N}Z_{\bar\mu}=Z_\mu$ by the co-area formula, we conclude (IV.1.51) from standard Laplace asymptotics.

Proof of (IV.1.52). Since μ_m is supported on $P^{-1}(m)$, we have that

$$P\mathrm{Cov}_{\mu_m}\left(f,\nabla\mathrm{H}\right)\;=\;\frac{1}{\varepsilon}\frac{1}{N}\,\mathrm{Cov}_{\mu_m}\left(f,\sum_{i=0}^{N-1}x_i^3\right), \tag{IV.1.56}$$

Then, using Hölder's inequality and (IV.1.44),

$$\begin{aligned}\left|P\mathrm{Cov}_{\mu_m}\left(f,\nabla\mathrm{H}\right)\right|^2\;&\leq\;\frac{1}{\varepsilon^2}\frac{1}{N^2}\mathrm{Var}_{\mu_m}\left(f\right)\;\mathrm{Var}_{\mu_m}\left(\sum_{i=0}^{N-1}x_i^3\right)\\&\leq\;\frac{1}{(J-1)^2N^2}\;\int\left|(\mathrm{id}-NP^tP)\nabla f\right|^2d\mu_m\;\int\left|(\mathrm{id}-NP^tP)\nabla\sum_{i=0}^{N-1}x_i^3\right|^2d\mu_m.\end{aligned} \tag{IV.1.57}$$

It remains to show that the second integral on the right-hand side of (IV.1.57) is bounded from above by $c'(\varepsilon N+\sqrt{N})$ for some constant $c'>0$, which is independent of ε and N.

First we observe that by symmetry,

$$\begin{aligned}\int \left|(\mathrm{id}-NP^tP)\nabla\sum_{i=0}^{N-1}x_i^3\right|^2 d\mu_m &= \int\sum_{i=0}^{N-1}\left|3x_i^2-\frac{1}{N}\sum_{j=0}^{N-1}3x_j^2\right|^2 d\mu_m\\ &= 9N\int\left|x_0^2-\frac{1}{N}\sum_{j=0}^{N-1}x_j^2\right|^2 d\mu_m\\ &= 9N\int x_0^4\,d\mu_m-18\sum_{j=0}^{N-1}\int x_0^2x_j^2d\mu_m+\frac{9}{N}\sum_{l=0}^{N-1}\int\sum_{j=0}^{N-1}x_l^2x_j^2d\mu_m\\ &= 9(N-1)\int x_0^4\,d\mu_m-9(N-1)\int x_0^2x_1^2\,d\mu_m.\end{aligned} \tag{IV.1.58}$$

Then, applying Proposition IV.11, the right-hand side of (IV.1.58) is lower or equal to

$$9(N-1)\int\left|z^2-\int z^2\,d\mu^{\varepsilon,\varphi'_\varepsilon(m)}\right|^2 d\mu^{\varepsilon,\varphi'_\varepsilon(m)} \;+\; O_K\left(\sqrt{N}\right). \tag{IV.1.59}$$

It remains to show that

$$\int\left|z^2-\int z^2\,d\mu^{\varepsilon,\varphi'_\varepsilon(m)}\right|^2 d\mu^{\varepsilon,\varphi'_\varepsilon(m)} \;=\; O_K(\varepsilon). \tag{IV.1.60}$$

In order to show (IV.1.60), we again apply the Bakry-Émery theorem (see e.g. [80, p. 305] and [102, Remark 1.2]) to observe that the measure $\mu^{\varepsilon,\varphi'_\varepsilon(m)}$ satisfies the Poincaré inequality (see (IV.0.1)) with constant $(J-1)/\varepsilon$. Hence,

$$\begin{aligned}\int\left|z^2-\int z^2\,d\mu^{\varepsilon,\varphi'_\varepsilon(m)}\right|^2 d\mu^{\varepsilon,\varphi'_\varepsilon(m)} &= \mathrm{Var}_{\mu^{\varepsilon,\varphi'_\varepsilon(m)}}\left(z^2\right)\\ &\le \frac{4\varepsilon}{J-1}\int z^2d\mu^{\varepsilon,\varphi'_\varepsilon(m)}.\end{aligned} \tag{IV.1.61}$$

A simple computation using Lemma IV.A.4 *(ii)* and Corollary IV.A.3 from the appendix shows that the integral on the right-hand side of (IV.1.61) is uniformly bounded in $m\in K$ and for ε small enough. This concludes the proof of (IV.1.52). □

Remark IV.10 *The proof of* (IV.1.52) *is the main reason for the assumption* (IV.1.3)*. Indeed, in this step, we use that, under* (IV.1.3)*, the (effective) single-site is strictly convex so that we can apply the Bakry-Émery theorem, which in turn yields that we have a good control on the covariance term* $P\mathrm{Cov}_{\mu_m}(f,\nabla\mathrm{H})$ *in* (IV.1.52) *for small* ε*. Note that, intuitively, the quantity* $P\mathrm{Cov}_{\mu_m}(f,\nabla\mathrm{H})$ *describes the microscopic fluctuation of the system around the hyperplane* $P^{-1}(m)$*.*

In (IV.1.59) we use that we can pass from expectations with respect to μ_m to expectations with respect to $\otimes_{i=1}^N\mu^{\varepsilon,\varphi'_\varepsilon(m)}$. Such a statement is known in the literature as the *equivalence of observables* (see [91]). The result in our setting is formulated in the following proposition. The proof is postponed to the appendix.

Proposition IV.11 (Equivalence of observables) *Let $K \subset \mathbb{R}$ be compact. Let $\ell \in \mathbb{N}$, and let $b : \mathbb{R}^\ell \to \mathbb{R}$ be such that*

$$\sup_{m\in K} \int_{\mathbb{R}^\ell} |b(z_0,\dots,z_\ell)|^2 \, d\mu^{\varepsilon,\varphi'_\varepsilon(m),\ell}(z_\varepsilon,\dots,z_\ell) \;<\; \infty, \tag{IV.1.62}$$

where $\mu^{\varepsilon,\varphi'_\varepsilon(m),\ell} = \otimes_{i=1}^{\ell} \mu^{\varepsilon,\varphi'_\varepsilon(m)}$. Then, there exist $C_{b,K,\ell}, \varepsilon_{b,K,\ell} > 0$, $N_{b,K,\ell} \in \mathbb{N}$ such that for all $N \geq N_{b,K,\ell}$,

$$\sup_{0<\varepsilon<\varepsilon_{b,K,\ell}} \sup_{m\in K} \left| \int_{P^{-1}(m)} b(x_0,\dots,x_\ell)\, d\mu_m - \int_{\mathbb{R}^\ell} b(z_0,\dots,z_\ell)\, d\mu^{\varepsilon,\varphi'_\varepsilon(m),\ell} \right| \;\leq\; C_b \frac{1}{\sqrt{N}}. \tag{IV.1.63}$$

Proof. The proof is postponed to Section IV.A.5. □

IV.1.5 The mass of the equilibrium potential

Proposition IV.12 *Consider the same setting as in Theorem IV.7. Then, for ε small enough,*

$$\int_{(B^+)^c} f^*_{B^-,B^+}(y)\, \mathrm{e}^{-\mathrm{H}(y)}\, dy \;=\; \frac{\mathrm{e}^{-N\bar H_\varepsilon(-m^\star_\varepsilon)}}{\sqrt{\bar H''_\varepsilon(-m^\star_\varepsilon)}} \sqrt{\varphi''_\varepsilon(-m^\star_\varepsilon)} \left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right). \tag{IV.1.64}$$

Proof. In this proof, C denotes a varying positive constant, which is independent of ε and N.

Step 1. [Splitting into four regions.]
Recall the definition of η in (IV.1.26). Let $R > 2$ be a positive number, which is independent of N, ε and $m \in K$, and whose precise value is chosen later in Step 3. Using that $f^*_{B^-,B^+}(y) = 1$ for $y \in B^-$, we split the left-hand side of (IV.1.64) according to this R in the following way.

$$\begin{aligned}
&\int_{(B^+)^c} f^*_{B^-,B^+}(y)\, \mathrm{e}^{-\mathrm{H}(y)}\, dy \\
&= \int_{\{P\in[-m^\star_\varepsilon-\eta,-m^\star_\varepsilon+\eta]\}} \mathrm{e}^{-\mathrm{H}(y)}\, dy \;+\; \int_{\{P\in[-m^\star_\varepsilon+\eta,m^\star_\varepsilon-\eta)\}} f^*_{B^-,B^+}(y)\, \mathrm{e}^{-\mathrm{H}(y)}\, dy \\
&\quad + \int_{\{P\in[-R,-m^\star_\varepsilon-\eta]\}} \mathrm{e}^{-\mathrm{H}(y)}\, dy \;+\; \int_{\{x\in\mathbb{R}^N \,|\, Px<-R\}} \mathrm{e}^{-\mathrm{H}(y)}\, dy \\
&=: \; I \;+\; II \;+\; III \;+\; IV.
\end{aligned} \tag{IV.1.65}$$

In Step 2 we compute the asymptotic value of the term I, and in Step 3 and 4 we show that the terms II, III and IV are of lower order than I.

Step 2. [Estimation of the term I.]
Note that, using the same arguments as in Step 4 and Step 5 of the proof of Proposition IV.8, for all $m \in [-m^\star_\varepsilon - \eta, -m^\star_\varepsilon + \eta]$,

$$\begin{aligned}
\sqrt{\varphi''_\varepsilon(m)} \;&=\; \sqrt{\varphi''_\varepsilon(-m^\star_\varepsilon)} \left(1 + O_K\left(\sqrt{\frac{\varepsilon\log(N\varepsilon^{-1})}{N}}\right)\right), \quad \text{and} \\
\bar H_\varepsilon(m) \;&=\; \bar H_\varepsilon(-m^\star_\varepsilon) + \frac{1}{2}(m+m^\star_\varepsilon)^2 \bar H''_\varepsilon(-m^\star_\varepsilon) + O_K\left(\sqrt{\varepsilon}\sqrt{\frac{\log(N\varepsilon^{-1})}{N}}^{\,3}\right).
\end{aligned} \tag{IV.1.66}$$

Then, using the co-area formula in the same way that we did in (IV.1.4)–(IV.1.7) and applying Proposition IV.1 for the compact set $[-R,R]$, we observe that

$$\begin{aligned} I &= \sqrt{N}\int_{-m_\varepsilon^\star-\eta}^{-m_\varepsilon^\star+\eta} \mathrm{e}^{-N\varphi_{N,\varepsilon}(m)+\frac{1}{\varepsilon}N\frac{J}{2}m^2}\,dm \\ &= \sqrt{N}\int_{-m_\varepsilon^\star-\eta}^{-m_\varepsilon^\star+\eta} \mathrm{e}^{-N\bar{H}_\varepsilon(m)}\,\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}\,dm\,\left(1+O\left(\frac{1}{\sqrt{N}}\right)\right). \end{aligned} \tag{IV.1.67}$$

Using (IV.1.66) and arguing as in the proof of Proposition IV.8, we have that for ε small enough,

$$\begin{aligned} I &= \frac{\sqrt{N}}{\sqrt{2\pi}}\,\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star)}\sqrt{\varphi_\varepsilon''(-m_\varepsilon^\star)}\int_{-\eta}^{\eta}\mathrm{e}^{-N\bar{H}_\varepsilon''(-m_\varepsilon^\star)\frac{m^2}{2}}\,dm\,\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right) \\ &= \frac{\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star)}}{\sqrt{\bar{H}_\varepsilon''(-m_\varepsilon^\star)}}\sqrt{\varphi_\varepsilon''(-m_\varepsilon^\star)}\,\left(1+O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right). \end{aligned} \tag{IV.1.68}$$

Step 3. [Estimation of the terms II and III.]
We only consider the term II. The term III can be estimated in the same way. By using that $|f^*_{B^-,B^+}|\le 1$ and by applying the co-area formula and Proposition IV.1 as in Step 1, we have that

$$|II| \;\le\; C\sqrt{N}\int_{-m_\varepsilon^\star+\eta}^{m_\varepsilon^\star-\eta}\mathrm{e}^{-N\bar{H}_\varepsilon(m)}\sqrt{\varphi_\varepsilon''(m)}\,dm. \tag{IV.1.69}$$

Note that, by (IV.1.16), we have that $|\bar{H}_\varepsilon''(-m_\varepsilon^\star)| = \Omega(1/\varepsilon)$. Together with (IV.1.38), this shows that $I=\Omega(\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star)})$. In the following we prove that $II = O(\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star)}\sqrt{N}^{-1})$, which shows that II is of lower order than I. Since $\bar{H}_\varepsilon$ is symmetric and has its two global minima at $\pm m_\varepsilon^\star$, we have that

$$\inf_{m\in[-m_\varepsilon^\star+\eta,m_\varepsilon^\star-\eta]}\bar{H}_\varepsilon(m) \;=\; \bar{H}_\varepsilon(-m_\varepsilon^\star+\eta). \tag{IV.1.70}$$

Then, by (IV.1.38), (IV.1.66) and the definition of η (see (IV.1.26)),

$$\begin{aligned} |I_2| &\le \frac{C\sqrt{N}}{\sqrt{\varepsilon}}\,\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star+\eta)} \le \frac{C\sqrt{N}}{\sqrt{\varepsilon}}\,\mathrm{e}^{-N(\bar{H}_\varepsilon(-m_\varepsilon^\star)+\bar{H}_\varepsilon''(-m_\varepsilon^\star)\frac{\eta^2}{2})} \\ &= \frac{C\sqrt{\varepsilon}}{\sqrt{N}}\,\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star)}. \end{aligned} \tag{IV.1.71}$$

Step 4. [Estimation of the term IV.]
Using Jensen's inequality, we have that $\sum_{i=0}^{N-1}x_i^4 \;\ge\; N(Px)^4$. Then, via the co-area formula,

$$\begin{aligned} |IV| &\le \int_{\{x\in\mathbb{R}^N\,|\,Px<-R\}}\mathrm{e}^{-\frac{1}{\varepsilon}\sum_{i=0}^{N-1}\frac{J-1}{2}y_i^2}\,\mathrm{e}^{-\frac{1}{\varepsilon}N\frac{1}{4}(Py)^4+\frac{1}{\varepsilon}N\frac{J}{2}(Py)^2}\,dy \\ &= \sqrt{N}\int_{-\infty}^{-R}\mathrm{e}^{-\frac{1}{\varepsilon}N\frac{1}{4}m^4+\frac{1}{\varepsilon}N\frac{J}{2}m^2}\int_{P^{-1}(m)}\mathrm{e}^{-\frac{1}{\varepsilon}\sum_{i=0}^{N-1}\frac{J-1}{2}y_i^2}\,d\mathcal{H}^{N-1}\,dm. \end{aligned} \tag{IV.1.72}$$

In Lemma IV.A.6, we show that for all $m\in\mathbb{R}$,

$$\int_{P^{-1}(m)}\mathrm{e}^{-\frac{1}{\varepsilon}\sum_{i=0}^{N-1}\frac{J-1}{2}y_i^2}\,d\mathcal{H}^{N-1} \;=\; \mathrm{e}^{-N\frac{1}{\varepsilon}\frac{J-1}{2}m^2+N\frac{1}{2}\log(2\pi\,\varepsilon\,(J-1)^{-1})}\sqrt{\frac{J-1}{\varepsilon\,2\pi}}. \tag{IV.1.73}$$

Therefore, by [28, 1.1], we have that for ε small enough,

$$\begin{aligned} |IV| \;&\leq\; \sqrt{N}\int_{-\infty}^{-R} \mathrm{e}^{-\frac{1}{\varepsilon}N\frac{1}{4}m^4+\frac{1}{\varepsilon}N\frac{1}{2}m^2}\,dm\,\sqrt{\frac{J-1}{\varepsilon\,2\pi}} \;\leq\; \sqrt{N}\int_{-\infty}^{-R} \mathrm{e}^{-\frac{1}{\varepsilon}N\frac{1}{2}(\frac{R^2}{2}-1)m^2}\,dm\,\sqrt{\frac{J-1}{\varepsilon\,2\pi}} \\ &=\; C\int_{-\infty}^{-R\sqrt{\frac{N}{\varepsilon}(\frac{R^2}{2}-1)}} \mathrm{e}^{-\frac{1}{2}m^2}\,dm \;\leq\; C\,\sqrt{\frac{\varepsilon}{N}}\,\mathrm{e}^{-\frac{1}{2}\frac{N}{\varepsilon}(\frac{R^2}{2}-1)R^2}. \end{aligned} \tag{IV.1.74}$$

Note that $\bar{H}_\varepsilon(-m_\varepsilon^\star) \leq \frac{c}{\varepsilon}$ for some $c > 0$. Indeed, by Lemma IV.A.4 *(ii)*, we have that for some bounded function τ_ε,

$$\begin{aligned} |\varphi_\varepsilon(-m_\varepsilon^\star)| \;&=\; \left|-\int_{-m_\varepsilon^\star}^{0}\varphi_\varepsilon'(m)\,dm \;+\; \varphi_\varepsilon(0)\right| \;=\; \left|-\frac{1}{\varepsilon}\int_{-m_\varepsilon^\star}^{0}\tau_\varepsilon(m)\,dm \;+\; \varphi_\varepsilon(0)\right| \\ &\leq\; \frac{1}{\varepsilon}\,\|\tau_\varepsilon\|_{\mathrm{L}^\infty(K\,;\,dm)}\,m_\varepsilon^\star \;+\; |\varphi_\varepsilon(0)|, \end{aligned} \tag{IV.1.75}$$

and by (IV.A.4) and (IV.A.7),

$$\varphi_\varepsilon(0) \;=\; -\varphi_\varepsilon^*(0) \;=\; \log\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}\psi_J(z)}\,dz \;\leq\; \frac{1}{2}\log\left(C\varepsilon\right). \tag{IV.1.76}$$

Combining (IV.1.75) and (IV.1.76) with the definition of $\bar{H}_\varepsilon$, shows that $\bar{H}_\varepsilon(-m_\varepsilon^\star) \leq \frac{c}{\varepsilon}$ for some $c > 0$. Then, choosing R large enough, the estimate (IV.1.74) implies that $IV = O(\mathrm{e}^{-N\bar{H}_\varepsilon(-m_\varepsilon^\star)}\sqrt{N}^{-1})$. This shows that the term IV is of lower order than I. □

IV.2 Rough estimates at high temperature

In this section, we consider the same system as in Chapter IV.1, but with two key differences. First, we do not consider the low-temperature regime here, that is, throughout this section, we suppose that $\varepsilon = 1$. The second difference is that, instead of $\psi(z) = z^4/4 - z^2/2$, we consider here a class of single-site potentials given by functions of the form $z \mapsto \Psi(z) - \frac{J}{2}z^2$, where $\Psi : \mathbb{R} \mapsto \mathbb{R}$ satisfies Assumption IV.13 below.

Hence, the microscopic Hamiltonian $\mathrm{H}^{N,1} : \mathbb{R}^N \mapsto \mathbb{R}$ in this section is given by

$$\begin{aligned} \mathrm{H}^{N,1}(x) \;&=\; \sum_{i=0}^{N-1}\left(\Psi(x_i) - \frac{J}{2}x_i^2\right) + \frac{J}{4N}\sum_{i,j=0}^{N-1}(x_i - x_j)^2 \\ &=\; \sum_{i=0}^{N-1}\Psi(x_i) - \frac{J}{2N}\sum_{i,j=0}^{N-1}x_i x_j, \end{aligned} \tag{IV.2.1}$$

where $J > 0$. We make the following assumptions on the single-site potential Ψ.

Assumption IV.13 *(1) There is a splitting $\Psi = \Psi_c + \Psi_b$ for some $\Psi_c, \Psi_b \in C^2(\mathbb{R})$, and there are constants $0 < c, c' < \infty$ such that $\Psi_c''(z) \geq c$ and $|\Psi_b|_{C^2} \leq c'$.*

(2) $\Psi(z) = \Psi(-z)$ for all $z \in \mathbb{R}$.

(3) $z \mapsto \Psi'(z)$ is convex on $[0, \infty)$.

(4) If Ψ_c is a quadratic function of the form $\Psi_c(x) = c_\Psi x^2 + c'_\Psi x + c''_\Psi$ for some $c_\Psi, c'_\Psi, c''_\Psi \in \mathbb{R}$, then we suppose that $c_\Psi > J$.

(5) $1/J < \int_{\mathbb{R}} z^2 \, \mathrm{e}^{-\Psi(z)} \, dz / (\int \mathrm{e}^{-\Psi(z)} \, dz)$.

(6) $\sigma \mapsto \int_{\mathbb{R}} (\Psi''(z))^2 \, \mathrm{e}^{-\Psi(z)+\sigma z} \, dz / (\int \mathrm{e}^{-\Psi(z)+\sigma z} \, dz)$ is locally bounded on $\mathbb{R}$.

Remark IV.14 *If $\Psi = \Psi_c$ is a quadratic function, then Assumption IV.13 is not fulfilled for any choice of J. However, we do not expect that Kramers' law holds true in this case, since the macroscopic Hamiltonian $\bar{H}_1$ is not of double-well form, where $\bar{H}_1$ is defined as in* (IV.1.13) *with ψ being replaced by the function $z \mapsto \Psi(z) - \frac{J}{2} z^2$ and with $\varepsilon = 1$. Indeed, from* (IV.A.60)*, we see that $\bar{H}_1$ is a quadratic function and hence not of double-well form.*

This section is organized similarly as Chapter IV.1. That is, in Subsection IV.2.1 we introduce the local Cramér theorem and show that the macroscopic Hamiltonian has a double-well structure. In Subsection IV.2.2 we formulate the main result of this section, which provides rough estimates on the average transition time between the metastable sets, where the metastable sets are defined analogously to Chapter IV.1. The only thing left to prove for this result is the lower bound on the capacity. This is done in Subsection IV.2.3.

IV.2.1 Preliminaries

Local Cramér Theorem. Replacing ψ by the function $z \mapsto \Psi(z) - \frac{J}{2} z^2$ and setting $\varepsilon = 1$, we define the Gibbs measure $\mu^{N,1}$ by (I.6.4), and introduce a disintegration of $\mu^{N,1}$ as $\mu^{N,1}(dx) = \mu_m^{N,1}(dx) \bar{\mu}^{N,1}(dm)$ as in (IV.1.4)–(IV.1.7). Analogously, we define the quantities $\varphi_{N,1}$, φ_1^*, φ_1 and $\mu^{1,\sigma}$ by (IV.1.6), (IV.1.8), (IV.1.9) and (IV.1.10), respectively, by replacing ψ by the function $z \mapsto \Psi(z) - \frac{J}{2} z^2$ and setting $\varepsilon = 1$. Then, the local Cramér theorem in this section is given as follows.

Proposition IV.15 (Local Cramér theorem) *Suppose Assumption IV.13. Then, for N large enough,*

$$\mathrm{e}^{-N\varphi_{N,1}(m)} = \mathrm{e}^{-N\varphi_1(m)} \, \frac{\sqrt{\varphi_1''(m)}}{\sqrt{2\pi}} \, \left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right). \tag{IV.2.2}$$

In particular,

$$d\bar{\mu}^{N,1}(m) = \frac{1}{Z_{\bar{\mu}^{N,1}}} \, \mathrm{e}^{-N\bar{H}_1(m)} \, \frac{\sqrt{\varphi_1''(m)}}{\sqrt{2\pi}} \, \left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right) dm. \tag{IV.2.3}$$

Proof. Using the same notation and proceeding as in the proof of Proposition IV.1, we observe that it suffices to show that

$$\left| g_{N,m}(0) - \frac{1}{\sqrt{2\pi}} \right| = O\left(\frac{1}{\sqrt{N}}\right). \tag{IV.2.4}$$

However, this was already shown in [101, Proposition 3.1 and Lemma 3.2]. □

IV.2.1.1 Analysis of the energy landscape. In the following lemma we show that the macroscopic Hamiltonian $\bar{H}_1$ has the form of a double-well function with at least quadratic growth at infinity.

Lemma IV.16 *Suppose Assumption IV.13. If Ψ_c is a quadratic function, then let c_Ψ denote the leading order coefficient. Otherwise, let $c_\Psi = \infty$. Then, we have that*

(i) $\liminf_{|t|\to\infty} \frac{\varphi_1(t)}{t^2} \;\geq\; c_\Psi$, $\liminf_{|t|\to\infty} \frac{\bar{H}_1(t)}{t^2} \;\geq\; c_\Psi - J/2$,

(ii) there exists $K_J > 0$ and $\delta > 0$ such that $\varphi_1'(t) \geq (J+\delta)t$ for all $t \geq K_J$ and $\varphi_1'(t) \leq (-J-\delta)t$ for all $t \leq -K_J$, and

(iii) $\bar{H}_1$ has exactly three critical points located at $-m_1^\star$, 0 and $m_1^\star$ for some $m_1^\star > 0$. Moreover, $\bar{H}_1''(0) < 0$, $\bar{H}_1''(m_1^\star) > 0$ and $\bar{H}_1''(-m_1^\star) > 0$. That is, $\bar{H}_1$ has a local maximum at 0, and the two global minima of $\bar{H}_1$ are located at $\pm m_1^\star$.

Proof. Since $\varphi_1(t) = \varphi_1(-t)$ for all $t \in \mathbb{R}$, it suffices to prove all claims only on $[0,\infty)$.

(i). As in Lemma IV.2, this statement follows from a simple argument given in [106, III.2.6].

(ii). From part *(i)* and Assumption IV.13 *(4)*, we know that there exist $K' > 0$ and $\delta' > 0$ such that $\varphi_1(t) \geq (J+\delta')t^2$ for all $t \geq K'$. Using that $t \mapsto \varphi_1'(t)$ is increasing (since φ_1 is strictly convex) we obtain that for all $t \geq K'$,

$$(J+\delta')\,t^2 \;\leq\; \varphi_1(t) \;=\; \int_0^t \varphi_1'(r)\,dr + \varphi_1(0) \;\leq\; \varphi_1'(t)t + \varphi_1(0), \tag{IV.2.5}$$

which concludes the claim.

(iii). Before we show the claims, note that the function $z \mapsto \varphi_1'(z)$ is convex on $[0,\infty)$. Indeed, from [58, Theorem 1.2 c)], we know that Assumption IV.13 yields that $z \mapsto (\varphi_1^*)'(z)$ is concave on $[0,\infty)$ (cf. [106, IV.0.4]). Hence, for $w > z$, we have that $(\varphi_1^*)''(\varphi_1'(w)) \leq (\varphi_1^*)''(\varphi_1'(z))$, since, due to the convexity of φ_1, we have that $\varphi_1'(w) \geq \varphi_1'(z)$. Therefore,

$$\varphi_1''(w) \;=\; \frac{1}{(\varphi_1^*)''(\varphi_1'(w))} \;\geq\; \frac{1}{(\varphi_1^*)''(\varphi_1'(z))} \;=\; \varphi_1''(z), \tag{IV.2.6}$$

which shows that $z \mapsto \varphi_1'(z)$ is convex.

To show that $\bar{H}_1$ admits a local maximum at 0, we observe that, since $\varphi_1'(0) = 0$, we have that $\bar{H}_1'(0) = 0$. Moreover, Assumption IV.13 implies that $(\varphi_1^*)''(0) > 1/J$. Therefore, $\varphi_1''(0) < J$ and $\bar{H}_1''(0) < 0$.

It remains to show that there exists a unique point $m_1^\star \in (0,\infty)$ such that $\bar{H}_1'(m_1^\star) = 0$ and $\bar{H}_1''(m_1^\star) > 0$. Using again that $\varphi_1''(0) < J$, we infer that for $z > 0$ small enough,

$$\varphi_1'(z) \;=\; \int_0^z \varphi_1''(r)\,dr \;<\; Jz. \tag{IV.2.7}$$

Moreover, by part *(ii)*, we know that there exists $m_1^\star > z > 0$ such that

$$\varphi_1'(m_1^\star) \;=\; Jm_1^\star \quad \text{and} \quad \varphi_1'(z) \;<\; Jz \;\text{ for all } z \in (0, m_1^\star). \tag{IV.2.8}$$

However, the mean value theorem implies that there exists $z' \in (0, m_1^\star)$ such that $\varphi_1''(z') > J$. Together with the fact that φ_1'' is non-decreasing, this implies that $\varphi_1''(z) > J$ for all $z \geq m_1^\star$. This is in turn yields that

$$\varphi_1'(z) \;>\; Jz \;\text{ for all } z > m_1^\star \qquad \text{and} \qquad \bar{H}_1''(m_1^\star) \;>\; 0. \tag{IV.2.9}$$

Combining (IV.2.8) and (IV.2.9) shows that, at $m_1^\star$, there is the unique global minimum of $\bar{H}_1$ on $[0, \infty)$. □

IV.2.2 Rough estimates on the average transition time

In this section we formulate the main result of this section. Most of its proof is omitted, since it is a straightforward adaptation from the proof of Theorem IV.7. However, the proof of the lower bound on the capacity is modified, since the (effective) single-site potential is not convex in this section. The new proof is given in Subsection IV.2.3.

Theorem IV.17 *Let $\pm m_1^\star$ be the two global minimisers of the macroscopic Hamiltonian $\bar{H}_1$. Let $\eta_1 > 0$ and $B_1^-, B_1^+ \subset \mathbb{R}^N$ be defined by* (IV.1.26) *and* (IV.1.27) *with $\varepsilon = 1$. Then, for some $a > 0$, which is independent of N, and for N large enough,*

$$\mathbb{E}_{\nu_{B_1^-,B_1^+}}[\mathcal{T}_{B_1^+}] \;\geq\; \frac{2\pi\,\sqrt{\varphi_1''(-m_1^\star)}\,\mathrm{e}^{N(\bar{H}_1(0)-\bar{H}_1(-m_1^\star))}}{\sqrt{\bar{H}_1''(-m_1^\star)\,|\bar{H}_1''(0)|\,\varphi_1''(0)}}\,\left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right), \quad \textit{and} \tag{IV.2.10}$$

$$\mathbb{E}_{\nu_{B_1^-,B_1^+}}[\mathcal{T}_{B_1^+}] \;\leq\; (1+a)\frac{2\pi\,\sqrt{\varphi_1''(-m_1^\star)}\,\mathrm{e}^{N(\bar{H}_1(0)-\bar{H}_1(-m_1^\star))}}{\sqrt{\bar{H}_1''(-m_1^\star)\,|\bar{H}_1''(0)|\,\varphi_1''(0)}}\,\left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right). \tag{IV.2.11}$$

Proof. As in the proof of Theorem IV.7, the starting point is the formula (IV.1.22). Then, proceeding exactly as in the proofs of Proposition IV.8 and Proposition IV.12, we can show that

$$\mathrm{Cap}(B_1^-, B_1^+) \;\leq\; \frac{1}{2\pi}\mathrm{e}^{-N\bar{H}(0)}\,\sqrt{|\bar{H}''(0)|}\sqrt{\varphi_1''(0)}\,\left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right), \quad \text{and} \tag{IV.2.12}$$

$$\int_{(B_1^+)^c} f^*_{B_1^-,B_1^+}(y)\,\mathrm{e}^{-\mathrm{H}(y)}\,dy \;=\; \frac{\mathrm{e}^{-N\bar{H}_1(-m_1^\star)}}{\sqrt{\bar{H}_1''(-m_1^\star)}}\,\sqrt{\varphi_1''(-m_1^\star)}\,\left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}}\right)\right), \tag{IV.2.13}$$

which yields (IV.2.10). Finally, (IV.2.11) follows from combining (IV.2.13) with Proposition IV.18. This concludes the proof of this theorem. □

IV.2.3 Rough lower bound on the capacity

In this section we prove the rough lower bound on the capacity. We proceed as in the proof of Proposition IV.9. Recall that the critical estimate in the proof of Proposition IV.9 is given by (IV.1.52), where we apply the Poincaré inequality for the fluctuation measure with a constant which is of order $1/\varepsilon$ (see (IV.1.44)). By using the strict convexity of the (effective) single-site potential, (IV.1.44) is a consequence of the Bakry-Émery theorem. Since the (effective) single-site potential is not assumed to be strictly convex in this section, the Bakry-Émery theorem

is not applicable here. However, instead, we can apply [101, 1.6], where it is shown that for all $N \in \mathbb{N}$ and $m \in \mathbb{R}$, $\mu_m^{N,1}$ satisfies the Poincaré inequality with a constant $\varrho > 0$, which is independent of N and m. That is, for all $N \in \mathbb{N}$ and $m \in \mathbb{R}$ and for all $f \in H^1(\mu_m^{N,1})$,

$$\mathrm{Var}_{\mu_m^{N,1}}(f) \;=\; \int \left| f - \int f d\mu_m^{N,1} \right|^2 d\mu_m^{N,1} \;\le\; \frac{1}{\varrho} \int \left| (\mathrm{id} - NP^tP)\nabla f \right|^2 d\mu_m^{N,1}, \tag{IV.2.14}$$

where $P^t m = (1/N)(m, \dots, m) \in \mathbb{R}^N$ for $m \in \mathbb{R}$. This is the main ingredient of the proof of the following proposition.

Proposition IV.18 *Consider the same setting as in Theorem IV.17. Let*

$$a \;=\; \frac{1}{\rho^2} \max_{m \in [-m_1^\star, m_1^\star]} \int \left| \Psi'' - \int \Psi'' d\mu^{1,\varphi_1'(m)} \right|^2 d\mu^{1,\varphi_1'(m)}, \tag{IV.2.15}$$

which is finite due to Assumption IV.13. Then, for N large enough,

$$\mathrm{Cap}(B_1^-, B_1^+) \;\ge\; \frac{1}{1+a} \frac{1}{2\pi} \mathrm{e}^{-N\bar{H}(0)} \sqrt{|\bar{H}''(0)|} \sqrt{\varphi_1''(0)} \left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}} \right) \right). \tag{IV.2.16}$$

Proof. Let $f = f^*_{B_1^-, B_1^+}(x)$. We proceed exactly as in the proof of Proposition IV.9, and obtain that for all $\tau \in [0,1]$,

$$\begin{aligned} \int \left| NP^tP\nabla f \right|^2 d\mu^{N,1} \;\ge\; & (1-\tau) \frac{\mathrm{e}^{-N\bar{H}(0)}}{2\pi\, Z_{\mu^{N,1}}} \sqrt{|\bar{H}''(0)|} \sqrt{\varphi_1''(0)} \left(1 + O\left(\sqrt{\frac{\log(N)^3}{N}} \right) \right) \\ & + \left(1 - \frac{1}{\tau} \right) N \int_{-m_1^\star + \eta_1}^{m_1^\star - \eta_1} \left| P\mathrm{Cov}_{\mu_m^{N,1}}(f, \nabla H) \right|^2 d\bar{\mu}^{N,1}(m). \end{aligned} \tag{IV.2.17}$$

Therefore, choosing $\tau = a/(1+a)$ it remains to show that

$$\int_{-m_1^\star + \eta_1}^{m_1^\star - \eta_1} \left| P\mathrm{Cov}_{\mu_m^{N,1}}(f, \nabla H) \right|^2 d\bar{\mu}^{N,1}(m) \le \frac{a}{N} \int \left| (\mathrm{id} - NP^tP)\nabla f \right|^2 d\mu^{N,1} \left(1 + O\left(\frac{1}{\sqrt{N}} \right) \right). \tag{IV.2.18}$$

In order to show (IV.2.18), note that as in (IV.1.57),

$$\begin{aligned} \left| P\mathrm{Cov}_{\mu_m^{N,1}}(f, \nabla H) \right|^2 \;&\le\; \frac{1}{N^2} \mathrm{Var}_{\mu_m^{N,1}}(f)\; \mathrm{Var}_{\mu_m^{N,1}} \left(\sum_{i=0}^{N-1} \Psi'(x_i) \right) \\ &\le\; \frac{1}{\varrho^2 N^2} \int \left| (\mathrm{id} - NP^tP)\nabla f \right|^2 d\mu_m^{N,1} \int \left| (\mathrm{id} - NP^tP)\nabla \sum_{i=0}^{N-1} \Psi'(x_i) \right|^2 d\mu_m^{N,1}. \end{aligned} \tag{IV.2.19}$$

Then, we proceed analogously to (IV.1.58) to observe that by the equivalence of ensembles (Proposition IV.19),

$$\begin{aligned} & \int \left| (\mathrm{id} - NP^tP)\nabla \sum_{i=0}^{N-1} \Psi'(x_i) \right|^2 d\mu_m^{N,1} \\ & \quad \le\; N \max_{m \in [-m_1^\star, m_1^\star]} \int \left| \Psi'' - \int \Psi'' d\mu^{1,\varphi_1'(m)} \right|^2 d\mu^{1,\varphi_1'(m)} \left(1 + O\left(\frac{1}{\sqrt{N}} \right) \right). \end{aligned} \tag{IV.2.20}$$

This concludes the proof of (IV.2.18). □

It remains to show the equivalence of observables, which was used in (IV.2.20). This is done in the following proposition.

Proposition IV.19 (Equivalence of observables) *Let $\ell \in \mathbb{N}$, and let $b : \mathbb{R}^\ell \to [0,\infty)$ be such that*

$$\sup_{m\in[-m_1^\star,m_1^\star]} \int_{\mathbb{R}^\ell} |b(z_1,\dots,z_\ell)|^2 \, d\mu^{1,\varphi_1'(m),\ell}(z_1,\dots,z_\ell) < \infty, \tag{IV.2.21}$$

where $\mu^{1,\varphi_1'(m),\ell} = \otimes_{i=1}^{\ell} \mu^{1,\varphi_1'(m)}$. Then there exists $C_b \in (0,\infty)$ such that for N large enough,

$$\sup_{m\in[-m_1^\star,m_1^\star]} \left[\int_{P^{-1}(m)} b(x_1,\dots,x_\ell)\, d\mu_m^{N,1} - \int_{\mathbb{R}^\ell} b(z_1,\dots,z_\ell)\, d\mu^{1,\varphi_1'(m),\ell} \right| \leq C_b \frac{1}{\sqrt{N}}. \tag{IV.2.22}$$

Proof. Proceeding exactly as in the proof of Proposition IV.11, we observe that the claim is proven once we show that

(i) the local Cramér theorem holds true in this setting,

(ii) $\sup_{m\in\mathbb{R}} \sum_{k=1}^{3} \int_{\mathbb{R}} \left| \frac{z-m}{s_1(m)} \right|^k d\mu^{1,\varphi_1'(m)}(z) < \infty$, where $s_1(m) = \varphi_1''(m)^{\frac{1}{2}}$, and

(iii) there exists $c > 0$ such that $\sup_{m\in\mathbb{R}} \left| \int_{\mathbb{R}} \mathrm{e}^{iz\xi} d\mu^{1,\varphi_1'(m)}(z) \right| \leq c|s_1(m)\xi|^{-1}$ for all $\xi \in \mathbb{R}$.

Claim (i) is shown in Proposition IV.15, and claim (ii) and (iii) are shown in [101, 3.2]. This concludes the proof of this proposition. □

IV.A Appendix

This appendix is organized as follows. In Subsection IV.A.1 we collect several properties of Cramér transforms and the cumulant generating functions. In Subsection IV.A.2 we derive asymptotic expressions for certain integrals by using standard Laplace asymptotics. Then, in Subsection IV.A.3 we apply these results to estimate the moments and the Fourier transforms of the measure $\mu^{\varepsilon,\varphi_\varepsilon'(m)}$ (see (IV.1.10)) for small ε. Finally, in Subsection IV.A.4 and Subsection IV.A.5 we prove *the local Cramér theorem* (Proposition IV.1) and *the equivalence of observables* (Proposition IV.11), respectively.

We note that the proofs in Subsection IV.A.4 and Subsection IV.A.5 remain true if we replace the effective single-site potentials ψ_J by some general strictly convex function.

IV.A.1 Properties of the Cramér transform

Lemma IV.A.1 *Let $W \in C^\infty(\mathbb{R})$ be such that $\liminf_{|z|\to\infty} W''(z) > 0$. Let*

$$\chi^*(\sigma) = \log \int_{\mathbb{R}} \mathrm{e}^{\sigma z - W(z)}\, dz, \quad \text{for } \sigma \in \mathbb{R}, \tag{IV.A.1}$$

and let χ denote its Legendre transform, i.e.

$$\chi(m) = \sup_{\sigma\in\mathbb{R}} \left(\sigma m - \chi^*(\sigma)\right). \tag{IV.A.2}$$

For all $\sigma \in \mathbb{R}$, define $\mu^\sigma \in \mathcal{M}_1(\mathbb{R})$ by

$$d\mu^\sigma(z) \;=\; \mathrm{e}^{-\chi^*(\sigma)+\sigma z - W(z)}\, dz \;=\; \frac{\mathrm{e}^{\sigma z - W(z)}}{\int_{\mathbb{R}} \mathrm{e}^{\sigma \bar{z} - W(\bar{z})}\, d\bar{z}}\, dz. \tag{IV.A.3}$$

Then, the following statements hold true.

(i) χ^* and χ are strictly convex and smooth. If W is even, then χ^* and χ are also even.

(ii) For $m \in \mathbb{R}$, we have that

$$\chi(m) \;=\; \chi'(m)m - \chi^*(\chi'(m)) \quad \text{and} \quad (\chi^*)'(\chi'(m)) \;=\; m. \tag{IV.A.4}$$

In particular,

$$\chi''(m) \;=\; \frac{1}{(\chi^*)''(\chi'(m))} \quad \text{and} \quad \chi'''(m) \;=\; \frac{-(\chi^*)'''(\chi'(m))}{(\chi^*)''(\chi'(m))^3}. \tag{IV.A.5}$$

(iii) For all $\sigma \in \mathbb{R}$,

$$\begin{aligned}
(\chi^*)'(\sigma) \;&=\; \frac{\int_{\mathbb{R}} z\, \mathrm{e}^{\sigma z - W(z)}\, dz}{\int_{\mathbb{R}} \mathrm{e}^{\sigma z - W(z)}\, dz} \;=\; \int_{\mathbb{R}} z\, d\mu^\sigma(z),\\
(\chi^*)''(\sigma) \;&=\; \int_{\mathbb{R}} \left(z - (\chi^*)'(\sigma)\right)^2 d\mu^\sigma(z),\\
(\chi^*)'''(\sigma) \;&=\; \int_{\mathbb{R}} \left(z - (\chi^*)'(\sigma)\right)^3 d\mu^\sigma(z),\\
(\chi^*)^{(4)}(\sigma) \;+\; 3\,(\chi^*)''(\sigma)^2 \;&=\; \int_{\mathbb{R}} \left(z - (\chi^*)'(\sigma)\right)^4 d\mu^\sigma(z).
\end{aligned} \tag{IV.A.6}$$

Proof. These are standard results that follow from some elementary computations. We refer to [106, III.2.5] and [80, Lemma 41] for more details. □

IV.A.2 Some asymptotic integrals

The main result in this subsection is the following lemma, which is based on Laplace asymptotics. In the proof we use the same strategy as in [82, A.3].

Lemma IV.A.2 *Let $\mathcal{K} \subset \mathbb{R}$ be a compact set. Let $U \in C^{0,\infty}(\mathcal{K}\times\mathbb{R})$, and for $m \in \mathcal{K}$, let $U_m(z) = U(m,z)$. Suppose that there exists $\alpha > 0$ and $R > 0$ such that, for all $m \in \mathcal{K}$, U_m admits a unique global minimum at some point $z_m \in \mathbb{R}$ with $U_m''(z_m) > R^{-1}$ and such that $U_m(z) \geq \alpha z^2$ for all $z \in [-R,R]^c$. Furthermore, we assume that the map $m \mapsto z_m$ is bounded on $\mathcal{K}$. Then, for each $k \in \mathbb{N}_0$ and for each $m \in \mathcal{K}$,*

$$\int_{\mathbb{R}} (z - z_m)^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz \;=\; \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)}\, \frac{\sqrt{2\pi}\,(2k-1)!!\,\varepsilon^{k+\frac{1}{2}}}{U_m''(z_m)^{k+\frac{1}{2}}} \left(1 \;+\; O_{\mathcal{K}}\left(\sqrt{\varepsilon\, \log(\varepsilon^{-1})^3}\right)\right), \tag{IV.A.7}$$

where for $n \in \mathbb{N}$, $n!!$ denotes the double factorial, and we make the convention that $(-1)!! := 1$. Moreover,

$$\begin{aligned}&\int_{\mathbb{R}} (z-z_m)^{2k+1}\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz \\ &\quad = -\mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)} \frac{\sqrt{2\pi}(2k+3)!!\, U_m'''(z_m)\varepsilon^{k+\frac{3}{2}}}{6U_m''(z_m)^{k+\frac{5}{2}}} \left(1 + O_{\mathcal{K}}\left(\sqrt{\varepsilon \log(\varepsilon^{-1})^3}\right)\right).\end{aligned} \tag{IV.A.8}$$

Proof. Fix $m \in \mathcal{K}$. In this proof, let C denote a varying positive constant, which is independent of ε and m.

Step 1. [Proof of (IV.A.7).]
Let $\rho = \sqrt{2(k+1)\,\varepsilon\, \log(\varepsilon^{-1})}/\sqrt{U_m''(z_m)}$ and $\bar{U}_m(z) = U_m(z+z_m)$. Let $\bar{R} \geq R + \sup_{m\in\mathcal{K}}(|z_m| + \sqrt{|U_m(z_m)|/\alpha})$ be such that, for some $\iota > 0$, $y^{2k} \leq e^{\iota U_m(y)}$ for all $y \in [-\bar{R}, \bar{R}]^c$. Then,

$$\begin{aligned}\int_{\mathbb{R}} (z-z_m)^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz \;&=\; \int_{\mathbb{R}} y^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}\bar{U}_m(y)}\, dy \\ &= \int_{-\rho}^{\rho} y^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}\bar{U}_m(y)}\, dy \;+\; \int_{B_{\bar{R}}(0)^c} y^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}\bar{U}_m(y)}\, dy \;+\; \int_{B_{\bar{R}}(0)\setminus B_\rho(0)} y^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}\bar{U}_m(y)}\, dy \\ &=: \; I \;+\; II \;+\; III.\end{aligned} \tag{IV.A.9}$$

In the following we show that I provides the main contribution and that II and III are negligible.

Step 1.1. [Estimation of the term I.]
Note that by Taylor's formula, for some $\theta \in [0,1]$,

$$\bar{U}_m(y) \;=\; U_m(z_m) \;+\; \frac{1}{2}\, y^2\, U_m''(z_m) \;+\; \frac{1}{6}\, y^3\, \bar{U}_m'''(\theta y). \tag{IV.A.10}$$

By using that $\bar{U}_m'''$ is locally bounded (uniformly in $m \in \mathcal{K}$), we see that there exists some $c > 0$ such that $|\bar{U}_m'''(\theta y)| \leq c$ for all $y \in [-\rho, \rho]$. Therefore,

$$\mathrm{e}^{\frac{-c\rho^3}{6\varepsilon}} \;\leq\; \frac{\int_{-\rho}^{\rho} y^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}\bar{U}_m(y)}\, dy}{\mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)} \int_{-\rho}^{\rho} y^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}\frac{1}{2}\, y^2\, U_m''(z_m)}\, dy} \;\leq\; \mathrm{e}^{\frac{c\rho^3}{6\varepsilon}}. \tag{IV.A.11}$$

Thus, by using the definition of ρ and by some standard Gaussian computations applied to the denominator in (IV.A.11), we infer that

$$I \;=\; \sqrt{\frac{2\pi\varepsilon}{U_m''(z_m)}}\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)} \left(\varepsilon^k\, \frac{(2k-1)!!}{U_m''(z_m)^k} \;+\; O_{\mathcal{K}}\left(\varepsilon^{k+\frac{1}{2}} \sqrt{\log(\varepsilon^{-1})^3}\right)\right). \tag{IV.A.12}$$

Step 1.2. [Estimation of the term II.]
We know that $\bar{U}_m(y) \geq \alpha y^2$ and $y^{2k} \leq e^{\iota U_m(y)}$ for all $y \in [-\bar{R}, \bar{R}]^c$. Hence, by [28, 1.1],

$$II \;\leq\; 2\int_{\bar{R}}^{\infty} \mathrm{e}^{-\left(\frac{\alpha}{\varepsilon} - \iota\right)y^2}\, dy \;\leq\; C\, \mathrm{e}^{-\frac{\alpha}{\varepsilon}\bar{R}^2}. \tag{IV.A.13}$$

Since $\alpha\bar{R}^2 > |U_m(z_m)|$, this shows that

$$II \;=\; \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)}\, O_{\mathcal{K}}\left(\varepsilon^{k+1}\sqrt{\log(\varepsilon^{-1})^3}\right). \tag{IV.A.14}$$

Step 1.3. [Estimation of the term III.]
Since $\bar{U}_m$ has its unique minimum in 0, we have that $\inf_{y\in B_{\bar{R}}(0)\setminus B_\rho(0)} \bar{U}_m(y) = \bar{U}_m(\rho)\wedge\bar{U}_m(-\rho)$ for ε small enough. Without restriction, we suppose that $\bar{U}_m(\rho) \le \bar{U}_m(-\rho)$. Then, by using (IV.A.10) and the arguments from Step 1.1,

$$|III| \;\le\; 2\bar{R}\,\bar{R}^{2k}\,\mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m+\rho)} \;\le\; C\,\mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)}\,\mathrm{e}^{-\frac{1}{\varepsilon}U_m''(z_m)\frac{1}{2}\rho^2}. \tag{IV.A.15}$$

Using the definition of ρ, we have shown that

$$III \;=\; \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)}\,O_{\mathcal{K}}\left(\varepsilon^{k+1}\sqrt{\log(\varepsilon^{-1})^3}\right). \tag{IV.A.16}$$

Step 2. [Proof of (IV.A.8).]
(IV.A.8) follows by proceeding exactly as in Step 1 (with $\bar{\rho} = \sqrt{2(k+2)\,\varepsilon\,\log(\varepsilon^{-1})}/\sqrt{U_m''(z_m)}$ replacing ρ) but with the only difference that here we estimate the leading order term I in the following way. The idea is based on Step 2.3 of the proof of [82, A.3]. First, by adding one more term in the Taylor expansion in (IV.A.10), we have that for some $\theta\in[0,1]$,

$$\bar{U}_m(y) \;=\; U_m^0 \;+\; \frac{1}{2}\,y^2\,U_m^2 \;+\; \frac{1}{6}\,y^3\,U_m^3 \;+\; \frac{1}{24}\,y^4\,\bar{U}_m^{(4)}(\theta y), \tag{IV.A.17}$$

where for $i=0,1,2,3$, we abbreviate $U_m^i := U_m^{(i)}(z_m)$. Then,

$$\begin{aligned}
\mathrm{e}^{\frac{1}{\varepsilon}U_m(z_m)}I \;&=\; \mathrm{e}^{\frac{1}{\varepsilon}U_m^0}\int_{-\bar{\rho}}^{\bar{\rho}} y^{2k+1}\,\mathrm{e}^{-\frac{1}{\varepsilon}\bar{U}_m(y)}\,dy \;=\; \int_{-\bar{\rho}}^{\bar{\rho}} y^{2k+1}\,\mathrm{e}^{-\frac{1}{\varepsilon}(\frac{y^2}{2}U_m^2+\frac{y^3}{6}U_m^3+\frac{y^4}{24}\bar{U}_m^{(4)}(\theta y))}\,dy\\
&=\; -\frac{1}{6\varepsilon}U_m^3\int_{-\bar{\rho}}^{\bar{\rho}} y^{2k+4}\,\mathrm{e}^{-\frac{1}{\varepsilon}\frac{1}{2}y^2U_m^2}\,dy \;-\; \frac{1}{24\varepsilon}\int_{-\bar{\rho}}^{\bar{\rho}} y^{2k+5}\,\bar{U}_m^{(4)}(\theta y)\,\mathrm{e}^{-\frac{1}{\varepsilon}\frac{y^2}{2}U_m^2}\,dy\\
&\quad+\; \int_{-\bar{\rho}}^{\bar{\rho}} y^{2k+1}\,\mathrm{e}^{-\frac{1}{\varepsilon}\frac{y^2}{2}U_m^2}\left(\mathrm{e}^{-\frac{1}{\varepsilon}(\frac{y^3}{6}U_m^3+\frac{y^4}{24}\bar{U}_m^{(4)}(\theta y))}-1+\frac{y^3}{6\varepsilon}U_m^3+\frac{y^4}{24\varepsilon}\bar{U}_m^{(4)}(\theta y)\right)dy\\
&=:\; I_1 \;+\; I_2 \;+\; I_3.
\end{aligned} \tag{IV.A.18}$$

We now show that the term I_1 provides the dominant contribution and that I_2 and I_3 are of lower order than I_1. Concerning I_1, simple Gaussian computations as in Step 1.1 yield that

$$I_1 \;=\; -\frac{1}{6}U_m^3\sqrt{\frac{2\pi\varepsilon}{U_m''(z_m)}}\;\left(\varepsilon^{k+1}\,\frac{(2k+3)!!}{U_m''(z_m)^{k+2}} \;+\; O_{\mathcal{K}}\left(\varepsilon^{k+1+\frac{1}{2}}\sqrt{\log(\varepsilon^{-1})^3}\right)\right). \tag{IV.A.19}$$

For I_2 we use that $\bar{U}_m^{(4)}$ is locally bounded to obtain that

$$|I_2| \;\le\; C\frac{1}{\varepsilon}\int_{-\bar{\rho}}^{\bar{\rho}} |y|^{2k+5}\,\mathrm{e}^{-\frac{1}{\varepsilon}\frac{y^2}{2}U_m^2}\,dy \;\le\; C\varepsilon^{k+2}. \tag{IV.A.20}$$

Finally, to estimate the term I_3, note that $y^4\le y^3$ for $y\in[-\bar{\rho},\bar{\rho}]$, $\bar{U}_m'''$ and $\bar{U}_m^{(4)}$ are locally bounded, and that $\bar{\rho}^3/\varepsilon \le C\sqrt{\varepsilon\log(\varepsilon^{-1})}$. Then, by using the inequality $|\mathrm{e}^{-x}-1+x|\le|x|^2\mathrm{e}^{|x|}$,

$$\begin{aligned}
|I_3| \;&\le\; \int_{-\bar{\rho}}^{\bar{\rho}} |y|^{2k+1}\,\mathrm{e}^{-\frac{1}{\varepsilon}\frac{y^2}{2}U_m^2}\,\mathrm{e}^{|\frac{1}{\varepsilon}(\frac{y^3}{6}U_m^3+\frac{y^4}{24}\bar{U}_m^{(4)}(\theta y))|}\left(\frac{y^3}{6\varepsilon}U_m^3+\frac{y^4}{24\varepsilon}\bar{U}_m^{(4)}(\theta y)\right)^2 dy\\
&\le\; \frac{C}{\varepsilon^2}\,\mathrm{e}^{C\frac{\bar{\rho}^3}{\varepsilon}}\int_{-\bar{\rho}}^{\bar{\rho}} |y|^{2k+1}\,\mathrm{e}^{-\frac{1}{\varepsilon}\frac{y^2}{2}U_m^2}\,|y|^6\,dy \;\le\; C\varepsilon^{k+2}.
\end{aligned} \tag{IV.A.21}$$

This concludes the proof of (IV.A.8). □

Corollary IV.A.3 *Consider the same setting as in Lemma IV.A.2. Then,*

$$\frac{\int_{\mathbb{R}} (z-z_m)^{2k}\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz} \;=\; \varepsilon^k\, \frac{(2k-1)!!}{U_m''(z_m)^k} \;+\; O_{\mathcal{K}}\left(\varepsilon^{k+\frac{1}{2}}\sqrt{\log(\varepsilon^{-1})^3}\right), \tag{IV.A.22}$$

and

$$\frac{\int_{\mathbb{R}} (z-z_m)^{2k+1}\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz} \;=\; -\frac{(2k+3)!!\, U_m'''(z_m)\varepsilon^{k+1}}{6U_m''(z_m)^{k+2}} + O_{\mathcal{K}}\left(\varepsilon^{k+\frac{3}{2}}\sqrt{\log(\varepsilon^{-1})^3}\right). \tag{IV.A.23}$$

Moreover,

$$\frac{\int_{\mathbb{R}} \left(\bar z - \frac{\int_{\mathbb{R}} z\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}\right)^{2k} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(\bar z)}\, d\bar z}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz} = \varepsilon^k\, \frac{(2k-1)!!}{U_m''(z_m)^k} + O_{\mathcal{K}}\left(\varepsilon^{k+1}\sqrt{\log(\varepsilon^{-1})^3}\right), \tag{IV.A.24}$$

and

$$\begin{aligned}\frac{\int_{\mathbb{R}} \left(\bar z - \frac{\int_{\mathbb{R}} z\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}\right)^{2k+1} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(\bar z)}\, d\bar z}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz} \\ = -\frac{2k(2k+1)!!\, U_m'''(z_m)\varepsilon^{k+1}}{6U_m''(z_m)^{k+2}} + O_{\mathcal{K}}\left(\varepsilon^{k+\frac{3}{2}}\sqrt{\log(\varepsilon^{-1})^3}\right).\end{aligned} \tag{IV.A.25}$$

Proof. To show (IV.A.22), similarly as in Step 5 in the proof of [82, A.3], we apply (IV.A.7) both to the numerator and to the denominator on the left-hand side of (IV.A.22). Analogously, we apply (IV.A.8) to the numerator and (IV.A.7) to the denominator to show (IV.A.23).

To show (IV.A.24), we first introduce the measure $d\nu(z) = \mathrm{e}^{-\frac{1}{\varepsilon}U_m(\bar z)}/(\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz)\, d\bar z$. Then, the left-hand side of (IV.A.24) is equal to

$$\int_{\mathbb{R}} (z-z_m)^{2k}\, d\nu(z) \;+\; \sum_{\ell=0}^{2k-1} \binom{2k}{\ell} \left(\int_{\mathbb{R}} (z-z_m)\, d\nu(z)\right)^{2k-\ell} \int_{\mathbb{R}} (z-z_m)^{\ell}\, d\nu(z). \tag{IV.A.26}$$

Using (IV.A.22) and (IV.A.23), it is easy to see that for each $\ell = 0, \dots, 2k-1$,

$$\left(\int_{\mathbb{R}} (z-z_m)\, d\nu(z)\right)^{2k-\ell} \int_{\mathbb{R}} (z-z_m)^{\ell}\, d\nu(z) \;=\; O_{\mathcal{K}}\left(\varepsilon^{2k-\ell+\lceil\frac{\ell}{2}\rceil}\right) \;\le\; O_{\mathcal{K}}\left(\varepsilon^{k+1}\right). \tag{IV.A.27}$$

Combining (IV.A.26), (IV.A.27) and (IV.A.22) yields (IV.A.24).

It remains to show (IV.A.25). Similarly as in (IV.A.26), we have that the left-hand side of (IV.A.25) is equal to

$$\int_{\mathbb{R}} (z-z_m)^{2k+1}\, d\nu(z) \;+\; (2k+1)\int_{\mathbb{R}} (z-z_m)\, d\nu(z) \int_{\mathbb{R}} (z-z_m)^{2k}\, d\nu(z) \tag{IV.A.28}$$

$$+\; \sum_{\ell=0}^{2k-1} \binom{2k}{\ell} \left(\int_{\mathbb{R}} (z-z_m)\, d\nu(z)\right)^{2k+1-\ell} \int_{\mathbb{R}} (z-z_m)^{\ell}\, d\nu(z). \tag{IV.A.29}$$

As above, we observe that all the summands in (IV.A.29) are of lower order. Then, using (IV.A.22) and (IV.A.23) for the two terms in (IV.A.28), we infer (IV.A.25). □

IV.A.3 A priori estimates for the measure $\mu^{\varepsilon,\varphi'_\varepsilon(m)}$

For the proof of the local Cramér theorem and the equivalence of observables we need some estimates on certain moments and Fourier transforms of $\mu^{\varepsilon,\varphi'_\varepsilon(m)}$.

Lemma IV.A.4 *Recall the definition of ψ_J, φ^*_ε, φ_ε and $\mu^{\varepsilon,\sigma}$ given in* (IV.1.2), (IV.1.8), (IV.1.9) *and* (IV.1.10). *Notice that the inverse $(\psi'_J)^{-1}$ of ψ'_J exists.*

(i) Let $\tilde{K} \subset \mathbb{R}$ be compact. Then, for all $\lambda \in \tilde{K}$, $(\varphi^*_\varepsilon)'\left(\frac{1}{\varepsilon}\lambda\right) = (\psi'_J)^{-1}(\lambda) + \Omega_{\tilde{K}}(\varepsilon)$.

(ii) For all compact intervals $K \subset \mathbb{R}$ there exists a function $\tau_\varepsilon : K \to \mathbb{R}$ and $\varepsilon_K > 0$ such that $\sup_{0<\varepsilon<\varepsilon_K} \sup_{m\in K} |\tau_\varepsilon(m)| < \infty$ and $\varphi'_\varepsilon(m) = \frac{1}{\varepsilon}\tau_\varepsilon(m)$ for all $m \in K$ and $\varepsilon < \varepsilon_K$.

(iii) For $m \in \mathbb{R}$, let

$$s_\varepsilon(m) = (\varphi^*_\varepsilon)''(\varphi'_\varepsilon(m))^{\frac{1}{2}}. \tag{IV.A.30}$$

Note that s_ε is well-defined, since φ^*_ε is strictly convex (see Lemma IV.A.1). Then, for each compact interval $K \subset \mathbb{R}$, there exist $C_K > 0$ and $\varepsilon_K > 0$ such that for all $m \in K$ and for all $\varepsilon < \varepsilon_K$,

$$s_\varepsilon(m)^2 = \Omega_K(\varepsilon) \quad \text{and} \quad \sum_{k=1}^{4} \int_{\mathbb{R}} \left|\frac{z-m}{s_\varepsilon(m)}\right|^k \, d\mu^{\varepsilon,\varphi'_\varepsilon(m)}(z) \leq C_K. \tag{IV.A.31}$$

Proof. (i). Note that for all $\lambda \in \tilde{K}$, the function $U(\lambda, z) = \psi_J(z) - \lambda z$ satisfies the same conditions as the function U from Corollary IV.A.3. In particular, U_λ admits a unique global minimum at $(\psi'_J)^{-1}(\lambda)$. Thus, part *(i)* follows immediately from Lemma IV.A.1 and (IV.A.23).

(ii). Let $K = [a, b]$ for some $a, b \in \mathbb{R}$ with $a < b$. Set $F(m) = (\varphi^*_\varepsilon)'(\psi'_J(m)/\varepsilon)$. From part *(i)*, we know that for ε small enough,

$$\begin{aligned} F(a-1) &= a - 1 + \Omega_{[a-1,b+1]}(\varepsilon) < a, \quad \text{and} \\ F(b+1) &= b + 1 + \Omega_{[a-1,b+1]}(\varepsilon) > b. \end{aligned} \tag{IV.A.32}$$

Therefore, by the continuity of F and the mean value theorem, $F([a-1, b+1]) \supset K$. We also know that $F : [a-1, b+1] \to F([a-1, b+1])$ is bijective, since F is strictly increasing. Setting now $\tau_\varepsilon(m) = \psi'_J(F^{-1}(m))$ for $m \in K$ yields that

$$(\varphi^*_\varepsilon)'\left(\frac{1}{\varepsilon}\tau_\varepsilon(m)\right) = m \qquad \text{for all } m \in K. \tag{IV.A.33}$$

Since $\varphi'_\varepsilon = ((\varphi^*_\varepsilon)')^{-1}$ (cf. (IV.A.4)), this concludes the proof of part *(ii)*.

(iii). Let $U(m, z) = \psi_J(z) - \tau_\varepsilon(m)z$. Then, using part *(ii)*, Lemma IV.A.1 and (IV.A.24), we know that for $k = 2, 4$ and for all $m \in K$,

$$\begin{aligned} \int_{\mathbb{R}} |z-m|^k \, d\mu^{\varepsilon,\varphi'_\varepsilon(m)}(z) &= \frac{\int_{\mathbb{R}} \left(\bar{z} - \frac{\int_{\mathbb{R}} z\, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz}\right)^k \mathrm{e}^{-\frac{1}{\varepsilon}U_m(\bar{z})}\, d\bar{z}}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)}\, dz} \\ &= \varepsilon^{\frac{k}{2}} \frac{(k-1)!!}{\psi''_J((\psi'_J)^{-1}(\tau_\varepsilon(m)))^{\frac{k}{2}}} + O_K\left(\varepsilon^{\frac{k+1}{2}} \sqrt{\log(\varepsilon^{-1})^3}\right). \end{aligned} \tag{IV.A.34}$$

The dependence on ε of τ_ε is of no problem here due to its uniform boundedness stated in part *(ii)*. Then, for $k = 2$, the left-hand side of (IV.A.34) equals $s_\varepsilon(m)^2$ (cf. Lemma IV.A.1). Thus, (IV.A.34) proves the first claim in (IV.A.31), since the map $(m, \varepsilon) \mapsto \psi_J''((\psi_J')^{-1}(\tau_\varepsilon(m))$ is locally bounded. Moreover, due to Hölder's inequality, to show the second claim in (IV.A.31), it suffices to show that there exists $\varepsilon_K' > 0$ such that

$$\sup_{0<\varepsilon<\varepsilon_K'} \sup_{m\in K} \int_{\mathbb{R}} \left| \frac{z-m}{s(m)} \right|^4 \, d\mu^{\varepsilon,\varphi_\varepsilon'(m)}(z) \;<\; \infty. \tag{IV.A.35}$$

However, combining (IV.A.34) for $k = 4$ and the first claim in (IV.A.31), already implies (IV.A.35). This conclude the proof of part *(iii)*. □

Lemma IV.A.5 *Consider the same setting as in Lemma IV.A.4. Let $K \subset \mathbb{R}$ be compact, and abbreviate $\hat{z}(m) = (z-m)/s(m)$. Then, there exists $C_K, \varepsilon_K > 0$ such that for all $\hat{\xi} \in \mathbb{R}$,*

$$\sup_{0<\varepsilon<\varepsilon_K} \sup_{m\in K} \left| \int_{\mathbb{R}} \mathrm{e}^{i\hat{z}(m)\hat{\xi}} \, d\mu^{\varepsilon,\varphi_\varepsilon'(m)}(z) \right| \;\le\; \frac{C_K}{|\hat{\xi}|}. \tag{IV.A.36}$$

Proof. Fix $m \in K$. In this proof $C \in (0,\infty)$ denotes a constant, which is independent of ε and m, and may change every time it appears.

Let $U_m(z) = \psi_J(z) - \tau_\varepsilon(m)z$, where $\tau_\varepsilon(m)$ is introduced in Lemma IV.A.4. Then, by partial integration (as in [101, p. 37]) and by (IV.A.31),

$$\begin{aligned} \left| \int_{\mathbb{R}} \mathrm{e}^{i\hat{z}(m)\hat{\xi}} \, d\mu^{\varepsilon,\varphi_\varepsilon'(m)}(z) \right| \;&=\; \frac{s_\varepsilon(m)}{|\hat{\xi}|} \frac{1}{\varepsilon} \left| \frac{\int_{\mathbb{R}} \mathrm{e}^{i\hat{z}(m)\hat{\xi}} \, U_m'(z) \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)} \, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)} \, dz} \right| \\ &\le\; \frac{C}{|\hat{\xi}|\sqrt{\varepsilon}} \frac{\int_{\mathbb{R}} |U_m'(z)| \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)} \, dz}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)} \, dz}. \end{aligned} \tag{IV.A.37}$$

Let z_m be the unique global minimum of U_m, and let $\rho = C' \sqrt{\varepsilon \log(\varepsilon^{-1})}$ for some $C' > 0$ large enough. Then, using the same arguments as in the proof of Lemma IV.A.2, we see that the integral in the numerator on the right-hand side of (IV.A.37) is concentrated around $B_\rho(z_m)$, i.e.

$$\int_{\mathbb{R}} |U_m'(z)| \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z)} \, dz = \int_{-\rho}^{\rho} |U_m'(z_m+z)| \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m+z)} \, dz + O_K(\varepsilon^2 \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)}). \tag{IV.A.38}$$

Moreover, by Taylor's formula for some $\theta, \theta' \in [0,1]$ (cf. (IV.A.10)),

$$\begin{aligned} &\int_{-\rho}^{\rho} |U_m'(z_m+z)| \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m+z)} \, dz \\ &\quad=\; \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)} \int_{-\rho}^{\rho} |zU_m''(z_m+\theta z)| \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m''(z_m)\frac{1}{2}z^2 - \frac{1}{\varepsilon}U_m'''(z_m+\theta' z)\frac{1}{6}z^3} \, dz \\ &\quad\le\; C \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)} \int_{-\rho}^{\rho} |z| \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m''(z_m)\frac{1}{2}z^2} \, dz \;\le\; C \, \mathrm{e}^{-\frac{1}{\varepsilon}U_m(z_m)} \, \varepsilon. \end{aligned} \tag{IV.A.39}$$

Combining (IV.A.37), (IV.A.38) and (IV.A.39) and applying (IV.A.7) to the denominator in the right-hand side of (IV.A.37) yields (IV.A.36). This concludes the proof. □

IV.A.4 Proof of the local Cramér theorem

In this subsection we prove the local Cramér theorem (Proposition IV.1). The main ideas of the proof are the same as in [80, Proposition 31] or [101, Section 3]. The main difficulty here is to show that the estimates are uniform in $\varepsilon \ll 1$.

Proof of Proposition IV.1. Fix $m \in K$. In this proof $C \in (0, \infty)$ denotes a varying constant, which is independent of N, ε and m, but may depend on K.

Let $s_\varepsilon(m)$ be defined by (IV.A.30). In order to simplify the presentation here, for any function $f : \mathbb{R} \to \mathbb{R}$ and for all $z \in \mathbb{R}$, we abbreviate

$$\langle f \rangle \;=\; \int_{\mathbb{R}} f(z)\, d\mu^{\varepsilon,\varphi'_\varepsilon(m)}(z) \qquad \text{and} \qquad \hat{z} \;=\; \frac{z-m}{s_\varepsilon(m)}. \tag{IV.A.40}$$

Step 1. [New representation of $\mathrm{e}^{-N\varphi_{N,\varepsilon}(m)}$.]
Let $(X_i)_i$ be a sequence of random variables that are independent and identically distributed with common law $\mu^{\varepsilon,\varphi'_\varepsilon(m)}$. Let

$$\tilde{S}_{\varepsilon,m,N} \;=\; \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} (X_i - m), \tag{IV.A.41}$$

and let $\tilde{g}_{\varepsilon,m,N}$ denote the Lebesgue density of the distribution of $\tilde{S}_{\varepsilon,m,N}$. As in [101, (31)], using the co-area formula, we have that

$$\tilde{g}_{\varepsilon,m,N}(0) \;=\; \mathrm{e}^{N\varphi_\varepsilon(m) - N\varphi_{N,\varepsilon}(m)}. \tag{IV.A.42}$$

Moreover, let $g_{\varepsilon,m,N}$ be the Lebesgue density of the distribution of

$$S_{\varepsilon,m,N} = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \frac{X_i - m}{s_\varepsilon(m)}. \tag{IV.A.43}$$

Then, by Lemma IV.A.1,

$$g_{\varepsilon,m,N}(0) \;=\; \tilde{g}_{\varepsilon,m,N}(0)\, s_\varepsilon(m) \;=\; \tilde{g}_{\varepsilon,m,N}(0)\, \varphi''_\varepsilon(m)^{-\frac{1}{2}}. \tag{IV.A.44}$$

Therefore, it suffices to show that for ε small enough,

$$\left| g_{\varepsilon,m,N}(0) - \frac{1}{\sqrt{2\pi}} \right| \;=\; O_K\left(\frac{1}{\sqrt{N}}\right). \tag{IV.A.45}$$

We show (IV.A.45) by mimicking the arguments of the proof of [101, 3.1]. Therefore, as in [101, (44)], we apply the inverse Fourier transform to obtain that

$$2\pi\, g_{\varepsilon,m,N}(0) \;=\; \int_{\mathbb{R}} \left\langle \mathrm{e}^{i\frac{1}{\sqrt{N}}\hat{z}\hat{\xi}} \right\rangle^N d\hat{\xi}, \tag{IV.A.46}$$

and we split this integral according to some $\delta > 0$ (which is chosen in Step 2) as

$$\begin{aligned} \int_{\mathbb{R}} \left\langle \mathrm{e}^{i\frac{1}{\sqrt{N}}\hat{z}\hat{\xi}} \right\rangle^N d\hat{\xi} \;&=\; \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}| \le \delta\}} \left\langle \mathrm{e}^{i\frac{1}{\sqrt{N}}\hat{z}\hat{\xi}} \right\rangle^N d\hat{\xi} \;+\; \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}| > \delta\}} \left\langle \mathrm{e}^{i\frac{1}{\sqrt{N}}\hat{z}\hat{\xi}} \right\rangle^N d\hat{\xi} \\ &=:\; I \;+\; II. \end{aligned} \tag{IV.A.47}$$

In the following we compute the asymptotic value of I, and show that II is of lower order than I.

Step 2. [Estimation of the term I.]
From Lemma IV.A.4 we know that there exists $\varepsilon_K > 0$ such that

$$\sup_{0<\varepsilon<\varepsilon_K} \sup_{m\in K} \sum_{k=1}^{3} \langle |\hat{z}|^k \rangle \;\le\; C. \tag{IV.A.48}$$

Then, as in [101, (46)], applying Taylor's formula to the functions $\hat{\xi} \mapsto h(\hat{\xi})$ and $\hat{\xi} \mapsto \left\langle \mathrm{e}^{i\hat{z}\hat{\xi}} \right\rangle$ shows that there exist $\hat{\delta}$, $c_K > 0$, and a complex-valued function h such that for all $|\hat{\xi}| \le \hat{\delta}$ and all $\varepsilon < \varepsilon_K$,

$$\left\langle \mathrm{e}^{i\hat{z}\hat{\xi}} \right\rangle \;=\; \mathrm{e}^{-h(\hat{\xi})} \quad \text{and} \quad \left| h(\hat{\xi}) - \frac{1}{2}\hat{\xi}^2 \right| \;\le\; c_K |\hat{\xi}|^3. \tag{IV.A.49}$$

As a consequence, by choosing $\delta < \hat{\delta}$, we have that

$$I \;=\; \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}| \le \delta\}} \mathrm{e}^{-Nh\left(\frac{\hat{\xi}}{\sqrt{N}}\right)} d\hat{\xi}. \tag{IV.A.50}$$

Moreover, by arguing similarly as in [101, (69)], (IV.A.49) yields that for $\delta < \hat{\delta}$ small enough,

$$\mathrm{Re}\left(Nh\left(\frac{\hat{\xi}}{\sqrt{N}}\right)\right) \;\ge\; \frac{|\hat{\xi}|^2}{2} - c_K\, \delta\, |\hat{\xi}|^2 \;\ge\; \frac{|\hat{\xi}|^2}{4}. \tag{IV.A.51}$$

This in turn implies that, by proceeding as in [101, p. 32],

$$\left| \mathrm{e}^{-Nh\left(\frac{\hat{\xi}}{\sqrt{N}}\right)} - \mathrm{e}^{-\frac{1}{2}\hat{\xi}^2} \right| \;\le\; \mathrm{e}^{-\frac{1}{4}\hat{\xi}^2} c_K\, \frac{|\hat{\xi}|^3}{\sqrt{N}}, \tag{IV.A.52}$$

which yields, as in [101, p. 32], to the estimate

$$\left| I - \sqrt{2\pi} \right| \;\le\; \frac{C}{\sqrt{N}}. \tag{IV.A.53}$$

Step 3. [Estimation of the term II.]
It remains to show that the term II is negligible. Recall from Lemma IV.A.4 and Lemma IV.A.5 that there exist $\varepsilon'_K, c'_K > 0$ such that for all $\hat{\xi} \in \mathbb{R}$,

$$\sup_{0<\varepsilon<\varepsilon'_K} \sup_{m\in K} \langle |\hat{z}| \rangle \;\le\; c'_K \quad \text{and} \quad \sup_{0<\varepsilon<\varepsilon'_K} \sup_{m\in K} \left| \left\langle \mathrm{e}^{i\hat{z}\hat{\xi}} \right\rangle \right| \;\le\; \frac{c'_K}{|\hat{\xi}|}. \tag{IV.A.54}$$

Then, following the proof of [101, 3.4], the estimates in (IV.A.54) (which are the analogues of [101, (52)] and [101, (53)]) imply that for all $\delta < \hat{\delta}$ there exists $\lambda_{K,\delta} < 1$ (which depends only on c'_K and δ) such that

$$\sup_{0<\varepsilon<\varepsilon'_K} \sup_{m\in K} \left| \left\langle \mathrm{e}^{i\hat{z}\hat{\xi}} \right\rangle \right| \;\le\; \lambda_{K,\delta} \qquad \text{for all } |\hat{\xi}| \;\ge\; \delta. \tag{IV.A.55}$$

Finally, applying the same arguments as in [101, p. 32] shows that

$$|II| \;\leq\; C\,N\lambda_{K,\delta}^{N-2}. \tag{IV.A.56}$$

Hence, $|II| \;\leq\; C/\sqrt{N}$ for N large enough. This concludes the proof. □

As a simple consequence of the ideas from the proof of Proposition IV.1, we can state the result in a more precise way in the trivial case that the (effective) single-site potential is a quadratic function. The result is given in the following lemma.

Lemma IV.A.6 *Let $V(z) = \frac{\alpha}{2}z^2$ for some $\alpha > 0$. Let χ_ε^*, χ_ε, $\mu^{\varepsilon,\chi_\varepsilon'(m)}$ be defined by* (IV.A.1), (IV.A.2) *and* (IV.A.3)*, respectively, with W replaced by $\frac{1}{\varepsilon}V$. Let $\chi_{N,\varepsilon} : \mathbb{R} \mapsto \mathbb{R}$ be defined by*

$$\varphi_{N,\varepsilon}(m) \;=\; -\frac{1}{N}\log\int_{P^{-1}(m)} \mathrm{e}^{-\frac{1}{\varepsilon}\sum_{i=0}^{N-1}V(x_i)}\,d\mathcal{H}^{N-1}(x). \tag{IV.A.57}$$

Then, for all $m \in \mathbb{R}$,

$$\mathrm{e}^{-N\chi_{N,\varepsilon}(m)} \;=\; \mathrm{e}^{-N\chi_\varepsilon(m)}\,\frac{\sqrt{\varphi_\varepsilon''(m)}}{\sqrt{2\pi}}. \tag{IV.A.58}$$

Proof. Using the same notation and the same arguments as in Step 1 of the proof of Proposition IV.1, we see that it suffices to show that

$$g_{\varepsilon,m,N}(0) \;=\; \frac{1}{\sqrt{2\pi}}. \tag{IV.A.59}$$

Note that by a simple computation, for all $\sigma, m \in \mathbb{R}$,

$$\chi_\varepsilon(m) \;=\; \frac{\alpha}{2\varepsilon}m^2 \;-\; \frac{1}{2}\log\left(2\pi\frac{\varepsilon}{\alpha}\right) \quad \text{and} \quad \mu^{\varepsilon,\chi_\varepsilon'(m)}(z) \;=\; \mathrm{e}^{-\frac{\alpha}{2\varepsilon}(z-m)^2}\,\frac{1}{\sqrt{2\pi\frac{\varepsilon}{\alpha}}}\,dz. \tag{IV.A.60}$$

In particular, $\mu^{\varepsilon,\chi_\varepsilon'(m)}(z)$ is a Gaussian measure. Therefore, the claim (IV.A.59) is a simple consequence of the stability of Gaussian measures under convolution. □

IV.A.5 Proof of the equivalence of observables

In this subsection we prove the equivalence of observables, which is stated in Proposition IV.11. The proof is similar to the proof of Proposition IV.1 and combines the ideas from [91] and [101].

Proof of Proposition IV.11. For simplicity, we only consider the case $\ell = 1$. A straightforward modification of the following proof yields the claim also in the case $\ell \in \mathbb{N}$.

Fix $m \in K = [-2, 2]$. In this proof, let C denote a varying positive constant, which does not depend on N, ε and m, but may depend on b and K.

Step 1. [Cramér's representation.]
Proceeding as in [91], we use the so-called *Cramér representation* in order to rewrite the left-hand side of (IV.1.63) in terms of the density of a certain random variable.

Let $\mu^{\varepsilon,\varphi_\varepsilon'(m),N} \;=\; \otimes_{i=1}^N \mu^{\varepsilon,\varphi_\varepsilon'(m)}$, and let, for $\sigma \in \mathbb{R}$, the measure $\mu^{\sigma,\varepsilon,\varphi_\varepsilon'(m),N} \in \mathcal{M}_1(\mathbb{R}^N)$ be defined by

$$\mu^{\sigma,\varepsilon,\varphi_\varepsilon'(m),N}(dx) \;=\; \frac{1}{Z}\exp\left(\varphi_\varepsilon'(m)\sum_{i=0}^{N-1}x_i + \sigma b(x_0) - \sum_{i=0}^{N-1}\psi_J(x_i)\right)dx, \tag{IV.A.61}$$

where Z denotes the normalization constant. Note that $\mu^{0,\varepsilon,\varphi'_\varepsilon(m),N} = \mu^{\varepsilon,\varphi'_\varepsilon(m),N}$. Let $(Y_i)_{i=1,\dots,N}$ be a random vector distributed according to $\mu^{\sigma,\varepsilon,\varphi'_\varepsilon(m),N}$, and let

$$S_{\sigma,\varepsilon,m,N} = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} (Y_i - m). \tag{IV.A.62}$$

Let $\tilde{g}_{\sigma,\varepsilon,m,N}$ denote the Lebesgue density of the distribution of $S_{\sigma,\varepsilon,m,N}$. Note that $\tilde{g}_{0,\varepsilon,m,N} = \tilde{g}_{\varepsilon,m,N}$, where $\tilde{g}_{\varepsilon,m,N}$ is defined in Step 1 of the proof of Proposition IV.1. Using the same arguments as in [91, Lemma 5 and Lemma 6], we observe that

$$\left| \int_{\mathbb{R}^N} b(x_0)\ d\mu^{\varepsilon,\varphi'_\varepsilon(m),N} - \int_{P^{-1}(m)} b(x_0)\ d\mu_m \right| = \left| \frac{\frac{d}{d\sigma}\big|_{\sigma=0} \tilde{g}_{\sigma,\varepsilon,m,N}(0)}{\tilde{g}_{0,\varepsilon,m,N}(0)} \right|. \tag{IV.A.63}$$

Hence, in order to show (IV.1.63), It suffices to show that there exist $\varepsilon_{b,K} > 0$ and $N_{b,K} \in \mathbb{N}$ such that for all $N \geq N_{b,K}$, $\varepsilon < \varepsilon_{b,K}$ and $m \in K$,

$$\left| \frac{d}{d\sigma}\Big|_{\sigma=0} \tilde{g}_{\sigma,\varepsilon,m,N}(0) \right| \ \leq\ \frac{C}{s_\varepsilon(m)\sqrt{N}} \tag{IV.A.64}$$

$$|\tilde{g}_{0,\varepsilon,m,N}(0)| \ \geq\ \frac{1}{s_\varepsilon(m)\sqrt{2\pi}} \left(1 + O_K\left(\frac{1}{\sqrt{N}}\right)\right), \tag{IV.A.65}$$

where $s_\varepsilon(m)$ is defined in (IV.A.30).

Step 2. [Proof of (IV.A.65).]
Using the same arguments as in Step 1 of the proof of Proposition IV.1, we observe that

$$\tilde{g}_{0,\varepsilon,m,N}(0) \ =\ \mathrm{e}^{N\varphi_\varepsilon(m) - N\varphi_{N,\varepsilon}(m)}. \tag{IV.A.66}$$

Then, Proposition IV.1 yields (IV.A.65).

Step 3. [Proof of (IV.A.64).]
Recall the abbreviations from (IV.A.40). Let $(X_i)_{i=1,\dots,N}$ be a random vector distributed according to $\mu^{\varepsilon,\varphi'_\varepsilon(m),N}$, and let X be a random variable distributed according to $\mu^{\varepsilon,\varphi'_\varepsilon(m)}$. By [91, Lemma 7], we have that

$$\begin{aligned} 2\pi \frac{d}{d\sigma}\Big|_{\sigma=0} \tilde{g}_{\sigma,\varepsilon,m,N}(0) &= \int_{\mathbb{R}} \mathbb{E}_{\mu^{\varepsilon,\varphi'_\varepsilon(m),N}} \left[(b(X_0) - \langle b \rangle) \mathrm{e}^{i\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} (X_i - m)\xi} \right] d\xi \\ &= \int_{\mathbb{R}} \mathbb{E}_{\mu^{\varepsilon,\varphi'_\varepsilon(m)}} \left[(b(X) - \langle b \rangle) \mathrm{e}^{i\frac{1}{\sqrt{N}}(X-m)\xi} \right] \mathbb{E}_{\mu^{\varepsilon,\varphi'_\varepsilon(m)}} \left[\mathrm{e}^{i\frac{1}{\sqrt{N}}(X-m)\xi} \right]^{N-1} d\xi \\ &= s_\varepsilon(m)^{-1} \int_{\mathbb{R}} \left\langle (b - \langle b \rangle) \mathrm{e}^{i\frac{1}{\sqrt{N}}\hat{z}\hat{\xi}} \right\rangle \left\langle \mathrm{e}^{i\frac{1}{\sqrt{N}}\hat{z}\hat{\xi}} \right\rangle^{N-1} d\hat{\xi}. \end{aligned} \tag{IV.A.67}$$

It remains to show that for N large enough,

$$\left| \int_{\mathbb{R}} \left\langle (b - \langle b \rangle) \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \left\langle \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle^{N-1} d\hat{\xi} \right| \ \leq\ \frac{C}{\sqrt{N}}. \tag{IV.A.68}$$

In order to show (IV.A.68), we proceed as in the proof of [101, 3.1] and Proposition IV.1. Let $\hat{\delta} > 0$ and h be given as in Step 2 of the proof of Proposition IV.1. We split the integral

on the left-hand side in (IV.A.68) according to some $\delta < \hat{\delta}$ (which is chosen in Step 3.1) as

$$\int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|\le\delta\}} \left\langle (b-\langle b\rangle) \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \mathrm{e}^{-(N-1)h\left(\frac{\hat{\xi}}{\sqrt{N}}\right)} d\hat{\xi} + \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|>\delta\}} \left\langle (b-\langle b\rangle) \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \left\langle \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle^{N-1} d\hat{\xi}$$
$$=: I + II. \tag{IV.A.69}$$

We now show that $|I| + |II| \le C/\sqrt{N}$.

Step 3.1. [Estimation of the term I.]
This step is very similar to Step 2 of the proof of Proposition IV.1. Using (IV.A.49), we have that there exists $c_K > 0$ such that

$$\left| (N-1)h\left(\frac{\hat{\xi}}{\sqrt{N}}\right) - \frac{1}{2}\hat{\xi}^2 \right| \;\le\; c_K \frac{|\hat{\xi}|^3}{\sqrt{N}} + \frac{|\hat{\xi}|^2}{2N}. \tag{IV.A.70}$$

Similarly as in (IV.A.51), this inequality yields that for $N \ge 4$ and for δ small enough,

$$\mathrm{Re}\left((N-1)h\left(\frac{\hat{\xi}}{\sqrt{N}}\right)\right) \;\ge\; \frac{|\hat{\xi}|^2}{2} - \left(c_K\,\delta + \frac{1}{8}\right)|\hat{\xi}|^2 \;\ge\; \frac{|\hat{\xi}|^2}{4}, \tag{IV.A.71}$$

and hence, as in (IV.A.52),

$$\left| \mathrm{e}^{-(N-1)h\left(\frac{\hat{\xi}}{\sqrt{N}}\right)} - \mathrm{e}^{-\frac{1}{2}\hat{\xi}^2} \right| \;\le\; \mathrm{e}^{-\frac{1}{4}\hat{\xi}^2} \left| c_K \frac{|\hat{\xi}|^3}{\sqrt{N}} + \frac{|\hat{\xi}|^2}{N} \right|. \tag{IV.A.72}$$

This implies that

$$\begin{aligned} \left| I - \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|\le\delta\}} \left\langle (b-\langle b\rangle) \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \mathrm{e}^{-\frac{\hat{\xi}^2}{2}} d\hat{\xi} \right| \;&\le\; C \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|\le\delta\}} \left| \frac{|\hat{\xi}|^3}{\sqrt{N}} + \frac{|\hat{\xi}|^2}{N} \right| \mathrm{e}^{-\frac{\hat{\xi}^2}{4}} d\hat{\xi} \\ &\le\; \frac{C}{\sqrt{N}}, \end{aligned} \tag{IV.A.73}$$

since, in view of (IV.1.62),

$$\left| \left\langle (b-\langle b\rangle) \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \right| \;\le\; \sup_{m\in K} \langle |b-\langle b\rangle| \rangle \;<\; \infty. \tag{IV.A.74}$$

Moreover, by Taylor's formula, for some $\theta \in [0,1]$,

$$\begin{aligned} &\int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|\le\delta\}} \left\langle (b-\langle b\rangle) \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \mathrm{e}^{-\frac{|\hat{\xi}|^2}{2}} d\hat{\xi} \\ &\quad\le\; 0 + \left| \frac{1}{\sqrt{N}} \int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|\le\delta\}} \left\langle \hat{z}(b-\langle b\rangle) \mathrm{e}^{i\theta\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle i\hat{\xi} \mathrm{e}^{-\frac{|\hat{\xi}|^2}{2}} d\hat{\xi} \right| \\ &\quad\le\; \frac{1}{\sqrt{N}} \left| \left\langle \hat{z}(b-\langle b\rangle) \mathrm{e}^{i\theta\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle \right| \int_{\mathbb{R}} |\hat{\xi}|\, \mathrm{e}^{-\frac{|\hat{\xi}|^2}{2}} d\hat{\xi} \;\le\; \langle |\hat{z}|^2 \rangle \langle |b|^2 \rangle \frac{C}{\sqrt{N}}. \end{aligned} \tag{IV.A.75}$$

Using (IV.1.62) and that $\sum_{k=1}^3 \langle |\hat{z}|^k \rangle \le C$ implies that $|I| \;\le\; C/\sqrt{N}$.

Step 3.2. [Estimation of the term II.]
First note that by using (IV.A.74) it only remains to show that

$$\int_{\{|\frac{\hat{\xi}}{\sqrt{N}}|\geq\delta\}} \left\langle \mathrm{e}^{i\frac{\hat{\xi}}{\sqrt{N}}\hat{z}} \right\rangle^{N-1} d\hat{\xi} \;\leq\; \frac{C}{\sqrt{N}}. \tag{IV.A.76}$$

Then, a straightforward adaptation of the arguments in Step 3 of the proof of Proposition IV.1 yields the claim. □

Chapter V

On the basin of attraction of McKean-Vlasov paths

The results of the present chapter are contained in the preprint [12].

Recall Section I.7, where we provide a motivation and a first formulation of the main results of this chapter. This chapter is organized as follows. First, we state the main assumption of this chapter, and introduce some notation. Then, in Section V.1, we provide three ingredients that we need for the proof of Theorem V.8. Namely, the relation between the functional $\mathcal{F}$ from Chapter III and the functional $\bar{H}_1$ from Chapter IV (see Lemma V.2), a symmetry property of the gradient flows for $\mathcal{F}$ (Lemma V.3), and a useful characterization of the corresponding stationary measures (Lemma V.4). In Section V.2 we first show some compactness property of the gradient flows for $\mathcal{F}$, and then use this property to prove Proposition I.25. In Section V.3 we prove the main part of Proposition I.26. Then, we state and prove the main result of this chapter in Section V.4. We conclude this chapter with some immediate consequences for the basin of μ^0.

The results of this chapter are subject to the following assumption.

Assumption V.1 *Suppose Assumption IV.13 and suppose that Assumption III.33 (ii) is true with $\ell \geq 2$.*

As an immediate consequence of this assumption, we observe that the *McKean-Vlasov functional* from Chapter III, $\mathcal{F} : \mathcal{P}_2(\mathbb{R}) \to (-\infty, \infty]$, which is given in the present setting by

$$\mathcal{F}(\mu) \;=\; \begin{cases} \int_{\mathbb{R}} \log(\rho) d\mu + \int_{\mathbb{R}} \Psi \, d\mu - \frac{J}{2} \left(\int_{\mathbb{R}} z d\mu(z) \right)^2 & \text{if } \mu \in \mathcal{P}_2(\mathbb{R}) \text{ has a Lebesgue density } \rho, \\ \infty & \text{else,} \end{cases} \tag{V.0.1}$$

is strongly λ-convex in the sense of Definition III.19 for some $\lambda < 0$, and that there exists $c > 0$ such that

$$\mathcal{F}(\mu) \;\geq\; c \left(\int_{\mathbb{R}} |x|^4 \, d\mu(x) - 1 \right) \qquad \text{for all } \mu \in \mathcal{P}_2(\mathbb{R}). \tag{V.0.2}$$

See Lemma III.34 and Theorem III.35 for more details on these facts.

Notation

- We use the same notation here as in Chapter IV. In particular, recall the definition of the objects φ_1^*, φ_1, $\bar{H}_1$ and $\mu^{1,\sigma}$.
- For all $\mu \in \mathcal{P}_2(\mathbb{R})$ and $\delta > 0$, let $B_\delta(\mu) = \{\nu \in \mathcal{P}_2(\mathbb{R}) \,|\, W_2(\mu,\nu) < \delta\}$.
- We denote by $\mathrm{m}[\mu] = \int_{\mathbb{R}} z d\mu(z)$ the mean of a probability measure $\mu \in \mathcal{P}_2(\mathbb{R})$.
- The *stationary measures* μ^-, μ^0 and μ^+ are defined by

$$\mu^- := \mu^{1,\varphi_1'(-m_1^\star)}, \qquad \mu^0 := \mu^{1,\varphi_1'(0)} \quad \text{and} \quad \mu^+ := \mu^{1,\varphi_1'(m_1^\star)}. \tag{V.0.3}$$

- We know from [3, 11.2.8] that, for all $\mu \in \overline{D(\mathcal{F})} = \mathcal{P}_2(\mathbb{R})$, there exists a unique Wasserstein gradient flow for $\mathcal{F}$ (see also Section I.2 and Theorem III.35), and we denote it by $(S[\mu](t))_{t\in(0,\infty)}$. This curve is often called *McKean-Vlasov path* in the literature.

V.1 Preliminaries

We have the following relation[1] between the free energy functionals $\mathcal{F}$ and $\bar{H}$.

Lemma V.2 *Suppose Assumption V.1. Then, for all $m \in \mathbb{R}$, we have that*

$$\mathcal{F}(\mu) \;>\; \mathcal{F}(\mu^{1,\varphi_1'(m)}) \quad \textit{for all } \mu \in \mathcal{P}_2(\mathbb{R}) \textit{ such that } \mathrm{m}[\mu] = m \textit{ and } \mu \neq \mu^{1,\varphi_1'(m)}. \tag{V.1.1}$$

Moreover,

$$\bar{H}_1(m) \;=\; \min_{\mu\in\mathcal{P}_2(\mathbb{R}),\, \mathrm{m}[\mu]=m} \mathcal{F}(\mu) \;=\; \mathcal{F}(\mu^{1,\varphi_1'(m)}). \tag{V.1.2}$$

In particular, $\mathcal{F}$ admits exactly two global minima, one at $\mu^- = \mu^{1,\varphi_1'(-m^\star)}$ and one at $\mu^+ = \mu^{1,\varphi_1'(m^\star)}$, and we have that $\mathcal{F}(\mu^-) = \mathcal{F}(\mu^+) < \mathcal{F}(\mu^0)$.

Proof. If $\mathcal{F}(\mu) = \infty$, then (V.1.1) is trivially satisfied. So assume that $\mathcal{F}(\mu) < \infty$. Recall the definition of the relative entropy in (III.0.3). Then, by denoting the Lebesgue density of μ by ρ,

$$\begin{aligned} \mathcal{F}(\mu) \;&=\; \int_{\mathbb{R}} \log(\rho\, \mathrm{e}^{\psi}) d\mu - \frac{J}{2}m^2 \;=\; \mathcal{H}(\mu \,|\, \mu^{1,\varphi_1'(m)}) + \varphi_1'(m)m - \varphi_1^*(\varphi_1'(m)) - \frac{J}{2}m^2 \\ &>\; \bar{H}_1(m) \;=\; \mathcal{F}(\mu^{1,\varphi_1'(m)}), \end{aligned} \tag{V.1.3}$$

since $\mathcal{H}(\mu \,|\, \mu^{1,\varphi_1'(m)}) > 0$ if $\mu \neq \mu^{1,\varphi_1'(m)}$, and $\varphi_1'(m)m - \varphi_1^*(\varphi_1'(m)) = \varphi_1(m)$ (see Lemma IV.A.1). This shows (V.1.1). Finally, a simple computation shows that $\bar{H}_1(m) = \mathcal{F}(\mu^{1,\varphi_1'(m)})$. This concludes the proof. □

The following lemma shows that gradient flows for $\mathcal{F}$ admit a useful symmetry property.

Lemma V.3 *Let $\varsigma : \mathbb{R} \to \mathbb{R}$ be defined by $\varsigma(z) = -z$, and let $\mu \in \mathcal{P}_2(\mathbb{R})$. Then,*

$$S[\varsigma_\# \mu](t) = \varsigma_\# S[\mu](t) \quad \textit{for all } t \in (0,T). \tag{V.1.4}$$

[1] See also [106, Section IV.2] for a more general result.

Proof. First note that

$$\mathcal{F}(\nu) = \mathcal{F}(\varsigma_{\#}\nu) \quad \text{for all } \nu \in \mathcal{P}_2(\mathbb{R}), \tag{V.1.5}$$

and therefore,

$$|\partial\mathcal{F}|(\nu) = |\partial\mathcal{F}|(\varsigma_{\#}\nu) \quad \text{for all } \nu \in \mathcal{P}_2(\mathbb{R}). \tag{V.1.6}$$

Moreover, for all $\nu \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R}))$ and $0 < s < t < T$,

$$W_2(\nu_s,\nu_t) \;=\; W_2(\varsigma_{\#}(\varsigma_{\#}\nu_s),\varsigma_{\#}(\varsigma_{\#}\nu_t)) \;\le\; W_2(\varsigma_{\#}\nu_s,\varsigma_{\#}\nu_t) \;\le\; W_2(\nu_s,\nu_t). \tag{V.1.7}$$

Therefore, $W_2(\nu_s,\nu_t) = W_2(\varsigma_{\#}\nu_s,\varsigma_{\#}\nu_t)$, and we have that the metric derivatives coincide, i.e.,

$$|\nu'|(t) = |(\varsigma_{\#}\nu)'|(t) \quad \text{for almost every } t \in (0,T) \text{ and for all } \nu \in \mathcal{AC}((0,T);\mathcal{P}_2(\mathbb{R})). \tag{V.1.8}$$

Suppose first that $\mu \in D(\mathcal{F})$. Then, using the characterization of $(S[\mu](t))_t$ as a curve of maximal slope (see Lemma I.8) and (V.1.5), (V.1.6) and (V.1.8), we have that

$$\begin{aligned} 0 \;&=\; \mathcal{F}(S[\mu](T)) - \mathcal{F}(\mu) + \frac{1}{2}\int_0^T \big(|\partial\mathcal{F}|^2(S[\mu](t)) + |(S[\mu])'|^2(t)\big)\,dt \\ &=\; \mathcal{F}(\varsigma_{\#}S[\mu](T)) - \mathcal{F}(\varsigma_{\#}\mu) + \frac{1}{2}\int_0^T \big(|\partial\mathcal{F}|^2(\varsigma_{\#}S[\mu](t)) + |(\varsigma_{\#}S[\mu])'|^2(t)\big)\,dt \end{aligned} \tag{V.1.9}$$

for all $T \in (0,\infty)$. Hence, $(\varsigma_{\#}S[\mu](t))_t$ is the unique gradient flow for $\mathcal{F}$ with initial value $\varsigma_{\#}\mu$. This shows (V.1.4) for all $\mu \in D(\mathcal{F})$. Combined with the regularization estimate ((III.1.79) or [3, 4.3.2]), this also yields (V.1.4) for all $\mu \in \mathcal{P}_2(\mathbb{R}) \setminus D(\mathcal{F})$. □

We next characterize the *stationary points* of the McKean-Vlasov evolution[2], where we say that $\mu \in \mathcal{P}_2(\mathbb{R})$ is *stationary* if

$$S[\mu](t) = \mu \quad \text{for all } t \in (0,T), \tag{V.1.10}$$

or equivalently,

$$|(S[\mu])'|(t) \;=\; 0 \quad \text{for almost every } t \in (0,\infty). \tag{V.1.11}$$

Lemma V.4 *Suppose Assumption V.1. Let $\mu \in \mathcal{P}_2(\mathbb{R})$. Then, the following statements are equivalent.*

(i) *μ is stationary.*

(ii) $|\partial\mathcal{F}|(\mu) = 0$.

(iii) $\mu \in \{\mu^-, \mu^0, \mu^+\}$.

Proof. *(i)* ⇒ *(ii)*. Suppose that μ is stationary. Recall from [3, 2.4.15] that $|(S[\mu])'|(t) = |\partial\mathcal{F}|(S[\mu](t))$ for almost every $t \in (0,\infty)$. Then, (V.1.11) implies part *(ii)*.

[2] See also [82] for similar results.

(ii) ⇒ *(iii)*. Using Proposition III.38 (or [3, 10.4.13]), we have that the Lebesgue density, ρ, of μ belongs to the Sobolev space $W^{1,1}_{loc}(\mathbb{R})$. Suppose that ρ is continuous[3], and let $m = \mathrm{m}[\mu]$. Then, by using again Proposition III.38,

$$|\partial\mathcal{F}|(\mu) \;=\; \int_{\mathbb{R}} \left| \frac{\partial_z \rho(z)}{\rho(z)} + \Psi'(z) - Jm \right|^2 d\mu(z) \;=\; \int_{\mathbb{R}} \left| \frac{\partial_z \left(\rho(z) \mathrm{e}^{\Psi(z)-Jmz}\right)}{\rho(z)\mathrm{e}^{\Psi(z)-Jmz}} \right|^2 d\mu(z). \quad \text{(V.1.12)}$$

Since $|\partial\mathcal{F}|(\mu) = 0$, (V.1.12) implies that for $\mu-$a.e. $z \in \mathbb{R}$,

$$\rho(z) \;=\; \rho(0)\, \mathrm{e}^{\Psi(0)}\, \mathrm{e}^{-\Psi(z)+Jmz} \;=\; \mathrm{e}^{-\Psi(z)+Jmz-\varphi_1^*(Jm)}, \quad \text{(V.1.13)}$$

where we used in the last step the definition of φ_1^* and that μ is a probability measure. In particular, (V.1.13) yields that $m = (\varphi_1^*)'(Jm)$, or equivalently, by Lemma IV.A.1, that $\bar{H}_1'(m) = 0$. However, in Lemma IV.16 we have seen that there are only three solutions to this equation. This implies that

$$m \;\in\; \{-m^\star, 0, m^\star\}. \quad \text{(V.1.14)}$$

Combining (V.1.13) and (V.1.14) yields part *(iii)*.

(iii) ⇒ *(ii)*. Combining the representation (V.1.12) with the definition of the measures μ^-, μ^0 and μ^+ yields part *(ii)*.

(ii) ⇒ *(i)*. From [3, 2.4.15], we have that for all $t > 0$,

$$|\partial\mathcal{F}|(S[\mu](t)) \;\le\; \mathrm{e}^{-\lambda t}|\partial\mathcal{F}|(\mu) \;=\; 0. \quad \text{(V.1.15)}$$

Again, using that $|(S[\mu])'|(t) = |\partial\mathcal{F}|(S[\mu](t))$ for almost every $t \in (0,\infty)$, (V.1.15) yields part *(i)*. □

V.2 Convergence in the valleys

In this section we first show some compactness property of the McKean-Vlasov paths in Lemma V.5. Then we use this result to show in Proposition V.6 that inside the valleys of the set $\{\mu \in \mathcal{P}_2(\mathbb{R}) \,|\, \mathcal{F}(\mu) \le \mathcal{F}(\mu^0)\}$ the convergence of $(S[\mu](t))_{t\in[0,\infty)}$ is determined by the sign of $\mathrm{m}[\mu]$.

Lemma V.5 *Suppose Assumption V.1. Let $\mu \in D(\mathcal{F})$. Then, there exist a sequence $(t_k)_k$ and $\mu^* \in \{\mu^-, \mu^0, \mu^+\}$ such that $\lim_{k\to\infty} t_k = \infty$,*

$$\lim_{k\to\infty} W_2(S[\mu](t_k), \mu^*) \;=\; 0 \quad \textit{and} \quad \lim_{t\to\infty} \mathcal{F}(S[\mu](t)) \;=\; \mathcal{F}(\mu^*). \quad \text{(V.2.1)}$$

Proof. In the following let $\mu_t = S[\mu](t)$. We prove this lemma in three steps.

Step 1. [There exists a subsequence $(t_n)_n$ such that $\lim_{n\to\infty} |\partial\mathcal{F}|(\mu_{t_n}) = 0$.]

Note that the sequence $(\mathcal{F}(\mu_t))_{t\in[0,\infty)}$ is a continuous, monotone and bounded sequence of real numbers by (V.0.2) and Lemma III.24 (or [3, 2.4.15]). Therefore, it converges, as $t \to \infty$, to a number $L^* \in \mathbb{R}$. In particular, by Lemma III.24,

$$\int_0^\infty |\partial\mathcal{F}|(\mu_r)\, dr \;=\; -\int_0^\infty \frac{d}{dr}\mathcal{F}(\mu_r)\, dr \;=\; -L^* + \mathcal{F}(\mu) \;<\; \infty. \quad \text{(V.2.2)}$$

[3] Recall that there exists a continuous representative for each element in $W^{1,1}_{loc}(\mathbb{R})$.

This implies the claim of Step 1.

Step 2. [$\lim_{k\to\infty} W_2(\mu_{t_{n_k}}, \mu^*)$ for some $\mu^* \in \{\mu^-, \mu^0, \mu^+\}$ and a subsubsequence $(t_{n_k})_k$.] By (V.0.2), the monotonicity of $t \mapsto \mathcal{F}(\mu_t)$ and the fact that $\mu_0 = \mu \in D(\mathcal{F})$, we have that

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}} |x|^4 \, d\mu_{t_n}(x) \;\leq\; \sup_{n\in\mathbb{N}} \left(\frac{1}{c}\mathcal{F}(\mu_{t_n}) + 1 \right) \;\leq\; \frac{1}{c}\mathcal{F}(\mu) + 1 \;<\; \infty. \tag{V.2.3}$$

Using (I.2.18) (or [127, 6.8 (iii)]), this implies that there exist a further subsequence $(t_{n_k})_k$ and $\mu^* \in \mathcal{P}_2(\mathbb{R})$ such that $\lim_{k\to\infty} W_2(\mu_{t_{n_k}}, \mu^*)$. It remains to show that $\mu^* \in \{\mu^-, \mu^0, \mu^+\}$. In order to do this, we use the lower semi-continuity of $|\partial\mathcal{F}|$ ([3, 4.3.2]]) and Step 1 to observe that

$$|\partial\mathcal{F}|(\mu^*) \;\leq\; \liminf_{k\to\infty} |\partial\mathcal{F}|(\mu_{t_{n_k}}) \;=\; 0. \tag{V.2.4}$$

Combining this with Lemma V.4 yields the claim of Step 2.

Step 3. [$\lim_{t\to\infty} \mathcal{F}(\mu_t) = \mathcal{F}(\mu^*)$.]
First note that by the lower semi-continuity of $\mathcal{F}$ (Theorem III.35 or [3, Section 9.3]]), we have that

$$L^* = \lim_{t\to\infty} \mathcal{F}(\mu_t) = \lim_{k\to\infty} \mathcal{F}(\mu_{t_{n_k}}) \geq \mathcal{F}(\mu^*). \tag{V.2.5}$$

To show the other inequality, we use [3, 2.4.9], and observe that for all $k \in \mathbb{N}$,

$$|\partial\mathcal{F}|(\mu_{t_{n_k}}) \;\geq\; \left(\frac{\mathcal{F}(\mu_{t_{n_k}}) - \mathcal{F}(\mu^*)}{W_2(\mu_{t_{n_k}}, \mu^*)} + \frac{\lambda}{2} W_2(\mu_{t_{n_k}}, \mu^*) \right)^+, \tag{V.2.6}$$

which is equivalent to

$$W_2(\mu_{t_{n_k}}, \mu^*)\, |\partial\mathcal{F}|(\mu_{t_{n_k}}) \;\geq\; \left(\mathcal{F}(\mu_{t_{n_k}}) - \mathcal{F}(\mu^*) + \frac{\lambda}{2} W_2^2(\mu_{t_{n_k}}, \mu^*) \right)^+. \tag{V.2.7}$$

Taking the limit as $k \to \infty$ on both sides, and using Step 1 and Step 2, this implies that

$$0 \;\geq\; \left(L^* - \mathcal{F}(\mu^*)\right)^+. \tag{V.2.8}$$

We conclude that $L^* \leq \mathcal{F}(\mu^*)$. □

Proposition V.6 *Suppose Assumption V.1. Let $\mu \in \mathcal{P}_2(\mathbb{R})$ be such that $\int_{\mathbb{R}} z\, d\mu(z) \neq 0$ and $\mathcal{F}(\mu) \leq \mathcal{F}(\mu^0)$. Then,*

$$\lim_{t\to\infty} \mathcal{F}(S[\mu](t)) \;=\; \mathcal{F}(\mu^-) \;=\; \mathcal{F}(\mu^+), \tag{V.2.9}$$

and

$$\lim_{t\to\infty} W_2(S[\mu](t), \mu^-) \;=\; 0 \quad \text{if } \int_{\mathbb{R}} z\, d\mu(z) < 0 \quad \text{and} \tag{V.2.10}$$

$$\lim_{t\to\infty} W_2(S[\mu](t), \mu^+) \;=\; 0 \quad \text{if } \int_{\mathbb{R}} z\, d\mu(z) > 0. \tag{V.2.11}$$

Proof. In the following let $\mu_t = S[\mu](t)$. It suffices to consider only the case that $\mathrm{m}[\mu] < 0$. We know from Lemma V.5 that there exists a subsequence $(\mu_{t_k})_k$ such that

$$\lim_{k\to\infty} W_2(\mu_{t_k}, \mu^*) = 0 \quad \text{and} \quad \lim_{t\to\infty} \mathcal{F}(\mu_t) = \mathcal{F}(\mu^*) \quad \text{for some } \mu^* \in \{\mu^-, \mu^0, \mu^+\}. \quad \text{(V.2.12)}$$

We first show that $\mu^* = \mu^-$ (which implies (V.2.9)), and then show that $\lim_{t\to\infty} W_2(\mu_t, \mu^-) = 0$ (which implies (V.2.10)).

Step 1. [$\mu^* = \mu^-$.]

We show that the cases $\mu^* = \mu^+$ or $\mu^* = \mu^0$ lead to contradictions. First suppose that $\mu^* = \mu^+$. Since the map $t \mapsto \mathrm{m}[\mu_t]$ is continuous and since $\mathrm{m}[\mu_0] = \mathrm{m}[\mu] < 0$, we have that there exists $t' \in (0, \infty)$ such that $\mathrm{m}[\mu_{t'}] = 0$. Then, by the monotonicity of $t \mapsto \mathcal{F}(\mu_t)$ and by Lemma V.2,

$$\mathcal{F}(\mu^0) \;\geq\; \mathcal{F}(\mu) \;\geq\; \mathcal{F}(\mu_{t'}) \;\geq\; \mathcal{F}(\mu^0). \quad \text{(V.2.13)}$$

Hence, $\mathcal{F}(\mu_{t'}) = \mathcal{F}(\mu)$, which implies that μ is stationary. Moreover, (V.2.13) yields that $\mathcal{F}(\mu^0) = \mathcal{F}(\mu)$. By Lemma V.2, we infer that $\mu = \mu^0$. This contradicts the fact that $\mathrm{m}[\mu] < 0$. The case $\mu^* = \mu^0$ is treated analogously.

Step 2. [$\lim_{t\to\infty} W_2(\mu_t, \mu^-) = 0$.]

Let $(\mu_{s_n})_{n\in\mathbb{N}}$ be any subsequence of $(\mu_t)_{t\in[0,\infty)}$. Using the same compactness argument from Step 2 of the proof of Lemma V.5, we know that there exists a further subsequence $(\mu_{s_{n_k}})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} W_2(\mu_{s_{n_k}}, \mu') = 0$ for some $\mu' \in \mathcal{P}_2(\mathbb{R})$. In order to show the claim of Step 2, it remains to show that $\mu' = \mu^-$. If $\mathrm{m}[\mu'] \geq 0$, we infer a contradiction by repeating the same arguments from Step 1. So we have that $\mathrm{m}[\mu'] < 0$. Moreover, we have that

$$\mathcal{F}(\mu') \;\leq\; \liminf_{k\to\infty} \mathcal{F}(\mu_{s_{n_k}}) \;=\; \lim_{t\to\infty} \mathcal{F}(\mu_t) \;=\; \mathcal{F}(\mu^-). \quad \text{(V.2.14)}$$

In view of Lemma V.2, this implies that $\mu' = \mu^-$ or $\mu' = \mu^+$. The latter case is not possible, since $\mathrm{m}[\mu'] < 0$. This yields the claim of Step 2. □

V.3 Basin of attraction

Proposition V.7 *Suppose Assumption V.1. Recall the definition of $\mathcal{B}^-$ and $\mathcal{B}^+$ from* (I.7.7). *Then, $\mathcal{B}^-$ and $\mathcal{B}^+$ are open subsets of $\mathcal{P}_2(\mathbb{R})$.*

Proof. In view of Lemma V.3, it suffices to show the claim only for $\mathcal{B}^-$. Let $\nu \in \mathcal{B}^-$. We abbreviate $\Delta := \mathcal{F}(\mu^0) - \mathcal{F}(\mu^-)$. Let $t' > 0$ be such that for all $t \geq t'$,

- $W_2(S[\nu](t), \mu^-) \;\leq\; \frac{1}{4} m^\star$,
- $\mathcal{F}(S[\nu](t)) \;\leq\; \mathcal{F}(\mu^-) + \frac{1}{4}\Delta$, and
- $\mathrm{e}^{\lambda t} \;=\; \mathrm{e}^{-|\lambda| t} \;\leq\; \frac{1}{2}$.

It is easy to see that such a number t' exists by using Lemma V.5 and the fact that $\nu \in \mathcal{B}^-$. Set

$$\delta \;:=\; \min\left\{ \mathrm{e}^{2\lambda t'} \frac{m^\star}{4} \;,\; \sqrt{\mathrm{e}^{2\lambda t'} \frac{1}{|\lambda|} \frac{\Delta}{4}} \right\}. \quad \text{(V.3.1)}$$

We now show that $B_\delta(\nu) \subset \mathcal{B}^-$. Let $\mu \in B_\delta(\nu)$. We have to show that $\lim_{t\to\infty} S[\mu](t) = \mu^-$. In view of Proposition V.6, it suffices to show that

(i) $\mathrm{m}[S[\mu](2t')] < 0$, and that

(ii) $\mathcal{F}(S[\mu](2t')) \leq \mathcal{F}(\mu^0)$.

In order to show (i), note that by the contraction property ((III.1.76) or [3, 11.2.1]) and the definition of t' and δ,

$$W_2(S[\mu](2t'), \mu^-) \leq W_2(S[\nu](2t'), \mu^-) + \mathrm{e}^{-2\lambda t'}\delta \leq \frac{m^\star}{2}. \tag{V.3.2}$$

This implies claim (i). To show claim (ii), we use the regularization estimate ((III.1.79) or [3, 4.3.2]), and obtain that

$$\begin{aligned} \mathcal{F}(S[\mu](2t')) &\leq \mathcal{F}(S[\nu](t')) + |\lambda|\, W_2(S[\nu](t'), S[\mu](t'))^2 \\ &\leq \mathcal{F}(\mu^-) + \frac{1}{4}\Delta + \frac{1}{4}\Delta < \mathcal{F}(\mu^0). \end{aligned} \tag{V.3.3}$$

This concludes the proof of claim (ii). □

V.4 The ergodic theorem

Theorem V.8 *Suppose Assumption V.1. Let $\mu \in \mathcal{P}_2(\mathbb{R})$. Then, there exists a measure $\mu^* \in \{\mu^-, \mu^0, \mu^+\}$ such that*

$$\lim_{t\to\infty} W_2(S[\mu](t), \mu^*) = 0 \quad \textit{and} \quad \lim_{t\to\infty} \mathcal{F}(S[\mu](t)) = \mathcal{F}(\mu^*). \tag{V.4.1}$$

Proof. Applying the semigroup property of the McKean-Vlasov path and the regularization estimate ((III.1.79) or [3, 4.3.2]), we can assume without restriction that $\mu \in D(\mathcal{F})$.

We know from Lemma V.5 that there exists a subsequence $(\mu_{t_k})_k$ and $\mu^* \in \{\mu^-, \mu^0, \mu^+\}$ such that

$$\lim_{k\to\infty} W_2(\mu_{t_k}, \mu^*) = 0 \quad \text{and} \quad \lim_{t\to\infty} \mathcal{F}(\mu_t) = \mathcal{F}(\mu^*). \tag{V.4.2}$$

Let $(\mu_{s_n})_{n\in\mathbb{N}}$ be a subsequence of $(\mu_t)_{t\in[0,\infty)}$. As in Step 2 of the proof of Lemma V.5, we infer the existence of a further subsequence, still denoted by $(\mu_{s_n})_{n\in\mathbb{N}}$, such that

$$\lim_{n\to\infty} W_2(\mu_{s_n}, \nu^*) = 0 \quad \text{for some } \nu^* \in \mathcal{P}_2(\mathbb{R}). \tag{V.4.3}$$

It remains to show that $\nu^* = \mu^*$. We divide the proof into the three cases $\mu^* = \mu^-$, $\mu^* = \mu^0$ and $\mu^* = \mu^+$.

Case 1. [$\mu^* = \mu^-$.]

As in (V.2.14), we infer that $\mathcal{F}(\nu^*) \leq \mathcal{F}(\mu^-)$. By Lemma V.2, this implies that either $\nu^* = \mu^- = \mu^*$ or $\nu^* = \mu^+$. It remains to show that the latter case leads to a contradiction. Note that by (V.4.2) and (V.4.3),

- there exists $T > 0$ such that $\mathcal{F}(\mu_t) \in [\mathcal{F}(\mu^-), \mathcal{F}(\mu^0))$ for all $t \geq T$,
- there exists $N \in \mathbb{N}$ such that $s_n \geq T$ and $\mathrm{m}[\mu_{s_n}] > 0$ for all $n \geq N$, and

- there exists $K \in \mathbb{N}$ such that $t_k > s_N$ and $\mathrm{m}[\mu_{t_k}] < 0$ for all $k \geq K$.

In particular, we have that

$$\mathcal{F}(\mu_t) < \mathcal{F}(\mu^0) \text{ for all } t \in [s_N, t_K], \qquad \mathrm{m}[\mu_{s_N}] > 0, \quad \text{and} \quad \mathrm{m}[\mu_{t_K}] < 0. \tag{V.4.4}$$

Hence, there exists $t' \in [s_N, t_K]$ such that $\mathcal{F}(\mu_{t'}) < \mathcal{F}(\mu^0)$ and $\mathrm{m}[\mu_{t'}] = 0$. This contradicts Lemma V.2.

Case 2. [$\mu^* = \mu^+$.]

This case is treated in the same way as Case 1.

Case 3. [$\mu^* = \mu^0$.]

In this case we have that $\mathcal{F}(\nu^*) \leq \mathcal{F}(\mu^0)$. There are three subcases given by $\mathrm{m}[\nu^*] = 0$, $\mathrm{m}[\nu^*] > 0$ and $\mathrm{m}[\nu^*] < 0$.

Case 3.1. [$\mathrm{m}[\nu^*] = 0$.]

By Lemma V.2, the combination of $\mathcal{F}(\nu^*) \leq \mathcal{F}(\mu^0)$ and $\mathrm{m}[\nu^*] = 0$ yields that $\nu^* = \mu^0 = \mu^*$.

Case 3.2. [$\mathrm{m}[\nu^*] < 0$.]

From Proposition V.6 we know that $\nu^* \in \mathcal{B}^-$. Hence, by Proposition V.7, there exists $\delta > 0$ such that $B_\delta(\nu^*) \subset \mathcal{B}^-$. In particular, by (V.4.3), there exists $N \in \mathbb{N}$ such that $\mu_{s_N} \in \mathcal{B}^-$. This contradicts (V.4.2). Indeed, the fact that $\mu_{s_N} \in \mathcal{B}^-$ implies that

$$\lim_{t\to\infty} \mu_{s_N+t} \;=\; \lim_{t\to\infty} S[\mu_{s_N}](t) \;=\; \mu^- \quad \text{ in } \mathcal{P}_2(\mathbb{R}), \tag{V.4.5}$$

which contradicts the fact that

$$\lim_{k\to\infty} \mu_{t_k} \;=\; \mu^* \;=\; \mu^0 \quad \text{ in } \mathcal{P}_2(\mathbb{R}). \tag{V.4.6}$$

Case 3.3. [$\mathrm{m}[\nu^*] > 0$.]

This case is treated in the same way as Case 3.2. □

We conclude this chapter with some immediate consequences of Theorem V.8 on certain properties of the set $\mathcal{B}^0$.

Corollary V.9 *(i)* $\mathcal{B}^0$ *is closed,*

(ii) $\mathcal{B}^0 \supset \{\ \mu \in \mathcal{P}_2(\mathbb{R}) \mid \mu$ *is symmetric, i.e.* $\varsigma_\# \mu = \mu\ \}$, *and*

(iii) $\mu^0 \in \partial\mathcal{B}^0$.

Proof. To show part *(i)*, we simply use Proposition V.7 and that $\mathcal{P}_2(\mathbb{R}) = \mathcal{B}^- \cup \mathcal{B}^0 \cup \mathcal{B}^+$. Part *(ii)* is a straightforward consequence of Theorem V.8 and Lemma V.3. Finally, to show part *(iii)*, we use that

- by Proposition V.6 and Lemma V.2, $\mu^{1,\varphi_1'(-\eta)} \in \mathcal{B}^-$ and $\mu^{1,\varphi_1'(\eta)} \in \mathcal{B}^+$ for all $\eta > 0$, and that
- $\lim_{\eta\to 0} \mu^{1,\varphi_1'(-\eta)} = \mu^0$ and $\lim_{\eta\to 0} \mu^{1,\varphi_1'(\eta)} = \mu^0$ in $\mathcal{P}_2(\mathbb{R})$.

This concludes the proof. □

Bibliography

[1] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Comm. Math. Phys.*, 307(3):791–815, 2011.

[2] L. Ambrosio and N. Gigli. A users guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[4] L. Ambrosio and G. Savaré. Gradient flows of probability measures. In *Handbook of differential equations: evolutionary equations. Vol. III*, Handb. Differ. Equ., pages 1–136. Elsevier/North-Holland, Amsterdam, 2007.

[5] S. Arnrich, A. Mielke, M. A. Peletier, G. Savaré, and M. Veneroni. Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction. *Calc. Var. Partial Differential Equations*, 44(3-4):419–454, 2012.

[6] S. Arrhenius. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Zeitschrift für physikalische Chemie*, 4(1):226–248, 1889.

[7] K. B. Athreya and S. N. Lahiri. *Measure theory and probability theory*. Springer Texts in Statistics. Springer, New York, 2006.

[8] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*, volume 17 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Applications to PDEs and optimization.

[9] F. Barret. Sharp asymptotics of metastable transition times for one dimensional SPDEs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):129–166, 2015.

[10] F. Barret, A. Bovier, and S. Méléard. Uniform estimates for metastable transition times in a coupled bistable system. *Electron. J. Probab.*, 15:no. 12, 323–345, 2010.

[11] K. Bashiri. On the metastability in three modifications of the Ising model. *Markov Proc. Rel. Fields*, 25(3):483–532, 2019.

[12] K. Bashiri. On the basin of attraction of McKean-Vlasov paths. Preprint, arXiv:2001.09106, 2020.

[13] K. Bashiri and A. Bovier. Gradient flow approach to local mean-field spin systems. *Stochastic Process. Appl.*, https://doi.org/10.1016/j.spa.2019.05.006, 2019.

[14] K. Bashiri and G. Menz. Metastability in a continuous mean-field model at low temperature and strong interaction. Preprint, arXiv:1910.11828, 2019.

[15] G. Basile, D. Benedetto, and L. Bertini. A gradient flow approach to linear Boltzmann equations. Preprint, arxiv:1707.09204, 2017.

[16] J. Beltrán and C. Landim. Tunneling and metastability of continuous time Markov chains. *J. Stat. Phys.*, 140(6):1065–1114, 2010.

[17] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[18] N. Berglund. Kramers' law: validity, derivations and generalisations. *Markov Process. Related Fields*, 19(3):459–490, 2013.

[19] N. Berglund, G. Di Gesù, and H. Weber. An Eyring-Kramers law for the stochastic Allen-Cahn equation in dimension two. *Electron. J. Probab.*, 22:Paper No. 41, 27, 2017.

[20] N. Berglund, B. Fernandez, and B. Gentz. Metastability in interacting nonlinear stochastic differential equations. II. Large-N behaviour. *Nonlinearity*, 20(11):2583–2614, 2007.

[21] N. Berglund and B. Gentz. The Eyring-Kramers law for potentials with nonquadratic saddles. *Markov Process. Related Fields*, 16(3):549–598, 2010.

[22] N. Berglund and B. Gentz. Sharp estimates for metastable lifetimes in parabolic SPDEs: Kramers' law and beyond. *Electron. J. Probab.*, 18:no. 24, 58, 2013.

[23] J. Bewersdorff. *Algebra für Einsteiger*. Friedr. Vieweg & Sohn, Wiesbaden, second edition, 2004. Von der Gleichungsauflösung zur Galois-Theorie.

[24] A. Bianchi, A. Bovier, and D. Ioffe. Sharp asymptotics for metastability in the random field Curie-Weiss model. *Electron. J. Probab.*, 14:no. 53, 1541–1603, 2009.

[25] A. Bianchi, A. Bovier, and D. Ioffe. Pointwise estimates and exponential laws in metastable systems via coupling methods. *Ann. Probab.*, 40(1):339–371, 2012.

[26] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[27] F. Bouchet and J. Reygner. Generalisation of the Eyring-Kramers transition rate formula to irreversible diffusion processes. *Ann. Henri Poincaré*, 17(12):3499–3532, 2016.

[28] A. Bovier. *Gaussian processes on trees*, volume 163 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2017. From spin glasses to branching Brownian motion.

[29] A. Bovier and F. den Hollander. *Metastability*, volume 351 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2015. A potential-theoretic approach.

[30] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theory Related Fields*, 119(1):99–161, 2001.

[31] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and low lying spectra in reversible Markov chains. *Comm. Math. Phys.*, 228(2):219–255, 2002.

[32] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4):399–424, 2004.

[33] A. Bovier, D. Ioffe, and P. Müller. The Hydrodynamic Limit for Local Mean-Field Dynamics with Unbounded Spins. *J. Stat. Phys.*, 172(2):434–457, 2018.

[34] A. Bovier and F. Manzo. Metastability in Glauber dynamics in the low-temperature limit: beyond exponential asymptotics. *J. Statist. Phys.*, 107(3-4):757–779, 2002.

[35] A. Bovier, S. Marello, and E. Pulvirenti. Metastability for the dilute Curie-Weiss model with Glauber dynamics. Preprint, arXiv:1912.10699, 2019.

[36] J. A. Carrillo, M. G. Delgadino, and G. A. Pavliotis. A proof of the mean-field limit for -convex potentials by -Convergence. Preprint, arXiv:1906.04601, 2019.

[37] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[38] M. Cassandro, A. Galves, E. Olivieri, and M. E. Vares. Metastable behavior of stochastic dynamics: a pathwise approach. *J. Statist. Phys.*, 35(5-6):603–634, 1984.

[39] R. Cerf and F. Manzo. Nucleation and growth for the Ising model in d dimensions at very low temperatures. *Ann. Probab.*, 41(6):3697–3785, 2013.

[40] E. N. M. Cirillo and F. R. Nardi. Relaxation height in energy landscapes: an application to multiple metastable states. *J. Stat. Phys.*, 150(6):1080–1114, 2013.

[41] E. N. M. Cirillo, F. R. Nardi, and J. Sohier. Metastability for general dynamics with rare transitions: escape time and critical configurations. *J. Stat. Phys.*, 161(2):365–403, 2015.

[42] E. B. Davies. Dynamical stability of metastable states. *J. Functional Analysis*, 46(3):373–386, 1982.

[43] E. B. Davies. Metastability and the Ising model. *J. Statist. Phys.*, 27(4):657–675, 1982.

[44] E. B. Davies. Metastable states of symmetric Markov semigroups. I. *Proc. London Math. Soc. (3)*, 45(1):133–150, 1982.

[45] E. B. Davies. Metastable states of symmetric Markov semigroups. II. *J. London Math. Soc. (2)*, 26(3):541–556, 1982.

[46] D. A. Dawson and J. Gärtner. Large deviations and tunnelling for particle systems with mean field interaction. *C. R. Math. Rep. Acad. Sci. Canada*, 8(6):387–392, 1986.

[47] D. A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[48] D. A. Dawson and J. Gärtner. Large deviations, free energy functional and quasi-potential for a mean field model of interacting diffusions. *Mem. Amer. Math. Soc.*, 78(398):iv+94, 1989.

[49] E. De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 81–98. Masson, Paris, 1993.

[50] P. Dehghanpour and R. H. Schonmann. A nucleation-and-growth model. *Probab. Theory Related Fields*, 107(1):123–135, 1997.

[51] L. Dello Schiavo. The Dirichlet-Ferguson Diffusion on the Space of Probability Measures over a Closed Riemannian Manifold. Preprint, arXiv:1811.11598, 2019.

[52] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[53] F. den Hollander and O. Jovanovski. Glauber dynamics on the Erdös-Rényi random graph. Preprint, arXiv:1912.10591, 2019.

[54] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[55] M. H. Duong, V. Laschos, and M. Renger. Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM Control Optim. Calc. Var.*, 19(4):1166–1188, 2013.

[56] M. H. Duong, M. A. Peletier, and J. Zimmer. GENERIC formalism of a Vlasov-Fokker-Planck equation and connection to large-deviation principles. *Nonlinearity*, 26(11):2951–2971, 2013.

[57] E. B. Dynkin. *Markov processes. Vols. I, II*, volume 122 of *Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121*. Academic Press Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg, 1965.

[58] R. S. Ellis, J. L. Monroe, and C. M. Newman. The ghs and other correlation inequalities for a class of even ferromagnets. *Comm. Math. Phys.*, 46(2):167–182, 1976.

[59] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):1–23, 2010.

[60] M. Erbar. Gradient flows of the entropy for jump processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(3):920–945, 2014.

[61] M. Erbar. A gradient flow approach to the Boltzmann equation. Preprint, arxiv:1603.0540v4, 2016.

[62] M. Erbar and M. Fathi. Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. *J. Funct. Anal.*, 274(11):3056–3089, 2018.

[63] M. Erbar, M. Fathi, V. Laschos, and A. Schlichting. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.*, 36(12):6799–6833, 2016.

[64] M. Erbar, M. Fathi, and A. Schlichting. Entropic curvature and convergence to equilibrium for mean-field dynamics on discrete spaces. Preprint, arXiv:1908.03397, 2019.

[65] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.

[66] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.

[67] M. Erbar, J. Maas, and D. R. M. Renger. From large deviations to Wasserstein gradient flows in multiple dimensions. *Electron. Commun. Probab.*, 20:no. 89, 12, 2015.

[68] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[69] H. Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935.

[70] M. Fathi. A gradient flow approach to large deviations for diffusion processes. *J. Math. Pures Appl. (9)*, 106(5):957–993, 2016.

[71] M. Fathi and M. Simon. The gradient flow approach to hydrodynamic limits for the simple exclusion process. In *From particle systems to partial differential equations. III*, volume 162 of *Springer Proc. Math. Stat.*, pages 167–184. Springer, [Cham], 2016.

[72] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[73] R. Fernandez, F. Manzo, F. R. Nardi, E. Scoppola, and J. Sohier. Conditioned, quasi-stationary, restricted measures and escape from metastable states. *Ann. Appl. Probab.*, 26(2):760–793, 2016.

[74] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs.

[75] J. Gärtner. On the McKean-Vlasov limit for interacting diffusions. *Math. Nachr.*, 137:197–248, 1988.

[76] B. Gaveau and M. Moreau. Metastable relaxation times and absorption probabilities for multidimensional stochastic systems. *J. Phys. A*, 33(27):4837–4850, 2000.

[77] B. Gaveau and L. S. Schulman. Theory of nonequilibrium first-order phase transitions for stochastic dynamics. *J. Math. Phys.*, 39(3):1517–1533, 1998.

[78] S. Glasstone, K. J. Laidler, and H. Eyring. The theory of rate processes; the kinetics of chemical reactions, viscosity, diffusion and electrochemical phenomena. Technical report, McGraw-Hill Book Company,, 1941.

[79] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.

[80] N. Grunewald, F. Otto, C. Villani, and M. G. Westdickenberg. A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(2):302–351, 2009.

[81] M. Herrmann and B. Niethammer. Kramers' formula for chemical reactions in the context of Wasserstein gradient flows. *Commun. Math. Sci.*, 9(2):623–635, 2011.

[82] S. Herrmann and J. Tugaut. Non-uniqueness of stationary measures for self-stabilizing processes. *Stochastic Process. Appl.*, 120(7):1215–1246, 2010.

[83] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[84] L. V. Kantorovich. On a problem of Monge. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):15–16, 2004.

[85] L. V. Kantorovich. On mass transportation. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):11–14, 2004.

[86] J. L. Kelley. *General topology*. Springer-Verlag, New York-Berlin, 1975. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27.

[87] C. Kipnis and C. Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[88] R. Kotecký and E. Olivieri. Droplet dynamics for asymmetric Ising model. *J. Statist. Phys.*, 70(5-6):1121–1148, 1993.

[89] R. Kotecký and E. Olivieri. Shapes of growing droplets—a model of escape from a metastable phase. *J. Statist. Phys.*, 75(3-4):409–506, 1994.

[90] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.

[91] Y. Kwon and G. Menz. Decay of correlations and uniqueness of the infinite-volume Gibbs measure of the canonical ensemble of 1d-lattice systems. *J. Stat. Phys.*, 176(4):836–872, 2019.

[92] C. Landim. Metastable Markov chains. *Probab. Surv.*, 16:143–227, 2019.

[93] C. Landim, M. Mariani, and I. Seo. Dirichlet's and Thomson's principles for non-selfadjoint elliptic operators with application to non-reversible metastable diffusion processes. *Arch. Ration. Mech. Anal.*, 231(2):887–938, 2019.

[94] C. Landim and I. Seo. Metastability of one-dimensional, non-reversible diffusions with periodic boundary conditions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(4):1850–1889, 2019.

[95] M. Ledoux. Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités, XXXV*, volume 1755 of *Lecture Notes in Math.*, pages 167–194. Springer, Berlin, 2001.

[96] S. Lisini. Characterization of absolutely continuous curves in Wasserstein spaces. *Calc. Var. Partial Differential Equations*, 28(1):85–120, 2007.

[97] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[98] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.

[99] F. Manzo, F. R. Nardi, E. Olivieri, and E. Scoppola. On the essential features of metastability: tunnelling time and critical configurations. *J. Statist. Phys.*, 115(1-2):591–642, 2004.

[100] M. Mariani. A Γ-convergence approach to large deviations. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 18(3):951–976, 2018.

[101] G. Menz and F. Otto. Uniform logarithmic Sobolev inequalities for conservative spin systems with super-quadratic single-site potential. *Ann. Probab.*, 41(3B):2182–2224, 2013.

[102] G. Menz and A. Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.*, 42(5):1809–1884, 2014.

[103] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations*, 48(1-2):1–31, 2013.

[104] G. Monge. *Mémoire sur la théorie des déblais et des remblais.* De l'Imprimerie Royale, 1781.

[105] P. E. Müller. Path large deviations for interacting diffusions with local mean-field interactions in random environment. *Electron. J. Probab.*, 22:Paper No. 76, 56, 2017.

[106] P. E. Müller. Limiting properties of a continuous local mean-field interacting spin system: Hydrodynamic limit, propagation of chaos, energy landscape and large deviations. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2016.

[107] F. R. Nardi and E. Olivieri. Low temperature stochastic dynamics for an Ising model with alternating field. volume 2, pages 117–166. 1996. Disordered systems and statistical physics: rigorous results (Budapest, 1995).

[108] E. J. a. Neves and R. H. Schonmann. Critical droplets and metastability for a Glauber dynamics at very low temperatures. *Comm. Math. Phys.*, 137(2):209–230, 1991.

[109] E. Olivieri and M. E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.

[110] H. C. Öttinger. *Stochastic processes in polymeric fluids*. Springer-Verlag, Berlin, 1996. Tools and examples for developing simulation algorithms.

[111] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[112] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361 – 400, 2000.

[113] M. A. Peletier, D. R. M. Renger, and M. Veneroni. Variational formulation of the Fokker-Planck equation with decay: a particle approach. *Commun. Contemp. Math.*, 15(5):1350017, 43, 2013.

[114] M. A. Peletier, G. Savaré, and M. Veneroni. Chemical reactions as Γ-limit of diffusion [revised reprint of mr2679596]. *SIAM Rev.*, 54(2):327–352, 2012.

[115] E. Sandier and S. Serfaty. Gamma-convergence of gradient flows with applications to Ginzburg-Landau. *Comm. Pure Appl. Math.*, 57(12):1627–1672, 2004.

[116] R. H. Schonmann. Slow droplet-driven relaxation of stochastic Ising models in the vicinity of the phase coexistence region. *Comm. Math. Phys.*, 161(1):1–49, 1994.

[117] I. Seo and P. Tabrizian. Asymptotics for scaled Kramers–Smoluchowski equations in several dimensions with general potentials. *Calc. Var. Partial Differential Equations*, 59(1):Paper No. 11, 2020.

[118] S. Serfaty. Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *Discrete Contin. Dyn. Syst.*, 31(4):1427–1451, 2011.

[119] M. Slowik. Contributions to the potential theoretic approach to metastability with applications to the random field Curie-Weiss-Potts model. PhD thesis, Technische Universität Berlin, 2012.

[120] M. Slowik and A. Schlichting. Poincaré and logarithmic Sobolev constants for metastable Markov chains via capacitary inequalities. Preprint, arXiv:1705.05135, 2017.

[121] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[122] K.-T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.

[123] M. Sugiura. Metastable behaviors of diffusion processes with small parameter. *J. Math. Soc. Japan*, 47(4):755–788, 1995.

[124] M. Sugiura. Sharp asymptotics of diffusion processes with small parameter and applications to metastable behavior. In *Nonlinear stochastic PDEs (Minneapolis, MN, 1994)*, volume 77 of *IMA Vol. Math. Appl.*, pages 127–136. Springer, New York, 1996.

[125] J. Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *Ann. Probab.*, 41(3A):1427–1460, 2013.

[126] A. D. Ventcel' and M. I. Freĭdlin. Small random perturbations of dynamical systems. *Uspehi Mat. Nauk*, 25(1 (151)):3–55, 1970.

[127] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[128] M.-K. von Renesse and K.-T. Sturm. Entropic measure and Wasserstein diffusion. *Ann. Probab.*, 37(3):1114–1191, 2009.

[129] H.-T. Yau. Relative entropy and hydrodynamics of Ginzburg-Landau models. *Lett. Math. Phys.*, 22(1):63–80, 1991.