



Universiteit  
Leiden  
The Netherlands

## Metastability for Glauber dynamics on the complete graph with coupling disorder

Bovier, A.; Hollander, W.T.F. den; Mareello, S

### Citation

Bovier, A., Hollander, W. T. F. den, & Mareello, S. (2022).  
Metastability for Glauber dynamics on the complete graph with  
coupling disorder. *Communications In Mathematical Physics*, 392,  
307-345. doi:10.1007/s00220-022-04351-8

Version: Publisher's Version  
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)  
Downloaded from: <https://hdl.handle.net/1887/3589999>

**Note:** To cite this publication please use the final published version (if applicable).



# Metastability for Glauber Dynamics on the Complete Graph with Coupling Disorder

Anton Bovier<sup>1</sup> , Frank den Hollander<sup>2</sup> , Saeda Mareello<sup>1</sup>

<sup>1</sup> Institut für Angewandte Mathematik, Rheinische Friedrich-Wilhelms-Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany. E-mail: [mareello@iam.uni-bonn.de](mailto:mareello@iam.uni-bonn.de); [bovier@uni-bonn.de](mailto:bovier@uni-bonn.de)

<sup>2</sup> Mathematical Institute, Leiden University, P. O. Box 9512, 2300 RA Leiden, The Netherlands. E-mail: [denholla@math.leidenuniv.nl](mailto:denholla@math.leidenuniv.nl)

Received: 9 July 2021 / Accepted: 17 February 2022  
Published online: 17 March 2022 – © The Author(s) 2022

**Abstract:** Consider the complete graph on  $n$  vertices. To each vertex assign an Ising spin that can take the values  $-1$  or  $+1$ . Each spin  $i \in [n] = \{1, 2, \dots, n\}$  interacts with a magnetic field  $h \in [0, \infty)$ , while each pair of spins  $i, j \in [n]$  interact with each other at coupling strength  $n^{-1}J(i)J(j)$ , where  $J = (J(i))_{i \in [n]}$  are i.i.d. non-negative random variables drawn from a probability distribution with finite support. Spins flip according to a Metropolis dynamics at inverse temperature  $\beta \in (0, \infty)$ . We show that there are critical thresholds  $\beta_c$  and  $h_c(\beta)$  such that, in the limit as  $n \rightarrow \infty$ , the system exhibits metastable behaviour if and only if  $\beta \in (\beta_c, \infty)$  and  $h \in [0, h_c(\beta))$ . Our main result is a sharp asymptotics, up to a multiplicative error  $1 + o_n(1)$ , of the average crossover time from any metastable state to the set of states with lower free energy. We use standard techniques of the potential-theoretic approach to metastability. The leading order term in the asymptotics does not depend on the realisation of  $J$ , while the correction terms do. The leading order of the correction term is  $\sqrt{n}$  times a centred Gaussian random variable with a complicated variance depending on  $\beta, h$ , on the law of  $J$  and on the metastable state. The critical thresholds  $\beta_c$  and  $h_c(\beta)$  depend on the law of  $J$ , and so does the number of metastable states. We derive an explicit formula for  $\beta_c$  and identify some properties of  $\beta \mapsto h_c(\beta)$ . Interestingly, the latter is not necessarily monotone, meaning that the metastable crossover may be re-entrant.

## 1. Introduction and Main Results

**1.1. Background.** Interacting particle systems evolving according to a Metropolis dynamics associated with an energy functional called the Hamiltonian, may be trapped for a long time near a state that is a local minimum of the free energy, but not a global minimum. The deepest local minima are called *metastable states*, the global minimum is called the *stable state*. The transition from a metastable state to the stable state marks the relaxation of the system to equilibrium. To describe this relaxation, one needs to identify the set of critical configurations the system must attain in order to achieve this

transition and to compute the crossover time. These critical configurations correspond to saddle points in the free energy landscape.

Metastability for interacting particle systems on *lattices* has been studied intensively in the past. For a summary, we refer the reader to the monographs by Olivieri and Vares [13], and Bovier and den Hollander [6]. Successful attempts towards understanding metastable behaviour in random environments were made for the random field Curie–Weiss model, by Mathieu and Picco [12], Bovier et al. [3] and Bianchi et al. [1, 2]. Recently, there has been interest in metastability for interacting particle systems on *random graphs*. This is challenging, because the crossover times typically depend on the realisation of the graph. In den Hollander and Jovanovski [11] and Bovier et al. [7], Glauber dynamics on dense *Erdős–Rényi* random graphs was analysed. Earlier work on metastability for Glauber dynamics on sparse random graphs can be found in Dommers [8] (random regular graph) and Dommers et al. [10] (configuration model). The present paper is a first step towards the study of metastability for Glauber dynamics on *Chung–Lu*-like random graphs.

To the best of our knowledge, Tindemans and Capel [14] and Dommers et al. [9] are the only references where the model with the interaction Hamiltonian in (1.2) below has been studied in detail. Both focus on equilibrium properties only.

**1.2. Glauber dynamics on the complete graph with coupling disorder.** Let  $\mathcal{K}_n$  be the complete graph on  $n$  vertices. Each vertex carries an Ising spin that can take the values  $-1$  or  $+1$ . Let  $S_n = \{-1, +1\}^{[n]}$  denote the set of spin configurations on  $\mathcal{K}_n$ , where  $[n] = \{1, 2, \dots, n\}$ . Let  $(\Omega, \mathcal{F}, \mathcal{P})$  be an abstract probability space, and let  $J = (J(i))_{i \in [n]}$  be a sequence of i.i.d. random variables on this probability space taking values in a *finite* set  $\{a_1, \dots, a_k\} \subset [0, \infty)$  of cardinality  $k \in \mathbb{N}$ . The distribution of these random variables is given by

$$\mathcal{P}(J(i) = a_\ell) = \omega_\ell \in (0, 1), \quad i \in [n], \ell \in [k], \quad (1.1)$$

with  $\sum_{\ell \in [k]} \omega_\ell = 1$ .

Let  $H_n: S_n \rightarrow \mathbb{R}$  be the interaction Hamiltonian defined by

$$H_n(\sigma) \equiv -\frac{1}{n} \sum_{\substack{i, j \in [n] \\ i < j}} J(i)J(j) \sigma(i)\sigma(j) - h \sum_{i \in [n]} \sigma(i), \quad \sigma \in S_n, \quad (1.2)$$

where  $h \in [0, \infty)$  is the magnetic field. We consider *Glauber dynamics* on  $S_n$ , defined as the continuous-time Markov process with transition rates

$$r_n(\sigma, \sigma') = \begin{cases} e^{-\beta[H_n(\sigma') - H_n(\sigma)]_+}, & \text{if } \sigma' \sim \sigma, \\ 0, & \text{otherwise,} \end{cases} \quad \sigma, \sigma' \in S_n, \quad (1.3)$$

where  $\beta \in (0, \infty)$  is the inverse temperature,  $\sigma' \sim \sigma$  means that  $\sigma'$  differs from  $\sigma$  by a single spin-flip and  $[\cdot]_+$  is the positive part. This dynamics is *reversible* with respect to the *Gibbs measure*

$$\mu_n(\sigma) \equiv \frac{1}{Z_n} e^{-\beta H_n(\sigma)}, \quad \sigma \in S_n, \quad (1.4)$$

where the normalising constant  $Z_n$  is called the partition sum. Note that the reference measure for (1.4) is the *counting measure* on  $S_n$ . We write

$$(\sigma_t)_{t \geq 0}, \quad \sigma_t \in S_n, \quad (1.5)$$

to denote a path of the Glauber dynamics on  $S_n$ , and  $\mathbb{P}_\sigma$  and  $\mathbb{E}_\sigma$  to denote probability and expectation on path space given  $\sigma_0 = \sigma$  (we suppress  $J, h, \beta$  and  $n$  from the notation).

For fixed  $n$ , if  $h = 0$  the Hamiltonian in (1.2) has two global minima, at  $\sigma \equiv +1$  and  $\sigma \equiv -1$ , while if  $h > 0$  it achieves a global minimum at  $\sigma \equiv +1$  and a local minimum at  $\sigma \equiv -1$ . The latter is the deepest local minimum not equal to the global minimum (at least for  $h$  small enough). However, in the limit as  $n \rightarrow \infty$ , these do *not* form a metastable pair of configurations because *entropy* comes into play.

**1.3. Metastability on the complete graph with coupling disorder.** In this section we state our main results.

**1.3.1. Empirical magnetisations** The relevant quantity to monitor in order to characterise the metastable behaviour is the *disorder weighted magnetisation*

$$K_n(\sigma) = \frac{1}{n} \sum_{i \in [n]} J(i) \sigma(i), \quad \sigma \in S_n. \quad (1.6)$$

The following quantities will be essential for *coarse-graining*. Define the *level sets*

$$A_{\ell,n} \equiv \{i \in [n] : J(i) = a_\ell\}, \quad \ell \in [k], \quad (1.7)$$

and the *level magnetisations*

$$m_{\ell,n}(\sigma) \equiv \frac{1}{|A_{\ell,n}|} \sum_{i \in A_{\ell,n}} \sigma(i), \quad \ell \in [k], \sigma \in S_n. \quad (1.8)$$

Put

$$m_n(\sigma) = (m_{\ell,n}(\sigma))_{\ell \in [k]} \in [-1, 1]^k, \quad \sigma \in S_n, \quad (1.9)$$

and note that  $K_n(\sigma) = \frac{1}{n} \sum_{\ell \in [k]} a_\ell |A_{\ell,n}| m_{\ell,n}(\sigma)$  depends on  $\sigma$  only through  $m_n(\sigma)$ . Thus, with abuse of notation, we may define

$$K_n(m) \equiv \frac{1}{n} \sum_{\ell \in [k]} a_\ell |A_{\ell,n}| m_\ell, \quad m = (m_\ell)_{\ell \in [k]} \in [-1, 1]^k, \quad (1.10)$$

so that  $K_n(\sigma) = K_n(m_n(\sigma))$ .

*1.3.2. Thermodynamic limit* As  $n \rightarrow \infty$ , by the law of large numbers the random function  $K_n$  converges uniformly in probability to a deterministic function  $K$  given by

$$K(m) = \sum_{\ell \in [k]} a_\ell \omega_\ell m_\ell, \quad m = (m_\ell)_{\ell \in [k]} \in [-1, 1]^k. \quad (1.11)$$

Similarly, the random free energy function  $F_n$  converges uniformly in probability to a deterministic function  $F_{\beta,h}$  (see (2.15) and (2.26) below for explicit formulas). In Sect. 3, we show that the stationary points of  $F_{\beta,h}$  are given by  $\mathbf{m} = (\mathbf{m}_\ell)_{\ell \in [k]}$ , where

$$\mathbf{m}_\ell = \tanh(\beta[a_\ell K(\mathbf{m}) + h]), \quad \ell \in [k]. \quad (1.12)$$

Note that, via (1.12), the  $k$ -dimensional vector  $\mathbf{m}$  is fully determined by the real number  $K(\mathbf{m})$ . Therefore, finding the stationary points of  $F_{\beta,h}$  reduces to finding the solutions of the equation

$$K = T_{\beta,h}(K), \quad T_{\beta,h}(K) = \sum_{\ell \in [k]} a_\ell \omega_\ell \tanh(\beta[a_\ell K + h]). \quad (1.13)$$

*1.3.3. Metastable regime* It turns out that the critical inverse temperature  $\beta_c$  is given by

$$\beta_c = \left[ \sum_{\ell \in [k]} a_\ell^2 \omega_\ell \right]^{-1}. \quad (1.14)$$

Namely, if  $\beta \in (0, \beta_c]$ , then the system is not in the metastable regime for any  $h \in [0, \infty)$ , while if  $\beta \in (\beta_c, \infty)$ , then, for  $h \in [0, \infty)$  small enough, it is in the metastable regime (i.e., (1.13) has more than one solution at which  $T_{\beta,h}$  is not tangent to the diagonal). Given  $\beta \in (\beta_c, \infty)$ , the critical magnetic field  $h_c(\beta)$  is the minimal value of  $h$  for which the system is not metastable. The *metastable regime* is thus

$$\beta \in (\beta_c, \infty), \quad h \in [0, h_c(\beta)). \quad (1.15)$$

In Sect. 3, we show that  $\beta \mapsto h_c(\beta)$  is continuous on  $(\beta_c, \infty)$ , with

$$\lim_{\beta \downarrow \beta_c} h_c(\beta) = 0, \quad \lim_{\beta \rightarrow \infty} h_c(\beta) = C \in (0, \infty), \quad (1.16)$$

where the explicit value of  $C$  is given in (3.12) below. Interestingly,  $\beta \mapsto h_c(\beta)$  is not necessarily monotone, i.e., the metastable crossover may be *re-entrant*.

It turns out that there exists an  $\ell \in [k]$  (depending on  $\beta, h$  and on the law of the components of  $J$ ), such that  $F_{\beta,h}$  has  $2\ell + 1$  stationary points.

*1.3.4. Metastable crossover* Let  $\mathcal{M}_n$  be the set of minima of  $F_n$ . Given  $\mathbf{m} \in \mathcal{M}_n$ , define

$$\mathcal{M}_n(\mathbf{m}) \equiv \{m \in \mathcal{M}_n \setminus \mathbf{m} : F_n(m) \leq F_n(\mathbf{m})\}. \quad (1.17)$$

Let  $\mathcal{G}(A, B)$  be the gate between two disjoint subsets  $A$  and  $B$  of  $\mathcal{M}_n$ . We refer to [6, Section 10.1] for a precise definition of the gate.

Fix  $\mathbf{m}_n \in \mathcal{M}_n$  as the initial magnetisation. Throughout the paper we assume that the following hypotheses hold for  $\mathbf{m}_n$ .

**Hypothesis 1.**

1.  $\mathcal{M}_n(\mathbf{m}_n)$  is non-empty.
2. The Hessian of  $F_n$  has only non-zero eigenvalues at  $\mathbf{m}_n$  and at all the points in  $\mathcal{G}(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))$ .
3. There is a unique point  $\mathbf{t}_n$  in  $\mathcal{G}(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))$ , which will often be called simply saddle point.
4. The saddle point  $\mathbf{t}_n$  is such that  $r_\ell [ |A_{\ell,n}| (1 - \mathbf{t}_{\ell,n}^2) ]^{-1}$  takes distinct values for different  $\ell \in [k]$ , where  $r_\ell$  is defined in (4.9) below.

Hypothesis 1(2) and (3) are made to avoid complications. Hypothesis 1(4) is needed in the proof of Lemma 4.1 below (as in [6, Lemma 14.9]). Neither is very restrictive: if for some parameter choice they fail, then after an infinitesimal parameter change they hold. Moreover, if Hypothesis 1(3) fails, it is sufficient to compute separately the contribution to the crossover time of the various saddle points in the gate.

Let  $S_n[\mathbf{m}_n]$  and  $S_n[\mathcal{M}_n(\mathbf{m}_n)]$  denote the sets of configurations in  $S_n$  for which the level magnetisations are  $\mathbf{m}_n$  and are contained in  $\mathcal{M}_n(\mathbf{m}_n)$ , respectively. For  $A \subset S_n$ , write

$$\tau_A = \{t \geq 0: \sigma_t \in A, \sigma_{t-} \notin A\} \quad (1.18)$$

to denote the first hitting time or return time of  $A$ .

We next state our main results for the crossover time. Theorem 1.1 provides a sharp asymptotics for the average crossover time from any metastable state to the set of states with lower free energy. Theorem 1.2 shows that asymptotically the crossover time is exponential on the scale of its mean, a property that is standard for metastable behaviour.

**Theorem 1.1** (Average crossover time with coupling disorder). *Let  $\mathbb{A}_n(\cdot)$  be the  $k \times k$  Hessian matrix defined in (4.2) below, and  $\gamma_n$  the unique negative solution of the equation in (4.20) below. For every  $\mathbf{m}_n \in \mathcal{M}_n$  satisfying Hypothesis 1 and within the metastable regime (1.15), uniformly in  $\sigma \in S_n[\mathbf{m}_n]$ , and with  $\mathcal{P}$ -probability tending to 1,*

$$\mathbb{E}_\sigma [\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]}] = [1 + o_n(1)] \sqrt{\frac{-\det(\mathbb{A}_n(\mathbf{t}_n))}{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left( \frac{\pi}{2\beta(-\gamma_n)} \right) e^{\beta n[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)]}. \quad (1.19)$$

**Theorem 1.2** (Exponential law with coupling disorder). *For every  $\mathbf{m}_n \in \mathcal{M}_n$  satisfying Hypothesis 1 and within the metastable regime (1.15), uniformly in  $\sigma \in S_n[\mathbf{m}_n]$  and with  $\mathcal{P}$ -probability tending to 1,*

$$\mathbb{P}_\sigma (\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]} > t \mid \mathbb{E}_\sigma [\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]}]) = [1 + o_n(1)] e^{-t}, \quad t \geq 0. \quad (1.20)$$

As the average crossover time estimated in Theorem 1.1 is a random variable, we next provide more information on the randomness of the quantity in the right-hand side of (1.19), which depends on the realisation of the random variable  $J$ . The prefactor in (1.19) converges with  $\mathcal{P}$ -probability tending to 1 to a deterministic limit, which depends on the law of  $J$  but not on the realisation of  $J$ . However, the exponent does not converge to a deterministic limit. In Theorem 1.3 we compute the exponent up to order  $O(1)$ . Recall that  $F_n \rightarrow F_{\beta,h}$ ,  $\mathbf{m}_n \rightarrow \mathbf{m}$  and  $\mathbf{t}_n \rightarrow \mathbf{t}$  as  $n \rightarrow \infty$ .

**Theorem 1.3** (Randomness of the exponent) *For every  $\mathbf{m}_n \in \mathcal{M}_n$  satisfying Hypothesis 1 and within the metastable regime (1.15), in distribution,*

$$n[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] = n[F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] + Z\sqrt{n} + O(1), \quad (1.21)$$

where  $Z$  is a normal random variable with mean zero and variance in  $(0, \infty)$ , defined on  $(\Omega, \mathcal{F}, \mathcal{P})$  and independent of  $J$ .

The variance of  $Z$  turns out to be a complicated function of  $\beta$ ,  $h$  and the distribution of  $J$ . We refer to Sect. 6.3 for further details. Computing the exponent up to order 1 is in principle possible, but the formulas become rather complicated. Without this precision the prefactor in (1.19) is asymptotically negligible. Still, knowing this prefactor allows us to determine what the leading order behaviour of the randomness is.

*1.4. Discussion on the continuous case.* Bianchi et al. [1,2] study the Curie–Weiss model with a *random magnetic field* whose distribution is continuous. Lumping techniques work for discrete distributions but not for continuous distributions. The latter require coarse-graining techniques to approximate the continuous distribution by a sequence of discrete distributions. In the present paper we consider pair interaction random variables with a discrete distribution only. It seems hard to obtain results with a similar precision for continuous distributions. The techniques employed in [1,2] do not carry over, because the error introduced by the coarse-graining turns out to be quadratic rather than linear.

*1.5. Techniques and outline.* In order to prove Theorems 1.1–1.3 we use the potential-theoretic approach to metastability developed in Bovier et al. [4,5]. More specifically, we first find a sharp approximation of the Dirichlet form associated with the coarse-grained dynamics. We use these results, together with lumpability properties and well-known variational principles, to obtain sharp capacity estimates that are key quantities in the proof. For a more detailed overview of the methods, we refer the reader to the monograph by Bovier and den Hollander [6].

The remainder of the paper is organised as follows. Section 2 provides quantities and notations that are needed throughout the paper. Section 3 identifies the metastable regime. Section 4 provides a sharp approximation of the Dirichlet form associated with the Glauber dynamics in the presence of the disorder. Section 5 provides estimates on capacity and on the metastable valley measure. Section 6 proves Theorems 1.1–1.3. Appendix A contains a brief overview on known results for the standard CW model, which corresponds to the setting without disorder. Appendix B gives numerical evidence for the presence of multiple metastable states for suitable choices of  $\beta$ ,  $h$  and of the law of the components of  $J$ . Appendix C contains an example in which  $\beta \mapsto h_c(\beta)$  is not increasing, implying the possibility of a re-entrant metastable crossover. Appendix D provides the limit as  $n \rightarrow \infty$  of the prefactor in (1.19).

## 2. Preparations

Section 2.1 introduces further notation and writes the Hamiltonian in terms of the level magnetisations. Section 2.2 introduces the Dirichlet form associated with the Glauber dynamics and rewrites this in terms of the level magnetisations. Section 2.3 computes gradients and Hessians of the free energy as a function of the level magnetisations. Section 2.4 closes with an approximation of the free energy that will be needed later on.

2.1. *Hamiltonian.* Recall (1.7). Abbreviate

$$\omega_{\ell,n} = \frac{|A_{\ell,n}|}{n}. \quad (2.1)$$

Since, by the law of large numbers,  $(\omega_{\ell,n})_{\ell \in [k]} \rightarrow (\omega_\ell)_{\ell \in [k]} \in (0, \infty)^k$  as  $n \rightarrow \infty$  with  $\mathcal{P}$ -probability tending to 1, we may and will assume that  $A_{\ell,n} \neq \emptyset$  for all  $\ell \in [k]$  and all  $n$  large enough. Recall (1.8)–(1.9). Note that  $m_{\ell,n}(\sigma)$  takes values in the set

$$\Gamma_{\ell,n} = \left\{ -1, -1 + \frac{2}{|A_{\ell,n}|}, \dots, 1 - \frac{2}{|A_{\ell,n}|}, 1 \right\}. \quad (2.2)$$

Hence  $m_n(\sigma)$  takes values in the set

$$\Gamma_n = \bigtimes_{\ell \in [k]} \Gamma_{\ell,n}. \quad (2.3)$$

The configurations corresponding to  $M \subseteq \Gamma_n$  are denoted by

$$S_n[M] = \{\sigma \in S_n : m_n(\sigma) \in M\}. \quad (2.4)$$

For singletons  $M = \{m\}$  we write  $S_n[m]$  instead of  $S_n[\{m\}]$ .

Let

$$H_n(\sigma) = -\frac{1}{2n} \sum_{i,j \in [n]} J(i)J(j) \sigma(i)\sigma(j) - h \sum_{i \in [n]} \sigma(i), \quad \sigma \in S_n, \quad (2.5)$$

which is the Hamiltonian in (1.2), except for the diagonal term  $-\frac{1}{2n} \sum_{i \in [n]} J^2(i)$ , which is a constant shift. Using the notation above, we can write the Hamiltonian in (2.5) as

$$H_n(\sigma) = -n \left[ \frac{1}{2} \left( \sum_{\ell \in [k]} a_\ell \omega_{\ell,n} m_{\ell,n}(\sigma) \right)^2 + h \sum_{\ell \in [k]} \omega_{\ell,n} m_{\ell,n}(\sigma) \right] = n E_n(m_n(\sigma)), \quad (2.6)$$

where we abbreviate

$$E_n(m) = -\frac{1}{2} \left( \sum_{\ell \in [k]} a_\ell \omega_{\ell,n} m_\ell \right)^2 - h \sum_{\ell \in [k]} \omega_{\ell,n} m_\ell, \quad m = (m_\ell)_{\ell \in [k]} \in \Gamma_n. \quad (2.7)$$

2.2. *Dirichlet form and mesoscopic dynamics.* By (1.3)–(1.4), the Dirichlet form associated with the Glauber dynamics equals

$$\begin{aligned} \mathcal{E}_{S_n}(h, h) &= \frac{1}{2} \sum_{\sigma, \sigma' \in S_n} \mu_n(\sigma) r_n(\sigma, \sigma') [h(\sigma) - h(\sigma')]^2 \\ &= \frac{1}{2Z_n} \sum_{\sigma \in S_n} \sum_{\substack{\sigma' \in S_n, \\ \sigma' \sim \sigma}} e^{-\beta H_n(\sigma)} e^{-\beta [H_n(\sigma') - H_n(\sigma)]_+} [h(\sigma) - h(\sigma')]^2, \end{aligned} \quad (2.8)$$

where  $h$  is a test function on  $S_n$  taking values in  $[0, 1]$ . Because of (2.6), for any  $h$  such that  $h(\sigma) = \bar{h}(m_n(\sigma))$ , with  $\bar{h}$  a test function on  $\Gamma_n$ , we have

$$\begin{aligned} \mathcal{E}_{S_n}(h, h) &= \frac{1}{2Z_n} \sum_{m \in \Gamma_n} \sum_{m' \in \Gamma_n} e^{-\beta n E_n(m)} e^{-\beta n [E_n(m') - E_n(m)]_+} [\bar{h}(m) - \bar{h}(m')]^2 \\ &\times \sum_{\substack{\sigma \in S_n, \\ m_n(\sigma) = m}} \sum_{\substack{\sigma' \in S_n, \sigma' \sim \sigma, \\ m_n(\sigma') = m'}} 1, \end{aligned} \quad (2.9)$$

where  $m = (m_\ell)_{\ell \in [k]}$ . If  $\sigma' \sim \sigma$ , then  $\sigma' = \sigma^i$  for some  $i \in [n]$ , with  $\sigma^i$  obtained from  $\sigma$  by flipping the spin with label  $i$ . Let  $\ell' \in [k]$  be such that  $i \in A_{\ell', n}$ . If  $\sigma(i) = \pm 1 = -\sigma^i(i)$ , then

$$m_{\ell, n}(\sigma^i) = \begin{cases} m_{\ell', n}(\sigma) \mp \frac{2}{|A_{\ell', n}|}, & \ell = \ell', \\ m_{\ell, n}(\sigma), & \ell \neq \ell'. \end{cases} \quad (2.10)$$

For  $m, m' \in \Gamma_n$ , we write  $m \sim m'$  when there exists an  $\ell' \in [k]$  such that  $m' = m^{\ell', +}$  or  $m' = m^{\ell', -}$ , where

$$m_{\ell}^{\ell', \pm} = \begin{cases} m_{\ell'} \pm \frac{2}{|A_{\ell', n}|}, & \ell = \ell', \\ m_{\ell}, & \ell \neq \ell'. \end{cases} \quad (2.11)$$

Moreover, for  $\ell \in [k]$  and  $\sigma \in S_n$  with  $m_n(\sigma) = m$ , the cardinality of the set  $\{\sigma' \in S_n : \sigma' \sim \sigma, m_n(\sigma') = m^{\ell, \pm}\}$  equals  $\frac{1 \mp m_\ell}{2} |A_{\ell, n}|$ , namely, the number of  $(\mp 1)$ -spins in  $\sigma$  with index in  $A_{\ell, n}$ . Furthermore,

$$|\{\sigma \in S_n : m_n(\sigma) = m\}| = \prod_{\ell \in [k]} \left( \frac{|A_{\ell, n}|}{\frac{1+m_\ell}{2} |A_{\ell, n}|} \right), \quad m \in \Gamma_n, \quad (2.12)$$

as is seen by counting the number of  $(-1)$ -spins with label in  $A_{\ell, n}$  of a configuration with  $\ell$ -th level magnetisation  $m_\ell$ . Using these observations, we can rewrite (2.9) as

$$\begin{aligned} \mathcal{E}_{S_n}(h, h) &= \frac{1}{2Z_n} \sum_{m \in \Gamma_n} e^{-\beta n E_n(m)} \sum_{m' \in \Gamma_n} e^{-\beta n [E_n(m') - E_n(m)]_+} [\bar{h}(m) - \bar{h}(m')]^2 \\ &\times \prod_{\ell \in [k]} \left( \frac{|A_{\ell, n}|}{\frac{1+m_\ell}{2} |A_{\ell, n}|} \right) \sum_{\ell \in [k]} |A_{\ell, n}| \left[ \frac{1-m_\ell}{2} \mathbb{1}(m' = m^{\ell, +}) + \frac{1+m_\ell}{2} \mathbb{1}(m' = m^{\ell, -}) \right]. \end{aligned} \quad (2.13)$$

Next, abbreviate

$$I_n(m) = -\frac{1}{n} \log \left[ \prod_{\ell \in [k]} \left( \frac{|A_{\ell, n}|}{\frac{1+m_\ell}{2} |A_{\ell, n}|} \right) \right], \quad m \in \Gamma_n, \quad (2.14)$$

and put

$$F_n(m) = E_n(m) + \frac{1}{\beta} I_n(m) = -\frac{1}{2} \left( \sum_{\ell \in [k]} a_\ell \omega_{\ell,n} m_\ell \right)^2 - h \sum_{\ell \in [k]} \omega_{\ell,n} m_\ell + \frac{1}{\beta} I_n(m), \quad m \in \Gamma_n, \quad (2.15)$$

where  $E_n(m)$  is defined in (2.7). Moreover, define

$$\bar{r}_n(m, m') = e^{-\beta n [E_n(m') - E_n(m)]_+} \sum_{\ell \in [k]} |A_{\ell,n}| \left[ \frac{1 - m_\ell}{2} \mathbb{1}(m' = m^{\ell,+}) + \frac{1 + m_\ell}{2} \mathbb{1}(m' = m^{\ell,-}) \right]. \quad (2.16)$$

With this notation, we can write the *mesoscopic measure*  $Q_n(\cdot) = \mu_n \circ m_n^{-1}(\cdot)$  on  $\Gamma_n$ , with  $\mu_n$  defined in (1.4), as

$$Q_n(m) = \mu_n(S_n[m]) = \frac{1}{Z_n} e^{-\beta n F_n(m)}, \quad m \in \Gamma_n, \quad (2.17)$$

and so the Dirichlet form in (2.13) becomes

$$\mathcal{E}_{S_n}(h, h) = \frac{1}{2} \sum_{m \in \Gamma_n} Q_n(m) \sum_{m' \in \Gamma_n} \bar{r}_n(m, m') [\bar{h}(m) - \bar{h}(m')]^2. \quad (2.18)$$

**2.3. Gradients and Hessians.** Denote the Cramér entropy by

$$I_C(x) = \frac{1-x}{2} \log \left( \frac{1-x}{2} \right) + \frac{1+x}{2} \log \left( \frac{1+x}{2} \right). \quad (2.19)$$

Define

$$\bar{I}_n(m) = \sum_{\ell \in [k]} \omega_{\ell,n} I_C(m_\ell). \quad (2.20)$$

Since  $|A_{\ell,n}| = [1 + o_n(1)] \omega_{\ell,n}$ , we can use Stirling's formula  $N! = [1 + o_N(1)] N^N e^{-N} \sqrt{2\pi N}$  to obtain

$$I_n(m) = \bar{I}_n(m) + \sum_{\ell \in [k]} \frac{1}{2n} \log \left( \frac{\pi(1 - m_\ell^2) |A_{\ell,n}|}{2} \right) + o(n^{-1}) = \bar{I}_n(m) + O(n^{-1} \log n), \quad (2.21)$$

where the error term is *uniform* in  $m \in \Gamma_n$ . For  $\ell, \bar{\ell} \in [k]$ , we compute

$$\frac{\partial \bar{I}_n(m)}{\partial m_\ell} = \frac{\omega_{\ell,n}}{2} \log \left( \frac{1 + m_\ell}{1 - m_\ell} \right) \quad (2.22)$$

and

$$\begin{aligned} \frac{\partial^2 \bar{I}_n(m)}{\partial m_\ell \partial m_{\bar{\ell}}} &= 0, \quad \ell \neq \bar{\ell}, \\ \frac{\partial^2 \bar{I}_n(m)}{\partial m_\ell^2} &= \frac{\omega_{\ell,n}}{1 - m_\ell^2}. \end{aligned} \quad (2.23)$$

Recalling (2.7), we compute

$$\frac{\partial E_n(m)}{\partial m_\ell} = -a_\ell \omega_{\ell,n} \left( \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} m_{\ell'} \right) - \omega_{\ell,n} h. \quad (2.24)$$

Define

$$\bar{F}_n(m) = E_n(m) + \frac{1}{\beta} \bar{I}_n(m) = -\frac{1}{2} \left( \sum_{\ell \in [k]} a_\ell \omega_{\ell,n} m_\ell \right)^2 - h \sum_{\ell \in [k]} \omega_{\ell,n} m_\ell + \frac{1}{\beta} \bar{I}_n(m). \quad (2.25)$$

*Remark 2.1.* By (2.21),  $F_n(m) = \bar{F}_n(m) + O(n^{-1} \log n)$ , where  $F_n$  is defined in (2.15).♠

For  $m \in [-1, 1]^k$ , define

$$F_{\beta,h}(m) = -\frac{1}{2} \left( \sum_{\ell \in [k]} a_\ell \omega_\ell m_\ell \right)^2 - h \sum_{\ell \in [k]} \omega_\ell m_\ell + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell I_{\mathbf{C}}(m_\ell), \quad (2.26)$$

which corresponds to the uniform limit in probability of  $F_n$  as  $n \rightarrow \infty$ . Compute

$$\frac{\partial \bar{F}_n(m)}{\partial m_\ell} = \omega_{\ell,n} \left[ \frac{1}{2\beta} \log \left( \frac{1+m_\ell}{1-m_\ell} \right) - a_\ell \left( \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} m_{\ell'} \right) - h \right] \quad (2.27)$$

and

$$\begin{aligned} \frac{\partial^2 \bar{F}_n(m)}{\partial m_\ell \partial m_{\ell'}} &= -a_\ell \omega_{\ell,n} a_{\ell'} \omega_{\ell',n}, \quad \ell \neq \ell', \\ \frac{\partial^2 \bar{F}_n(m)}{\partial m_\ell^2} &= \frac{\omega_{\ell,n}}{\beta} \frac{1}{1-m_\ell^2} - a_\ell^2 \omega_{\ell,n}^2. \end{aligned} \quad (2.28)$$

The same formulas apply for  $I_n, F_n$ , with an error term  $O(n^{-1})$ .

**2.4. Additional computation.** We conclude with a computation that will be useful later on. Recalling (2.11), we write

$$\begin{aligned} & n[\bar{I}_n(m^{\ell,\pm}) - \bar{I}_n(m)] \\ &= n \omega_{\ell,n} \left[ \frac{1+m_\ell}{2} \log \left( 1 \pm \frac{2}{|A_{\ell,n}|(1+m_\ell)} \right) + \frac{1-m_\ell}{2} \log \left( 1 \mp \frac{2}{|A_{\ell,n}|(1-m_\ell)} \right) \right. \\ & \quad \left. \pm \frac{1}{|A_{\ell,n}|} A_{\ell,n}^\pm \right] \\ &= n \omega_{\ell,n} \left[ \pm \frac{1}{|A_{\ell,n}|} \mp \frac{1}{|A_{\ell,n}|} + O(n^{-2}) \pm \frac{1}{|A_{\ell,n}|} \Delta_{\ell,n}^\pm \right] = \Delta_{\ell,n}^\pm + O(n^{-1}), \end{aligned} \quad (2.29)$$

where

$$\Delta_{\ell,n}^{\pm} = \log \left( 1 + \frac{2m_{\ell} \pm \frac{4}{|A_{\ell,n}|}}{1 - m_{\ell} \mp \frac{2}{|A_{\ell,n}|}} \right). \quad (2.30)$$

The same formula applies for  $I_n$  with an error term of order  $O(n^{-1})$ , and hence

$$n[I_n(m^{\ell,\pm}) - I_n(m)] = \Delta_{\ell,n}^{\pm} + O(n^{-1}). \quad (2.31)$$

Note that  $\Delta_{\ell,n}^{\pm} = O(1)$ . Therefore, using (2.15), we get

$$\begin{aligned} n[E_n(m^{\ell,\pm}) - E_n(m)] &= n[F_n(m^{\ell,\pm}) - F_n(m)] - \frac{1}{\beta} n[I_n(m^{\ell,\pm}) - I_n(m)] \\ &= n[F_n(m^{\ell,\pm}) - F_n(m)] - \frac{1}{\beta} \Delta_{\ell,n}^{\pm} + O(n^{-1}). \end{aligned} \quad (2.32)$$

### 3. Metastable Regime

Section 3.1 identifies the stationary points of  $\bar{F}_n$ . Section 3.2 identifies the metastable regime. Section 3.3 provides details on the 1-dimensional metastable landscape.

*3.1. Stationary points of  $\bar{F}_n$  and  $F_{\beta,h}$ .* By (2.27), the critical points  $m = (m_{\ell})_{\ell \in [k]}$  of  $\bar{F}_n$  solve the system of equations (with  $\omega_{\ell,n} \neq 0$ )

$$0 = \frac{\partial \bar{F}_n(m)}{\partial m_{\ell}} = \omega_{\ell,n} \left[ \frac{1}{2\beta} \log \left( \frac{1+m_{\ell}}{1-m_{\ell}} \right) - a_{\ell} \left( \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} m_{\ell'} \right) - h \right], \quad \ell \in [k]. \quad (3.1)$$

Hence

$$\frac{1}{2} \log \left( \frac{1+m_{\ell}}{1-m_{\ell}} \right) = \beta \left[ a_{\ell} \left( \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} m_{\ell'} \right) + h \right]. \quad (3.2)$$

Since  $\operatorname{arctanh} x = \frac{1}{2} \log \frac{1+x}{1-x}$ ,  $x \in (-1, +1)$ , (3.2) can be rewritten as

$$m_{\ell} = \tanh \left( \beta \left[ a_{\ell} \left( \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} m_{\ell'} \right) + h \right] \right), \quad \ell \in [k]. \quad (3.3)$$

Similarly, the critical points  $m = (m_{\ell})_{\ell \in [k]}$  of  $F_{\beta,h}$  solve the deterministic equation

$$m_{\ell} = \tanh \left( \beta \left[ a_{\ell} \left( \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} m_{\ell'} \right) + h \right] \right), \quad \ell \in [k]. \quad (3.4)$$

Note that this can also be obtained directly from (3.3) after replacing  $\omega_{\ell,n}$  by its mean value  $\omega_{\ell}$ .

**3.2. Metastable regime.** We are interested in identifying the metastable regime, i.e., the set of pairs  $(\beta, h)$  for which  $F_{\beta,h}$  has more than one minimum. Put

$$K = K(m) = \sum_{\ell \in [k]} a_\ell \omega_\ell m_\ell. \quad (3.5)$$

From the characterisation of the critical points of  $F_{\beta,h}$  in (3.4) it follows that

$$K = T_{\beta,h}(K), \quad T_{\beta,h}(K) = \sum_{\ell \in [k]} a_\ell \omega_\ell \tanh(\beta[a_\ell K + h]). \quad (3.6)$$

Note that any critical point  $m = (m_\ell)_{\ell \in [k]} \in [-1, 1]^k$  of  $F_{\beta,h}$  is uniquely determined by  $K(m) \in \mathbb{R}$ . Consequently, the problem of solving the  $k$ -dimensional system in (3.4) can be reduced to solving the 1-dimensional Eq. (3.6). Recalling Hypothesis 1(2), the system is in the metastable regime if and only if (3.6) has more than one solution that is not tangent to the diagonal.

Compute

$$\begin{aligned} T'_{\beta,h}(K) &= \beta \sum_{\ell \in [k]} a_\ell^2 \omega_\ell (1 - \tanh^2(\beta[a_\ell K + h])), \\ T''_{\beta,h}(K) &= -2\beta^2 \sum_{\ell \in [k]} a_\ell^3 \omega_\ell \tanh(\beta[a_\ell K + h]) (1 - \tanh^2(\beta[a_\ell K + h])). \end{aligned} \quad (3.7)$$

For  $h = 0$ , the system is metastable when

$$\beta > \frac{1}{\sum_{\ell \in [k]} a_\ell^2 \omega_\ell}, \quad (3.8)$$

in which case  $T_{\beta,h}$  has a unique inflection point at  $K = 0$ , implying that (3.6) has three solutions  $K \in \{-K^*, 0, +K^*\}$  with  $K^* > 0$ . Otherwise (3.6) has only one solution  $K = 0$ .

We proceed with the more interesting case  $h > 0$ .

### 3.2.1. Number of solutions

**Lemma 3.1** (Number of solutions). *For  $h > 0$ , the number of critical points of  $F_{\beta,h}$ , i.e., solutions of (3.6), varies in  $\{1, 3, \dots, 2\ell + 1\}$ , where  $\ell \in [k]$  and  $2\ell - 1$  is the number of inflection points of  $T_{\beta,h}$ .*

*Proof.* For  $h > 0$  and  $K$  positive and large enough,  $T'_{\beta,h}(K) < 0$ . Moreover, for  $h > 0$  and  $K$  negative with  $|K|$  large enough,  $T'_{\beta,h}(K) > 0$ . Therefore,  $T_{\beta,h}$  has at least one inflection point and the number of inflection points of  $T_{\beta,h}$  cannot be even: it takes values in  $\{1, 3, \dots, 2k - 1\}$  depending on  $\beta, h$  and on the law of the components of  $J$ . Consequently, if  $2\ell - 1$  ( $\ell \in [k]$ ) is the number of inflection points, then the cardinality of the solutions of (3.6) takes values in  $\{1, 3, \dots, 2\ell + 1\}$  depending on  $\beta, h$  and on the law of the components of  $J$ .  $\square$

We conjecture that for any finite  $k$  there exist  $\beta, h$  and a law of the components of  $J$  such that (3.6) has any number of solutions in the set  $\{1, 3, \dots, 2k + 1\}$ . We found numerical evidence for this fact for  $k \in \{2, 3, 4\}$ . See Appendix B.

**Lemma 3.2** (Unique strictly positive solution). *For every  $\beta > 0$  and  $h > 0$ , (3.6) has exactly one strictly positive solution.*

*Proof.* Put  $W(K) = T_{\beta,h}(K) - K$ . The solutions of (3.6) are the roots of  $W$ . Clearly,  $W(0) > 0$ . Moreover,  $\lim_{K \rightarrow \infty} W(K) = -\infty$  because  $\lim_{K \rightarrow \infty} T_{\beta,h}(K) = \sum_{\ell \in [k]} a_\ell \omega_\ell > 0$  is finite. Therefore, by continuity, a root of  $W(K)$  exists in  $(0, \infty)$ .

Let  $\tilde{K}$  be the smallest positive root of  $W$ . Next we will prove that this root is unique. Indeed,  $W(K)' < 0$  when  $K \in [0, \infty)$ , meaning that  $K \mapsto W(K)'$  is strictly decreasing. By continuity, since  $W(K) > 0$  for all  $K \in [0, \tilde{K})$ , we have  $W(\tilde{K})' \leq 0$  and  $\lim_{K \rightarrow \infty} W(K)' = -1$ . Therefore,  $W(K)' < 0$  for all  $K \in (\tilde{K}, \infty)$ , and so  $W$  is strictly decreasing. Moreover,  $W(K) < W(\tilde{K}) = 0$  for all  $K \in (\tilde{K}, \infty)$ . Thus,  $\tilde{K}$  is the only positive root of  $W$ .  $\square$

### 3.2.2. Metastable regime

**Lemma 3.3** (Characterisation of the metastable regime). *Equation (3.6) has at least three solutions not tangent to the diagonal if and only if there exists  $\bar{K} < 0$  such that  $\bar{K} > T_{\beta,h}(\bar{K})$ , i.e.,*

$$\bar{K} > \sum_{\ell \in [k]} a_\ell \omega_\ell \tanh(\beta[a_\ell \bar{K} + h]). \quad (3.9)$$

*Proof.* Using Lemma 3.2, we see that (3.6) has at least three solutions if and only if it has at least two strictly negative solutions. As above, we define  $W(K) = T_{\beta,h}(K) - K$ . The solutions of (3.6) are the roots of  $W$ . Now, assume that there exists a  $\bar{K} < 0$  such that  $\bar{K} > T_{\beta,h}(\bar{K})$ . Since  $W(\bar{K}) < 0$  and  $W(0) > 0$ ,  $W(K)$  has a root in  $(\bar{K}, 0)$ , implying that (3.6) has at least one solution in  $(\bar{K}, 0)$ . Moreover, since  $\lim_{K \rightarrow -\infty} T_{\beta,h}(K) = -\sum_{\ell \in [k]} a_\ell \omega_\ell$  is finite, we have  $\lim_{K \rightarrow -\infty} W(K) = \infty$ . Because  $W(\bar{K}) < 0$ , it follows that  $W$  has at least one root in  $(-\infty, \bar{K})$ . With the same argument it can be shown that the negative roots of  $W$  are always even. The opposite implication is trivial.  $\square$

*Remark 3.1.* Applying the intermediate value theorem to the derivative of  $W(K) = T_{\beta,h}(K) - K$ , we get that if the condition in Lemma 3.3 is satisfied, then there exists a  $\bar{K} < 0$  such that  $T'_{\beta,h}(\bar{K}) = 1$  and  $\bar{K} > T_{\beta,h}(\bar{K})$ .  $\spadesuit$

**Theorem 3.1** (Metastable regime). *Define, as in (1.14),*

$$\beta_c = \frac{1}{\sum_{\ell \in [k]} a_\ell^2 \omega_\ell}. \quad (3.10)$$

*The metastable regime is*

$$\beta \in (\beta_c, \infty), \quad h \in [0, h_c(\beta)), \quad (3.11)$$

*with  $\beta \mapsto \beta h_c(\beta)$  non-decreasing on  $[\beta_c, \infty)$ . Furthermore, if the support of the law of the components of  $J$  is put into increasing order, i.e.,  $a_1 < a_2 < \dots < a_k$ , then*

$$\lim_{\beta \rightarrow \infty} h_c(\beta) = \min_{\ell \in [k]^*} \left( \sum_{\ell'=\ell}^k a_\ell a_{\ell'} \omega_{\ell'} - \sum_{\ell'=1}^{\ell-1} a_\ell a_{\ell'} \omega_{\ell'} \right), \quad (3.12)$$

*where the minimum is over all  $\ell \in [k]$  such that the quantity between brackets is positive.*

*Proof.* Recalling Lemma 3.3, we look for conditions for the existence of a  $K < 0$  satisfying (3.9). If such a  $K$  exists, then by Remark 3.1 there exists a  $\bar{K} < 0$  satisfying (3.9) such that  $T'_{\beta,h}(\bar{K}) = 1$ , which reads

$$\sum_{\ell \in [k]} a_\ell^2 \omega_\ell \tanh^2(\beta[a_\ell \bar{K} + h]) = \sum_{\ell \in [k]} a_\ell^2 \omega_\ell - \frac{1}{\beta}. \quad (3.13)$$

Since the left-hand side of (3.13) is positive, it admits solutions only if

$$\frac{1}{\beta} < \sum_{\ell \in [k]} a_\ell^2 \omega_\ell = \frac{1}{\beta_c}. \quad (3.14)$$

Therefore, (3.14) is a necessary condition for the metastable regime.

Now assume (3.14). Since  $\tanh x \sim x$ ,  $x \rightarrow 0$ , for  $|K| \ll \beta(\max_{\ell \in [k]} a_\ell)^{-1}$  and  $h \downarrow 0$ , we have

$$K = T_{\beta,h}(K) = \sum_{\ell \in [k]} a_\ell \omega_\ell \tanh(\beta[a_\ell K + h]) \sim \sum_{\ell \in [k]} a_\ell \omega_\ell \beta[a_\ell K + h], \quad (3.15)$$

which reads

$$K \sim - \left( \sum_{\ell \in [k]} a_\ell \omega_\ell \right) \left( \frac{1}{\beta_c} - \frac{1}{\beta} \right)^{-1} h \quad (3.16)$$

and proves the existence of a negative solution. A positive solution is guaranteed by Lemma 3.2. The existence of a third (strictly negative) solution of (3.4), for every  $\beta > \beta_c$  and for  $h \downarrow 0$ , follows as in the proof of Lemma 3.3. Therefore, the lower bound on  $\beta_c$  is sharp.

Since  $h \mapsto T_{\beta,h}(K)$  is strictly increasing for every fixed  $\beta > 0$  and  $K \in \mathbb{R}$ , there exists a unique critical curve  $\beta \mapsto h_c(\beta)$  such that the system is metastable for  $0 \leq h < h_c(\beta)$  and not metastable for  $h \geq h_c(\beta)$ . We know that  $h_c(\beta) > 0$  for  $\beta > \beta_c$ . By passing to the parametrisation  $g = h\beta$ , we get that  $\beta \mapsto T_{\beta,g}(K)$  is strictly decreasing for every  $g$  and for every  $K < 0$ , from which it follows that  $\beta \mapsto g_c(\beta) = \beta h_c(\beta)$  is non-decreasing.

We next focus on the limit of  $h_c(\beta)$  as  $\beta \rightarrow \infty$ . By Lemma 3.3, we may focus on the existence of  $\bar{K}$  satisfying (3.9). In the limit as  $\beta \rightarrow \infty$ ,  $\tanh(\beta[a_\ell \bar{K} + h]) \rightarrow 2\Theta_{-h/a_\ell}(\bar{K}) - 1$ , where  $\Theta_x(\cdot)$  is the Heaviside function centred in  $x$ . Thus, for all  $\ell \in [k+1]$ ,

$$\lim_{\beta \rightarrow \infty} \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} \tanh(\beta[a_{\ell'} K + h]) = - \sum_{\ell'=\ell}^k a_{\ell'} \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \omega_{\ell'}, \quad K \in \left( -\frac{h}{a_{\ell-1}}, -\frac{h}{a_\ell} \right), \quad (3.17)$$

and, for all  $\ell \in [k]$ ,

$$\lim_{\beta \rightarrow \infty} \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} \tanh(\beta[a_{\ell'} K + h]) = - \sum_{\ell'=\ell+1}^k a_{\ell'} \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \omega_{\ell'}, \quad K = -\frac{h}{a_\ell}, \quad (3.18)$$

where we set  $-\frac{h}{a_0} = -\infty$  and  $-\frac{h}{a_{k+1}} = \infty$ . Thus, for  $\bar{K} \in \left(-\frac{h}{a_{\ell-1}}, -\frac{h}{a_\ell}\right)$ , (3.9) can be written as

$$\bar{K} > -\sum_{\ell'=\ell}^k a_{\ell'} \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \omega_{\ell'}. \quad (3.19)$$

Therefore, (3.9) has a solution if and only if there exists an  $\ell \in [k]$  such that

$$-\sum_{\ell'=\ell}^k a_{\ell'} \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \omega_{\ell'} < -\frac{h}{a_\ell}, \quad (3.20)$$

in which case a solution  $\bar{K}$  of (3.9) exists in

$$\left(-\sum_{\ell'=\ell}^k a_{\ell'} \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \omega_{\ell'}, -\frac{h}{a_\ell}\right). \quad (3.21)$$

Note that the quantity between brackets in (3.12) is always positive for  $\ell = 1$ . Thus, the minimum is always finite.

The proof is complete after we show why we may drop the case where  $\bar{K} = -\frac{h}{a_\ell}$  for some  $\ell \in [k]$ . In this case the condition for  $\bar{K}$  to satisfy (3.9) is

$$-\sum_{\ell'=\ell+1}^k a_{\ell'} \omega_{\ell'} + \sum_{\ell'=1}^{\ell-1} a_{\ell'} \omega_{\ell'} < -\frac{h}{a_\ell}, \quad (3.22)$$

which implies (3.20). Thus, if  $\bar{K} = -\frac{h}{a_\ell}$  satisfies (3.9), then also some other  $K$  in (3.21) satisfies (3.9). Therefore, the condition in (3.20) is equivalent to having metastability.  $\square$

**Lemma 3.4** (Re-entrant crossover). *The function  $\beta \mapsto h_c(\beta)$  is not necessarily non-decreasing.*

*Proof.* In Appendix C we provide an example of  $\beta \mapsto h_c(\beta)$  that is not increasing.  $\square$

### 3.2.3. Bounds on the inflection points and on the critical curve

**Lemma 3.5** (Bounds on inflection points). *All solutions of  $T''_{\beta,h}(K) = 0$  are contained in the interval*

$$\left[-\frac{h}{\min_{\ell \in [k]} a_\ell}, -\frac{h}{\max_{\ell \in [k]} a_\ell}\right]. \quad (3.23)$$

*In particular, they are all strictly negative.*

*Proof.* If  $K > -\frac{h}{\max_{\ell \in [k]} a_\ell}$ , then  $\tanh(\beta[a_\ell K + h]) > 0$  for all  $\ell \in [k]$ , which implies  $T''_{\beta,h}(K) < 0$ . If  $K < -\frac{h}{\min_{\ell \in [k]} a_\ell}$ , then  $\tanh(\beta[a_\ell K + h]) < 0$  for all  $\ell \in [k]$ , which implies  $T''_{\beta,h}(K) > 0$ .  $\square$

**Lemma 3.6** (Upper bound on  $h_c$ ).  $\sup_{\beta \in (\beta_c, \infty)} h_c(\beta) < (\max_{\ell \in [k]} a_\ell) \sum_{\ell \in [k]} a_\ell \omega_\ell$ .

*Proof.* Use Lemma 3.3 to characterise the metastable regime and Remark 3.1. We claim that if a solution  $\tilde{K}$  of (3.9) with  $T'_{\beta,h}(\tilde{K}) = 1$  exists, then it must be negative and strictly less than an inflection point. Using this fact, together with Lemma 3.5 and the inequality in (3.9), we obtain a necessary upper bound on  $h$ :

$$\sum_{\ell \in [k]} a_\ell \omega_\ell \tanh(\beta[a_\ell \tilde{K} + h]) < -\frac{h}{\max_{\ell \in [k]} a_\ell}. \quad (3.24)$$

Using that  $\tanh(\beta[a_\ell \tilde{K} + h]) > -1$ , we conclude the proof.

We are left to prove the claim. By Lemma 3.5, all inflection points are negative, and  $T''_{\beta,h}(K) < 0$  for  $K \geq 0$ . Assume, by contradiction, that  $T''_{\beta,h}(K) < 0$  for all  $K \in (\tilde{K}, \infty)$ . Then  $T'_{\beta,h}$  is strictly decreasing. Therefore,  $T'_{\beta,h}(K) < 1$  for all  $K \in (\tilde{K}, \infty)$ , which implies  $T_{\beta,h}(K) - T_{\beta,h}(0) < K$ . Since  $T_{\beta,h}(0) > 0$ , there exists a  $\tilde{K} \in (\tilde{K}, 0)$  such that  $T_{\beta,h}(\tilde{K}) > 0 > \tilde{K}$ . Thus,  $T_{\beta,h}(\tilde{K}) - T_{\beta,h}(0) > \tilde{K}$ , which contradicts what we have proved for all  $K \in (\tilde{K}, \infty)$ .  $\square$

**3.3. Quasi 1-dimensional landscape.** Given  $K \in \mathbb{R}$ , by standard saddle point approximation, the leading order of

$$-\frac{1}{\beta n} \log \mu_n(\{\sigma : K_n(m_n(\sigma)) = K\}) \quad (3.25)$$

turns out to be the function  $G_n : \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$G_n(K) = \inf_{m : K_n(m) = K} \tilde{F}_n(m). \quad (3.26)$$

Recalling definitions (2.25) and (3.5), using Lagrange multipliers and integrating the condition  $K_n(m) = K$ , we obtain

$$G_n(K) = -\frac{1}{2}K^2 - \frac{\log 2}{\beta} - \inf_{t \in \mathbb{R}} \left( Kt + \sum_{\ell \in [k]} \frac{\omega_{\ell,n}}{\beta} \log \cosh[\beta(h - ta_\ell)] \right). \quad (3.27)$$

**Lemma 3.7** (Alternative characterisation for the critical points).

1. If  $m^*$  is a (not maximal) critical point for  $F_n$ , then  $K_n(m^*)$  is a critical point for  $G_n$ .
2. If  $K$  is a critical point for  $G_n$ , then  $m^* = (m^*)_{\ell \in [k]}$  with  $m^*_\ell = \tanh(\beta[a_\ell K + h])$  (recall (3.3)) is a critical point for  $F_n$ .
3.  $F_n(m^*) = G_n(K_n(m^*))$  for any (not maximal) critical point  $m^*$ .

*Proof.* Similar to [3, Lemma 7.4].  $\square$

We have already seen that  $K_n(m)$  fully determines any critical value  $m$  of  $F_n$ , and is useful to order them. Lemma 3.7 exhibits the one-dimensional structure underlying the metastable landscape and provides a tool to describe the nature of the critical points of  $F_n$ .

**Remark 3.2.** The above results extend to the limit  $n \rightarrow \infty$ : replace  $F_n$  by  $F_{\beta,h}$  and  $G_n$  by  $G_{\beta,\ell}$ , obtained after replacing  $\omega_{\ell,n}$  by  $\omega_\ell$  in (3.27), and  $K_n(\cdot)$  by  $K(\cdot)$ .  $\spadesuit$

#### 4. Approximation of the Dirichlet form Near the Saddle Point

In this section we approximate the Dirichlet form associated with the coarse-grained dynamics near the saddle point. This is a key step to obtain capacity estimates in the following section. Further details and examples on the techniques we use here can be found in [6, Chapters 9, 10 and 14].

Section 4.1 introduces some key quantities that are needed to express the mesoscopic measure. Section 4.2 introduces an approximate mesoscopic measure that leads to an approximate dynamics. Section 4.3 approximates the harmonic functions associated with this dynamics. Section 4.4 computes an approximate Dirichlet form. Section 4.5 uses the latter to approximate the full Dirichlet form.

**4.1. Key quantities.** Let  $\mathbf{m}_n = (\mathbf{m}_{\ell,n})_{\ell \in [k]}$  and  $\mathbf{t}_n = (\mathbf{t}_{\ell,n})_{\ell \in [k]}$  in  $\Gamma_n$  be a local minimum of  $F_n$  and the correspondent saddle point, respectively, as defined in Sect. 1.3.4. Note that both  $\mathbf{m}_n$  and  $\mathbf{t}_n$  satisfy (3.3). Consider the neighbourhood of  $\mathbf{t}_n$  defined by

$$\mathcal{D}_n = \left\{ m \in \Gamma_n : d(m, \mathbf{t}_n) \leq C' n^{-1/2} \log^{1/2} n \right\}, \quad (4.1)$$

where  $d$  is the Euclidean distance and  $C' \in (0, \infty)$  is a constant. Abbreviate the Hessian of  $F_n$

$$\mathbb{A}_n(m) = (\nabla^2 F_n)(m), \quad m \in \Gamma_n, \quad (4.2)$$

and put

$$\mathbb{A}_n = \mathbb{A}_n(\mathbf{t}_n). \quad (4.3)$$

By (2.28),

$$\begin{aligned} (\mathbb{A}_n(m))_{\ell, \ell'} &= -a_\ell \omega_{\ell,n} a_{\ell'} \omega_{\ell',n} + O(n^{-1}), \quad \ell \neq \ell', \\ (\mathbb{A}_n(m))_{\ell, \ell} &= \frac{\omega_{\ell,n}}{\beta} \frac{1}{1 - m_\ell^2} - a_\ell^2 \omega_{\ell,n}^2 + O(n^{-1}) = \frac{1}{\beta} \frac{\partial^2 \bar{I}_n(m)}{\partial m_\ell^2} - a_\ell^2 \omega_{\ell,n}^2 + O(n^{-1}). \end{aligned} \quad (4.4)$$

Note that  $\mathbb{A}_n(m)$  is a diagonal matrix minus a rank one matrix. Compute

$$\det \mathbb{A}_n(m) = \left( 1 - \sum_{\ell \in [k]} \beta a_\ell^2 \omega_{\ell,n} [1 - m_\ell^2] \right) \prod_{\ell' \in [k]} \frac{1}{\beta} \frac{\omega_{\ell',n}}{1 - m_{\ell'}^2} [1 + O(n^{-1})]. \quad (4.5)$$

**4.2. Approximate dynamics and Dirichlet form.** For any two vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^k$ , let  $\langle \mathbf{v}, \mathbf{w} \rangle$  denote their scalar product. For any  $k \times k$  matrix  $\mathbf{M}$  and any  $\mathbf{v} \in \mathbb{R}^k$ , let  $\mathbf{M} \cdot \mathbf{v}$  denote their matrix product, as  $\mathbf{v}$  was in  $\mathbb{R}^k \times 1$ .

For  $m \in \mathcal{D}_n$ , define

$$\tilde{Q}_n(m) = \frac{1}{Z_n} \exp \left[ -\frac{\beta n}{2} \langle [m - \mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n] \rangle \right] \exp [-\beta n F_n(\mathbf{t}_n)], \quad (4.6)$$

and

$$\tilde{r}_n(m, m') = \begin{cases} \bar{r}_n(\mathbf{t}_n, \mathbf{t}_n^{\ell,+}), & m' = m^{\ell,+}, \\ \bar{r}_n(\mathbf{t}_n^{\ell,-}, \mathbf{t}_n) \frac{\tilde{Q}_n(m^{\ell,-})}{\tilde{Q}_n(m)}, & m' = m^{\ell,-}, \\ 0, & \text{else,} \end{cases} \quad (4.7)$$

where  $\bar{r}_n$  is defined in (2.16). The transition rates  $\tilde{r}_n$  define a random dynamics on  $\mathcal{D}_n$  that is reversible with respect to the mesoscopic measure  $\tilde{Q}_n$ . The corresponding *Dirichlet form* is

$$\tilde{\mathcal{E}}_{\mathcal{D}_n}(u, u) = \sum_{m \in \mathcal{D}_n} \tilde{Q}_n(m) \sum_{\ell \in [k]} \tilde{r}_n(m, m^{\ell,+}) \left[ u(m) - u(m^{\ell,+}) \right]^2, \quad (4.8)$$

where  $u$  is a test function on  $\mathcal{D}_n$ . Put

$$r_\ell = \tilde{r}_n(m, m^{\ell,+}) = \bar{r}_n(\mathbf{t}_n, \mathbf{t}_n^{\ell,+}). \quad (4.9)$$

Using (2.7) and (2.16), we get

$$r_\ell = |A_{\ell,n}| \frac{1 - \mathbf{t}_{\ell,n}}{2} \exp \left[ -2\beta \left( -h - a_\ell \left( \frac{a_\ell}{n} + \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} \mathbf{t}_{\ell',n} \right) \right) \right]_+. \quad (4.10)$$

**4.2.1. Approximation estimates** Next we estimate how close the pairs  $(\bar{r}_n, \tilde{r}_n)$  and  $(Q_n, \tilde{Q}_n)$  are. By Taylor expansion around  $\mathbf{t}_n$ , we have

$$F_n(m) - F_n(\mathbf{t}_n) = \frac{1}{2} [m - \mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n] + O(d(m, \mathbf{t}_n)^3). \quad (4.11)$$

In particular,

$$\begin{aligned} F_n(\mathbf{t}_n^{\ell,\pm}) - F_n(\mathbf{t}_n) &= \frac{1}{2} \frac{4}{|A_{\ell,n}|^2} (\mathbb{A}_n)_{\ell,\ell} + O(|A_{\ell,n}|^{-3}) \\ &= \frac{2}{n^2 \omega_{\ell,n}^2} \left[ \frac{\omega_{\ell,n}}{\beta} \frac{1}{1 - \mathbf{t}_{\ell,n}^2} - a_\ell^2 \omega_{\ell,n}^2 + o((n \omega_{\ell,n})^{-1}) \right] \\ &\quad + O((n \omega_{\ell,n})^{-3}) \\ &= \frac{2}{n^2} \left( \frac{1}{\beta \omega_{\ell,n} (1 - \mathbf{t}_{\ell,n}^2)} - a_\ell^2 \right) + O((n \omega_{\ell,n})^{-3}), \end{aligned} \quad (4.12)$$

where the second equality uses (4.4). Moreover, for  $m \in \mathcal{D}_n$  ( $\mathbf{e}_\ell$  is the unitary vector in  $\mathbb{R}^k$  whose  $\ell$ -th component is non-zero),

$$\begin{aligned} F_n(m^{\ell,\pm}) - F_n(m) &= \left\langle \left[ \pm \frac{2}{|A_{\ell,n}|} \mathbf{e}_\ell \right], \mathbb{A}_n \cdot [m - \mathbf{t}_n] \right\rangle + \frac{1}{2} \left\langle \left[ \pm \frac{2}{|A_{\ell,n}|} \mathbf{e}_\ell \right], \mathbb{A}_n \cdot \left[ \pm \frac{2}{|A_{\ell,n}|} \mathbf{e}_\ell \right] \right\rangle \\ &\quad + O(d(m, \mathbf{t}_n)^3) \end{aligned}$$

$$\begin{aligned}
&= \pm \frac{2}{|A_{\ell,n}|} \sum_{\ell' \in [k]} (\mathbb{A}_n)_{\ell,\ell'} (m_{\ell'} - \mathbf{t}_{\ell',n}) + \frac{2}{|A_{\ell,n}|^2} (\mathbb{A}_n)_{\ell,\ell} + O(d(m, \mathbf{t}_n)^3) \\
&= \left( \pm \frac{2}{n\omega_{\ell,n}} (m_\ell - \mathbf{t}_{\ell,n}) + \frac{2}{n^2\omega_{\ell,n}^2} \right) \left( \frac{\omega_{\ell,n}}{\beta} \frac{1}{1 - \mathbf{t}_{\ell,n}^2} - a_\ell^2 \omega_{\ell,n}^2 + o(n^{-1}) \right) \\
&\quad \pm \frac{2}{n\omega_{\ell,n}} \sum_{\ell' \in [k], \ell' \neq \ell} (-a_\ell \omega_{\ell,n} a_{\ell'} \omega_{\ell',n}) (m_{\ell'} - \mathbf{t}_{\ell',n}) + O(n^{-3/2} \log^{3/2} n) \\
&= \mp \frac{2}{n} \sum_{\ell' \in [k]} a_\ell a_{\ell'} \omega_{\ell',n} (m_{\ell'} - \mathbf{t}_{\ell',n}) \pm \frac{2(m_\ell - \mathbf{t}_{\ell,n})}{\beta n(1 - \mathbf{t}_{\ell,n}^2)} + O(n^{-3/2} \log^{3/2} n), \quad (4.13)
\end{aligned}$$

where the third equality uses (4.4). For  $m \in \mathcal{D}_n$ , we have  $d(m, \mathbf{t}_n)^3 = O(n^{-3/2} \log^{3/2} n)$ . Therefore, combining (2.17), (4.6) and (4.11), we have

$$\left| \frac{Q_n(m)}{\tilde{Q}_n(m)} - 1 \right| \leq C'' n^{-1/2} \log^{3/2} n, \quad m \in \mathcal{D}_n, \quad (4.14)$$

for some  $C'' \in (0, \infty)$  constant. Using (2.16) and (2.32), we can write

$$\bar{r}_n(m, m^{\ell,\pm}) = \exp \left[ -\beta \left[ n \left[ F_n(m^{\ell,\pm}) - F_n(m) \right] - \frac{1}{\beta} \Delta_{\ell,n}^\pm + O(n^{-1}) \right] \right] \frac{1 \mp m_\ell}{2}, \quad (4.15)$$

where  $\Delta_{\ell,n}^\pm$  is defined in (2.30).

Using (4.7), (4.12), (4.13) and (4.15), we find that, for all  $m \in \mathcal{D}_n$ ,

$$\begin{aligned}
\left| \frac{\bar{r}_n(m, m^{\ell,+})}{\bar{r}_n(m, m^{\ell,+})} - 1 \right| &= \left| \frac{\bar{r}_n(m, m^{\ell,+})}{\bar{r}_n(\mathbf{t}_n, \mathbf{t}_n^{\ell,+})} - 1 \right| \\
&= \left| \frac{(1 - m_\ell) \exp \left\{ - \left[ I_1 + O(n^{-1/2} \log^{3/2} n) - \Delta_{\ell,n}^\pm + o_n(1) \right]_+ \right\}}{(1 - \mathbf{t}_{\ell,n}) \exp \left\{ - \left[ I_2 + O(n^{-2} \omega_{\ell,n}^{-3}) - \Delta_{\ell,n}^\pm + o_n(1) \right]_+ \right\}} - 1 \right| \\
&= \left| \frac{(1 - m_\ell) \exp \left\{ - \left[ I_1 - \Delta_{\ell,n}^\pm + o_n(1) \right]_+ \right\}}{(1 - \mathbf{t}_{\ell,n}) \exp \left\{ - \left[ -\Delta_{\ell,n}^\pm + o_n(1) \right]_+ \right\}} - 1 \right| \\
&\leq C''' n^{-1/2} \log^{1/2} n, \quad (4.16)
\end{aligned}$$

where  $C''' \in (0, \infty)$  is a constant and we abbreviate

$$\begin{aligned}
I_1 &= -2\beta \sum_{\ell' \in [k]} a_\ell a_{\ell'} \omega_{\ell',n} (m_{\ell'} - \mathbf{t}_{\ell',n}) + \frac{2(m_\ell - \mathbf{t}_{\ell,n})}{1 - \mathbf{t}_{\ell,n}^2}, \\
I_2 &= \frac{2}{n} \left( \frac{1}{\omega_{\ell,n}(1 - \mathbf{t}_{\ell,n}^2)} - \beta a_\ell^2 \right). \quad (4.17)
\end{aligned}$$

Equations (4.14) and (4.16) are relevant for the following approximation.

4.3. *Approximate harmonic function.* Let  $\mathbb{B}_n$  be the  $k \times k$  matrix defined by

$$(\mathbb{B}_n)_{\ell\ell'} = \frac{\sqrt{r_\ell r_{\ell'}}}{n \omega_{\ell,n} \omega_{\ell',n}} (\mathbb{A}_n)_{\ell\ell'}, \quad (4.18)$$

where  $\mathbb{A}_n$  is defined in (4.3). Note that

$$\det \mathbb{B}_n = (\det \mathbb{A}_n) \prod_{\ell \in [k]} \frac{r_\ell}{n \omega_{\ell,n}^2}. \quad (4.19)$$

Let  $\gamma_n^{(\ell)}$ ,  $\ell \in [k]$ , be the eigenvalues of  $\mathbb{B}_n$ , ordered in increasing order. Let  $\gamma_n = \gamma_n^{(1)}$  denote the unique negative eigenvalue of  $\mathbb{B}_n$ , and  $\hat{v}$  the corresponding unitary eigenvector.

Define  $v = (v_\ell)_{\ell \in [k]}$  by  $v_\ell = \hat{v}_\ell \frac{\omega_{\ell,n} \sqrt{n}}{\sqrt{r_\ell}}$ .

*Remark 4.1.* As in [6, Remark 10.4], it follows by Hypothesis 1 that  $\mathbb{A}_n$  has all strictly positive eigenvalues but one strictly negative. It can be seen that the same property holds for the eigenvalues of  $\mathbb{B}_n$ . ♠

**Lemma 4.1** (Eigenvalue). *The eigenvalue  $\gamma_n$  is the unique solution of the equation*

$$\frac{1}{n} \sum_{\ell \in [k]} \frac{a_\ell^2}{\frac{1}{n\beta \omega_{\ell,n}(1-\mathbf{t}_{\ell,n}^2)} - \frac{\gamma_n}{r_\ell}} = 1 + O(n^{-1}). \quad (4.20)$$

*Proof.* We follow the line of proof of [6, Lemma 14.9], using the last point in Hypothesis 1. In our case, [6, Eq. (14.7.12)] reads

$$-\frac{1}{n} a_\ell \sqrt{r_\ell} \sum_{\ell' \in [k]} a_{\ell'} \sqrt{r_{\ell'}} u_{\ell'} + \left( r_\ell \frac{1}{n\beta \omega_{\ell,n}(1-\mathbf{t}_{\ell,n}^2)} - \gamma_n \right) u_\ell + O(n^{-1}) = 0, \quad \ell \in [k]. \quad (4.21)$$

□

*Remark 4.2.* As in [6, Lemma 14.9], since the left-hand side of (4.20) is increasing in  $\gamma_n$  for  $\gamma_n \geq 0$ , a negative solution of (4.20) exists if and only if

$$\beta \sum_{\ell \in [k]} a_\ell^2 \omega_{\ell,n}(1-\mathbf{t}_{\ell,n}^2) > 1. \quad (4.22)$$

Using (4.5), (4.22) holds if and only if  $\det \mathbb{A}_n < 0$ . By Remark 4.1 the latter holds true. ♠

Define  $f: \mathbb{R} \rightarrow [0, 1]$  as

$$f(x) = \sqrt{\frac{(-\gamma_n)\beta n}{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}(-\gamma_n)\beta n u^2} du \quad (4.23)$$

and  $g: \mathbb{R}^k \rightarrow [0, 1]$  as

$$g(m) = f(\langle v, m - \mathbf{t}_n \rangle). \quad (4.24)$$

Recall the definition of  $\mathcal{M}_n(\mathbf{m}_n)$  given in (1.17).

Let  $W_0$  be a strip in  $\Gamma_n$  of width  $Cn^{-1/2} \log^{1/2} n$  such that  $\mathbf{t}_n \in W_0$ ,  $\mathcal{M}_n(\mathbf{m}_n) \cap W_0$  is empty and  $W_0^c$  consists in two non-neighbouring parts:  $W_1$  containing  $\mathbf{m}_n$  and

$W_2$  containing  $\mathcal{M}_n(\mathbf{m}_n)$ . Moreover, we require that, for some fixed constant  $c > 1$ ,  $W_0 \cap \mathcal{D}_n^c \subseteq \{m \in \Gamma_n : F_n(m) > F_n(\mathbf{t}_n) + cn^{-1} \log n\}$ . Define

$$\tilde{g}(m) = \begin{cases} 0, & m \in W_1, \\ 1, & m \in W_2, \\ g(x), & m \in W_0 \cap \mathcal{D}_n, \\ 0, & m \in W_0 \cap \mathcal{D}_n^c. \end{cases} \quad (4.25)$$

By choosing  $W_0$  and  $\mathcal{D}_n$  suitably we have, for  $m \sim m'$  (i.e.,  $\bar{r}_n(m, m') > 0$ ) and  $c \in (0, \infty)$  large enough (coming from the definition of  $W_0$ ),

$$Q_n(m) \leq Q_n(\mathbf{t}_n)n^{-c\beta}, \quad m \in W_0 \cap \mathcal{D}_n^c, \quad (4.26)$$

$$(\tilde{g}(m) - \tilde{g}(m'))^2 \bar{r}_n(m, m') Q_n(m) \leq Q_n(\mathbf{t}_n)n^{-c\beta}, \quad m \in W_0 \cap \mathcal{D}_n, m' \in W_0^c. \quad (4.27)$$

**4.4. Computation of the approximate Dirichlet form.** In this section we follow [6, Sections 10.2.2–10.2.3] to approximate  $\tilde{\mathcal{E}}_{\mathcal{D}_n}(g, g)$  defined in (4.8). As in [6, Eq. (10.2.24)], for  $m \in \mathcal{D}_n$  and  $\ell \in [k]$  such that  $m^{\ell,+} \in D_n$ , compute

$$\begin{aligned} g(m^{\ell,+}) - g(m) &= \frac{2}{|A_{\ell,n}|} v_\ell f'(\langle v, m - \mathbf{t}_n \rangle) \\ &\quad + \frac{2}{|A_{\ell,n}|^2} v_\ell^2 f''(\langle v, m - \mathbf{t}_n \rangle) + \frac{4}{3|A_{\ell,n}|^3} v_\ell^3 f'''(\langle v, \tilde{m} - \mathbf{t}_n \rangle) \\ &= v_\ell \sqrt{\frac{2(-\gamma_n)\beta}{\pi n \omega_{\ell,n}^2}} \exp\left(-\frac{\beta n}{2}(-\gamma_n) \langle v, m - \mathbf{t}_n \rangle^2\right) \\ &\quad \times \left(1 - \frac{1}{\omega_{\ell,n}} v_\ell(-\gamma_n)\beta \langle v, m - \mathbf{t}_n \rangle + O(\omega_{\ell,n}^{-2} n^{-1} \log n)\right). \end{aligned} \quad (4.28)$$

Recalling (4.8)–(4.9), we have

$$\begin{aligned} \tilde{\mathcal{E}}_{\mathcal{D}_n}(g, g) &= \sum_{m \in \mathcal{D}_n} \tilde{Q}_n(m) \sum_{\ell \in [k]} r_\ell \left[ g(m^{\ell,+}) - g(m) \right]^2 \\ &= \frac{1}{Z_n} \sum_{m \in \mathcal{D}_n} \exp\left[-\frac{\beta n}{2} \langle [m - \mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n] \rangle\right] e^{-\beta n F_n(\mathbf{t}_n)} \\ &\quad \times \sum_{\ell \in [k]} r_\ell v_\ell^2 \frac{2(-\gamma_n)\beta}{\pi n \omega_{\ell,n}^2} \exp\left(-\beta n(-\gamma_n) \langle v, m - \mathbf{t}_n \rangle^2\right) \\ &\quad \times \left(1 - \frac{v_\ell}{\omega_{\ell,n}}(-\gamma_n)\beta \langle v, m - \mathbf{t}_n \rangle + O(\omega_{\ell,n}^{-2} n^{-1} \log n)\right)^2 \\ &= \frac{1}{Z_n} \frac{2(-\gamma_n)\beta}{\pi} \sum_{m \in \mathcal{D}_n} \exp\left[-\frac{\beta n}{2} \langle [m - \mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n] \rangle\right] e^{-\beta n F_n(\mathbf{t}_n)} \\ &\quad \times \exp\left(-\beta n(-\gamma_n) \langle v, m - \mathbf{t}_n \rangle^2\right) \left[1 + O\left(\omega_{\ell,n}^{-1} n^{-1/2} \log^{1/2} n\right)\right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{Z_n} \frac{2(-\gamma_n)\beta}{\pi} \left[ 1 + O\left(\omega_{\ell,n}^{-1} n^{-1/2} \log^{1/2} n\right) \right] e^{-\beta n F_n(\mathbf{t}_n)} \left( \prod_{\ell \in [k]} \frac{|A_{\ell,n}|}{2} \right) \\
&\quad \times \int_{\mathcal{D}_n} dm \exp \left[ -\frac{\beta n}{2} \langle [m - \mathbf{t}_n], \mathbb{A}_n \cdot [m - \mathbf{t}_n] \rangle \right] \\
&\quad \times \exp \left( -\beta n (-\gamma_n) \langle v, m - \mathbf{t}_n \rangle^2 \right) \\
&= \frac{1}{Z_n} e^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n)n}{\sqrt{[-\det \mathbb{A}_n]}} \left( \frac{\pi n}{2\beta} \right)^{\frac{k}{2}-1} \left( \prod_{\ell \in [k]} \omega_{\ell,n} \right) \\
&\quad \times \left[ 1 + O\left(\omega_{\ell,n}^{-1} n^{-1/2} \log^{1/2} n\right) \right], \tag{4.29}
\end{aligned}$$

where we use [6, Eq. (10.2.33)] with  $\varepsilon = \frac{1}{\beta n}$  and  $d = k$ . Here  $\frac{1}{2} |A_{\ell,n}|$  is the inverse of the step in the  $\ell$ -direction, while in [6, Eq. (10.2.33)] the step is  $\varepsilon$ .

*Remark 4.3.* Note that

$$\tilde{\mathcal{E}}_{\mathcal{D}_n}(g, g) = \tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g}, \tilde{g}) [1 + o(1)] \tag{4.30}$$

because  $\tilde{g}(m) = g(m) [1 + o(1)]$  for all  $m \in W_0^c \cap \mathcal{D}_n$ . The latter can be proved by approximating the Gaussian integral by 0 or 1 when  $\langle v, m - \mathbf{t}_n \rangle$  is proportional to  $-n^{-1/2} \log^{1/2} n$  or  $n^{-1/2} \log^{1/2} n$ , respectively. ♠

*4.5. Final Dirichlet form approximation.* We are now ready to compare  $\mathcal{E}_{S_n}$  with  $\tilde{\mathcal{E}}_{\mathcal{D}_n}$ . Let  $h: S_n \rightarrow [0, 1]$  be such that  $h(\sigma) = \tilde{g}(m_n(\sigma))$ ,  $\sigma \in S_n$ . We split the sum in (2.18) into four subsets of  $\Gamma_n \times \Gamma_n$ :  $m \in W_0 \cap \mathcal{D}_n^c$ ,  $m' \in \Gamma_n$ ;  $m \in W_0 \cap \mathcal{D}_n$ ,  $m' \in W_1$ ;  $m \in W_0 \cap \mathcal{D}_n$ ,  $m' \in W_2$ ;  $m \in W_0 \cap \mathcal{D}_n$ ,  $m' \in W_0 \cap \mathcal{D}_n$ . Then, using (4.25)–(4.27), we obtain

$$\mathcal{E}_{S_n}(h, h) = O(n^{-c\beta}) + \frac{1}{2} \sum_{m \in W_0 \cap \mathcal{D}_n} \sum_{m' \in W_0 \cap \mathcal{D}_n} Q_n(m) \tilde{r}_n(m, m') [\tilde{g}(m) - \tilde{g}(m')]^2. \tag{4.31}$$

Using (4.14) and (4.16), we obtain

$$\begin{aligned}
\mathcal{E}_{S_n}(h, h) &= O(n^{-c\beta}) + \frac{1}{2} \sum_{m \in W_0 \cap \mathcal{D}_n} \left[ 1 + O(n^{-1/2} \log^{3/2} n) \right] \tilde{Q}_n(m) \\
&\quad \times \sum_{m' \in W_0 \cap \mathcal{D}_n} \left( 1 + O(n^{-1/2} \log^{1/2} n) \right) \tilde{r}_n(m, m') [\tilde{g}(m) - \tilde{g}(m')]^2 \\
&= \left[ 1 + O(n^{-1/2} \log^{1/2} n) \right] \frac{1}{2} \sum_{m, m' \in W_0 \cap \mathcal{D}_n} \tilde{Q}_n(m) \tilde{r}_n(m, m') [\tilde{g}(m) - \tilde{g}(m')]^2 \\
&= \left[ 1 + O(n^{-1/2} \log^{1/2} n) \right] \frac{1}{2} \sum_{m, m' \in \mathcal{D}_n} \tilde{Q}_n(m) \tilde{r}_n(m, m') [\tilde{g}(m) - \tilde{g}(m')]^2 \\
&= \tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g}, \tilde{g}) \left[ 1 + O(n^{-1/2} \log^{1/2} n) \right]
\end{aligned}$$

$$= [1 + o_n(1)] \frac{1}{Z_n} \exp[-\beta n F_n(\mathbf{t}_n)] \frac{(-\gamma_n)n}{\sqrt{[-\det \mathbb{A}_n]}} \left( \frac{\pi n}{2\beta} \right)^{\frac{k}{2}-1} \left( \prod_{\ell \in [k]} \omega_{\ell,n} \right), \quad (4.32)$$

where the third equality follows from (4.25)–(4.27) together with (4.14), and the last equality follows from (4.29)–(4.30).

## 5. Capacity and Valley Estimates

Section 5.1 provides sharp asymptotic upper bounds and lower bounds on the capacity of the metastable pair between which the crossover is being considered. These estimates use the results of the Sect. 4 together with the Dirichlet principle and the Berman–Konsowa principle, which are variational representations of capacity. Section 5.2 provides a sharp asymptotic estimate for the mesoscopic measure of the valleys of the minima of  $F_n$ , which leads to a sharp asymptotic estimate for  $F_n$  inside this valley.

**5.1. Capacity estimates.** Given a Markov process  $(x_t)_{t \geq 0}$  with state space  $S$ , a key quantity in the potential-theoretic approach to metastability is the *capacity*  $\text{cap}(A, B)$  of two disjoint subsets  $A, B$  of  $S$ . This is defined by (see [6, Eq. (7.1.39)])

$$\text{cap}(A, B) = \sum_{x \in A} \mu(x) \mathbb{P}_x(\tau_B < \tau_A), \quad (5.1)$$

where  $\mu$  is the invariant measure and  $\mathbb{P}_x$  is the probability distribution of the Markov process starting in  $x$ .

Recall that  $\mathcal{M}_n$  is the set of local minima of  $F_n$ .

**Proposition 5.1** (Asymptotics of the capacity). *Let  $\mathbf{m}_n = (\mathbf{m}_{\ell,n})_{\ell \in [k]} \in \mathcal{M}_n$  and  $M_n \subset \mathcal{M}_n \setminus \mathbf{m}_n$ , such that the gate  $\mathcal{G}(\mathbf{m}_n, M_n)$  consists of a unique point  $\mathbf{t}_n = (\mathbf{t}_{\ell,n})_{\ell \in [k]}$ . Suppose that  $\beta \in (\beta_c, \infty)$  and  $h \in [0, h_c(\beta))$ . Then, as  $n \rightarrow \infty$ ,*

$$\begin{aligned} & \text{cap}(S_n[\mathbf{m}_n], S_n[M_n]) \\ &= [1 + o_n(1)] \frac{1}{Z_n} e^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n)n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]} } \left( \frac{\pi n}{2\beta} \right)^{\frac{k}{2}-1} \left( \prod_{\ell \in [k]} \omega_{\ell,n} \right). \end{aligned} \quad (5.2)$$

**Remark 5.1.** Proposition 5.1 holds for any subset  $M_n \subseteq \mathcal{M}_n \setminus \mathbf{m}_n$ , separated from  $\mathbf{m}_n$  by  $\mathbf{t}_n$ , independently on the values of  $F_n$  on  $M_n$ . ♠

**5.1.1. Upper bound: Dirichlet principle** An important characterisation of the capacity between two disjoint sets is given by the *Dirichlet principle*. For our quantity of interest this states that

$$\text{cap}(S_n[\mathbf{m}_n], S_n[M_n]) = \inf_{u \in \tilde{\mathcal{H}}} \mathcal{E}_{S_n}(u, u), \quad (5.3)$$

where  $\tilde{\mathcal{H}}$  is the set of functions from  $S_n$  to  $[0, 1]$  that are equal to 1 on  $S_n[\mathbf{m}_n]$  and 0 on  $S_n[M_n]$ .

Given that, by assumption,  $\mathcal{G}(\mathbf{m}_n, M_n) = \{\mathbf{t}_n\}$ , we use the Dirichlet principle in (5.3) to obtain an upper bound on the capacity. We take as test function  $h \in \mathcal{H}$  defined in Sect. 4.5 and, using (4.32), we obtain

$$\begin{aligned} \text{cap}(S_n[\mathbf{m}_n], S_n[M_n]) &\leq \mathcal{E}_{S_n}(h, h) \\ &= [1 + o_n(1)] \frac{1}{Z_n} e^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n)n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]} } \left( \frac{\pi n}{2\beta} \right)^{\frac{k}{2}-1} \left( \prod_{\ell \in [k]} \omega_{\ell, n} \right). \end{aligned} \quad (5.4)$$

**5.1.2. Lower bound: Berman–Konsowa principle** We first note that the process  $(\sigma_t)_{t \geq 0}$  is lumpable. Indeed, the process  $(m_n(\sigma_t))_{t \geq 0}$  is Markovian because the Hamiltonian  $H_n(\sigma)$  depends on  $m_n(\sigma)$  only (see (2.6)). Therefore, for  $\mathbf{A} = S_n[A]$  and  $\mathbf{B} = S_n[B]$  with  $A$  and  $B$  disjoint subsets of  $\Gamma_n$ ,

$$\text{cap}(\mathbf{A}, \mathbf{B}) = \text{cap}_\Gamma(A, B), \quad (5.5)$$

where  $\text{cap}_\Gamma$  denotes the capacity for the process  $(m_n(\sigma_t))_{t \geq 0}$ , i.e., the projection of the process  $(\sigma_t)_{t \geq 0}$  on the magnetisation space  $\Gamma_n$ . We write  $\mathbb{P}^\Gamma$  and  $\mathbb{E}^\Gamma$  to denote the law of  $(m_n(\sigma_t))_{t \geq 0}$  induced by the law  $\mathbb{P}$  of  $(\sigma_t)_{t \geq 0}$ , and its expectation, respectively. By the lumpability, we can focus on the dynamics on  $\Gamma_n$ .

Following the line of argument in [6, Section 10.3] (with  $\varepsilon = \frac{2}{n}$  and  $d = k$ ), we obtain the lower bound

$$\begin{aligned} \text{cap}(S_n[\mathbf{m}_n], S_n[M_n]) &= \text{cap}_\Gamma(\mathbf{m}_n, M_n) \geq \tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g}, \tilde{g}) \left[ 1 + O(n^{-1/2} \log^{1/2} n) \right] \\ &= \frac{1}{Z_n} e^{-\beta n F_n(\mathbf{t}_n)} \frac{(-\gamma_n)n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]} } \left( \frac{\pi n}{2\beta} \right)^{\frac{k}{2}-1} \left( \prod_{\ell \in [k]} \omega_{\ell, n} \right) [1 + o_n(1)], \end{aligned} \quad (5.6)$$

where we use (4.29) and (4.30).

We sketch the proof. The main idea is to use the Berman–Konsowa principle for a suitable defective flow. More precisely, given disjoint subsets  $A, B$  of the state space, for any *defective loop-free unit flow*  $f_{A,B}$  from  $A$  to  $B$  with defect function  $\delta$  (as defined in [6, Definition 9.2]), we can estimate (see [6, Lemma 9.4], and notation therein)

$$\text{cap}(A, B) \geq \prod_{i=1}^M \left( 1 + \left[ \max_{y \in A_i} \frac{\delta(y)}{\mathcal{F}(y)} \right]_+ \right)^{-1} \sum_{\gamma} \mathbb{P}^{f_{A,B}}(\gamma) \left[ \left( \sum_{(x,y) \in \gamma} \frac{f_{A,B}((x,y))}{\mu(x)p(x,y)} \right)^{-1} \right], \quad (5.7)$$

where  $[\cdot]_+$  denotes the positive part and  $\gamma$  is a self-avoiding path from  $A$  to  $B$ . It turns out that, with a suitable choice of the flow  $f$ , the product in the right-hand side of (5.7) is bounded from below by  $1 + O(n^{-1/2} \log^{1/2} n)$ , and the sum over  $\gamma$  from below by  $\tilde{\mathcal{E}}_{\mathcal{D}_n}(\tilde{g}, \tilde{g})[1 + o_n(1)]$ . This proves (5.6).

We give a sketch of the test flow definition in our setting. Here  $A = \{\mathbf{m}_n\}$  and  $B = M_n$ . Let  $v^*$  be the eigenvector corresponding to the unique negative eigenvalue of the Hessian of  $F_n$  at the saddle point  $\mathbf{t}_n$  (unique gate point in  $\mathcal{G}(\{\mathbf{m}_n\}, M_n)$ ). Let  $G_n$  be the cylinder in  $\mathbb{R}^k$  intersected with  $\Gamma_n$ , centred at  $\mathbf{t}_n$ , with axis  $v^*$ , radius  $\rho = C n^{-1/2} \log^{1/2} n$  and length  $\rho' = C' n^{-1/2} \log^{1/2} n$ . We will denote by  $\partial_B G_n$  the base facing  $B$  and by  $\partial_A G_n$  the central part of radius  $C'' n^{-1/2} \log^{1/2} n$  of the base facing  $A$ , with  $C'' < C$ . Choose the constants so that  $G_n$  is contained in  $\mathcal{D}_n$  defined in (4.1).

We define a defective flow  $f_{A,B}$  from  $A$  to  $B$  consisting of three parts:  $f_A$ , a unitary flow from  $A$  to  $\partial_A G_n$ ;  $f$ , a defective loop-free unit flow from  $\partial_A G_n$  to  $\partial_B G_n$  inside  $G_n$ ;  $f_B$ , a unitary flow from  $\partial_B G_n$  to  $B$ . This choice implies that the sum over  $\gamma$  in (5.7) is relevant only on the paths entering  $G_n$  in  $\partial_A G_n$ , exiting  $G_n$  in  $\partial_B G_n$ , and afterwards reaching  $B$  without going back to  $G_n$ . For this purpose we choose  $f_A$  and  $f_B$  such that  $f_A((x, y))$  and  $f_B((x, y))$  are proportional to  $Q_n(x)$ . For  $m \in G_n$  such that  $m^{\ell,+} \in G_n$ , define

$$f((m, m^{\ell,+})) = \frac{\tilde{Q}_n(m)r_\ell [g(m^{\ell,+}) - g(m)]_+}{N(g)}, \quad (5.8)$$

where  $g$  is defined in (4.24),  $\tilde{Q}_n$  in (4.6),  $r_\ell$  in (4.9) and

$$N(g) = \sum_{m \in \partial_A G_n} \sum_{\substack{\ell \in [k]: \\ m^{\ell,+} \in G_n}} \tilde{Q}_n(m)r_\ell [g(m^{\ell,+}) - g(m)]_+. \quad (5.9)$$

The contribution to the sum in brackets in (5.7) turns out to be negligible outside  $G_n$ . Therefore, no further conditions on the flows  $f_A$  and  $f_B$  are necessary, provided the total flow out of  $A$  is 1 and the total flow  $f_{A,B}$  is defective and loop-free.

**5.2. Measure of the valley.** In order to prove Theorem 1.1, we need the following estimate on the measure of the valley of the minima of  $F_n$ . For  $\mathbf{m}_n \in \mathcal{M}_n$ , let  $A(\mathbf{m}_n) \subset \Gamma_n$  be the valley of  $\mathbf{m}_n$  as defined in [6, Eq. (8.2.10)].

**Lemma 5.1** (Gibbs weight of the valley). *Given  $\mathbf{m}_n \in \mathcal{M}_n$ ,*

$$Q_n(A(\mathbf{m}_n)) = \frac{1}{Z_n} \frac{\exp(-\beta n F_n(\mathbf{m}_n))}{\sqrt{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left( \frac{n\pi}{2\beta} \right)^{\frac{k}{2}} \left( \prod_{\ell \in [k]} \omega_{\ell,n} \right) \left[ 1 + O(n^{-1/2} \log^{3/2} n) \right], \quad (5.10)$$

where  $Q_n$  is the mesoscopic measure defined in (2.17), and  $\mathbb{A}_n(\mathbf{m}_n)$  is the  $k \times k$  Hessian matrix defined in (4.2).

*Proof.* The proof follows that of [6, Lemma 10.12 and (10.2.33)]. The relevant contribution to  $Q_n(A(\mathbf{m}_n))$  is given by the measure of a ball  $B_\rho$  of radius  $\rho = C n^{-1/2} \log^{1/2} n$  centred in  $\mathbf{m}_n$ , with  $C$  constant, contained in  $A(\mathbf{m}_n)$ . Indeed, if  $y \in A(\mathbf{m}_n)$  and  $d(\mathbf{m}_n, y) > \rho$ , then by Taylor expansion of  $F_n$  around  $\mathbf{m}_n$  we have

$$\begin{aligned} Q_n(y) &= \frac{1}{Z_n} \exp[-\beta n F_n(y)] = \frac{1}{Z_n} \exp[-\beta n [F_n(\mathbf{m}_n) + c d(\mathbf{m}_n, y)^2]] \\ &\leq \frac{1}{Z_n} \exp[-\beta n [F_n(\mathbf{m}_n) + c \rho^2]] = \frac{n^{-\beta c C^2}}{Z_n} \exp[-\beta n F_n(\mathbf{m}_n)], \end{aligned} \quad (5.11)$$

where  $c$  is a constant. The condition  $y \in A(\mathbf{m}_n)$  is needed to ensure that  $F_n(y) > F_n(\mathbf{m}_n)$ , implying that  $c$  is positive. Therefore, we obtain the rough estimate

$$Q_n(A(\mathbf{m}_n) \setminus B_\rho) \leq n^k \frac{n^{-\beta c C^2}}{Z_n} \exp[-\beta n F_n(\mathbf{m}_n)], \quad (5.12)$$

where we use that  $|\Gamma_n| \leq n^k$ . The bound in (5.12) is sufficient to show that  $Q_n(A(\mathbf{m}_n) \setminus B_\rho)$  is negligible in  $Q_n(A(\mathbf{m}_n))$ .

Compute

$$\begin{aligned}
& Z_n Q_n(A(\mathbf{m}_n) \cap B_\rho) \\
&= Z_n Q_n(B_\rho) = Z_n \sum_{y \in B_\rho} Q_n(y) = \sum_{y \in B_\rho} e^{-\beta n F_n(y)} \\
&= e^{-\beta n F_n(\mathbf{m}_n)} \sum_{y \in B_\rho} \exp \left[ -\frac{\beta n}{2} \langle y - \mathbf{m}_n, (\mathbb{A}_n(\mathbf{m}_n)) \cdot (y - \mathbf{m}_n) \rangle + O(n\rho^3) \right] \\
&= e^{-\beta n F_n(\mathbf{m}_n)} [1 + O(n\rho^3)] \sum_{y \in B_\rho} \exp \left[ -\frac{\beta n}{2} \langle y - \mathbf{m}_n, (\mathbb{A}_n(\mathbf{m}_n)) \cdot (y - \mathbf{m}_n) \rangle \right] \\
&= e^{-\beta n F_n(\mathbf{m}_n)} \left( \prod_{\ell \in [k]} \frac{|A_{\ell,n}|}{2} \right) [1 + O(n\rho^3)] \\
&\quad \times \int_{B_\rho} dy \exp \left[ -\frac{\beta n}{2} \langle y - \mathbf{m}_n, (\mathbb{A}_n(\mathbf{m}_n)) \cdot (y - \mathbf{m}_n) \rangle \right] \\
&= e^{-\beta n F_n(\mathbf{m}_n)} \left( \frac{n}{2} \right)^k \left( \prod_{\ell \in [k]} \omega_{\ell,n} \right) [1 + O(n\rho^3)] \left( \frac{2\pi}{n\beta} \right)^{\frac{k}{2}} \sqrt{\frac{1}{\det(\mathbb{A}_n(\mathbf{m}_n))}} \\
&= \frac{e^{-\beta n F_n(\mathbf{m}_n)}}{\sqrt{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left( \frac{n\pi}{2\beta} \right)^{\frac{k}{2}} \left( \prod_{\ell \in [k]} \omega_{\ell,n} \right) [1 + O(n\rho^3)], \tag{5.13}
\end{aligned}$$

where we use the Taylor expansion

$$F_n(y) = F_n(\mathbf{m}_n) + \frac{1}{2} \langle y - \mathbf{m}_n, (\nabla^2 F_n) \cdot (\mathbf{m}_n)(y - \mathbf{m}_n) \rangle + O(\rho^3), \quad y \in B_\rho, \tag{5.14}$$

and the approximation of the sum by an integral is correct up to an error  $1 + O(\rho)$ . In the last lines we approximated the Gaussian integral on intervals  $[-\rho, \rho]$  by the Gaussian integral on  $\mathbb{R}$ , with an error  $1 + O(n^{-c})$ . We conclude by looking at (5.12) and (5.13), and noting that for  $C$  large enough  $Q_n(A(\mathbf{m}_n) \setminus B_\rho)$  is negligible compared to  $Q_n(A(\mathbf{m}_n) \cap B_\rho)$ .  $\square$

## 6. Proof of the Theorems

In this section we prove Theorems 1.1–1.3. Section 6.1 uses the asymptotics for the capacity of the metastable pair from Sect. 5.1 and the asymptotics for the mesoscopic measure from Sect. 5.2 to prove Theorem 1.1. Section 6.2 proves Theorem 1.2. Section 6.3 proves Theorem 1.3.

**6.1. Average crossover time.** Let us return to the notation of Theorem 1.1, where  $\mathbf{m}_n \in \mathcal{M}_n$  and  $\mathcal{M}_n(\mathbf{m}_n) = \{m \in \mathcal{M}_n \setminus \mathbf{m}_n : F_n(m) \leq F_n(\mathbf{m}_n)\}$ . To prove Theorem 1.1 we use the relation

$$\mathbb{E}_{\mathbf{m}_n}^{\Gamma}(\tau_{\mathcal{M}_n(\mathbf{m}_n)}) = [1 + o_n(1)] \frac{\mu(A(\mathbf{m}_n))}{\text{cap}_{\Gamma}(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))}, \quad (6.1)$$

Recall notation introduced in Sect. 5.1.2. Because  $F_n(m) \leq F_n(\mathbf{m}_n)$  for all  $m \in \mathcal{M}_n(\mathbf{m}_n)$ , (6.1) follows from [6, Theorem 8.15] after proving that  $\mathcal{M}_n$  is a set of metastable points in the sense of [6, Definition 8.2]. The latter follows along the lines of the proof of [6, Theorem 10.6], where similar values of capacities and invariant measures occur.

Using (6.1) in combination with Proposition 5.1 and Lemma 5.1, we obtain that, for all  $\sigma \in S_n[\mathbf{m}_n]$ ,

$$\begin{aligned} & \mathbb{E}_{\sigma}(\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)])} \\ &= \mathbb{E}_{\mathbf{m}_n}^{\Gamma}(\tau_{\mathcal{M}_n(\mathbf{m}_n)}) = [1 + o_n(1)] \frac{Q_n(A(\mathbf{m}_n))}{\text{cap}_{\Gamma}(\mathbf{m}_n, \mathcal{M}_n(\mathbf{m}_n))} \\ &= [1 + o_n(1)] \frac{Q_n(A(\mathbf{m}_n))}{\text{cap}(S_n[\mathbf{m}_n], S_n[\mathcal{M}_n(\mathbf{m}_n)])} \\ &= [1 + o_n(1)] \frac{\frac{1}{Z_n} \exp(-\beta n F_n(\mathbf{m}_n))}{\sqrt{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left( \frac{n\pi}{2\beta} \right)^{\frac{k}{2}} (\prod_{\ell \in [k]} \omega_{\ell, n}) \\ &= [1 + o_n(1)] \frac{\frac{1}{Z_n} \exp[-\beta n F_n(\mathbf{t}_n)]}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]} } \frac{(-\gamma_n)n}{\sqrt{[-\det(\mathbb{A}_n(\mathbf{t}_n))]} } \left( \frac{\pi n}{2\beta} \right)^{\frac{k}{2}-1} (\prod_{\ell \in [k]} \omega_{\ell, n}) \\ &= [1 + o_n(1)] \sqrt{\frac{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}{\det(\mathbb{A}_n(\mathbf{m}_n))}} \left( \frac{\pi}{2\beta(-\gamma_n)} \right) \exp[\beta n(F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n))], \end{aligned} \quad (6.2)$$

where we use that the dynamics depends on the starting configuration  $\sigma \in S_n[\mathbf{m}_n]$  only, through its level magnetisations  $m_n(\sigma) = \mathbf{m}_n$  (see (2.6)), and also use the lumpability.

**6.2. Exponential law.** In this section we prove Theorem 1.2. Since the dynamics depends on the starting configuration  $\sigma \in S_n[\mathbf{m}_n]$  through its level magnetisation  $m_n(\sigma) = \mathbf{m}_n$  only (see (2.6)), we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}_{\sigma}(\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]} > t) \mathbb{E}_{\sigma}[\tau_{S_n[\mathcal{M}_n(\mathbf{m}_n)]}] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_{\mathbf{m}_n}^{\Gamma}(\bar{\tau}_{\mathcal{M}_n(\mathbf{m}_n)} > t) \mathbb{E}_{\mathbf{m}_n}^{\Gamma}[\bar{\tau}_{\mathcal{M}_n(\mathbf{m}_n)}], \end{aligned} \quad (6.3)$$

where  $\bar{\tau}$  is the hitting time of the process projected on  $\Gamma_n$ . Given the non-degeneracy hypothesis (Hypothesis 1 in Sect. 1.3.4) and the one-dimensional landscape analysis (in Sect. 3.3), we can apply [6, Theorem 8.45] to the right-hand side of (6.3) and conclude the proof.

**6.3. Randomness of the exponent.** In this section we prove Theorem 1.3. In particular, we compute  $F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n) - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})]$  to leading order.

Recalling definitions (2.26) and (3.5), we have

$$F_{\beta,h}(m) = -\frac{1}{2}K(m)^2 - h \sum_{\ell \in [k]} \omega_\ell m_\ell + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell I_C(m_\ell). \quad (6.4)$$

Let  $\mathbf{m} = (\mathbf{m}_\ell)_{\ell \in [k]}$ ,  $\mathbf{t} = (\mathbf{t}_\ell)_{\ell \in [k]} \in [-1, 1]^k$  be the critical points of  $F_{\beta,h}$  closest to  $\mathbf{m}_n$ ,  $\mathbf{t}_n$  (i.e., the critical points of  $F_n$  defined above), respectively. Note that  $\mathbf{m}$  and  $\mathbf{t}$  satisfy (3.4), while  $\mathbf{m}_n$  and  $\mathbf{t}_n$  satisfy (3.3). Using (2.21), we get

$$\begin{aligned} F_n(\mathbf{t}_n) - F_{\beta,h}(\mathbf{t}_n) &= -\frac{1}{2}[K_n(\mathbf{t}_n)^2 - K(\mathbf{t}_n)^2] \\ &\quad - h \sum_{\ell \in [k]} [\omega_{\ell,n} - \omega_\ell] \mathbf{t}_{\ell,n} \\ &\quad + \frac{1}{\beta} \left[ \sum_{\ell \in [k]} [\omega_{\ell,n} - \omega_\ell] I_C(\mathbf{t}_{\ell,n}) + \sum_{\ell \in [k]} \frac{1}{2n} \log \left( \frac{\pi(1 - \mathbf{t}_{\ell,n}^2)}{2} \right) \omega_{\ell,n} \right. \\ &\quad \left. - \frac{k}{2n} + o(n^{-1}) \right] \end{aligned} \quad (6.5)$$

and

$$F_{\beta,h}(\mathbf{t}_n) - F_{\beta,h}(\mathbf{t}) = -\frac{1}{2}[K(\mathbf{t}_n)^2 - K(\mathbf{t})^2] + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell [I_C(\mathbf{t}_{\ell,n}) - I_C(\mathbf{t}_\ell)]. \quad (6.6)$$

By (3.2), we have

$$\begin{aligned} \frac{1}{2} \log \left( \frac{1 + \mathbf{t}_{\ell,n}}{1 - \mathbf{t}_{\ell,n}} \right) &= \beta [a_\ell K_n(\mathbf{t}_n) + h], \\ \frac{1}{2} \log \left( \frac{1 + \mathbf{t}_\ell}{1 - \mathbf{t}_\ell} \right) &= \beta [a_\ell K(\mathbf{t}) + h]. \end{aligned} \quad (6.7)$$

Thus,

$$\begin{aligned} I_C(\mathbf{t}_{\ell,n}) - I_C(\mathbf{t}_\ell) &= (\mathbf{t}_{\ell,n} - \mathbf{t}_\ell) I'_C(\mathbf{t}_\ell) + O((\mathbf{t}_{\ell,n} - \mathbf{t}_\ell)^2) \\ &= (\mathbf{t}_{\ell,n} - \mathbf{t}_\ell) \frac{1}{2} \log \left( \frac{1 + \mathbf{t}_\ell}{1 - \mathbf{t}_\ell} \right) + O((\mathbf{t}_{\ell,n} - \mathbf{t}_\ell)^2) \\ &= (\mathbf{t}_{\ell,n} - \mathbf{t}_\ell) \beta [a_\ell K(\mathbf{t}) + h] + O((\mathbf{t}_{\ell,n} - \mathbf{t}_\ell)^2). \end{aligned} \quad (6.8)$$

Moreover,

$$\begin{aligned} K(\mathbf{t}_n)^2 - K(\mathbf{t})^2 &= \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \omega_\ell \omega_{\ell'} [\mathbf{t}_{\ell,n} \mathbf{t}_{\ell',n} - \mathbf{t}_\ell \mathbf{t}_{\ell'}] \\ &= \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \omega_\ell \omega_{\ell'} (\mathbf{t}_\ell [\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}] + \mathbf{t}_{\ell'} [\mathbf{t}_{\ell,n} - \mathbf{t}_\ell] \\ &\quad + [\mathbf{t}_{\ell,n} - \mathbf{t}_\ell][\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}]) \end{aligned} \quad (6.9)$$

and

$$\begin{aligned}
 K_n(\mathbf{t}_n)^2 - K(\mathbf{t}_n)^2 &= \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} [\omega_{\ell, n} \omega_{\ell', n} - \omega_\ell \omega_{\ell'}] \mathbf{t}_{\ell, n} \mathbf{t}_{\ell', n} \\
 &= \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \mathbf{t}_{\ell, n} \mathbf{t}_{\ell', n} (\omega_\ell [\omega_{\ell', n} - \omega_{\ell'}] + \omega_{\ell'} [\omega_{\ell, n} - \omega_\ell] \\
 &\quad + [\omega_{\ell, n} - \omega_\ell] [\omega_{\ell', n} - \omega_{\ell'}]) .
 \end{aligned} \tag{6.10}$$

Similar equalities hold after we replace  $\mathbf{t}$  by  $\mathbf{m}$  and  $\mathbf{t}_n$  by  $\mathbf{m}_n$ . Using the previous computations, we obtain

$$\begin{aligned}
 F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n) - [F_{\beta, h}(\mathbf{t}) - F_{\beta, h}(\mathbf{m})] \\
 &= F_n(\mathbf{t}_n) - F_{\beta, h}(\mathbf{t}_n) + F_{\beta, h}(\mathbf{t}_n) - F_{\beta, h}(\mathbf{t}) \\
 &\quad - [F_n(\mathbf{m}_n) - F_{\beta, h}(\mathbf{m}_n) + F_{\beta, h}(\mathbf{m}_n) - F_{\beta, h}(\mathbf{m})] \\
 &= -\frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} [\mathbf{t}_{\ell, n} \mathbf{t}_{\ell', n} - \mathbf{m}_{\ell, n} \mathbf{m}_{\ell', n}] (\omega_\ell [\omega_{\ell', n} - \omega_{\ell'}] + \omega_{\ell'} [\omega_{\ell, n} - \omega_\ell] \\
 &\quad + [\omega_{\ell, n} - \omega_\ell] [\omega_{\ell', n} - \omega_{\ell'}]) \\
 &\quad - \frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \omega_\ell \omega_{\ell'} [\mathbf{t}_{\ell, n} \mathbf{t}_{\ell', n} - \mathbf{t}_\ell \mathbf{t}_{\ell'} + \mathbf{m}_\ell \mathbf{m}_{\ell'} - \mathbf{m}_{\ell, n} \mathbf{m}_{\ell', n}] \\
 &\quad - h \sum_{\ell \in [k]} [\omega_{\ell, n} - \omega_\ell] [\mathbf{t}_{\ell, n} - \mathbf{m}_{\ell, n}] \\
 &\quad + \frac{1}{\beta} \sum_{\ell \in [k]} [\omega_{\ell, n} - \omega_\ell] [I_C(\mathbf{t}_{\ell, n}) - I_C(\mathbf{m}_{\ell, n})] + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{1}{2n} \log \left( \frac{1 - \mathbf{t}_{\ell, n}^2}{1 - \mathbf{m}_{\ell, n}^2} \right) \\
 &\quad + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell [I_C(\mathbf{t}_{\ell, n}) - I_C(\mathbf{t}_\ell) + I_C(\mathbf{m}_\ell) - I_C(\mathbf{m}_{\ell, n})] + o(n^{-1}) .
 \end{aligned} \tag{6.11}$$

Using (6.8), we find

$$\begin{aligned}
 [F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] - [F_{\beta, h}(\mathbf{t}) - F_{\beta, h}(\mathbf{m})] \\
 &= -\frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} [\mathbf{t}_{\ell, n} \mathbf{t}_{\ell', n} - \mathbf{m}_{\ell, n} \mathbf{m}_{\ell', n}] (\omega_\ell [\omega_{\ell', n} - \omega_{\ell'}] + \omega_{\ell'} [\omega_{\ell, n} - \omega_\ell] \\
 &\quad + [\omega_{\ell, n} - \omega_\ell] [\omega_{\ell', n} - \omega_{\ell'}]) \\
 &\quad - \frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \omega_\ell \omega_{\ell'} [\mathbf{t}_{\ell, n} \mathbf{t}_{\ell', n} - \mathbf{t}_\ell \mathbf{t}_{\ell'} + \mathbf{m}_\ell \mathbf{m}_{\ell'} - \mathbf{m}_{\ell, n} \mathbf{m}_{\ell', n}] \\
 &\quad - h \sum_{\ell \in [k]} [\omega_{\ell, n} - \omega_\ell] [\mathbf{t}_{\ell, n} - \mathbf{m}_{\ell, n}] \\
 &\quad + \frac{1}{\beta} \sum_{\ell \in [k]} [\omega_{\ell, n} - \omega_\ell] [I_C(\mathbf{t}_{\ell, n}) - I_C(\mathbf{m}_{\ell, n})] + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{1}{2n} \log \left( \frac{1 - \mathbf{t}_{\ell, n}^2}{1 - \mathbf{m}_{\ell, n}^2} \right) \\
 &\quad + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell [(\mathbf{t}_{\ell, n} - \mathbf{t}_\ell) \beta [a_\ell K(\mathbf{t}) + h] + O((\mathbf{t}_{\ell, n} - \mathbf{t}_\ell)^2)]
 \end{aligned}$$

$$\begin{aligned}
& -(\mathbf{m}_{\ell,n} - \mathbf{m}_\ell)\beta \left[ a_\ell K(\mathbf{m}) + h \right] + O((\mathbf{m}_{\ell,n} - \mathbf{m}_\ell)^2) \\
& + o\left(n^{-1}\right).
\end{aligned} \tag{6.12}$$

Since

$$\mathbf{t}_{\ell,n}\mathbf{t}_{\ell',n} - \mathbf{t}_\ell\mathbf{t}_{\ell'} = (\mathbf{t}_\ell[\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}] + \mathbf{t}_{\ell'}[\mathbf{t}_{\ell,n} - \mathbf{t}_\ell] + [\mathbf{t}_{\ell,n} - \mathbf{t}_\ell][\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}]), \tag{6.13}$$

we focus on estimating  $\mathbf{t}_{\ell,n} - \mathbf{t}_\ell$ .

From Taylor expansion, we get

$$\begin{aligned}
\mathbf{t}_{\ell,n} - \mathbf{t}_\ell &= \tanh\left(\beta \left[ a_\ell \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} \mathbf{t}_{\ell',n} + h \right]\right) - \tanh\left(\beta \left[ a_\ell \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} \mathbf{t}_{\ell'} + h \right]\right) \\
&= \beta a_\ell \sum_{\ell' \in [k]} a_{\ell'} [\omega_{\ell',n} \mathbf{t}_{\ell',n} - \omega_{\ell'} \mathbf{t}_{\ell'}] \left[ 1 - \tanh\left(\beta \left[ a_\ell \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} \mathbf{t}_{\ell'} + h \right]\right)^2 \right] \\
&\quad - \beta^2 a_\ell^2 \left( \sum_{\ell' \in [k]} a_{\ell'} [\omega_{\ell',n} \mathbf{t}_{\ell',n} - \omega_{\ell'} \mathbf{t}_{\ell'}] \right)^2 \times \tanh\left(\beta \left[ a_\ell \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} \mathbf{t}_{\ell'} + h \right]\right) \\
&\quad \times \left[ 1 - \tanh\left(\beta \left[ a_\ell \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell'} \mathbf{t}_{\ell'} + h \right]\right)^2 \right] \\
&\quad + O\left(a_\ell^3 \left( \sum_{\ell' \in [k]} a_{\ell'} [\omega_{\ell',n} \mathbf{t}_{\ell',n} - \omega_{\ell'} \mathbf{t}_{\ell'}] \right)^3\right).
\end{aligned} \tag{6.14}$$

Since

$$\omega_{\ell',n} \mathbf{t}_{\ell',n} - \omega_{\ell'} \mathbf{t}_{\ell'} = (\omega_{\ell',n} - \omega_{\ell'}) \mathbf{t}_{\ell'} + \omega_{\ell',n} (\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'}), \tag{6.15}$$

we have

$$\begin{aligned}
\mathbf{t}_{\ell,n} - \mathbf{t}_\ell &= \beta a_\ell \left[ 1 - \mathbf{t}_\ell^2 \right] \sum_{\ell' \in [k]} a_{\ell'} [(\omega_{\ell',n} - \omega_{\ell'}) \mathbf{t}_{\ell'} + \omega_{\ell',n} (\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'})] \\
&\quad - \beta^2 a_\ell^2 \mathbf{t}_\ell \left[ 1 - \mathbf{t}_\ell^2 \right] \left( \sum_{\ell' \in [k]} a_{\ell'} [(\omega_{\ell',n} - \omega_{\ell'}) \mathbf{t}_{\ell'} + \omega_{\ell',n} (\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'})] \right)^2 \\
&\quad + O\left(a_\ell^3 \left( \sum_{\ell' \in [k]} a_{\ell'} [(\omega_{\ell',n} - \omega_{\ell'}) \mathbf{t}_{\ell'} + \omega_{\ell',n} (\mathbf{t}_{\ell',n} - \mathbf{t}_{\ell'})] \right)^3\right).
\end{aligned} \tag{6.16}$$

Suppose that  $\mathbf{t}_{\ell,n} - \mathbf{t}_\ell \sim \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}}$ . By the Central Limit Theorem,  $\omega_{\ell,n} - \omega_\ell \sim \frac{Z_\ell}{\sqrt{n}}$ , where  $Z_\ell$  is the normal random variable  $N(0, \omega_\ell(1 - \omega_\ell))$ . Hence

$$\begin{aligned}
 \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} &= \beta a_\ell [1 - \mathbf{t}_\ell^2] \sum_{\ell' \in [k]} a_{\ell'} \left[ \frac{Z_{\ell'}}{\sqrt{n}} \mathbf{t}_{\ell'} + \left( \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \right) \frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}} \right] \\
 &\quad - \beta^2 a_\ell^2 \mathbf{t}_\ell [1 - \mathbf{t}_\ell^2] \left( \sum_{\ell' \in [k]} a_{\ell'} \left[ \frac{Z_{\ell'}}{\sqrt{n}} \mathbf{t}_{\ell'} + \left( \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \right) \frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}} \right] \right)^2 \\
 &\quad + O \left( a_\ell^3 \left( \sum_{\ell' \in [k]} a_{\ell'} \left[ \frac{Z_{\ell'}}{\sqrt{n}} \mathbf{t}_{\ell'} + \left( \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \right) \frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}} \right] \right)^3 \right) \quad (6.17) \\
 &= \frac{1}{\sqrt{n}} \beta a_\ell [1 - \mathbf{t}_\ell^2] \sum_{\ell' \in [k]} a_{\ell'} (\mathbf{t}_{\ell'} Z_{\ell'} + \omega_{\ell'} Y_{\ell'}^{\mathbf{t}}) \\
 &\quad + \frac{1}{n} \beta a_\ell [1 - \mathbf{t}_\ell^2] \sum_{\ell' \in [k]} a_{\ell'} Z_{\ell'} \left( Y_{\ell'}^{\mathbf{t}} - \beta a_\ell \mathbf{t}_\ell \mathbf{t}_{\ell'} \sum_{\ell'' \in [k]} a_{\ell''} \omega_{\ell''} Y_{\ell''}^{\mathbf{t}} \right) + o(n^{-1})
 \end{aligned}$$

and so

$$Y_\ell^{\mathbf{t}} = \beta a_\ell [1 - \mathbf{t}_\ell^2] \frac{\sum_{\ell' \in [k]} a_{\ell'} \mathbf{t}_{\ell'} Z_{\ell'}}{1 - \beta \sum_{\ell' \in [k]} a_{\ell'}^2 \omega_{\ell'} [1 - \mathbf{t}_{\ell'}^2]} + O(n^{-\frac{1}{2}}), \quad (6.18)$$

where the denominator does not vanish because of Remark 4.2. Thus, up to a factor  $O(n^{-\frac{1}{2}})$ ,  $Y_\ell^{\mathbf{t}}$  is a normal random variable with mean 0 and variance

$$\beta^2 a_\ell^2 [1 - \mathbf{t}_\ell^2]^2 \frac{\sum_{\ell' \in [k]} a_{\ell'}^2 \mathbf{t}_{\ell'}^2 \omega_{\ell'} (1 - \omega_{\ell'})}{(1 - \beta \sum_{\ell' \in [k]} a_{\ell'}^2 \omega_{\ell'} [1 - \mathbf{t}_{\ell'}^2])^2}. \quad (6.19)$$

Similar results hold after we replace  $\mathbf{t}$  by  $\mathbf{m}$ .

Going back to (6.12), using (6.13) and (6.18), and inserting  $\mathbf{t}_{\ell,n} - \mathbf{t}_\ell \sim \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}}$  and  $\mathbf{m}_{\ell,n} - \mathbf{m}_\ell \sim \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}}$  and  $\omega_{\ell,n} - \omega_\ell \sim \frac{Z_\ell}{\sqrt{n}}$ , we obtain

$$\begin{aligned}
 &[F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] \\
 &\sim -\frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \left[ \left( \mathbf{t}_\ell + \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} \right) \left( \mathbf{t}_{\ell'} + \frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}} \right) - \left( \mathbf{m}_\ell + \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} \right) \left( \mathbf{m}_{\ell'} + \frac{Y_{\ell'}^{\mathbf{m}}}{\sqrt{n}} \right) \right] \\
 &\quad \times \left( \omega_\ell \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \frac{Z_\ell}{\sqrt{n}} + \frac{Z_\ell Z_{\ell'}}{n} \right) \\
 &\quad - \frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \omega_\ell \omega_{\ell'} \left( \mathbf{t}_\ell \frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}} + \mathbf{t}_{\ell'} \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} + \frac{Y_\ell^{\mathbf{t}} Y_{\ell'}^{\mathbf{t}}}{n} - \mathbf{m}_\ell \frac{Y_{\ell'}^{\mathbf{m}}}{\sqrt{n}} - \mathbf{m}_{\ell'} \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} - \frac{Y_\ell^{\mathbf{m}} Y_{\ell'}^{\mathbf{m}}}{n} \right) \\
 &\quad - h \sum_{\ell \in [k]} \frac{Z_\ell}{\sqrt{n}} \left( \mathbf{t}_\ell + \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} - \mathbf{m}_\ell - \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} \right)
 \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{Z_\ell}{\sqrt{n}} \left[ I_{\mathbf{C}} \left( \mathbf{t}_\ell + \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} \right) - I_{\mathbf{C}} \left( \mathbf{m}_\ell + \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} \right) \right] \\
& + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{1}{2n} \log \left( \frac{1 - \left( \mathbf{t}_\ell + \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} \right)^2}{1 - \left( \mathbf{m}_\ell + \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} \right)^2} \right) \\
& + \frac{1}{\beta} \sum_{\ell \in [k]} \omega_\ell \left[ \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} \beta [a_\ell K(\mathbf{t}) + h] + O \left( \frac{(Y_\ell^{\mathbf{t}})^2}{n} \right) - \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} [a_\ell K(\mathbf{m}) + h] \right. \\
& \left. + O \left( \frac{(Y_\ell^{\mathbf{m}})^2}{n} \right) \right] + o(n^{-1}). \tag{6.20}
\end{aligned}$$

Thus,

$$\begin{aligned}
& [F_n(\mathbf{t}_n) - F_n(\mathbf{m}_n)] - [F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})] \\
& = -\frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} [\mathbf{t}_\ell \mathbf{t}_{\ell'} - \mathbf{m}_\ell \mathbf{m}_{\ell'}] \left( \omega_\ell \frac{Z_{\ell'}}{\sqrt{n}} + \omega_{\ell'} \frac{Z_\ell}{\sqrt{n}} \right) \\
& \quad - \frac{1}{2} \sum_{\ell, \ell' \in [k]} a_\ell a_{\ell'} \omega_\ell \omega_{\ell'} \left( \mathbf{t}_\ell \frac{Y_{\ell'}^{\mathbf{t}}}{\sqrt{n}} + \mathbf{t}_{\ell'} \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} - \mathbf{m}_\ell \frac{Y_{\ell'}^{\mathbf{m}}}{\sqrt{n}} - \mathbf{m}_{\ell'} \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} \right) \\
& \quad - h \sum_{\ell \in [k]} [\mathbf{t}_\ell - \mathbf{m}_\ell] \frac{Z_\ell}{\sqrt{n}} + \frac{1}{\beta} \sum_{\ell \in [k]} \frac{Z_\ell}{\sqrt{n}} \left[ I_{\mathbf{C}} \left( \mathbf{t}_\ell + \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} \right) - I_{\mathbf{C}} \left( \mathbf{m}_\ell + \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} \right) \right] \\
& \quad + \sum_{\ell \in [k]} \omega_\ell \left[ \frac{Y_\ell^{\mathbf{t}}}{\sqrt{n}} [a_\ell K(\mathbf{t}) + h] - \frac{Y_\ell^{\mathbf{m}}}{\sqrt{n}} [a_\ell K(\mathbf{m}) + h] \right] + O(n^{-1}). \tag{6.21}
\end{aligned}$$

Since the random variables  $Y_\ell^{\mathbf{t}}, Y_\ell^{\mathbf{m}}, Z_\ell$  are centred normal, this concludes the proof of Theorem 1.3.

From (6.21) it is possible to compute explicitly the variance of  $Z$  defined in Theorem 1.3, because the variances of all the random variables involved are known (at least to leading order).

**Funding** Open Access funding enabled and organized by Projekt DEAL. AB and SM were supported through DFG Excellence Cluster GZ 2047/1, Projekt-id 390685813, and DFG SFB 1060, Projektnummer 211504053. FdH was supported through NWO Gravitation Grant 024.002.003-NETWORKS, and by the Alexander von Humboldt Foundation during sabbatical leaves in Bonn and Erlangen in the Fall of 2019 and 2020.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory

regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## A. Metastability on the Complete Graph Without Disorder

We give a brief overview of well-known results for the standard Curie–Weiss model. We refer to [6, Chapter 13] for more details.

The Glauber dynamics is defined as in Sect. 1.2, but with  $J \equiv 1$ . For convenience we write the Curie–Weiss Hamiltonian as

$$H_n(\sigma) = -\frac{1}{2n} \sum_{i,j \in [n]} \sigma(i)\sigma(j) - h \sum_{i \in [n]} \sigma(i), \quad \sigma \in S_n, \quad (\text{A.1})$$

which is as (2.5) when  $J \equiv 1$ . What makes this case easier than the one with disorder is that the interaction is *mean-field*. Indeed, we may write

$$H_n(\sigma) = n \left[ -\frac{1}{2} m_n(\sigma)^2 - h m_n(\sigma) \right], \quad (\text{A.2})$$

with

$$m_n(\sigma) = \frac{1}{n} \sum_{i \in [n]} \sigma(i) \in [-1, 1] \quad (\text{A.3})$$

the magnetisation. In this case the magnetisation process  $(m_n(t))_{t \geq 0}$ , defined by

$$m_n(t) = m_n(\sigma_t), \quad (\text{A.4})$$

is Markovian. More specifically, it is a nearest-neighbour random walk on the grid

$$\Gamma_n = \left\{ -1, -1 + \frac{2}{n}, \dots, +1 - \frac{2}{n}, +1 \right\}. \quad (\text{A.5})$$

In the limit as  $n \rightarrow \infty$ , (A.4) converges to a Brownian motion on  $[-1, +1]$  in the potential  $F_{\beta,h}$  given by

$$F_{\beta,h}(m) = -\frac{1}{2} m^2 - h m + \frac{1}{\beta} I(m), \quad (\text{A.6})$$

with

$$I(m) = \frac{1-m}{2} \log \left( \frac{1-m}{2} \right) + \frac{1+m}{2} \log \left( \frac{1+m}{2} \right) \quad (\text{A.7})$$

the relative entropy of the Bernoulli measure on  $\{-1, +1\}$  with parameter  $m$  with respect to the counting measure on  $\{-1, +1\}$ .  $F_{\beta,h}(m)$  is the *free energy* at magnetisation  $m$ , consisting of an *energy term*  $-\frac{1}{2} m^2 - h m$  and an *entropy term*  $\frac{1}{\beta} I(m)$ . See [6, Chapter 13] for more details.

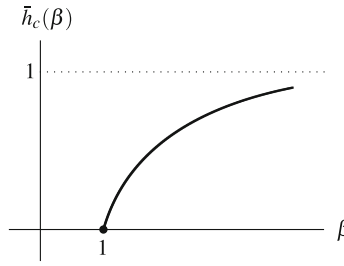


Fig. 1. Plot of  $\beta \mapsto \bar{h}_c(\beta)$

Since

$$F'_{\beta,h}(m) = -m - h + \frac{1}{2\beta} \log \left( \frac{1+m}{1-m} \right), \quad F''_{\beta,h}(m) = -1 - \frac{1}{\beta} \frac{m}{1-m^2}, \quad (\text{A.8})$$

the stationary points of  $F_{\beta,h}$  are the solutions to the equation

$$m = T_{\beta,h}(m), \quad T_{\beta,h}(m) = \tanh[\beta(m+h)]. \quad (\text{A.9})$$

Since

$$T'_{\beta,h}(m) = \beta[1 - T_{\beta,h}^2(m)], \quad (\text{A.10})$$

$T_{\beta,h}$  is strictly increasing and has a unique inflection point at  $m = -h$ . Consequently, (A.9) has either one or three solutions. The latter occurs if and only if

$$\beta \in (\bar{\beta}_c, \infty) \quad \text{and} \quad h \in (0, h_c(\beta)), \quad (\text{A.11})$$

where  $\bar{\beta}_c = 1$  is the *critical inverse temperature* and  $\bar{h}_c(\beta)$  is the *critical magnetic field*, i.e., the unique value of  $h$  for which  $T_{\beta,h}$  touches the diagonal at a unique value of the magnetisation, say  $-m(\beta)$ . Clearly,  $1 = \beta(1 - m^2(\beta))$ , i.e.

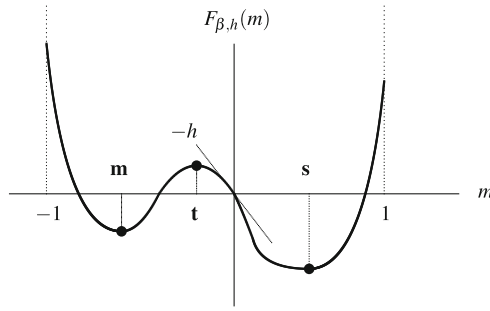
$$m(\beta) = \sqrt{1 - \beta^{-1}}, \quad (\text{A.12})$$

and so  $\bar{h}_c(\beta)$  solves the equation  $T_{\beta,\bar{h}_c(\beta)}(-m(\beta)) = -m(\beta)$ . Hence (see Fig. 1)

$$\bar{h}_c(\beta) = m(\beta) - \frac{1}{2\beta} \log \left( \frac{1+m(\beta)}{1-m(\beta)} \right), \quad \beta \geq 1. \quad (\text{A.13})$$

The range of parameters in (A.11) represents the *metastable regime* in which  $F_{\beta,h}$  has a *double-well* shape and, in the limit as  $n \rightarrow \infty$ , the Gibbs measure  $\mu_n$  in (1.4) has two phases given by the two minima of  $F_{\beta,h}$ : the *metastable phase* with magnetisation  $\mathbf{m} < 0$  and the *stable phase* with magnetisation  $\mathbf{s} > 0$ . The unique *saddle point* in the gate  $\mathcal{G}(\mathbf{m}, \mathbf{s})$  has magnetisation  $\mathbf{t} < 0$  (see Fig. 2).

Theorems A.1–A.2 can be found in Bovier and den Hollander [6, Chapter 13]. Here the notation is the same as the one in Sect. 1. Let  $S_n[\mathbf{m}]$ ,  $S_n[\mathbf{s}]$  denote the sets of configurations in  $S_n$  for which the magnetisation is closest to  $\mathbf{m}$ ,  $\mathbf{s}$ , respectively.



**Fig. 2.** Plot of  $m \mapsto F_{\beta,h}(m)$  for  $\beta, h$  in the metastable regime

**Theorem A.1** (Average crossover time). *Subject to (A.11), uniformly in  $\sigma \in S_n[\mathbf{m}]$ ,*

$$\mathbb{E}_\sigma [\tau_{S_n[s]}] = [1 + o_n(1)] \frac{\pi}{1-t} \sqrt{\frac{1-t^2}{1-m^2}} \frac{1}{\beta \sqrt{F''_{\beta,h}(\mathbf{m})[-F''_{\beta,h}(\mathbf{t})]}} e^{\beta n[F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})]}. \quad (\text{A.14})$$

**Theorem A.2** (Exponential law). *Subject to (A.11), uniformly in  $\sigma \in S_n[\mathbf{m}]$ ,*

$$\mathbb{P}_\sigma (\tau_{S_n[s]} > t \mathbb{E}_\sigma [\tau_{S_n[s]}]) = [1 + o_n(1)] e^{-t}, \quad t \geq 0. \quad (\text{A.15})$$

Figure 2 illustrates the setting: the average crossover time from  $S_n[\mathbf{m}]$  to  $S_n[s]$  depends on the energy barrier  $F_{\beta,h}(\mathbf{t}) - F_{\beta,h}(\mathbf{m})$  and on the curvature of  $F_{\beta,h}$  at  $\mathbf{m}$  and  $\mathbf{t}$ . The crossover time is exponential on the scale of its average.

## B. Examples with Multiple Metastable States

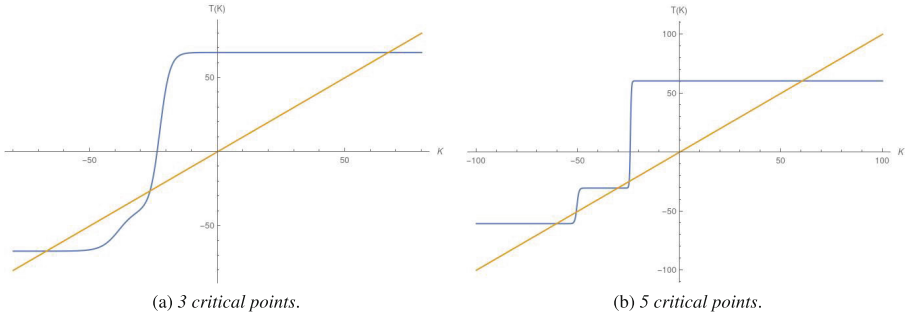
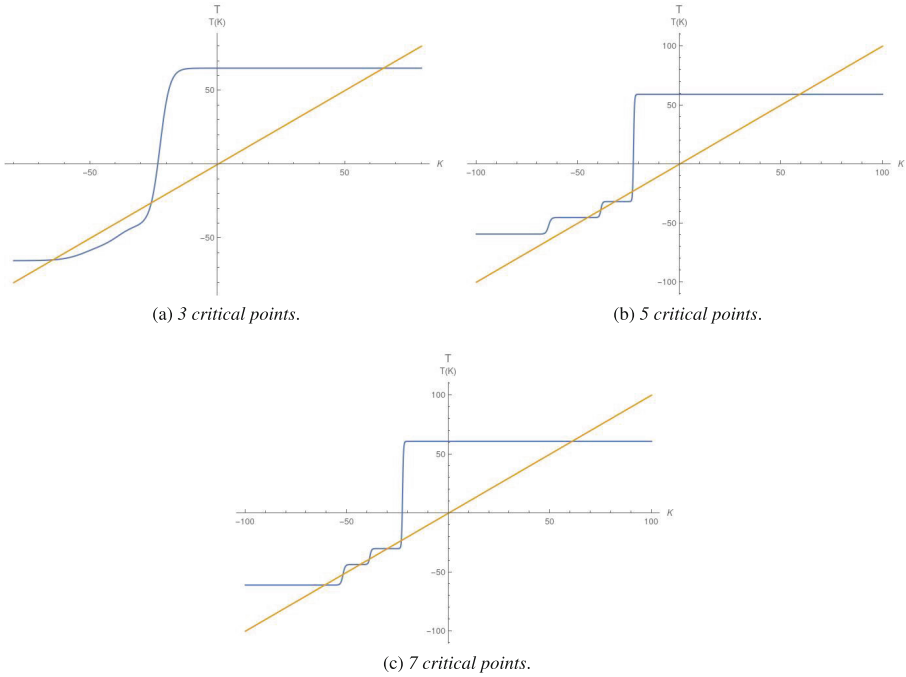
We provide examples of distributions and parameter choices (in the metastable regime) for which the model with disorder has multiple critical points. More specifically, we provide numerical evidence that, for  $k \in \{2, 3, 4\}$ , (3.6) can have any number of solutions in the set  $\{3, 5, \dots, 2k+1\}$ . The cases with strictly more than 3 solutions present multiple minimal critical points, i.e. multiple metastable states.

### B.1. Case $k = 2$ .

- Figure 3a: 3 critical points, parameters  $a_1 = 77, a_2 = 45, \omega_1 = 0.688, h = 1740, \beta = 113 \beta_c$ .
- Figure 3b: 5 critical points, parameters  $a_1 = 774, a_2 = 36.84, \omega_1 = 0.59, h = 1740, \beta = 131 \beta_c$ .

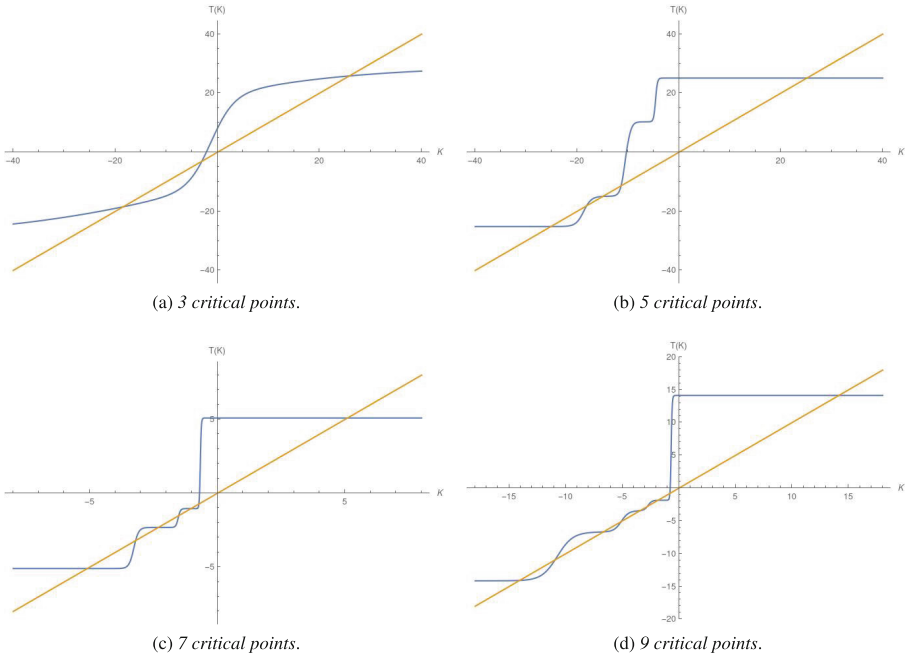
### B.2. Case $k = 3$ .

- Figure 4a: 3 critical points, parameters  $a_1 = 77, a_2 = 45, a_3 = 33.5, \omega_1 = 0.688, \omega_2 = 0.15, h = 1740, \beta = 113 \beta_c$ .
- Figure 4b: 5 critical points, parameters  $a_1 = 77, a_2 = 45, a_3 = 27, \omega_1 = 0.59, \omega_2 = 0.15, h = 1740, \beta = 113 \beta_c$ .
- Figure 4c: 7 critical points, parameters  $a_1 = 77, a_2 = 45, a_3 = 33.5, \omega_1 = 0.59, \omega_2 = 0.15, h = 1740, \beta = 113 \beta_c$ .

**Fig. 3.**  $T_{\beta,h}, k = 2$ **Fig. 4.**  $T_{\beta,h}, k = 3$ 

### B.3. Case $k = 4$ .

- Figure 5a: 3 critical points, parameters  $a_1 = 12, a_2 = 16, a_3 = 139.5, a_4 = 24.5, \omega_1 = 0.474, \omega_2 = 0.22, \omega_3 = 0.111, h = 178, \beta = 3.8 \beta_c$ .
- Figure 5b: 5 critical points, parameters  $a_1 = 14, a_2 = 27, a_3 = 57, a_4 = 24.5, \omega_1 = 0.366, \omega_2 = 0.1, \omega_3 = 0.13, h = 262, \beta = 38.4 \beta_c$ .
- Figure 5c: 7 critical points, parameters  $a_1 = 2.32, a_2 = 4.92, a_3 = 5, a_4 = 11.32, \omega_1 = 0.6, \omega_2 = 0.096, \omega_3 = 0.033, h = 7.6, \beta = 95.2 \beta_c$ .
- Figure 5d: 9 critical points, parameters  $a_1 = 12, a_2 = 16, a_3 = 50.5, a_4 = 24.5, \omega_1 = 0.474, \omega_2 = 0.22, \omega_3 = 0.111, h = 178, \beta = 63.2 \beta_c$ .

Fig. 5.  $T_{\beta,h}, k = 4$ 

### C. Example of $h_c(\beta)$ Not Increasing

We provide here an example of choice of the law of  $J$  for which the critical threshold  $\beta \mapsto h_c(\beta)$  is not monotone increasing. This implies the possibility of a re-entrant metastable crossover.

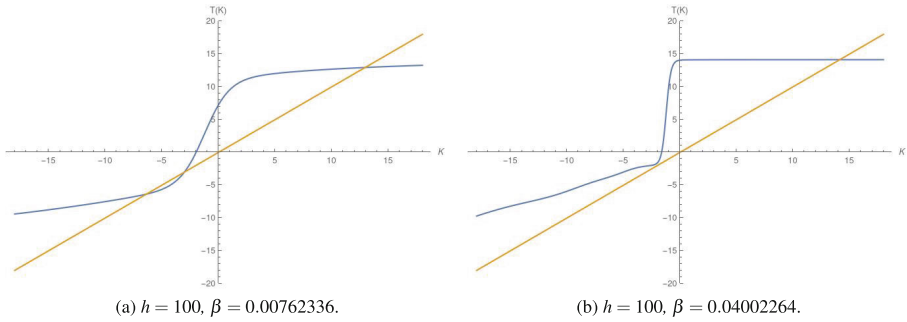
For  $k = 4$ , pick  $a_1 = 12$ ,  $a_2 = 16$ ,  $a_3 = 50.5$ ,  $a_4 = 24.5$  and  $\omega_1 = 0.474$ ,  $\omega_2 = 0.22$ ,  $\omega_3 = 0.111$ . Take  $h = 100$ , and plot the function  $K \mapsto T_{\beta,h}(K)$  varying  $\beta$ . For  $\beta_1 = 4$   $\beta_c = 0.00762336$  the system is metastable:  $T_{\beta,h}$  intersects the diagonal three times (see Fig. 6a), which implies that  $h < h_c(\beta_1)$ . For  $\beta_2 = 21$   $\beta_c = 0.04002264 > \beta_1$  the system is not metastable:  $T_{\beta,h}$  intersects the diagonal only once (see Fig. 6b), which implies that  $h > h_c(\beta_2)$ . This shows that  $h_c(\beta)$  is not necessarily an increasing function of  $\beta$ .

### D. Limit of the Prefactor

Below Theorem 1.2 we stated that the prefactor in (1.19) converges. For completeness, in this Appendix we compute its limit, although, as we mentioned after Theorem 1.3, it is negligible because of the order of approximation of the exponent.

We focus first on  $\gamma_n$ . Recall notation in (1.10), (1.11) and (2.1). Then (4.20) can be written as

$$\begin{aligned}
 & 1 + O(n^{-1}) \\
 &= \sum_{\ell \in [k]} \frac{a_{\ell}^2 \omega_{\ell,n} (1 - \mathbf{t}_{\ell,n}) \exp \left[ -2\beta \left( -a_{\ell} \left( \frac{a_{\ell}}{n} + K_n(\mathbf{t}_n) \right) - h \right)_+ \right]}{\frac{\exp \left[ -2\beta \left( -a_{\ell} \left( \frac{a_{\ell}}{n} + K_n(\mathbf{t}_n) \right) - h \right)_+ \right]}{\beta(1 + \mathbf{t}_{\ell,n})} - 2\gamma_n}
 \end{aligned}$$



**Fig. 6.**  $T_{\beta,h}$ , fixed  $h$  and law of the components of  $J$ , varying  $\beta$

$$= \sum_{\ell \in [k]} \frac{a_\ell^2 \omega_{\ell,n} (1 - \tanh(\beta [a_\ell K_n(\mathbf{t}_n) + h])) \exp[-2\beta (-a_\ell (\frac{a_\ell}{n} + K_n(\mathbf{t}_n)) - h)_+]}{\frac{\exp[-2\beta (-a_\ell (\frac{a_\ell}{n} + K_n(\mathbf{t}_n)) - h)_+]}{\beta(1+\tanh(\beta [a_\ell K_n(\mathbf{t}_n) + h]))} - 2\gamma_n}. \quad (\text{D.1})$$

In the first equality we use (3.3) for  $\mathbf{t}_n$ , i.e., the approximation of the stationary points of  $F_n$  by the stationary points of  $\bar{F}_n$ . This makes  $\mathbf{t}_{\ell,n}$  independent of  $\ell$ , so that we can use the law of large numbers in the limit as  $n \rightarrow \infty$ . Thus, we obtain that  $\gamma_n$  converges to  $\gamma$ , the solution of the equation

$$\mathfrak{E} \left( \frac{J(1)^2 (1 + \tanh U) e^{-2U_+}}{\frac{1}{\beta(1 - \tanh U)} e^{-2U_+} - 2\gamma} \right) = 1, \quad (\text{D.2})$$

where  $\mathfrak{E}$  denotes expectation with respect to  $\mathcal{P}$  and  $U = -\beta[J(1)K(\mathbf{t}) + h]$ , with  $\mathbf{t}$  solving (3.4). Note that (D.2) is similar to [6, Eq. (14.4.14)].

We are left to find the limit of the determinants ratio. By (4.5),

$$\det \mathbb{A}_n(m) = \left( 1 - \sum_{\ell \in [k]} \beta a_\ell^2 \omega_{\ell,n} [1 - (m_\ell)^2] \right) \prod_{\ell' \in [k]} \frac{1}{\beta} \frac{\omega_{\ell',n}}{1 - (m_{\ell'})^2} [1 + O(n^{-1})]. \quad (\text{D.3})$$

Using (3.3) for  $m \in \{\mathbf{t}_n, \mathbf{m}_n\}$ , we have

$$\begin{aligned} & \sum_{\ell \in [k]} \beta a_\ell^2 \omega_{\ell,n} [1 - (m_{\ell,n})^2] \\ &= \sum_{\ell \in [k]} \beta a_\ell^2 \omega_{\ell,n} \left[ 1 - \tanh^2 \left( \beta \left[ a_\ell \sum_{\ell' \in [k]} a_{\ell'} \omega_{\ell',n} m_{\ell',n} + h \right] \right) \right]. \end{aligned} \quad (\text{D.4})$$

Using the law of large numbers as above and with the same notation, we find

$$\lim_{n \rightarrow \infty} \frac{[-\det(\mathbb{A}_n(\mathbf{t}_n))]}{\det(\mathbb{A}_n(\mathbf{m}_n))} = \frac{-1 + \mathfrak{E}(\beta J(1)^2 [1 - \tanh^2 [U(\mathbf{t})]])}{1 - \mathfrak{E}(\beta J(1)^2 [1 - \tanh^2 [U(\mathbf{m})]])} \prod_{\ell' \in [k]} \frac{1 - (\mathbf{m}_{\ell'})^2}{1 - (\mathbf{t}_{\ell'})^2}, \quad (\text{D.5})$$

where  $U(\mathbf{x}) = -\beta(J(1)K(\mathbf{x}) + h)$ .

## References

1. Bianchi, A., Bovier, A., Ioffe, D.: Sharp asymptotics for metastability in the random field Curie–Weiss model. *Electron. J. Probab.* **14**, 1541–1603 (2009). <https://doi.org/10.1214/EJP.v14-673>
2. Bianchi, A., Bovier, A., Ioffe, D.: Pointwise estimates and exponential laws in metastable systems via coupling methods. *Ann. Probab.* **40**(1), 339–371 (2012). <https://doi.org/10.1214/10-AOP622>
3. Bovier, A., Eckhoff, M., Gayraud, V., Klein, M.: Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theory Relat. Fields* **119**(1), 99–161 (2001). <https://doi.org/10.1007/PL00012740>
4. Bovier, A., Eckhoff, M., Gayraud, V., Klein, M.: Metastability and low lying spectra in reversible Markov chains. *Commun. Math. Phys.* **228**(2), 219–255 (2002). <https://doi.org/10.1007/s002200200609>
5. Bovier, A., Eckhoff, M., Gayraud, V., Klein, M.: Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc.* **6**(4), 399–424 (2004). <https://doi.org/10.4171/JEMS/14>
6. Bovier, A., den Hollander, F.: Metastability—A Potential-Theoretic Approach, *Grundlehren der mathematischen Wissenschaften*, vol. 351. Springer, Cham (2015). <https://doi.org/10.1007/978-3-319-24777-9>
7. Bovier, A., Marella, S., Pulvirenti, E.: Metastability for the dilute Curie–Weiss model with Glauber dynamics. *Electron. J. Probab.* **26**, 1–38 (2021). <https://doi.org/10.1214/21-ejp610>
8. Dommers, S.: Metastability of the Ising model on random regular graphs at zero temperature. *Probab. Theory Relat. Fields* **167**, 305–324 (2017). <https://doi.org/10.1007/s00440-015-0682-0>
9. Dommers, S., Giardinà, C., Giberti, C., van der Hofstad, R., Prioriello, M.L.: Ising critical behavior of inhomogeneous Curie–Weiss models and annealed random graphs. *Commun. Math. Phys.* **348**(1), 221–263 (2016). <https://doi.org/10.1007/s00220-016-2752-2>
10. Dommers, S., den Hollander, F., Jovanovski, O., Nardi, F.R.: Metastability for Glauber dynamics on random graphs. *Ann. Appl. Probab.* **27**, 2130–2158 (2017). <https://doi.org/10.1214/16-AAP1251>
11. den Hollander, F., Jovanovski, O.: Glauber dynamics on the Erdős–Rényi random graph. In: *In and Out of Equilibrium. 3. Celebrating Vidas Sidoravicius. Progress in Probability*, vol. 77, pp. 519–589. Birkhäuser/Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-60754-8\\_24](https://doi.org/10.1007/978-3-030-60754-8_24)
12. Mathieu, P., Picco, P.: Metastability and convergence to equilibrium for the random field Curie–Weiss model. *J. Stat. Phys.* **91**(3–4), 679–732 (1998). <https://doi.org/10.1023/A:1023085829152>
13. Olivieri, E., Vares, M.E.: Large Deviations and Metastability. *Encyclopedia of Mathematics and its Applications*, vol. 100. Cambridge University Press, Cambridge (2005). <https://doi.org/10.1017/CBO9780511543272>
14. Tindemans, P.A.J., Capel, H.W.: An exact calculation of the free energy in systems with separable interactions. *Physica* **72**(3), 433–464 (1974). [https://doi.org/10.1016/0031-8914\(74\)90209-2](https://doi.org/10.1016/0031-8914(74)90209-2)

Communicated by H. Spohn