Chapter 2
Aims and Scopes

What a convenient thing it would be if all thieves had the same shape! It's so confusing to have some of them quadrupeds and others bipeds! (Lewis Carroll, Sylvie and Bruno)

While *classical mechanics* is concerned with deterministic equations of motion, *statistical mechanics* adopts the view that many-particle systems can best be described with the help of probabilistic techniques that do justice to the intrinsic complexity of these systems and to our incomplete knowledge of the precise microscopic state they are in. The aim of *equilibrium statistical mechanics* is to describe many-particle systems through Gibbs distributions, i.e., probability distributions on configuration spaces given by Boltzmann weight factors based on interaction Hamiltonians.

A key target is the computation of the *free energy* of the system as a function of macroscopic or mesoscopic parameters. (Gibbs distributions minimise the free energy according to the Gibbs variational principle.) The idea is that the free energy, which captures the *equilibrium* (= static) properties of the system, implicitly contains information that is pertinent to the *non-equilibrium* (= dynamic) properties of the system as well, since both depend on the energy landscape encoded in the interaction Hamiltonian, with the Gibbs distribution being invariant under the dynamics. A guiding principle of the present book is to make this idea as precise as possible within the context of metastability. On the intuitive level, we deal with Gibbs distributions that put *relatively* large weight on disjoint sets with different macroscopic properties ("metastable states"), separated by regions of small weight through which the system cannot move easily ("saddles"). To make this intuition precise we must make certain assumptions on the dynamics. Typically we must assume that the dynamics is either *local* or *diffusive*, meaning that long-range jumps are either rare or are excluded.

In Sect. 2.1 we describe two paradigmatic models for metastability that serve as a *red thread* through much of the book. In Sect. 2.2 we explain the importance of model reduction, linking complex realistic models to simple toy models that capture the metastable behaviour on an aggregate level. This reduction is necessary to understand the universality behind metastable phenomena. In Sect. 2.3, we give a brief outline of the *variational point of view* that is central to the potential-theoretic approach to metastability—the subject of the present monograph—thereby expanding further on what was already written in Sect. 1.3.4. In Sect. 2.4 we provide a list

A. Bovier, F. den Hollander, *Metastability*,
Grundlehren der mathematischen Wissenschaften 351,
DOI 10.1007/978-3-319-24777-9_2

of the models to be considered in Parts IV–VIII, with a brief indication of what we prove about them and how we organise them. In Sect. 2.5 we mention a number of related topics that are not treated in this book but are slowly coming within reach of mathematical theory.

2.1 Two paradigmatic models

In Sects. 2.1.1–2.1.2 we describe two paradigmatic models for metastability: the Kramers model for Brownian motion in a double-well potential and finite-state Markov chains with exponentially small transition probabilities.

2.1.1 Kramers model: Brownian motion in a double-well

One of the first *mathematical models* for metastability was proposed in 1940 by Kramers [157]. It consists of the one-dimensional *diffusion equation* (or Langevin equation in physics terminology)

$$dX_t = b(X_t)\,dt + \sqrt{2\varepsilon}\,dB_t, \tag{2.1.1}$$

where X_t denotes the position at time t of a "particle" diffusing in a drift field $b = -W'$, with $W\colon \mathbb{R} \to \mathbb{R}$ a *double-well potential*, i.e., a function with two local minima and two steep walls (see Fig. 2.1), and B_t denotes the position at time t of a standard Brownian motion.[1]

Equation (2.1.1) has become the *paradigm of metastability*. Kramers was able to settle essentially all the interesting questions related to this model. In particular, he derived the so-called *Kramers formula* for the average transition time from a local minimum at u to a global minimum at v via a saddle point at z^* (see Fig. 2.1):

$$\mathbb{E}_u[\tau_v] = \big[1 + o(1)\big]\frac{2\pi}{\sqrt{[-W''(z^*)]W''(u)}}\exp\big[\big(W(z^*) - W(u)\big)/\varepsilon\big]. \tag{2.1.2}$$

This formula fits the classical Arrhenius law with *activation energy* $E = W(z^*) - W(u)$, *amplitude* $A = 2\pi/\sqrt{[-W''(z^*)]W''(u)}$ and inverse temperature $\beta = 1/kT = 1/\varepsilon$. Note that the flatter W is near z^* and u, the larger is the amplitude: flatness slows down the crossover at z^* and increases the number of returns to u.

[1] In fact, (2.1.1) emerges as a special case of the more general equation considered by Kramers, namely, the Ornstein-Uhlenbeck equation $dX_t = V_t dt$, $\mu^{-1} dV_t = -dX_t + b(X_t)dt + \sqrt{2\varepsilon}dB_t$, where V_t denotes the velocity at time t and μ is a friction parameter. This equation gives rise to (2.1.1) in the limit as $\mu \to \infty$. Thus, (2.1.1) can be seen as the equation of motion of a particle moving under the influence of a friction force, a gradient force and a random force in the limit where the friction becomes infinitely strong.

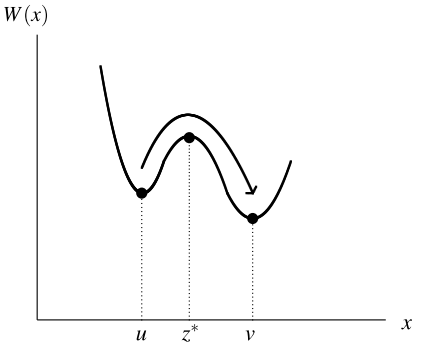

Fig. 2.1 A double-well potential with a local minimum at u, a global minimum at v and a saddle point at z^*

Formula (2.1.2) exhibits a structure that is typical for metastable systems. There is an *exponential term*, here given by $\exp[(W(z^*) - W(u))/\varepsilon]$, which provides the leading asymptotic behaviour. The pathwise approach to metastability, which is based on large deviation theory, typically is capable to identify this term, sometimes referred to as the *exponential asymptotics*, by showing that

$$\varepsilon \ln \mathbb{E}_u[\tau_v] = \big[1 + o(1)\big]\big(W(z^*) - W(u)\big), \quad \varepsilon \downarrow 0. \tag{2.1.3}$$

However, identifying the *prefactor*, here given by $2\pi/\sqrt{[-W''(z^*)]W''(u)}$, is in general a far more subtle problem. It is the ambition of the potential-theoretic approach exhibited in this book to provide a *unified framework* that allows to obtain rigorous asymptotic formulas as in (2.1.2) with an explicit *prefactor* for a wide class of metastable systems. We will see plenty of examples in Parts VI–VII.

The multi-dimensional generalisation of (2.1.2) is attributed to Eyring and is called the *Eyring-Kramers formula* (see Glasstone, Laidler and Eyring [127], Weidenmüller and Zhang [236], Maier and Stein [170]). Actually, Eyring's so-called transition-state theory [106] is based on quantum-mechanical considerations and is different from the classical theory of Kramers. It interprets the potential as a restricted quantum-mechanical free energy. For a historical discussion, see Pollak and Talkner [201].

2.1.2 Finite-state Markov processes

The model of Kramers describes the evolution of an *effective order parameter* of a metastable system driven by *diffusive noise*. It is clear that in this model the "particle" spends most of its time close to the two local minima of the double-well potential. This suggests a further simplification of the picture, namely, a reduction to a *two-state system*, with the two states u, v representing the two wells of the potential. The time at which this system jumps from state u to state v and backwards

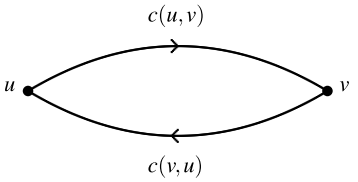

Fig. 2.2 Transition rates for the two-state Markov chain

can be reasonably approximated by the first hitting time τ_v of the local minimum v starting from the local minimum u, and vice versa for τ_u. As we will see in Parts IV–V, in the limit as $\varepsilon \downarrow 0$ the times τ_v and τ_u normalised by their expectations tend to exponentially distributed random variables. This means that a rough approximation of the long-term behaviour of the Kramers model is given by a *continuous-time Markov chain* with state space $\{u, v\}$ and transition rates (see Fig. 2.2)

$$\begin{aligned} c(u,v) &= \mathrm{e}^{-r(u,v)/\varepsilon}, \qquad & r(u,v) &= W\big(z^*\big) - W(u), \\ c(v,u) &= \mathrm{e}^{-r(v,u)/\varepsilon}, \qquad & r(v,u) &= W\big(z^*\big) - W(v). \end{aligned} \tag{2.1.4}$$

The average crossover times $\mathbb{E}_u[\tau_v] = 1/c(u,v)$ and $\mathbb{E}_v[\tau_u] = 1/c(v,u)$ capture the leading order asymptotics of (2.1.1), as expressed in (2.1.3).

The above setting can be easily generalised to systems with multiple metastable states. An effective model for such systems would be a continuous-time Markov chain with a *finite* state space $\mathscr{M} = \{m_1, \dots, m_n\}$ and transition rates $c(m_i, m_j) = \exp[-r(m_i, m_j)/\varepsilon]$, $i, j = 1, \dots, n$. The basic task of a theory of metastability is to determine these transition rates from first principles. This idea was properly formalised by Freidlin and Wentzell [115] in the context of small random perturbations of dynamical systems. In their theory the coefficients $r(m_i, m_j)$ are computed with the help of the theory of *large deviations on path space* (see Chap. 6).

Finite-state Markov chains with exponentially small transition rates have become a subject of interest by themselves. By allowing the transition rates to be either exponentially small *or* equal to one, the above picture is capable of describing models from statistical physics, in particular, *spin-flip systems* and *lattice gases* in finite volumes at low temperatures, with ε playing the role of temperature (see Part VI).

The analysis of the metastability properties of finite-state Markov chains is a non-trivial problem in itself. In the early 1990's an intense activity in this direction started with the work of Catoni and Cerf [53] and Olivieri and Scoppola [196, 197]. The methods used were, once again, large deviations on the path space of these Markov chains. The difficulties that arise in the analysis of specific models are essentially of a combinatorial nature: the optimal paths for transitions between metastable states need to be identified and to be counted. This leads to interesting problems, such as the *discrete isoperimetric inequalities* studied in Alonso and Cerf [4], which we will encounter in Part VI. Only later, in the 2000's, was it noted that potential theory is very well suited to simplify the analysis and to draw sharper results from the same input, as first pointed out by Bovier and Manzo [39] and later amplified in Bovier, den Hollander and Nardi [31].

2.2 Model reduction

Kramers model and finite-state Markov chains can both be seen as simple *toy models* that ought to be *derivable* from more complex *realistic models* of interest. Ideally, we would like to start with many-body systems of interacting quantum particles. This, however, is beyond present-day technology. The most complex models we will consider in this book are classical interacting particle systems, in particular spin-flip systems and lattice gases. These are Markov processes with a high-dimensional (sometimes even infinite-dimensional) state space. Typically, the noise on the level of the microscopic dynamics is not small, and the large-scale dynamics of the system depends on the interplay between energetic and entropic effects.

It is generally accepted in the physics and chemistry literature that reduced models, describing the time evolution of the system on an *intermediate aggregate level* of mesoscopic variables, provides a good description of metastable behaviour. Examples of such models are stochastic differential or partial differential equations with small noise. Ideally, such effective dynamics should be derived with the help of *coarse-graining techniques*, in the spirit of the renormalisation group theory in equilibrium statistical mechanics (see the monograph by Presutti [202]). However, this derivation is quite problematic, partly because renormalisation maps typically do not preserve the Markovian nature of the dynamics. An even more serious issue is that, while at least formally *deterministic* evolution equations (like the Allen-Cahn equation [3] treated in Chap. 12) can be derived as scaling limits (i.e., laws of large numbers in probabilistic language), a proper understanding of metastability requires that we move beyond the deterministic limit and retain at least part of the *random* perturbations of the dynamics. In the literature this goes by the name of *diffusion limits*. However, there are subtle and poorly understood issues regarding the proper choice of the noise term. In this book we will treat diffusion processes with small noise as interesting models in their own right in Part IV. The issue of the derivation of mesoscopic dynamics from microscopic dynamics in the mean-field setting will be touched upon in Part V.

2.3 Variational point of view

The focus of this book is on the potential-theoretic approach to metastability. The basic ideas are classical: many probabilistic quantities can be represented as solutions of Dirichlet problems. The usefulness of this observation may appear to be limited, as it amounts to having to solve partial differential equations or discrete analogues thereof. In general, no explicit analytic solutions of such problems are available. Two notable exceptions are one-dimensional diffusions (which is the reason for the solvability of the Kramers model) and one-dimensional nearest-neighbour random walks.

The power of the potential-theoretic approach arises from the fact that it avoids to solve the Dirichlet problem. Instead, it makes use of a representation formula

for the Green function in terms of *capacities*, the invariant measure and harmonic functions. Since renewal arguments can be used to control harmonic functions by capacities, the key objects of the theory are capacities and the invariant measure. The great advantage of this approach materialises in the context of *reversible* Markov processes, i.e., Markov processes whose semi-groups are self-adjoint operators in an L^2-space with respect to an invariant measure. This provides the main weapon of the method: the *Dirichlet principle* expresses capacities as infima of the *Dirichlet form* over classes of functions that are constrained by boundary conditions. The usefulness of this variational principle has long been recognised, e.g. in the analysis of finite-state Markov chains. The book by Doyle and Snell [96] is an excellent source for this material. For a more recent exposition, see Levin, Peres and Wilmer [163]. Part II of the book provides the background on potential theory of reversible Markov processes that is necessary to deal with problems of metastability.

As a variational problem, the Dirichlet principle is a simple instrument to turn physical intuition into upper bounds, and the sharpness of these upper bounds is limited by diligence and imagination only. A particularly nice aspect of the Dirichlet problem is that it satisfies certain monotonicity properties with respect to underlying parameters. In fact, on this basis Berman and Konsowa [23] derived a *dual variational principle* that expresses capacities (in the case of a discrete state space) in terms of suprema over flows (similar to, but different from the better known *Thomson principle*), which we call the *Berman-Konsowa principle*. As an upshot, the latter allows for the derivation of lower bounds that complement the upper bounds obtained via the Dirichlet principle. It is a rather remarkable fact that in many examples upper and lower bounds can be obtained for the metastable crossover time that differ by a multiplicative factor of the form $1 + o(1)$ only, where $o(1)$ tends to zero as the time scale of the metastable system tends to infinity. We will see these ideas at work in a variety of examples throughout the book. Part III outlines the basic techniques that are needed to implement these ideas.

A *key observation* is that the analysis of the Dirichlet principle and the Berman-Konsowa principle in essence is part of *equilibrium* statistical physics, since it deals with acquiring the relevant knowledge of the free energy landscape of the system. Potential theory links this knowledge to the metastable dynamics of the system, which is part of *non-equilibrium* statistical physics.

2.4 Specific models

The following models will be considered in Parts IV–VIII.

- In Part IV we study *diffusions with small noise*. Chapter 10 deals with diffusions on lattices with small spacings, the simplest setting in which the potential-theoretic approach to metastability can be applied. Under certain regularity assumptions on the transition probabilities (for discrete time) or transition rates (for continuous time), we carry out a detailed calculation of metastable crossover times. Chapter 11 considers finite-dimensional diffusions on subsets of $\mathbb{R}^d$ and

sharpens the classical results of Freidlin-Wentzell theory by using the potential-theoretic approach. The Kramers formula is generalised to a d-dimensional diffusion in a general potential satisfying minimal structural assumptions, a link is made with the principal eigenvalue of the generator of the diffusion, and the exponential distribution of the crossover time is established. Chapter 12 looks at stochastic partial differential equations, which are the infinite-dimensional analogues of the diffusions dealt with in Chap. 11, and shows that similar results apply for a particular example called the Allen-Cahn equation. The theory is complete in one dimension, but suffers from difficulties in higher dimensions, where the noise has to be "truncated properly".

- In Part V we deal with models that allow for *coarse-graining*, i.e., a lumping of states that leads to a simpler Markov process on a reduced state space. Chapter 13 analyses the Curie-Weiss model (the archetype model for ferromagnetism) subject to Glauber spin-flip dynamics. The metastable behaviour of the magnetisation can be fully computed in the limit as the volume tends to infinity, at any subcritical temperature, and turns out to be similar to that of the Kramers model. Chapters 14–15 extend the analysis to the random-field Curie-Weiss model. If the support of the distribution of the magnetic field is finite (Chap. 14), then this model behaves similarly as the Curie-Weiss model, with a large-volume metastable behaviour that is like the Kramers model when the dimension is equal to the size of the support. The computations become much more complicated when the support is infinite (Chap. 15), in which case delicate coupling techniques are required.
- Part VI looks at lattice models subject to a Metropolis dynamics in a *finite volume in the limit as the temperature tends to zero*. Chapter 16 explains how the potential-theoretic approach can be used to prove that these models have the same metastable behaviour as the two-state Markov chain, provided a number of minimal hypotheses are satisfied. Two other dynamics are briefly discussed as well, namely, heat-bath dynamics and probabilistic cellular automata, for which the same universal metastable behaviour can be derived under similar hypotheses. Chapter 17 settles the hypotheses in the case of Glauber dynamics for Ising spins. Chapter 18 in the case of Kawasaki dynamics for lattice gas particles.
- Part VII looks at nucleation in lattice systems that *grow to infinity as the temperature tends to zero*. Spatial entropy comes into play: in large volumes, even at low temperatures, entropy is competing with energy because the metastable state and the states that evolve from it under the dynamics have a non-trivial spatial structure. The main idea is that the system exhibits "homogeneous nucleation", i.e., after the large volume is divided up into smaller (but still large) subvolumes, the system is found to behave more or less independently in different subvolumes. Chapter 19 looks at Glauber dynamics, Chap. 20 at Kawasaki dynamics. For the latter, the computations are delicate and require a proof of "equivalence of ensembles" in a dynamical setting.
- Part VIII describes lattice systems at *high densities*. The focus in Chap. 21 is on the zero-range process, which consists of a collection of particles performing continuous-time simple random walks with on-site attraction and no on-site repulsion. We consider the limit where the particle density is high, show that the

process spends most of its time in a "condensed state", i.e., a configuration where most of the particles pile up on a single site, and prove that the process evolves via a "metastable hopping" of this pile from one site to another. Both the hopping time and the hopping distribution are computed.

The different parts on applications can essentially be read independently and have the following substructure:

- Part IV: (diffusions with small noise)

  (10) Discrete diffusions
  (11) Continuous diffusions $*$
  (12) Stochastic partial differential equations $*$

- Part V: (coarse-graining in large volumes at positive temperatures)

  (13) Curie-Weiss mean-field model
  (14) Curie-Weiss in discrete random magnetic field
  (15) Curie-Weiss in continuous random magnetic field $*$

- Part VI: (lattice systems in small volumes at low temperatures)

  (16) General theory
  (17) Glauber dynamics
  (18) Kawasaki dynamics

- Part VII: (lattice systems in large volumes at low temperatures)

  (19) Glauber dynamics $*$
  (20) Kawasaki dynamics $*$

- Part VIII: (lattice systems in small volumes at high densities)

  (21) Zero-range dynamics $*$

The chapters without $*$ concern models where the state space is *simple* (e.g. discrete and finite) and a complete description of the metastable behaviour is achieved. The chapters with $*$ concern models where the state space is *not simple* (e.g. continuous and infinite) and only partial results are obtained.

2.5 Related topics

Apart from being manifest in interacting particle systems, metastability is an important feature of complex systems in general. Topics within reach that will not be considered in this monograph include:

- *Ageing*: A random dynamics goes through a cascade of metastable states in which it gets trapped on increasingly larger space-time scales. As a consequence, the decay of space-time correlation functions depends on the age of the system. Examples are random walks in random environments, used to describe spin glass dynamics (Bouchaud, Cugliandolo, Kurchan and Mézard [30], Ben Arous, Bovier and Gayrard [18]).

- *Conformational dynamics*: Large (bio)-molecules undergo transitions between metastable states (= conformations) under the influence of thermal noise. There is a strong application-driven interest in the numerical identification of these states and of their lifetimes (Grassberger, Barkema, Nadler [129], E, Ren and Vanden-Eijnden [99], Schütte and Sarich [217]). It is a challenge to develop a rigorous mathematical framework to analyse such systems (Caputo, Lacoin, Martinelli, Simenhaus and Toninelli [50]).
- *Population dynamics*: Selective sweeps in genetic populations, triggered by mutations that drive the population from one dominant trait to another, can be viewed as transitions between metastable states (Dawson and Greven [75]). Viruses moving through a complex network may cause an epidemic. The epidemic may be interpreted as a metastable state of the network, which lasts until the virus disappears. This metastable state depends sensitively on the size and the architecture of the network (Chatterjee and Durrett [56], Mourrat and Valesin [187]).
- *Gene regulatory networks*: The genetic information encoded in DNA fixes the topology of the network. Transitions between the various phenotypic states can be understood as crossovers between metastable states of the corresponding dynamical system subject to noise (Kauffmann [149], Huang [141]).

http://www.springer.com/978-3-319-24775-5

Metastability
A Potential-Theoretic Approach
Bovier, A.; den Hollander, F.
2015, XXI, 581 p. 96 illus., 14 illus. in color., Hardcover
ISBN: 978-3-319-24775-5