Chapter 1
Background and Motivation

In Science—in fact, in most things—it is usually best to begin at the beginning. In some things, of course, it's better to begin at the other end. For instance, if you wanted to paint a dog green, it might be best to begin with the tail, as it doesn't bite at that end.
(Lewis Carroll, Sylvie and Bruno Concluded)

We begin with a brief description of the *phenomenon* of metastability (Sects. 1.1–1.2) and a brief historical perspective of the *mathematical theories* that were developed to obtain a quantitative understanding of this phenomenon (Sect. 1.3).

1.1 Phenomenology

Metastability is a widespread phenomenon that arises in a large variety of systems—physical, chemical, biological or economic. A simple experiment anyone can do at home goes as follows. Fill a plastic bottle with distilled water and put it into the freezer. After an hour or so, carefully take it out of the freezer. If you are lucky, then the water still is liquid, but the temperature is down to somewhere between minus 5 and minus 10 degrees centigrade. Now slowly pour the water out of the bottle and into a bowl. When done carefully, you should see a very fast freezing of the water as it hits the bowl. What happens is that the water is *undercooled*, i.e., the *stable state* of the water would have been ice, but you found it in the freezer in a *metastable state*. This state is very sensitive to perturbations, and the shaking you subject it to when pouring it into the bowl triggers an immediate transition to the stable state, which is ice. Should you have left the bottle in the freezer unperturbed, such a transition would eventually have happened *spontaneously*. In fact, if you were to watch the bottle in the freezer over a long time, then you would eventually see this quick freezing happen (you may need a lot of patience). Moreover, if you repeat this experiment many times, then you will observe that the time until freezing is rather variable, and is much longer than the time of the actual freezing itself. It is reported that similar phenomena occur in very still and clean mountain lakes in winter. The water cools well below the freezing point until suddenly the lake freezes over.

Similarly, in chemistry the mixing of two reactive compounds (like oxygen and hydrogen) may lead to a metastable state that can persist for a very long time, but when triggered (by a spark) transits very rapidly to the stable state (water). In economics, stock prizes may persist for a long time on high levels, in spite of economists warnings that the market is "overheated", until a "crash" occurs and prices drop within days or hours to much lower levels. In economics jargon, there was a "bubble" that has collapsed. Such phenomena are ubiquitous. The common features are: a large variability in the moment of the onset of some dramatic change in the properties of the system, a much shorter time for the actual transition (i.e., between the onset of a noticeable change and the moment a new state is reached), and *unpredictability* of the time of the onset of the transition.

What is behind all of this? A simple thought experiment reveals a possible mechanism. Suppose that in a mountain range there are two valleys, A and B. Reaching valley B from valley A requires climbing up 1 km on a steep slope. An experienced mountaineer will easily make this journey in a fairly predictable time, say, 4 hours. Now suppose that there is a tourist visiting valley A, a drunken tourist who wanders around in the valley without any particular purpose and occasionally climbs up the slope towards valley B. However, as he does so, he will encounter certain obstacles or will get tired, and just slurps back to the base of the valley (where, for the sake of the argument, he will have some drinks to retain his confused state of mind). As our tourist is not terribly interested in getting to the second valley anyway, we may assume that he does not learn anything from his excursions up the slope and after each visit to the local pub finds himself in just the same condition as before. Let us now assume that, after many days, we find our tourist in valley B. What has happened? Well, after many failed excursions uphill, and equally many returns to the pub, on a lucky day he just happened to climb straight up the slope and then tumbled down into valley B. Should anyone have observed this final successful climb, he might not have been able to distinguish the tourist from the experienced mountaineer, who would have taken the same path on purpose in the first place. A rough estimate reveals how long it took the tourist to get to valley B: the number of attempts (returns to the pub) will be on average $1/p$ when p is the probability to get over the edge before returning to the pub. If the average time Δ of such an unsuccessful excursion is not too tiny (say, 30 minutes), then the average time Δ/p until the final crossing can run into years when p is small (as is to be expected). Moreover, given the fact that our tourist is not learning anything (and given that no other conditions are changing over time, such as the weather), the fact that at each time back to the pub the tourist is back to where he started implies that the number of failed attempts, and hence the total time until the final crossing, are essentially unpredictable. Thus, in this simple example we recognise and understand all the features of the metastability phenomenon mentioned above. As we will see later, this little thought experiment indeed captures all the crucial features behind metastability.

The first main challenge of *mathematics* is of a *qualitative* nature, namely, to explain why in a large variety of systems the same type of metastable behaviour is observed. Many such systems can be described from first principles as many-body systems subject to classical or quantum dynamics. While the corresponding equations of motion are known, they are typically very hard to analyse, in particular,

over the extremely long time intervals in which metastable behaviour occurs. Also, metastability manifestly exhibits *randomness* (the unpredictable time of the occurrence of the transition), the source of which may be difficult to extract from the underlying deterministic dynamics. It may be due to quantum effects, or external perturbations of a (non-closed) system, or the effect of unresolved high-frequency degrees of freedom. A first simplification is to pass to a description of the system as a *stochastic dynamics*. The justification of such a description is an interesting topic in itself, which will not be addressed in the present book. Rather, a stochastic model of the dynamics of the systems we are interested in will be the starting point of our analysis of metastability. Even more restrictively, we will limit our analysis to *Markov processes*. Still, even within this restricted setting, there is a wide variety of different models where metastability emerges and where the explanation of the underlying universality is possible.

The second main challenge is of a *quantitative* nature. Given the parameters of some underlying model, we would like to be able to compute *as precisely as possible* the quantities controling the metastable phenomena, in particular, the distribution of the times of the transitions between metastable and stable states. Again, this is hard because most metastable systems of practical relevance are many-body systems whose dynamics is not easy to capture, neither analytically nor numerically, and because extremely long time scales may be involved. (See Newman and Barkema [194] for an overview on Monte Carlo methods in statistical physics.) Understanding metastability on the quantitative level is of considerable practical interest, as it affects the behaviour and functioning of many systems in nature.

1.2 Condensation and magnetisation: from gases to ferromagnets

From the point of view of *statistical mechanics*, metastability is the dynamical signature of a *first-order phase transition*. In *equilibrium* statistical physics, a first-order phase transition is said to occur if a system is sensitive to the change of a parameter (or a boundary condition), in the sense that certain *extensive variables* (such as density or magnetisation) show a discontinuity as functions of certain *intensive variables* (such as pressure or magnetic field). Dynamically, this sensitivity manifests itself in the fact that, as the parameter is varied across the phase transition curve, the system remains for a considerable amount of time (typically random) in the "old phase" before it suddenly changes to the "new phase", the true equilibrium phase. In other words, the extensive variables change their value as a function of time with a random delay. Thus, the study of metastability can be seen as part of *non-equilibrium* statistical physics. Let us discuss this in a bit more detail in an example.

The most commonly observed occurrence of metastability is the phenomenon of condensation of over-saturated water vapour (rainfall). The common explanation of what is going on can be found in elementary physics textbooks. If water vapour is
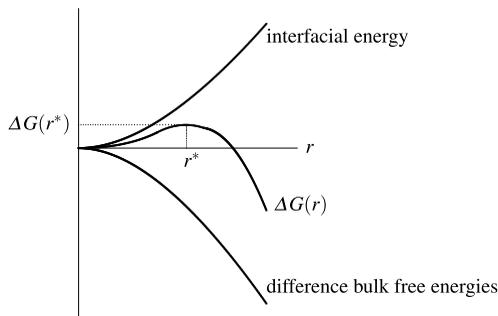
Fig. 1.1 Effective free energy $\Delta G(r)$ of a droplet as a function of its radius r (*middle curve*). The threshold for condensation is the critical radius r^*

cooled below the critical temperature, then the free energy of the gas-phase is larger than that of the liquid-phase. Therefore, thermodynamics predicts a transition from the gas-phase to the liquid-phase. However, this transition can only be achieved by an aggregation of water molecules. This aggregation has to start somewhere in the system with the formation of small droplets of liquid. The key point is that the effective free energy of such a droplet is made up of two terms: (1) the difference between the bulk free energies of the two phases; (2) the interfacial energy between the two phases. This leads to a formula of the type (see Fig. 1.1)

$$\Delta G(r) = \text{difference bulk free energies} + \text{interfacial energy} = -\Delta r^d + \sigma r^{d-1},$$
$$(1.2.1)$$

where $\Delta, \sigma > 0$ represent the effect of (1) and (2) per unit volume, respectively, per unit surface, and r is the radius of the droplet.

The function in (1.2.1) is increasing and positive up to a value r^*, the *critical radius*, and is decreasing afterwards. This means that it is unfavourable for the system to have small droplets and favourable to have large droplets: indeed droplets with a radius smaller than r^* tend to evaporate while droplets with a radius larger than r^* tend to grow. But how is it possible to create a large enough droplet of liquid-phase within the gas-phase? The answer is: by *thermal fluctuations*, i.e., the system on some small scale temporarily violates the laws of thermodynamics and evolves in directions that locally increase the free energy. In this way it can form critical droplets of liquid, and once this is done these droplets can continue to grow in full agreement with the laws of thermodynamics. If the parameters of the system are such that r^* is large, then the fluctuations that produce such supercritical droplets are very rare, which leads to a long lifetime of the *metastable* gas-phase. Thus, the crossover is triggered by the appearance of a *critical droplet* of the new phase inside the old phase, which subsequently grows and invades the system. Just as in the example of our drunken tourist, *the transition from the metastable state is characterised by many unsuccessful attempts of the system to create a critical droplet.*

The adhoc notion of "thermal fluctuations" may appear rather mysterious, and does require explanation and theory. We will see in Chaps. 18 and 20, in the context of *stochastic* models for the dynamics of lattice gases, that the above picture arises
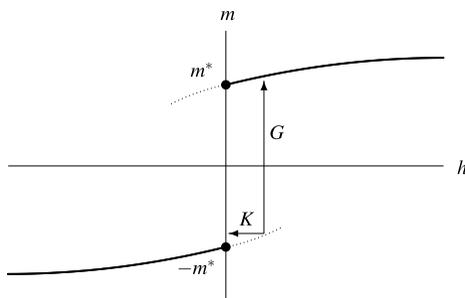
Fig. 1.2 Hysteresis in ferromagnets: plot of the magnetisation m versus the magnetic field h. The *dotted pieces* refer to the magnetisation of the metastable states. The *arrows* refer to the metastable crossovers. The *symbols G and K* stand for Glauber and Kawasaki dynamics (to be treated in Part VI), for which the magnetisation is not preserved, respectively, is preserved

very naturally and can be fully quantified. In this context, the excess free energy of a critical droplet is called the *free energy barrier* for the onset of the phase transition. The presence of such a barrier is the reason for the metastable behaviour, and thermal fluctuations are the driving force for transitions out of the metastable state. The formation of a critical droplet (i.e., a droplet of critical radius) is the minimal effort these fluctuations have to make to initiate the phase transition dynamically. Of course, the same explanation applies when a liquid freezes and—reversed in time—when a liquid evaporates or a solid melts. The fine details are different, but the overall picture is the same.

Another situation where the same principles are at work is magnetic hysteresis, which is treated in Chaps. 13, 17 and 19 (see Fig. 1.2). When a ferromagnetic material is placed in a magnetic field h it *magnetises*, i.e., the atomic magnetic moments ("spins") tend to align with the field. At temperatures below the so-called Curie temperature, this magnetised state persists (forever) even when the field is turned off: the spontaneous magnetisation is $m^\star$. This persistence is the sign of a first-order phase transition. Moreover, even when afterwards the direction of the field is inverted, the magnetisation will remain in the old direction and will only align with the new direction after some time. The reason is the same as for the supersaturation of a gas: the ferromagnetic material has to create local droplets with the opposite magnetisation, and these droplets become energetically favourable and hence start to grow only after they have acquired some minimal size. The creation of such critical droplets is again the work of thermal fluctuations.

Ferromagnets are particularly easy to manipulate and very precise measurements are possible. Figure 1.2 is the paradigmatic figure for metastable behaviour, and is held to be ubiquitous in all situations of metastability.

1.3 Historical perspective

The study of metastability has a long and rich history. In this section we give a brief summary of the most important developments.
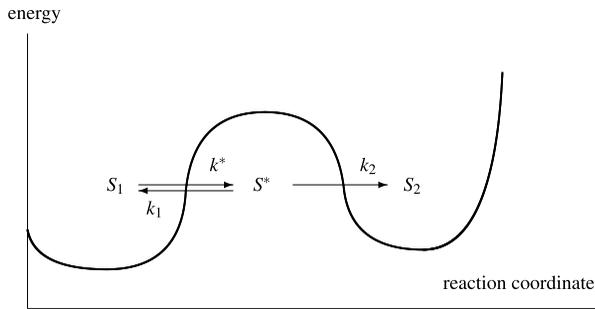
Fig. 1.3 Chemical reaction from state S_1 to state S_2 via transition state S^* with reaction rates k^*, k_1 and k_2

1.3.1 Early achievements

The earliest attempt at a *quantitative* description of metastability dates back to the work of van 't Hoff [229], within the context of chemical reaction-rate theory (see Fig. 1.3). In 1884 he proposed a formula for the temperature dependence of the rate constant R associated with a chemical reaction, of the form

$$R = \exp[-E/kT]. \qquad (1.3.1)$$

Here, E is the *activation energy* associated with the reaction (in joules per molecule), T is the absolute temperature (in degrees Kelvin), and k is the Boltzmann constant. (If the activation energy is measured in joules per mole, then k is to be replaced by what is called the gas constant.) In 1889 Arrhenius [8] proposed a refinement of (1.3.1), namely,

$$R = A \exp[-E/kT], \qquad (1.3.2)$$

where the prefactor A is called the *amplitude*. He also provided the following physical interpretation of (1.3.2). For molecules to react they must first acquire a minimum amount of energy, say E. At absolute temperature T, the fraction of molecules that have a kinetic energy larger than E is proportional to $\exp[-E/kT]$, according to the Maxwell-Boltzmann distribution of statistical mechanics. Hence, $\exp[-E/kT]$ is the probability that a single collision causes a reaction. If A is interpreted as the average number of collisions per time unit, then R is the average number of collisions per time unit that cause a reaction, and the inverse $1/R$ is the *average reaction time*.

Equation (1.3.2) goes under the name of *Arrhenius equation* or *Arrhenius law*. The same equation applies to other situations where an energy barrier is involved, such as the phenomena of condensation and magnetisation mentioned in Sect. 1.2. Still other examples are the motion of dislocations in crystals, the ageing of spin glasses and the folding of proteins, which underlines the *universal character* of the Arrhenius formula. In Part VI we will see that (1.3.2) provides an excellent

approximation of the average metastable crossover time for a large class of models with a stochastic dynamics in small volumes at low temperatures.

Several modifications of the Arrhenius equation have been proposed over the years. One modification is a temperature-dependence of the prefactor of the form $A(T/T_0)^\alpha$, with T_0 a reference temperature and $\alpha \in \mathbb{R}$ a dimensionless exponent. In Part VII we will encounter a model with a stochastic dynamics in large volume at low temperature where this form of the prefactor is needed, with A proportional to the volume and $\alpha = 1$. In general, however, this form of the prefactor is neither easy to explain theoretically nor easy to verify experimentally. Another modification is a stretched exponential of the form

$$R = A \, \exp\big[-(E/kT)^{\bar{\alpha}}\big], \qquad\qquad (1.3.3)$$

where $\bar{\alpha} \in (0, 1)$ is a dimensionless exponent. Such an equation appears when the reaction is controlled by a range of activation energies (occurring e.g. in disordered systems) or a range of space-time scales (occurring e.g. in Mott multi-range random hopping). The system is said to be "ageing" as it explores larger and larger activation energies and space-time scales. In this book we will not deal with models that require this modification.

In 1940 Kramers proposed a toy model of a chemical reaction based on Brownian motion in a double-well potential [157]. Using this model, he was able to derive explicit expressions for E and A in (1.3.2) in terms of the shape of the potential (see Sect. 2.1.1 for details). This work was the first to provide a *mathematical* verification of the Arrhenius equation based on a *mesoscopic model* that replaces the microscopic collisions of the molecules involved in the chemical reaction by a Brownian motion, in the spirit of Einstein's explanation of Brownian motion. Various refinements of the Kramers formula, e.g. to higher dimensions and to different choices of the noise, were obtained in the 1960's and 1970's. See the 1981 monograph *Stochastic Processes in Physics and Chemistry* by van Kampen [228] for an overview. These refinements in turn led to the theory of random perturbations of dynamical systems developed by Freidlin and Wentzell (see Chap. 2 for details), in which explicit expressions for E and A were derived in much greater generality. This line of research eventually led to the so-called *pathwise approach* to metastability, which will be discussed in more detail in Sect. 1.3.2 below.

In 1966, Lebowitz and Penrose [162] provided a mathematical explanation of the gas-liquid phase transition within the context of the so-called *van der Waals-Maxwell theory*. They proposed a spin-model with a long-range interaction called "Kac-potential" and showed that the free energy of this model correctly predicts the pressure-versus-volume phase diagram, including the line of coexistence constructed via the "Maxwell's equal area rule". In 1971, Penrose and Lebowitz [199] proposed a framework for a rigorous theory of metastability for particle systems. They characterised metastable states via three conditions: (1) the system has only one stable state (the thermodynamic phase); (2) the lifetime of the metastable state is very long; (3) the crossover from the metastable state to the stable state is an "irreversible" process, in the sense that the return time is much longer than the

decay time. The main tool in [199] is the *restricted ensemble*, which is defined to be the Gibbs measure *conditioned* on the particle configuration lying in a suitable subset R of the configuration space, representing the metastable state, e.g. corresponding to a supersaturated vapour whose density is conditioned to lie below the density of the liquid. The rate at which the stochastic dynamics brings the system outside R is maximal at time zero. This incipient rate plays the role of an escape rate λ. The lifetime of the metastable state is identified with $1/\lambda$, and is an inherently *dynamical* quantity. The choice of R must be such that: (1) the Gibbs measure conditioned on R describes a pure phase; (2) λ is very small; (3) R has a very small weight under the unconditional Gibbs measure. For the spin-model with Kac-potential, Penrose and Lebowitz were able to compute λ explicitly (on a rough scale) and show that $1/\lambda$ coincides with the *activation free energy* needed to move out of R. Based on these results, an early attempt to axiomatise metastability was made by Sewell [218, Chap. 6]. For further details we refer the reader to Penrose and Lebowitz [200] and to Olivieri and Vares [198, Sect. 4.1].

1.3.2 The pathwise approach

The pathwise approach to metastability was initiated in the late 1960's and early 1970's by Freidlin and Wentzell. They introduced the theory of *large deviations on path space* in order to analyse the long-term behaviour of *dynamical systems* under the influence of weak random perturbations. Their realisation that metastable behaviour is controlled by large deviations of the random processes driving the dynamics has permeated most of the mathematical literature on the subject since. A comprehensive account of this development is given in their 1984 monograph *Random Perturbations of Dynamical Systems* [115]. The application of these ideas in a statistical physics context was pioneered in 1984 by Cassandro, Galves, Olivieri and Vares [51]. They realised that the theory put forward by Freidlin and Wentzell could be applied to study metastable behaviour of *interacting particle systems*. This paper led to a flurry of results for a variety of Markovian lattice models, which are described at length in the 2005 monograph *Large Deviations and Metastability* [198] by Olivieri and Vares. This work provides the key elements of the symbiosis between statistical physics, large deviation theory and metastability.

The advantage of the pathwise approach is that it gives very detailed information on the metastable behaviour of the system. By identifying the most likely path between metastable states (typically, the global minimiser of some "action integral" that constitutes the large deviation rate function in path space), the time of the crossover can be determined and information can be obtained on what the system does before and after the crossover ("tube of typical trajectories"). The drawback of the pathwise approach is that it is generally hard to identify and control the rate function, especially for systems with a spatial interaction, for which the dynamics is non-local. Consequently, the pathwise approach typically leads to relatively crude results on the crossover time.

1.3.3   The spectral approach

In the 1980's, Davies [70–74] proposed an axiomatic approach to metastability based on *spectral properties* of generators of reversible Markov processes (in some L^2-space). He showed that metastable behaviour arises when the spectrum of the generator consists of a cluster of very small real eigenvalues, separated by a comparatively wide gap from the rest of the spectrum. Under additional assumptions on boundedness of the corresponding eigenfunctions, he showed that the eigenfunctions allow for a decomposition of the state space into "metastable" sets, and that the motion of the Markov process between these sets is slow, with time-scales that are given by the inverses of the corresponding eigenvalues. In the 1990's, these results were developed further by Gaveau and Schulman [124], and Gaveau and Moreau [123]. While the spectral approach to metastability is conceptually nice and natural, it is typically very difficult to verify the assumptions made on the spectrum.

1.3.4   The potential-theoretic approach

The potential-theoretic approach to metastability was initiated in 2001 in a paper by Bovier, Eckhoff, Gayrard and Klein [33]. Here, the pathwise view is largely discarded. Instead of aiming at identifying the most likely paths realising a metastable crossover and estimating their probabilities, it interprets the metastability phenomenon as a sequence of visits of the path to different metastable sets, and focuses on a precise analysis of the respective hitting probabilities and hitting times of these sets with the help of *potential theory*. Phrased differently, it translates the problem of understanding the metastable behaviour of Markov processes to the study of equilibrium potentials and capacities of *electric networks*.

   More precisely, the configurations of the system are viewed as the vertices of the network and the transitions between pairs of configurations as the edges of the network. The transition probabilities are represented by the conductances of the associated edges. In this language, the hitting probability of a set of configurations as a function of the starting configuration of the Markov process can be expressed in terms of the equilibrium potential on the network when the potential is set to 1 on the vertices of the target set and to 0 on the starting vertex. The average hitting time of the set can then be expressed in terms of the equilibrium potential and the capacity associated with the target set and the starting vertex. For metastable sets it turns out that *the average hitting time is essentially the inverse of the capacity*.

   A key observation in the potential-theoretic approach is the fact that capacities can be estimated by exploiting powerful *variational principles*. In fact, dual variational principles are available that express the capacity both as a supremum (over potentials) and as an infimum (over flows). This opens up the possibility to derive sharp lower bounds and upper bounds on the capacity via a judicious choice of test functions. In fact, with the proper physical insight, test functions can be found for which the lower bounds and the upper bounds are asymptotically equivalent (in an appropriate limit corresponding to a metastable regime). A second key obser-

vation is that the relevant equilibrium potentials can, to the extent necessary, be in turn bounded from above and below by capacities with the help of *renewal* equations. This is absolutely crucial, as it avoids the formidable problem of solving the boundary value problems through which the equilibrium potentials are defined. Effectively, it means that estimates of the average crossover time can be derived that are much sharper than those obtained via the pathwise approach.

Capacities are expressed with the help of *Dirichlet forms*, which are functionals of the space of potentials, respectively, flows, and correspond to the energy associated with the network. These Dirichlet forms have the dimension of the configuration space, and thus are typically very high-dimensional. However, it turns out that the ensuing high-dimensional variational principles for the capacity often can be *reduced* to low-dimensional variational principles when the system is metastable. This comes from the fact that metastable crossovers occur near saddles connecting metastable sets of configurations and, consequently, the equilibrium potential is very close to 1 or to 0 away from these saddles. As a result, the full variational principle *reduces* to a simpler variational principle, which only lives on the configurations close to the saddle and captures the fine details of the dynamics when it makes the crossover. In Parts IV–VIII we will see plenty of examples of this reduction. In some cases the simpler variational problem is so low-dimensional that it can be solved explicitly.

The quantitative success of the potential-theoretic approach, relying on tractable variational principles for capacities, also entails its effective limitation to the case of *reversible* Markov processes. While variational characterisations of capacities are known also for *non-reversible* Markov processes (see Sect. 7.3), they are far more complicated and difficult to use than their reversible counterparts. Some attempts in this direction have been made by Eckhoff [101, 102], and more recently by Gaudillière and Landim [121] and Slowik [220]. This area is wide open for future research.

Historically, the potential-theoretic approach has its roots in the early work by Kramers [157], who performed precise computations of metastable crossover times in the context of a Brownian motion in a double-well potential. Such explicit solutions of the Dirichlet problems involved are, however, possible only in the one-dimensional setting. There have been numerous computations in higher-dimensional settings, based on formal perturbation theory, which can be seen as precursors of the potential-theoretic approach (see e.g. Matkowsky and Schuss [181], Matkowsky, Schuss and Tier [182], Knessl, Matkowsky, Schuss and Tier [153], and the discussion in Maier and Stein [169]).

The potential-theoretic approach also connects nicely to the spectral approach. As we will see in Chaps. 8 and 11, in many cases the spectral assumptions of Davies are a consequence of metastability as characterised by capacities.

1.3.5 The computational approach

As mentioned above, there is great interest in quantitative numerical computations for specific systems that exhibit metastable phenomena. Since metastability is driven

by rare events and involves excessively long time-scales, doing a simulation is extremely challenging and requires highly sophisticated techniques. Some of the methods developed so far have relations to or are motivated by theoretical work, in particular, the spectral approach (see e.g. Schütte, Huisinga and Meyn [216]). The so-called *transition path theory*, developed in the 2000's, uses ideas similar to those appearing in the potential-theoretic approach, but relies on numerical methods to compute harmonic functions (see e.g. E and Vanden-Eijnden [100], Ren and Vanden-Eijnden [99], Metzner, Schütte and Vanden-Eijnden [184]). Covering this huge field is beyond the scope of the present book.