

## Chapter 1

# Introduction

This chapter presents the physical and mathematical framework to understand the basics of molecular simulation and computational statistical physics techniques. Section 1.1 recalls the aims of computational statistical physics, gives some historical landmarks, and provides the orders of magnitude of the quantities to be computed. Section 1.2 is a short summary of the most important concepts of statistical physics which will be of constant use throughout this book. It is decomposed into three parts: the static description of microscopic systems (unknowns, boundary conditions, interaction potentials), the dynamics of isolated systems (the Hamiltonian dynamics), and some elements on thermodynamic ensembles. We are then in a position to define free energies in Section 1.3, discuss their relationships with metastability issues, and finally classify the most common methods currently used to compute free energy differences in terms of the mathematical objects at hand. This classification is the basis of the construction of the book, see Section 1.4 for more details.

### 1.1 Computational statistical physics: some landmarks

Before giving a detailed mathematical framework of computational statistical physics, we first describe the scientific context, by recalling in Section 1.1.1 some order of magnitudes for the quantities under investigation, and by expliciting in Section 1.1.2 what we understand to be the current aims of molecular simulation.

Table 1.1 Some important physical constants or quantities in quantum and statistical physics.

Physical constant	Usual notation	Value
Avogadro number	$\mathcal{N}_A$	$6.02 \times 10^{23}$
Boltzmann constant	$k_B$	$1.381 \times 10^{-23}$ J/K
Reduced Planck constant	$\hbar$	$1.054 \times 10^{-34}$ Js
Elementary charge	$e$	$1.602 \times 10^{-19}$ C
Electron mass	$m_e$	$9.11 \times 10^{-31}$ kg
Proton mass	$m_p$	$1.67 \times 10^{-27}$ kg
Electron-Volt	eV	$1.602 \times 10^{-19}$ J

### 1.1.1 Some orders of magnitude

In the framework of statistical physics, matter is most often described at the atomic level, either in a quantum or classical framework. Some of the concepts developed in this introduction may however be used in other physical frameworks than molecular simulation (for instance, the Hamiltonian dynamics presented in Section 1.2.2 is the fundamental evolution equation in celestial mechanics).

In this book, only classical systems are considered. Some important physical constants are recalled in Table 1.1. From those constants, the orders of magnitudes of the classical description of matter at the microscopic level can be inferred. The typical distances are expressed in Å ( $10^{-10}$  m), the energies are of the order of  $k_B T \simeq 4 \times 10^{-21}$  J at room temperature, and the typical times are of the order of  $10^{-15}$  s when the proton mass is the reference mass.

The orders of magnitude used in the microscopic description of matter are far from the orders of magnitude of the macroscopic quantities we are used to. For instance, the number of particles under consideration in a macroscopic sample of material is of the order of the Avogadro number  $\mathcal{N}_A \sim 10^{23}$ . For practical numerical computations of matter at the microscopic level, following the dynamics of every atom would require simulating  $\mathcal{N}_A$  atoms and performing  $O(10^{15})$  time integration steps, which is of course impossible! These numbers should be compared with the current orders of magnitude of the problems that can be tackled with classical molecular simulation, such as the simulation of the complete satellite tobacco mosaic virus [Freddolino *et al.* (2006)], which involved 1 million atoms over 50 ns, or the folding simulations of the Villin headpiece,<sup>1</sup> where

<sup>1</sup>See the website of the Folding@Home project: <http://folding.stanford.edu/>

a trajectory of  $500 \mu\text{s}$  was integrated for  $2 \times 10^4$  atoms.

To give some insight into such large numbers, it is helpful to compute the number of moles of water on earth. Recall that one mole of water corresponds to 18 mL, so that a standard glass of water contains roughly 10 moles, and a typical bathtub contains  $10^5$  mol. On the other hand, there are approximately  $1.3 \times 10^{18} \text{ m}^3$  of water in the oceans, *i.e.*  $7.2 \times 10^{22}$  mol, a number comparable to the Avogadro number. This means that inferring the macroscopic behavior of physical systems described at the microscopic level by the dynamics of several millions of particles only is like inferring the ocean's dynamics from hydrodynamics in a bathtub...

Describing the macroscopic behavior of matter knowing its microscopic description therefore seems out of reach. Statistical physics allows us to bridge the gap between microscopic and macroscopic descriptions of matter, at least on a conceptual level. The question is whether the estimated quantities for a system of  $N$  particles correctly approximate the macroscopic property, formally obtained in the thermodynamic limit  $N \rightarrow +\infty$  (the density being kept fixed). In some cases, in particular for simple homogeneous systems, the macroscopic behavior is well approximated from small-scale simulations, see Section 1.1.2.1. However, the convergence of the estimated quantities as a function of the number of particles involved in the simulation should be checked in all cases.

### 1.1.2 Aims of molecular simulation

Despite its intrinsic limitations on spatial and timescales, molecular simulation, has been used and developed over the past 50 years, and its number of users keeps increasing. As we understand it, it has two major aims nowadays.

First, it can be used as a *numerical microscope*, which allows us to perform “computer” experiments. This was the initial motivation for simulations at the microscopic level: physical theories were tested on computers. This use of molecular simulation is particularly clear in its historic development, which was triggered and sustained by the physics of simple liquids. Indeed, there was no good analytical theory for these systems, and the observation of computer trajectories was very helpful to guide the physicists' intuition about what was happening in the system, for instance the mechanisms leading to molecular diffusion. In particular, the pioneering works on Monte-Carlo methods [Metropolis *et al.* (1953)], and the first molecular dynamics simulation [Alder and Wainwright (1956)] were performed because

of such motivations. Today, understanding the behavior of matter at the microscopic level can still be difficult from an experimental viewpoint (because of the high resolution required, both in time and in space), or because we simply do not know what to look for! Numerical simulations are then a valuable tool to test some ideas or obtain some data to process and analyze in order to help assessing experimental setups. This is particularly true for current nanoscale systems.

Another major aim of molecular simulation, maybe even more important than the previous one, is to compute macroscopic quantities or thermodynamic properties, typically through averages of some functionals of the system. In this case, molecular simulation is a way to obtain *quantitative* information on a system, instead of resorting to approximate theories, constructed for simplified models, and giving only qualitative answers. Sometimes, these properties are accessible through experiments, but in some cases only numerical computations are possible since experiments may be unfeasible or too costly (for instance, when high pressure or large temperature regimes are considered, or when studying materials not yet synthesized). More generally, molecular simulation is a tool to explore the links between the microscopic and macroscopic properties of a material, allowing to address modelling questions such as “Which microscopic ingredients are necessary (and which are not) to observe a given macroscopic behavior?”

### 1.1.2.1 An example: the equation of state of Argon

Let us detail to some extent the second approach, and illustrate it with a simple but realistic example. We consider microscopic systems composed of  $N$  particles (typically atoms, *i.e.* nuclei together with their electronic clouds), described by the positions of the particles  $q = (q_1, \dots, q_N) \in \mathcal{D}$  and the associated momenta  $p = (p_1, \dots, p_N) \in \mathbb{R}^{3N}$ . For physical and biological systems currently studied,  $N$  is typically between  $10^3$  and  $10^9$ . The vector  $(q, p)$  is called the *microscopic state* or the *configuration* of the system.

In the framework of statistical physics, macroscopic quantities of interest are written as averages over thermodynamic ensembles, which are probability measures on all the admissible microscopic configurations:

$$\mathbb{E}_\mu(A) = \int_{T^*\mathcal{D}} A(q, p) \mu(dq dp). \quad (1.1)$$

In this expression, the function  $A$  is called an *observable*. The position variable  $q = (q_1, \dots, q_N)$  belongs to  $\mathcal{D}$ , which is called the configuration space.

The set  $\mathcal{D}$  is an open subset (possibly the whole) of  $\mathbb{R}^n$  with  $n = 3N$ , or  $\mathcal{D} = \mathbb{T}^n$  (where  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  denotes the one-dimensional torus). The choice of  $\mathcal{D}$  depends on the boundary conditions at hand, see Section 1.2.1.1. For the two choices mentioned above, the momentum variable  $p = (p_1, \dots, p_N)$  belongs to  $\mathbb{R}^n$ , so that the cotangent space  $T^*\mathcal{D}$  used in (1.1) can be identified with  $\mathcal{D} \times \mathbb{R}^n$ . The set of all possible microscopic configurations  $(q, p)$  is called the *phase space*. The probability measure  $\mu$  has support on the phase space and depends on the thermodynamic ensemble used, see Section 1.2 for further precision on the most common choices.

**Remark 1.1 (Generalization to other configuration spaces).**

*All the results presented in this book may be generalized to the case when the configuration space  $\mathcal{D}$  is not  $\mathbb{R}^n$ , but some open subset of  $\mathbb{R}^n$ , with a potential energy function which goes sufficiently fast to  $\infty$  on  $\partial\mathcal{D}$  to prevent the dynamics from leaving the domain  $\mathcal{D}$ . For the case of molecular constraints, we refer to Section 3.3.6.2.*

A statistical description through a probability measure  $\mu$  is a convenient description since the whole microscopic information is both unimportant (what matters are average quantities, and not the positions of all particles composing the system) and too large to be processed.

An example of an observable is the bulk pressure  $P$  in a Lennard-Jones liquid. For particles of masses  $m_i$ , described by their positions  $q_i$  and their momenta  $p_i$ , it is given by  $P = \mathbb{E}_\mu(A)$  with

$$A(q, p) = \frac{1}{3|\mathcal{D}|} \sum_{i=1}^N \left( \frac{|p_i|^2}{m_i} - q_i \cdot \frac{\partial V}{\partial q_i}(q) \right),$$

where  $|\mathcal{D}|$  is the physical volume of the box occupied by the fluid, and the potential energy function  $V$  is made precise below, see (1.4)-(1.5).

In practice, such averages may yield results that are very close to experimental measurements, even for systems small in comparison to the actual sizes of macroscopic systems (provided the interaction potentials are short-ranged). For example, the equation of state of Figure 1.1 has been computed with a system of a few thousand particles only, a number which is 20 orders of magnitude lower than the Avogadro number. The computed results are compared with experimental measurements.<sup>2</sup> The agreement is very good in the case of Argon. Notice also that high-pressure results, not easily obtained with experimental setups, can be computed.

<sup>2</sup>See for instance the NIST webpage <http://webbook.nist.gov/chemistry/fluid/>

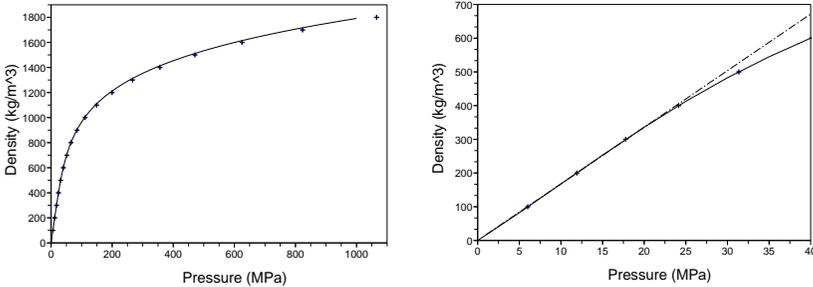


Fig. 1.1 Numerical equation of state of argon at  $T = 300$  K (“+”) and experimental reference curve (solid line). The picture on the right is a zoom on the low density/low pressure part of the curve, where the ideal gas regime is plotted in dash-dotted line.

We will restrict ourselves in this book to static equilibrium properties of the form of (1.1), and will not consider dynamical properties depending on the actual time evolution of the system (autocorrelation functions, transport coefficients such as thermal conductivity, exit times out of some region in phase space, ...).

## 1.2 Microscopic description of physical systems

The description of systems in statistical physics requires several ingredients: microscopic interaction laws between the constituents of matter and possibly the environment (see Section 1.2.1), time evolution equations for isolated systems (see Section 1.2.2), and the notion of thermodynamic ensembles, which are probability measures on the set of all possible microscopic configurations, consistent with the macroscopic state of the system (see Section 1.2.3).

### 1.2.1 Interactions

The interactions between the particles are taken into account through a potential function  $V$ , depending on the positions  $q$  only. The total energy of the system is given by the Hamiltonian

$$H(q, p) = E_{\text{kin}}(p) + V(q), \quad (1.2)$$

where the kinetic energy is

$$E_{\text{kin}}(p) = \frac{1}{2} p^T M^{-1} p, \quad M = \begin{pmatrix} m_1 \text{Id}_3 & & 0 \\ & \ddots & \\ 0 & & m_N \text{Id}_3 \end{pmatrix}.$$

The matrix  $M$  is called the mass matrix. A Hamiltonian such as (1.2) is said to be *separable* since the energetic contributions of the momentum and position variables can be added independently. An instance of a non-separable Hamiltonian is the case when the mass-matrix depends on the configuration  $q$  of the system.

Most Hamiltonians encountered in applications are separable, and we will in any case restrict ourselves to separable Hamiltonians in this book. Non-separable Hamiltonians may be considered for modelling purposes (when working with internal coordinates, for rigid body dynamics for instance), or for mathematical convenience (such as the modified Hamiltonians used in the backward analysis of Hamiltonian dynamics, see the references at the end of Section 1.2.2.4).

In order to describe more precisely the interactions between the elementary constituents of the system, several points have to be made precise. First, the boundary conditions of the system must be specified (see Section 1.2.1.1). Then, we give more detail on the interaction potential  $V$  in Section 1.2.1.2. This function is very important since it incorporates almost all the physics of the problem. It is therefore no surprise that obtaining reliable potential functions is still a very active research field.

### 1.2.1.1 *Boundary conditions*

Several boundary conditions can be imposed onto the system:

- (i) Many current simulations are performed using periodic boundary conditions, so that surface effects can be avoided and configurations typically encountered in the bulk of the system can be obtained. In this case, a particle interacts not only with all the particles in the systems, but also with their periodic images (see Figure 1.2). In practice, interactions are set to 0 when the distance between two or several particles exceeds a given cut-off radii  $r_{\text{cut}}$ . When cubic domains of length  $L$  are considered as in Figure 1.2, the domain length should be chosen so that  $r_{\text{cut}} < L/2$ . This ensures that a particle interacts either with the primitive particle, or at most one of its periodic images;

- (ii) In some simulations, the system is allowed to visit the entire physical space ( $\mathcal{D} = \mathbb{R}^{3N}$ ). This is the case for isolated systems, such as molecules *in vacuo*. It may be convenient however to quotient out rigid body motions in this case since the potential energy is usually invariant under translations and rotations of the system;
- (iii) It is sometimes necessary to consider confined systems. In this case, the positions of the particles are restricted to some predefined region of space, and some rules have to be set for reflections on the boundaries of the system (such as specular reflection of the momenta).

Let us finally mention that open systems with inflows or outflows of energy, particles etc., are sometimes considered. In this case, there may be some exchanges or forcing at the boundaries. Such situations are not considered in this book.

### 1.2.1.2 Potential functions

***Ab initio* interaction potentials.** Ideally, the interaction potentials between the particles should be obtained in a non-empirical approach by resorting to *ab initio* computations. Relying on the standard Born-Oppenheimer assumption, the positions  $q_i$  of the nuclei of charges  $Z_i$  are kept fixed, and the energy of the system is obtained by adding the Coulomb interaction energies between the nuclei, and the electronic ground-state energy:

$$V(q_1, \dots, q_N) = \sum_{1 \leq i < j \leq N} \frac{Z_i Z_j}{|q_i - q_j|} + V_{\text{elec}}(q_1, \dots, q_N). \quad (1.3)$$

Denote by  $M = Z_1 + \dots + Z_N$  the number of electrons. The system is assumed to be neutral. The electronic ground-state energy is obtained by minimizing the electronic problem over the Hilbert space  $\mathcal{H}$  of admissible wavefunctions, which is a subset of the space  $\bigwedge_{m=1}^M L^2(\mathbb{R}^3, \mathbb{C})$  of antisymmetric functions. We omit the spin variable for notational simplicity although this variable is very important for quantitative computations. The electronic problem then reads

$$V_{\text{elec}}(q_1, \dots, q_N) = \inf \left\{ \left\langle \psi, \hat{H}_{q_1, \dots, q_N} \psi \right\rangle_{\mathcal{H}} \mid \psi \in \mathcal{H}, \|\psi\|_{L^2} = 1 \right\},$$

where the electronic Hamiltonian operator reads

$$\hat{H}_{q_1, \dots, q_N} = - \sum_{m=1}^M \frac{1}{2} \Delta_{x_m} - \sum_{m=1}^M \sum_{i=1}^N \frac{Z_i}{|x_m - q_i|} + \sum_{1 \leq n < m \leq M} \frac{1}{|x_n - x_m|}.$$

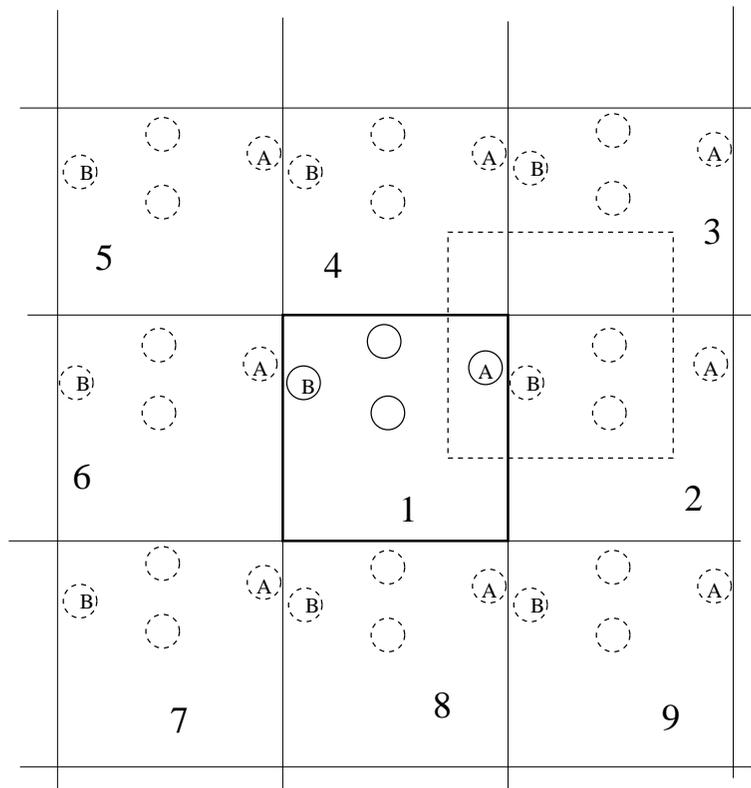


Fig. 1.2 System with periodic boundary conditions. The simulation cell is numbered “1”, and the other cells are obtained by translation. The particles inside the primitive cell have interactions with particles in all the other cells.

We refer for instance to [Cancès *et al.* (2003)] for further precision on the computation of *ab initio* interaction potentials. Such computations are however very time-consuming, so that only small systems can be simulated this way (using Born-Oppenheimer molecular dynamics [Niklasson *et al.* (2006)] or the Car-Parrinello approach [Car and Parrinello (1985)]).

**Empirical potentials.** In practice, empirical formulas for the potential energy function are used to study larger systems. These empirical formulae are obtained by assuming a functional form for the interaction potential, which depends on a set of parameters. These parameters may be chosen so that the potential energy function is as close as possible to the function (1.3) obtained from small *ab initio* computations. Alternatively, the parameters

may be such that average properties computed from molecular simulations match experimental thermodynamic properties such as the equation of state of the material, its heat capacity, etc.

A very simple example of an empirical potential is the potential function of a fluid composed of  $N$  particles, interacting through a pairwise additive potential depending only on the distance between the particles:

$$V(q_1, \dots, q_N) = \sum_{1 \leq i < j \leq N} \mathcal{V}(|q_i - q_j|). \quad (1.4)$$

For example, noble gases are well described using (1.4) when  $\mathcal{V}$  is the

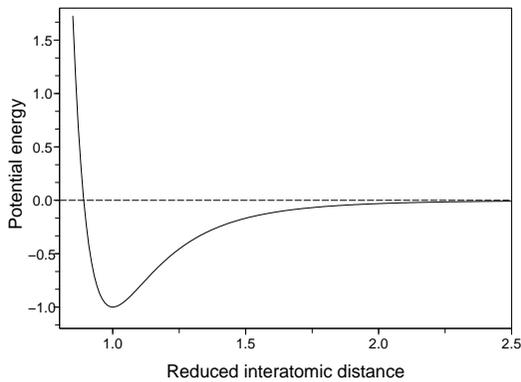


Fig. 1.3 Lennard-Jones potential (1.5) where the distance and the energy are expressed in terms of the equilibrium distance  $2^{1/6}\sigma$  and the reference energy  $\varepsilon$ .

Lennard-Jones potential (depicted in Figure 1.3)

$$\mathcal{V}(r) = 4\varepsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right). \quad (1.5)$$

This potential depends on two parameters: an energy  $\varepsilon$  and a distance  $\sigma$ . For argon for instance,  $\varepsilon = 1.66 \times 10^{-21}$  J, and  $\sigma = 3.405$  Å. The model (1.4)-(1.5) is suitable for noble gases since these systems are monatomic and the corresponding atoms have closed electronic shells. Therefore, the dominant physical interaction is the weakly attractive long-range van der Waals contribution, which scales as  $r^{-6}$ .

**Potential functions for molecules.** Many molecular systems contain molecules. Therefore, interaction potentials describing the existence of bonds between atoms are required. This is modelled through interactions

involving several atoms. To describe these potentials, it is convenient to introduce the vector  $r_{i,j} = q_j - q_i$ .

- (1) The interactions of two atoms involved in a covalent bond can be modelled *via* a harmonic potential energy

$$\mathcal{V}_2(q_i, q_{i+1}) = \frac{k_0}{2} (|r_{i,i+1}| - l_{\text{eq}})^2,$$

where  $l_{\text{eq}}$  is the equilibrium length;

- (2) Three atoms can interact *via* the three-body interaction potential energy

$$\mathcal{V}_3(q_i, q_{i+1}, q_{i+2}) = \frac{k_\theta}{2} (\theta_i - \theta_{\text{eq}})^2,$$

where the bond angle  $\theta_i$  is

$$\theta_i = \arccos \left( \frac{r_{i,i+1} \cdot r_{i+1,i+2}}{|r_{i,i+1}| |r_{i+1,i+2}|} \right),$$

while  $\theta_{\text{eq}}$  is the equilibrium bond angle;

- (3) Four atoms may experience the four-body interaction potential energy

$$\mathcal{V}_4(q_i, q_{i+1}, q_{i+2}, q_{i+3}) = u_{\text{tors}}(\cos \phi_i), \quad (1.6)$$

where the dihedral angle  $\phi_i$  is obtained from the relation

$$\cos \phi_i = - \frac{r_{i,i+1} \times r_{i+1,i+2}}{|r_{i,i+1}| |r_{i+1,i+2}|} \cdot \frac{r_{i+1,i+2} \times r_{i+2,i+3}}{|r_{i+1,i+2}| |r_{i+2,i+3}|}.$$

Local interactions have to be complemented by non-bonded interactions: van der Waals forces modelled by Lennard-Jones potentials, and Coulomb interactions, see [Schlick (2002)] for further precision.

A typical example of a simple molecular system is depicted in Figure 1.4 (left), which corresponds to the pentane molecule in the so-called united-atom representation (see [Ryckaert and Bellemans (1978)]). In this representation, the hydrogen atoms are not explicitly represented. We label by  $q_1, \dots, q_5$  the positions of the carbon atoms in the pentane molecule, while  $q_6, \dots, q_N$  are the positions of the solvent molecules. The solvent molecules are assumed to interact with all the other atoms through a pairwise potential  $\mathcal{V}_{\text{sol}}$  depending only on the relative distance. The total interaction energy then reads

$$V(q) = V_{\text{pentane}}(q_1, \dots, q_5) + V_{\text{solvent}}(q_6, \dots, q_N) + V_{\text{interaction}}(q),$$

with

$$V_{\text{solvent}}(q_6, \dots, q_N) = \sum_{6 \leq i < j \leq N} \mathcal{V}_{\text{sol}}(|q_i - q_j|),$$

and

$$V_{\text{interaction}}(q) = \sum_{i=1}^5 \sum_{6 \leq j \leq N} \mathcal{V}_{\text{sol}}(|q_i - q_j|).$$

The interactions within the molecule are

$$\begin{aligned} V_{\text{pentane}}(q_1, \dots, q_5) &= \sum_{i=1}^4 \mathcal{V}_2(q_i, q_{i+1}) + \sum_{i=1}^3 \mathcal{V}_3(q_i, q_{i+1}, q_{i+2}) \\ &\quad + \sum_{i=1}^2 \mathcal{V}_4(q_i, q_{i+1}, q_{i+2}, q_{i+3}), \end{aligned}$$

where the dihedral potential function  $u_{\text{tors}}$  in (1.6) is given by an expression of the form

$$u_{\text{tors}}(x) = c_1(1 - x) + 2c_2(1 - x^2) + c_3(1 + 3x - 4x^3).$$

The parameters  $c_i$  ( $i = 1, 2, 3$ ) used in the united-atom model of [Ryckaert and Bellemans (1978)] are such that there are three stable dihedral angles, the one at  $\phi = 0$  being energetically more favorable than the others (see Figure 1.4, right).

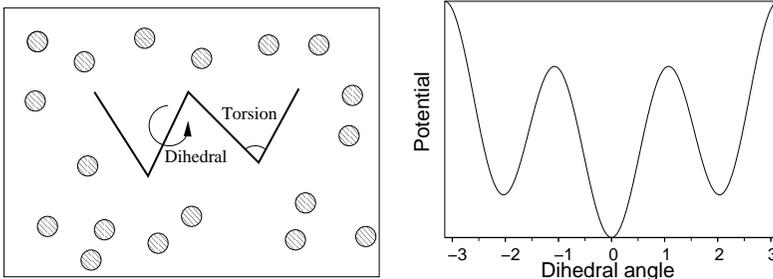


Fig. 1.4 Left: Schematic representation of a pentane molecule in a solvent (projected on a two-dimensional plane), and definition of the bond angles and dihedral angles. Right: Typical shape of the potential for the dihedral angle.

**More realistic force fields.** Pairwise additive potentials such as (1.4), and two-, three- or four-body bonded interactions may however not be a good approximation of the many-body *ab initio* potential function (1.3). Many studies aim at proposing better empirical potential functions (or force fields). Recent instances of such potentials are the (Modified) Embedded-Atom Model potentials [Baskes (1992)], or bond-order potentials of REBO [Tersoff (1989)] or ReaxFF [van Duin *et al.* (2001)] types, which contain term depending on the local coordination of the atoms. The latter potentials can even account for chemical reactions (*i.e.* bond breakings and bond formations).

**Non-dimensional units.** In practice, it is more convenient (and numerically more stable) to work with non-dimensional quantities. In this case, the manipulated numbers are all of order 1. In general, reduced units require the following reference quantities:

- a reference mass  $m_0$ , for instance the mass of the heaviest or the lightest atom in the system;
- a reference energy  $\varepsilon_0$ , given by the magnitude of a typical interaction energy, or alternatively by  $k_B T$ . This energy is therefore of the order of  $10^{-21}$  J;
- a reference length  $l_0$ , given by the typical interaction distance, for instance a covalent bond length when molecules are present in the system. Usually,  $l_0$  is of the order of several angströms.

Moreover, other reference quantities can be derived from the above fundamental reference quantities. For instance, a reference time  $t_0$  is obtained by requiring that the typical kinetic energy is of the order of magnitude of the reference energy:

$$t_0 = \frac{m_0^{1/2} l_0}{\varepsilon_0^{1/2}}. \quad (1.7)$$

This time is typically of the order of the pico-second.

### 1.2.2 Dynamics of isolated systems

We consider in this section the time evolution of isolated systems described at the microscopic level. After a general presentation of the Hamiltonian dynamics in its usual form in Section 1.2.2.1, some equivalent reformulations are proposed in Section 1.2.2.2. We then recall some important properties

of the dynamics in Section 1.2.2.3, and close the section with some elements on the numerical analysis of time-discretization schemes (Section 1.2.2.4).

### 1.2.2.1 The Hamiltonian dynamics

For separable Hamiltonians, the evolution of isolated systems is governed by the Hamiltonian dynamics

$$\begin{cases} \frac{dq(t)}{dt} = \nabla_p H(q(t), p(t)) = M^{-1}p(t), \\ \frac{dp(t)}{dt} = -\nabla_q H(q(t), p(t)) = -\nabla V(q(t)). \end{cases} \quad (1.8)$$

Initial conditions

$$(q(0), p(0)) = (q^0, p^0) \quad (1.9)$$

should be provided. Introducing the matrix

$$J = \begin{pmatrix} 0 & \text{Id}_{3N} \\ -\text{Id}_{3N} & 0 \end{pmatrix}, \quad (1.10)$$

and denoting  $x = (q, p) \in T^*\mathcal{D}$ , the Hamiltonian dynamics can be seen as the first-order ordinary differential equation:

$$\frac{dx}{dt} = J\nabla H(x) = J \begin{pmatrix} \nabla_q H(q, p) \\ \nabla_p H(q, p) \end{pmatrix}. \quad (1.11)$$

The existence and uniqueness of trajectories typically follows from the Cauchy-Lipschitz theorem. A sufficient condition is that  $\nabla H$  is locally Lipschitz continuous and  $H$  is bounded below. We will always assume in the sequel that the Cauchy problem (1.8)-(1.9) is well-posed.

We denote by  $\phi_t$  the flow of the Hamiltonian dynamics, *i.e.* the application which associates to some initial condition  $(q^0, p^0)$  the solution  $(q(t), p(t)) = \phi_t(q^0, p^0)$  to (1.8) at time  $t$ . Let us emphasize that (1.8) is an autonomous equation since the system is assumed to be isolated, so that the flow only depends on the duration time  $t$  of the trajectory, and not on the initial and final times separately.

The flow is a semi-group:  $\phi_{t+u} = \phi_t \circ \phi_u$  for all  $t, u \geq 0$ . Actually, it is possible to define the backward evolution  $\phi_{-t}$  for  $t \geq 0$ , using for instance the reversibility of the dynamics (see (1.19) below), so that  $\phi_{t+u} = \phi_t \circ \phi_u$  for all  $t, u \in \mathbb{R}$ .

### 1.2.2.2 Equivalent reformulations

When the Hamiltonian equation (1.8) is reformulated in terms of the positions only, it reads

$$M \frac{d^2 q(t)}{dt^2} = -\nabla V(q),$$

which is Newton's second law.

There are also more abstract reformulations of (1.8), which will be useful below. The Poisson bracket for two smooth observables  $A_1, A_2$  (i.e. functions of  $(q, p)$ ) is defined as

$$\{A_1, A_2\} = (\nabla_q A_1)^T \nabla_p A_2 - (\nabla_p A_1)^T \nabla_q A_2, \quad (1.12)$$

where  $C^T$  denotes the transpose of a matrix  $C$ . Notice that the Poisson bracket can be rewritten as

$$\{A_1, A_2\} = (\nabla A_1)^T J \nabla A_2.$$

Hamilton's equations of motion, (1.8), are then equivalent to the following transport equation: for any smooth observable  $A$ ,

$$\frac{d}{dt} [A(q(t), p(t))] = \{A, H\}(q(t), p(t)). \quad (1.13)$$

The concept of the Poisson bracket will be particularly useful to study generalized Hamiltonian dynamics for system with mechanical constraints (see Section 3.3). Some important properties of the Poisson bracket (1.12), which can be checked by simple computations, are the following:

- *Non-degeneracy*: if for any compactly-supported smooth test function  $\varphi_1$ ,  $\{\varphi_1, \varphi_2\} = 0$ , then  $\varphi_2$  is a constant function.

For any compactly supported smooth test functions  $\varphi_1, \varphi_2, \varphi_3$ ,

- *Skew-symmetry*:

$$\{\varphi_1, \varphi_2\} = -\{\varphi_2, \varphi_1\}.$$

- *Jacobi identity*:

$$\{\varphi_1, \{\varphi_2, \varphi_3\}\} + \{\varphi_2, \{\varphi_3, \varphi_1\}\} + \{\varphi_3, \{\varphi_1, \varphi_2\}\} = 0. \quad (1.14)$$

- *Leibniz' rule*:

$$\{\varphi_1 \varphi_2, \varphi_3\} = \varphi_1 \{\varphi_2, \varphi_3\} + \varphi_2 \{\varphi_1, \varphi_3\}. \quad (1.15)$$

- *Divergence formula*:

$$\int_{T^*\mathcal{D}} \{\varphi_1, \varphi_2\} dq dp = 0. \quad (1.16)$$

- *Integration by parts:*

$$\int_{T^*\mathcal{D}} \{\varphi_1, \varphi_2\} \varphi_3 dq dp = \int_{T^*\mathcal{D}} \{\varphi_2, \varphi_3\} \varphi_1 dq dp. \quad (1.17)$$

The transport equation (1.13) (or equivalently the Hamiltonian equation (1.8)) may also be restated as an evolution equation for the phase space density of the particles. Assume that the initial conditions  $(q^0, p^0)$  are distributed according to some measure with density  $\psi^0(q, p)$  with respect to the phase space Lebesgue measure, and that each initial phase space configuration is evolved according to the dynamics (1.8). Then the configurations  $\phi_t(q^0, p^0)$  at time  $t$  are distributed according to a measure with density  $\psi(t, q, p)$ , whose evolution is governed by the following partial differential equation (called the Liouville equation):

$$\partial_t \psi = \nabla_q H \cdot \nabla_p \psi - \nabla_p H \cdot \nabla_q \psi = \{H, \psi\}, \quad \psi(0, q, p) = \psi^0(q, p).$$

### 1.2.2.3 Properties of the Hamiltonian dynamics

The Hamiltonian dynamics has several interesting mathematical and structural properties:

- (1) *Symmetry.* Since  $\phi_t \circ \phi_{-t} = \text{Id}$ ,

$$\phi_{-t} = \phi_t^{-1}. \quad (1.18)$$

- (2) *Reversibility.* Consider the momentum reversal function

$$S(q, p) = (q, -p).$$

Then, the time-reversed evolution  $\phi_{-t}$  for  $t \geq 0$ , defined by (1.18), is easily seen to be equal to a forward evolution with momenta reversed (the so-called backward flow):

$$\phi_{-t} = S \circ \phi_t \circ S. \quad (1.19)$$

- (3) *Energy conservation.* The choice  $A = H$  in the Poisson bracket reformulation of the Hamiltonian dynamics (1.13) leads to  $\frac{dH(q(t), p(t))}{dt} = 0$ , which means that

$$H(q(t), p(t)) = H(q^0, p^0).$$

- (4) *Volume preservation.* For all measurable sets  $B \subset T^*\mathcal{D}$ , and for all  $t \in \mathbb{R}$ ,

$$\int_{\phi_t(B)} dq dp = \int_B dq dp. \quad (1.20)$$

This identity, often referred to as Liouville's theorem, is a consequence of the equality

$$|\text{jac } \phi_t(q, p)| = 1,$$

where  $\text{jac } \phi_t(q, p) = \det(\nabla \phi_t(q, p))$ . The proof of the latter assertion relies on the fact that the Hamiltonian vector field is divergence-free. Lemma VII.3.1 in [Hairer *et al.* (2006)] then allows us to conclude.

- (5) *Symplecticity.* The matrix  $J$  defined by (1.10) is antisymmetric and orthogonal ( $J^T = -J = J^{-1}$ ). For an open set  $U \in T^*\mathcal{D}$ , a mapping  $g : U \rightarrow \mathbb{R}^{2dN}$  is symplectic if  $\nabla g(q, p)$  satisfies

$$(\nabla g)^T J \nabla g = J \tag{1.21}$$

for all  $(q, p) \in U$ . It is easily shown that the flow  $\phi_t$  is symplectic for all  $t \in \mathbb{R}$  (this is a result due to Poincaré). Actually, any symplectic map is locally Hamiltonian (see Section VI.2 in [Hairer *et al.* (2006)] for further precision), which shows that the symplecticity of the flow is indeed a characteristic feature of Hamiltonian systems. Note that the volume preservation property (1.20) is recovered as a consequence of the symplecticity property since (1.21) with  $g = \phi_t$  implies that  $(\det \nabla \phi_t)^2 = 1$ . The symplecticity property is stronger, and can be understood as the conservation of oriented elementary parallelograms, see Section 3.5 in [Leimkuhler and Reich (2005)] for a pedagogical presentation.

#### 1.2.2.4 Numerical integration

We discuss in this section numerical schemes to integrate (1.8). Let us first mention a few reasons why it is both hopeless and useless to integrate precisely the Hamiltonian dynamics in the context of molecular simulation:

- (i) The Hamiltonian dynamics is known to be strongly sensitive to the initial conditions, or to numerical errors such as round-off errors: Small differences between two initially close configurations are exponentially magnified as time passes. Since the initial conditions can never be known exactly for physical reasons in molecular systems (in particular because there are too many atoms whose positions and momenta are required) and very long integration times are needed, this is a first reason not to try to integrate too precisely the Hamiltonian dynamics. The situation may of course be different in other application fields where Hamiltonian dynamics are used for systems with less degrees of freedom, such as celestial mechanics.

- (ii) Moreover, given the large number of particles in molecular simulations (hence the numerical cost of evaluating forces) and the very small time-steps that would be needed to integrate precisely the trajectory are prohibitive.
- (iii) Finally, the aim of many current computations in computational statistical physics is the evaluation of average properties along a long trajectory (see the ergodicity assumption (1.30) below). Therefore, it is sufficient to ensure a correct sampling rather than integrating precisely the trajectory. In particular, a basic requirement is the preservation of the energy over long trajectories.

The above arguments led to the development of numerical techniques devoted to Hamiltonian systems, fully taking into account the energy preservation as a basic first requirement. This requirement is more important than the scheme's order (*i.e.* the integer  $p$  such that the error between the exact solution over a time interval  $\Delta t$  and the numerical solution after one step of the numerical scheme is of order  $\Delta t^{p+1}$ ), which determines the convergence rate of the numerical approximation only on finite-time intervals.

A very convenient algorithm to approximately preserve the energy was proposed in [Verlet (1967)] (actually rediscovered by Verlet, since it was already known by Störmer in the context of celestial mechanics at the beginning of the 20th century, and even by Newton; see Section 1.3 in [Hairer *et al.* (2003)] for historical precisions). The Verlet algorithm is nowadays the standard integration scheme for Hamiltonian dynamics. Denoting by  $(q^n, p^n)$  an approximation of  $(q(t_n), p(t_n))$  at time  $t_n = n\Delta t$  (where  $\Delta t$  is the time step), it reads

$$\begin{cases} p^{n+1/2} = p^n - \frac{\Delta t}{2} \nabla V(q^n), \\ q^{n+1} = q^n + \Delta t M^{-1} p^{n+1/2}, \\ p^{n+1} = p^{n+1/2} - \frac{\Delta t}{2} \nabla V(q^{n+1}). \end{cases} \quad (1.22)$$

The numerical flow associated with this scheme is denoted by  $\Phi_{\Delta t}^{\text{Verlet}}$  in the sequel:  $(q^{n+1}, p^{n+1}) = \Phi_{\Delta t}^{\text{Verlet}}(q^n, p^n)$ . It is easy to check that the scheme is of order 2.

Stability requirements limit the time-step  $\Delta t$  which can be used in practice. We now detail the study of the *linear stability*, for the one-dimensional harmonic potential  $V(q) = \omega^2 q^2/2$  of mass  $m = 1$ . In this case, the Verlet

scheme reads

$$\begin{pmatrix} q^{n+1} \\ p^{n+1} \end{pmatrix} = A_{\Delta t} \begin{pmatrix} q^n \\ p^n \end{pmatrix}, \quad A_{\Delta t} = \begin{pmatrix} 1 - \frac{(\omega\Delta t)^2}{2} & \Delta t \\ -\omega^2\Delta t \left(1 - \frac{(\omega\Delta t)^2}{4}\right) & 1 - \frac{(\omega\Delta t)^2}{2} \end{pmatrix}.$$

The eigenvalues of the matrix  $A_{\Delta t}$  have modulus 1 if and only if  $\omega\Delta t < 2$ . In this case, the trajectory  $(q^n, p^n)_{n \geq 0}$  is bounded. Otherwise, one eigenvalue has a modulus strictly larger than 1, so that the trajectory  $(q^n, p^n)_{n \geq 0}$  is not bounded in general. Besides, it is easily shown that the modified energy

$$H_{\Delta t}(q, p) = H(q, p) - \frac{(\omega\Delta t)^2}{4} q^2$$

is preserved exactly:  $H_{\Delta t}(q^n, p^n) = H_{\Delta t}(q^0, p^0)$  for all  $n \geq 0$ . Therefore, when  $\omega\Delta t < 2$ , the boundedness of the trajectory implies

$$\sup_{n \in \mathbb{N}} \left| H(q^n, p^n) - H(q^0, p^0) \right| \leq C \Delta t^2.$$

For more general potentials, there is no simple rule to place an upper bound on the time-step. However, the linear stability requirement suggests that an admissible time-step should be a fraction of the fastest vibration period in the system.

Actually, the positions  $q^n$  obtained by the Verlet scheme (1.22) satisfy

$$M \frac{q^{n+1} - 2q^n + q^{n-1}}{\Delta t^2} = -\nabla V(q^n),$$

which is the simple centered finite-difference discretization for the equation  $M \frac{d^2 q}{dt^2}(t) = -\nabla V(q(t))$ . However, the very good properties of the numerical method cannot be understood from this equation. It is important to keep both variables  $q$  and  $p$ , and study the numerical flow  $\Phi_{\Delta t}^{\text{Verlet}}$  of (1.22). Indeed, this application shares many qualitative properties with the exact flow  $\phi_t$  of (1.8); in particular, it is time reversible:

$$S \circ \Phi_{\Delta t}^{\text{Verlet}} \circ S = \Phi_{-\Delta t}^{\text{Verlet}},$$

where  $S(q, p) = (q, -p)$  is the momentum reversal operator; symmetric:

$$(\Phi_{\Delta t}^{\text{Verlet}})^{-1} = \Phi_{-\Delta t}^{\text{Verlet}};$$

and symplectic:  $(\nabla \Phi_{\Delta t}^{\text{Verlet}})^T J \nabla \Phi_{\Delta t}^{\text{Verlet}} = J$ . The latter property is of paramount importance for the longtime integration of the Hamiltonian dynamics. A well-established result, recalled in the reference book [Hairer *et al.* (2006)] on geometric numerical integration (see in particular Chapters VIII and IX), is that, when  $\Delta t$  is small enough, the energy of the

system is conserved up to  $O(\Delta t^p)$  error terms over very long times when symplectic numerical schemes of order  $p$  are used (under some technical assumptions).

More generally, the longtime stability properties of symplectic numerical methods applied to symplectic flows can be studied with the help of the so-called backward analysis. Contrarily to standard error analysis where the numerical trajectory is considered as an approximation of the true trajectory of the exact problem, the backward analysis consists in interpreting the numerical trajectory as the *exact trajectory of some modified ordinary differential equation*, and then to study the properties of the modified problem. For symplectic methods approaching symplectic flows, the modified equation is still Hamiltonian. Therefore, *some modified energy is preserved exactly*. This property is finally used to show that the *exact energy is preserved approximately*. In fact, some rather involved analysis has to be used since the modified Hamiltonian is defined as a formal series, which does not converge in general. Some optimal truncations should then be considered, and the modified energy is therefore not strictly preserved, but the error terms are very small.

### 1.2.3 Thermodynamic ensembles

The macroscopic state of a system is described, within the framework of statistical physics, by a probability measure  $\mu$  on the phase space  $T^*\mathcal{D} = \mathcal{D} \times \mathbb{R}^{3N}$ . Macroscopic features of the system are then computed as averages of an observable  $A$  with respect to this measure, as given by (1.1):

$$\mathbb{E}_\mu(A) = \int_{T^*\mathcal{D}} A(q, p) \mu(dq dp).$$

We therefore call the measure  $\mu$  the *macroscopic state* of the system.

The practical computation of the ensemble average requires numerical techniques to sample configurations  $(q^n, p^n)$  according to the probability measure  $\mu$  (or possibly according to a measure  $\tilde{\mu}$  very close to  $\mu$ , the difference between  $\mu$  and  $\tilde{\mu}$  originating from errors in the numerical integration of a continuous dynamics for instance, see Section 2.3.1.1 for further precision). The ensemble average (1.1) is then approximated by

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N A(q^n, p^n). \quad (1.23)$$

The numerical techniques of course depend on the thermodynamic ensemble at hand. Most methods generate a sequence of microscopic configurations

$(q^n, p^n)_{n \geq 1}$  from a time-discrete dynamics, so that the successive configurations are not independent.

We present more thoroughly in this section two very commonly used thermodynamic ensembles, namely the microcanonical ensemble (Section 1.2.3.1) and the canonical ensemble (Section 1.2.3.2). These ensembles describe respectively isolated systems, and systems at a fixed temperature (in contact with a so-called thermostat or energy reservoir). We also mention some other thermodynamic ensembles in Section 1.2.3.3, for the sake of completeness.

### 1.2.3.1 The microcanonical ensemble

The thermodynamic ensemble naturally associated with the Hamiltonian dynamics (1.8) is the *microcanonical ensemble*, which describes isolated systems at constant energy. This ensemble is also often termed *NVE ensemble*, the capital letters referring to the invariants of the system, namely the number of particles  $N$ , the volume of the simulation box  $V$ , and the energy  $E$ .

The corresponding probability measure is the normalized uniform probability measure on the set  $\mathcal{S}(E)$  of configurations at the given energy level  $E$ :

$$\mathcal{S}(E) = \left\{ (q, p) \in T^*\mathcal{D} \mid H(q, p) = E \right\}.$$

We present three ways to understand this idea.

**An explicit construction.** The building block for the construction of the microcanonical measure is the measure  $\delta_{H(q,p)-E}(dq dp)$ , where the conditioning relies on level sets of constant total energy. This measure can be obtained by an explicit construction, using a limiting procedure. Consider a given energy level  $E$ , some small energy variation  $\Delta E > 0$ , and define

$$\mathcal{N}_{\Delta E}(E) = \left\{ (q, p) \in T^*\mathcal{D} \mid E \leq H(q, p) \leq E + \Delta E \right\}.$$

Then, the following integral of a given test function  $A$  expresses the fact that the set  $\mathcal{N}_{\Delta E}(E)$  is endowed with a uniform measure:

$$\Pi_{E, \Delta E}(A) = \frac{1}{\Delta E} \int_{\mathcal{N}_{\Delta E}(E)} A(q, p) dq dp.$$

In the limit  $\Delta E \rightarrow 0$ , a measure supported on the submanifold  $\mathcal{S}(E)$  is recovered. Notice that this measure is not normalized to 1 *a priori*. The measure  $\delta_{H(q,p)-E}(dq dp)$  is defined through the expectations of any observable  $A$  as

$$\int_{\mathcal{S}(E)} A(q, p) \delta_{H(q,p)-E}(dq dp) = \lim_{\Delta E \rightarrow 0} \frac{1}{\Delta E} \int_{\mathcal{N}_{\Delta E}(E)} A(q, p) dq dp. \quad (1.24)$$

The construction highlights the fact that the regions where  $|\nabla H|$  is large have a lower weight in the average since the volume of the infinitesimal domain included in  $\mathcal{N}_{\Delta E}(E)$  and centered at  $(q, p) \in \mathcal{S}(E)$  is proportional to  $|\nabla H(q, p)|^{-1}$ , see Figure 1.5. This observation is consistent with the result (1.26) below, obtained with the co-area formula, and motivates the factor  $|\nabla H(q, p)|^{-1}$  on the right-hand side of (1.26).

Once the measure  $\delta_{H(q,p)-E}(dq dp)$  is defined, the microcanonical measure is obtained by a suitable normalization:

$$\mu_{\text{mc}, E}(dq dp) = Z_E^{-1} \delta_{H(q,p)-E}(dq dp),$$

where the partition function used in the normalization

$$Z_E = \int_{\mathcal{S}(E)} \delta_{H(q,p)-E}(dq dp)$$

is assumed to be finite. See the discussion after (1.28) for some sufficient conditions to this end.

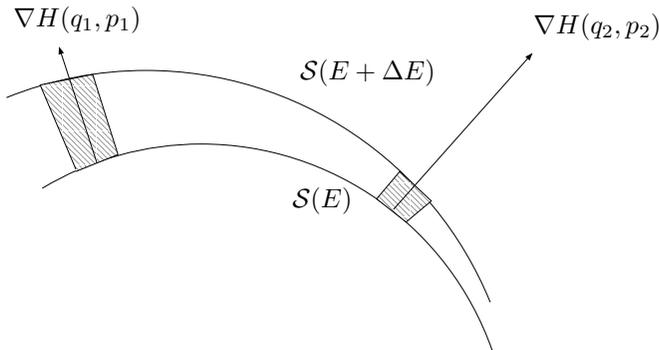


Fig. 1.5 Limiting procedure used to construct the microcanonical measure. The volume of the infinitesimal domain between  $\mathcal{S}(E)$  and  $\mathcal{S}(E + \Delta E)$  centered at a given point  $(q, p) \in T^*\mathcal{D}$  is proportional to  $|\nabla H|^{-1}$ .

**An alternative definition of the microcanonical measure.** The measure  $\delta_{H(q,p)-E}(dq dp)$  for a given energy level  $E$  has support in  $\mathcal{S}(E)$ , and is defined by the following relation: for all test functions  $f : T^*\mathcal{D} \rightarrow \mathbb{R}$

and  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\int_{T^*\mathcal{D}} g(H(q, p)) f(q, p) dq dp = \int_{\mathbb{R}} g(E) \int_{\mathcal{S}(E)} f(q, p) \delta_{H(q, p) - E}(dq dp) dE. \quad (1.25)$$

By the co-area formula (see (3.12)), an alternative expression of the measure  $\delta_{H(q, p) - E}(dq dp)$  is

$$\delta_{H(q, p) - E}(dq dp) = \frac{\sigma_{\mathcal{S}(E)}(dq dp)}{|\nabla H(q, p)|}, \quad (1.26)$$

where  $\sigma_{\mathcal{S}(E)}(dq dp)$  is the area measure induced by the Lebesgue measure on the manifold  $\mathcal{S}(E)$  when the phase space is endowed with the standard Euclidean scalar product (see Remark 3.4 and Section 3.3.2.1 for further precision on the definition of surface measures).

The microcanonical measure can then be rewritten as

$$\mu_{\text{mc}, E}(dq dp) = Z_E^{-1} \delta_{H(q, p) - E}(dq dp) = Z_E^{-1} \frac{\sigma_{\mathcal{S}(E)}(dq dp)}{|\nabla H(q, p)|}, \quad (1.27)$$

with

$$Z_E = \int_{\mathcal{S}(E)} \delta_{H(q, p) - E}(dq dp) = \int_{\mathcal{S}(E)} \frac{\sigma_{\mathcal{S}(E)}(dq dp)}{|\nabla H(q, p)|}. \quad (1.28)$$

The partition function  $Z_E$  is finite for instance when  $\mathcal{S}(E)$  is bounded and  $|\nabla H| \neq 0$  on this set. Since we consider only separable Hamiltonians, the condition  $|\nabla H(q, p)| = 0$  is equivalent to  $p = 0$  and  $\nabla V(q) = 0$ . Therefore,  $|\nabla H| \neq 0$  is ensured as soon as  $\nabla V(q) \neq 0$  for all configurations  $(q, 0) \in \mathcal{S}(E)$ .

**The microcanonical measure as an ergodic limit.** Practitioners often see microcanonical averages as ergodic limits over Hamiltonian trajectories. Notice first that  $\mu_{\text{mc}, E}(dq dp)$  is invariant by the Hamiltonian dynamics flow  $\phi_t$  for all energy levels  $E$ . Indeed, by the conditioning formula (1.25),

$$\begin{aligned} & \int_{\mathbb{R}} g(E) \int_{\mathcal{S}(E)} f(\phi_t(q, p)) \delta_{H(q, p) - E}(dq dp) dE \\ &= \int_{T^*\mathcal{D}} g(H(q, p)) f(\phi_t(q, p)) dq dp \\ &= \int_{T^*\mathcal{D}} g(H \circ \phi_{-t}(Q, P)) f(Q, P) dQ dP \\ &= \int_{T^*\mathcal{D}} g(H(Q, P)) f(Q, P) dQ dP \\ &= \int_{\mathbb{R}} g(E) \int_{\mathcal{S}(E)} f(q, p) \delta_{H(q, p) - E}(dq dp) dE, \end{aligned}$$

where we have used the change of variables  $(Q, P) = \phi_t(q, p)$  and the invariance of the Hamiltonian by the flow  $\phi_t$ . Therefore,

$$\int_{\mathcal{S}(E)} f(q, p) \delta_{H(q,p)-E}(dq dp) = \int_{\mathcal{S}(E)} f \circ \phi_t(q, p) \delta_{H(q,p)-E}(dq dp) \quad (1.29)$$

for all times  $t \in \mathbb{R}$ , which shows the claimed invariance. A more intuitive way to understand this equality is to realize that

$$\frac{1}{\Delta E} \int_{\mathcal{N}_{\Delta E}(E)} f(Q, P) dQ dP = \frac{1}{\Delta E} \int_{\mathcal{N}_{\Delta E}(E)} f \circ \phi_t(q, p) dq dp$$

by the same change of variables as above, and then to use (1.24) to obtain (1.29) in the limit  $\Delta E \rightarrow 0$ .

In view of the preservation of the microcanonical measure by the Hamiltonian flow, the following ergodicity assumption can therefore be considered: Thermodynamic integrals of the form (1.1) are computed as trajectorial averages

$$\int_{\mathcal{S}(E)} A(q, p) \mu_{\text{mc}, E}(dq dp) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T A(\phi_t(q, p)) dt, \quad (1.30)$$

where  $\phi_t$  is the flow of the Hamiltonian dynamics (1.8), and the initial condition  $(q^0, p^0)$  is such that  $H(q^0, p^0) = E$ .

Ergodicity can be rigorously shown for completely integrable systems and their perturbations (see for instance [Arnol'd (1989)]). In general however, no convergence result can be stated, and examples of non-ergodicity can be found. A simple instance of non-ergodicity is the following. Consider the one-dimensional double-well potential

$$V(q) = (q^2 - 1)^2. \quad (1.31)$$

The submanifolds  $\mathcal{S}(E)$  for  $E < 1$  are composed of two simply connected subdomains, and ergodicity can only be expected in a given connected component (see Figure 1.6). Other instances of non-ergodicity cases are situations when there are other invariants than the energy (such as the total momentum of the system, for instance). In those cases, ergodicity is possible only with respect to the Lebesgue measure conditioned to the set of all the invariants of the dynamics.

From a numerical viewpoint, the computation of averages according to the right-hand side of (1.30) requires very stable algorithms allowing a

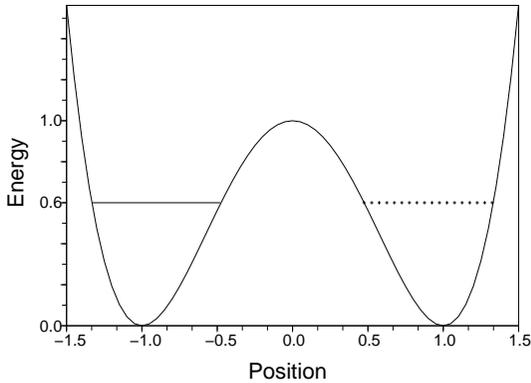


Fig. 1.6 Accessible positions in the energy surface  $H(q, p) = 0.6$  for the double-well potential (1.31). If the dynamics starts in one of the connected components, it remains there.

longtime integration of the Hamiltonian dynamics with a very good preservation of the energy, such as the Verlet algorithm (1.22). The numerical analysis of microcanonical sampling methods based on these properties (in the very particular case of completely integrable systems) can be read in [Cancès *et al.* (2004, 2005)]. There exist also stochastic methods based on constrained diffusion processes to sample the microcanonical measure, see [Faou (2006); Faou and Lelièvre (2009)]. The aim of these methods is to destroy all invariants of the dynamics, except the energy.

### 1.2.3.2 *The canonical ensemble*

In many physical situations, systems in contact with some energy thermostat are considered, rather than isolated systems with a fixed energy. In this case, the energy of the system fluctuates. It however has a fixed temperature. In this situation, the microscopic configurations are distributed according to the so-called *canonical measure*. The canonical ensemble is also often termed *NVT ensemble*, since the number of particles  $N$ , the volume  $V$  and the temperature  $T$  are fixed.

We first define the canonical measure, then give some elements on its derivation from a principle of entropy maximization under constraints, and close this section with a brief presentation of some techniques to sample the canonical measure.

**Definition of the canonical measure.** We assume in the sequel that  $e^{-\beta V} \in L^1(\mathcal{D})$ . The canonical probability measure  $\mu$  on  $T^*\mathcal{D}$  reads

$$\mu(dq dp) = Z_\mu^{-1} \exp(-\beta H(q, p)) dq dp, \quad (1.32)$$

where  $\beta = 1/(k_B T)$  ( $T$  denotes the temperature and  $k_B$  the Boltzmann constant). The normalization constant

$$Z_\mu = \int_{T^*\mathcal{D}} \exp(-\beta H(q, p)) dq dp$$

in (1.32) is called the *partition function*. When the Hamiltonian  $H$  is separable, the canonical measure is of the tensorized form

$$\mu(dq dp) = \nu(dq) \kappa(dp),$$

where  $\nu$  and  $\kappa$  are the two following probability measures:

$$\nu(dq) = Z_\nu^{-1} e^{-\beta V(q)} dq, \quad Z_\nu = \int_{\mathcal{D}} e^{-\beta V(q)} dq, \quad (1.33)$$

and

$$\kappa(dp) = \left(\frac{\beta}{2\pi}\right)^{3N/2} \prod_{i=1}^N m_i^{-3/2} \exp\left(-\frac{\beta}{2} p^T M^{-1} p\right) dp. \quad (1.34)$$

Under  $\mu$ , the position  $q$  and the momentum  $p$  are independent random variables. Therefore, sampling configurations  $(q, p)$  according to the canonical measure  $\mu(dq dp)$  can be performed by independently sampling positions according to  $\nu(dq)$  and momenta according to  $\kappa(dp)$ .

It is straightforward to sample from  $\kappa$  since the momenta are Gaussian random variables. The actual issue is therefore to sample from  $\nu$ . Appropriate methods are presented in Sections 2.1 and 2.2.

**Some elements on the derivation of the canonical measure.** The expression (1.32) of the canonical probability measure can be obtained by maximizing the entropy under the constraint that the energy is fixed *in average*. Such a derivation is performed in [Balian (2007)] for instance. The constraint that the average energy of the system is fixed formalizes the idea that the system under study exchanges energy with the thermostat or energy reservoir to which it is coupled. The energy is therefore not fixed, but it has nonetheless a well-defined average value.

Consider a measure which has a density  $\rho(q, p)$  with respect to the Lebesgue measure. The constraints on the admissible functions  $\rho(q, p)$  are

$$\rho \geq 0, \quad \int_{T^*\mathcal{D}} \rho(q, p) dq dp = 1, \quad \int_{T^*\mathcal{D}} H \rho(q, p) dq dp = E \quad (1.35)$$

for some energy level  $E$ . The first two conditions ensure that  $\rho$  is the density of a probability measure, while the last one expresses the conservation of the energy in average.

The statistical entropy is defined as

$$\mathfrak{S}(\rho) = - \int_{T^*\mathcal{D}} \rho(q, p) \ln \rho(q, p) dq dp. \quad (1.36)$$

It quantifies the amount of information missing, or the “degree of disorder” as is sometimes stated in a more physical language. The entropy is non-positive since  $x \ln x - x + 1 \geq 0$  for all  $x > 0$ . We refer to Chapter 3 in [Balian (2007)] for further precision on the properties of  $\mathfrak{S}$ . The statistical entropy allows us to give a rigorous meaning to the idea that a thermodynamic measure is (quoting [Balian (2007)], Section 4.1.3) “*the most disordered macrostate possible compatible with the data,*” or, equivalently, the measure which “*contains no more information than is strictly necessary to take the data into account.*” The amount of information or disorder is quantified by the entropy.

The canonical measure is recovered as the solution to the following optimization problem

$$\sup \left\{ \mathfrak{S}(\rho), \rho \in L^1(T^*\mathcal{D}), \rho \geq 0, \int_{T^*\mathcal{D}} \rho = 1, \int_{T^*\mathcal{D}} H\rho = E \right\}. \quad (1.37)$$

Formally, the Euler-Lagrange equation satisfied by an extremum reads

$$\mathfrak{S}'(\rho) + \lambda + \gamma H = 0,$$

where  $\lambda, \gamma$  are the Lagrange multipliers associated with the two constraints in (1.37) (normalization and average energy fixed). Since  $\mathfrak{S}'(\rho) = 1 + \ln \rho$ , a candidate maximizer in (1.37) is the measure with density

$$\exp(-1 - \lambda - \gamma H(q, p)).$$

Usually, the Lagrange multiplier  $\gamma$  associated with the energy constraint is denoted by  $\beta$ , and  $\exp(1 + \lambda) = Z$  is a normalization constant. The Lagrange multiplier  $\beta$  exists and is unique since

$$\beta \mapsto \mathcal{E}(\beta) = \frac{\int_{T^*\mathcal{D}} H e^{-\beta H}}{\int_{T^*\mathcal{D}} e^{-\beta H}}$$

is an increasing function. This is a consequence of the positivity of the derivative of the average energy

$$\mathcal{E}'(\beta) = \frac{\int_{T^*\mathcal{D}} (H - \mathcal{E}(\beta))^2 e^{-\beta H}}{\int_{T^*\mathcal{D}} e^{-\beta H}}$$

when  $H$  is not constant.

It is easy to verify that the canonical measure is indeed the unique maximizer of (1.37), as shown in Section 4.2 of [Balian (2007)]. For the sake of completeness, we sketch the proof of this statement. Consider an arbitrary function satisfying (1.35). Using the inequality  $\ln x \leq x - 1$  (with equality if and only if  $x = 1$ ):

$$\int_{T^*\mathcal{D}} \rho_1 \ln \rho_2 - \int_{T^*\mathcal{D}} \rho_1 \ln \rho_1 = \int_{T^*\mathcal{D}} \rho_1 \ln \left( \frac{\rho_2}{\rho_1} \right) \leq \int_{T^*\mathcal{D}} \rho_1 - \rho_2 = 0$$

when  $\rho_1$  and  $\rho_2$  satisfy the constraints (1.35). Equality holds if and only if  $\rho_1(q, p) = \rho_2(q, p)$  almost everywhere. Then, choosing  $\rho_2(q, p) = Z^{-1} \exp(-\beta H(q, p))$ , it holds, for any  $\rho$  satisfying the constraints (1.35):

$$-\int_{T^*\mathcal{D}} \rho \ln \rho \leq -\int_{T^*\mathcal{D}} \rho \ln (Z^{-1} e^{-\beta H}) \leq \ln Z + \beta \int_{T^*\mathcal{D}} H \rho.$$

In view of the energy constraint (last condition in (1.35)),

$$\mathfrak{S}(\rho) \leq \ln Z + \beta E = \mathfrak{S}(Z^{-1} e^{-\beta H}),$$

with equality if and only if  $\rho(q, p) = Z^{-1} \exp(-\beta H(q, p))$ . This shows that the canonical measure is indeed the unique maximizer of the entropy under the constraints (1.35).

**Sampling the canonical measure.** Let us now describe briefly some techniques to sample the canonical measure (1.32). We will rely on these methods in Section 1.3.4 when we present methods to compute free energy differences. The techniques we consider here are stochastic dynamics  $t \mapsto (q_t, p_t)$  which are ergodic for the canonical measure, in the sense that the expectation of a given observable

$$\mathbb{E}_\mu(A) = \int_{T^*\mathcal{D}} A(q, p) \mu(dq dp),$$

where  $\mu$  is the canonical measure (1.32), can be obtained as an ergodic limit

$$\mathbb{E}_\mu(A) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T A(q_t, p_t) dt \quad (1.38)$$

over one realization of the stochastic dynamics. The dynamics we use are motivated solely by the ergodicity property (1.38), and should therefore be seen as a sampling mean. We do not care whether the evolution is physically relevant since we are only interested in time-independent equilibrium properties.

As will be seen in more detail in Section 2.2.3, (1.38) holds under mild assumptions on the potential for the Langevin dynamics

$$\begin{cases} dq_t = M^{-1}p_t dt, \\ dp_t = -\nabla V(q_t) dt - \gamma M^{-1}p_t dt + \sigma dW_t, \end{cases} \quad (1.39)$$

where  $W_t$  is a standard  $dN$ -dimensional Brownian motion, and  $\gamma, \sigma > 0$  verify

$$\sigma^2 = \frac{2\gamma}{\beta}. \quad (1.40)$$

The relation (1.40) is called *fluctuation-dissipation relation* since it relates the magnitude of the dissipative term  $-\gamma M^{-1}p_t dt$  and the magnitude of the random term  $\sigma dW_t$ . The Langevin dynamics may be seen as the superposition of a Hamiltonian dynamics, which preserves the energy, and a stochastic process on the momenta which ensures that the energy levels are visited according to their weight in the canonical ensemble. The latter condition fixes the magnitude of the random term.

Actually, since the difficult task is the sampling of the configurational part  $\nu$  of the canonical measure, we could consider a dynamics on the configurational space only, such as the *overdamped Langevin dynamics* (see Section 2.2.2 for further precision, and Section 2.2.4 for a motivation of the terminology):

$$dq_t = -\nabla V(q_t) dt + \sqrt{\frac{2}{\beta}} dW_t, \quad (1.41)$$

where  $W_t$  is again a standard  $dN$ -dimensional Brownian motion. Under reasonable assumptions on the potential, this dynamics satisfies

$$\mathbb{E}_\nu(A) = \int_{\mathcal{D}} A(q) \nu(dq) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T A(q_t) dt,$$

where  $\nu$  is given in (1.33). Intuitively, each term in (1.41) can be motivated as follows: the gradient force  $-\nabla V(q_t)$  ensures that the energy decreases (which in turn, ensures that the visited configurations are likely enough), while the random noise term supplies some energy so that the temperature is correct. The precise balance between the drift term which removes energy in average and the stochastic term is determined by the condition that the canonical measure is preserved by the dynamics (1.41).

### 1.2.3.3 Other thermodynamic ensembles

We saw in Section 1.2.3.2 that the Boltzmann-Gibbs probability measure (1.32) can be seen as the phase space probability measure maximizing the statistical entropy among the set of phase space probability measures compatible with the observed macroscopic data (in this case, average energy given). The derivation performed for an average energy fixed may be performed for any average thermodynamic quantity, leading to other thermodynamic ensembles. The choice of the ensemble amounts to choosing which quantities are fixed exactly or in average.

We present in this section a general derivation of thermodynamic ensembles associated with a given set of constraints, and next focus on two examples, the isobaric-isothermal ensemble (NPT) where the number of particles, the pressure and the temperature are fixed, and the grand-canonical ensemble ( $\mu$ VT) where the chemical potential, the volume and the temperature are fixed. Many other cases could be treated in a similar fashion (fixed temperature and magnetization for a spin system, fixed temperature and average velocity for a fluid, etc.). This section is not necessary for understanding the remainder of the book, and can be omitted in a first reading.

**General derivation.** Assume that the microscopic state of the system is described by  $(q, p, x)$ , where  $(q, p)$  denotes as above a phase space configuration, and where  $x \in \mathcal{X}$  is some additional degree of freedom. We denote by  $\mathcal{D}_x$  and  $T^*\mathcal{D}_x$  the set of admissible positions  $q$  and configurations  $(q, p)$  for a given value of  $x$ , so that the set of admissible configurations  $(q, p, x)$  is the space

$$\mathcal{E} = \bigcup_{x \in \mathcal{X}} T^*\mathcal{D}_x \times \{x\}.$$

Denote by  $\lambda(dq dp dx)$  some reference measure on  $\mathcal{E}$ . This measure expresses the *a priori* information available on the system. Here, we will consider a reference measure of the form

$$\lambda(dq dp dx) = 1_{(q,p) \in T^*\mathcal{D}_x} dq dp \pi(dx).$$

The conditional measure with respect to the parameter  $x$  (*i.e.* the measure obtained in the  $(q, p)$  variables when the parameter  $x$  is kept fixed) is the usual Lebesgue measure on the set of admissible configurations. The reference measure  $\pi$  on the variable  $x$  depends on the problem at hand.

Consider then a measure  $\rho(dq dp dx)$  describing the macroscopic state of the system, and several observables  $A_1, \dots, A_M$ , functions of  $(q, p, x)$ , whose

averages are fixed. We assume that the measure  $\rho(dq dp dx)$  is absolutely continuous with respect to the reference measure  $\lambda(dq dp dx)$ , and denote, with an abuse of notation, by  $\rho(q, p, x)$  the corresponding density. In this setting, the entropy is defined as

$$\mathfrak{S}_\lambda(\rho) = - \int_{\mathcal{E}} \rho(q, p, x) \ln \rho(q, p, x) \lambda(dq dp dx),$$

and the probability measure describing the system is obtained as the solution of the following maximization problem:

$$\sup_{\rho \in \mathcal{S}(A_1^0, \dots, A_M^0)} \left\{ \mathfrak{S}_\lambda(\rho) \right\}, \quad (1.42)$$

with

$$\begin{aligned} & \mathcal{S}(A_1^0, \dots, A_M^0) \\ &= \left\{ \rho \in L^1(\lambda) \mid \rho \geq 0, \int_{\mathcal{E}} \rho d\lambda = 1, \int_{\mathcal{E}} A_i \rho d\lambda = A_i^0, \forall i \in \{1, \dots, M\} \right\}. \end{aligned}$$

The necessary condition to be satisfied by an extremum of (1.42) reads

$$\mathfrak{S}'_\lambda(\rho) + \alpha_0 + \sum_{i=1}^M \alpha_i A_i = 0.$$

Therefore,

$$\rho(q, p, x) = Z^{-1} \exp \left( \sum_{i=1}^M \alpha_i A_i(q, p, x) \right). \quad (1.43)$$

**Remark 1.2 (Nonequilibrium steady states).** *Let us stress that the above derivation is performed under the assumption that the system is at equilibrium. In particular, no notion of dynamics is required. For nonequilibrium systems in a steady state, the dynamics has to be made precise. It is not always clear whether a stationary probability measure exists, and, when it exists, whether it is unique and whether the distribution of the microscopic configurations converges to it. There are some positive results, see [Rey-Bellet (2006)] in the case of heat transport in one-dimensional atom chains. In general however, no explicit expression of the invariant measure is available, in contrast to formulas such as (1.43).*

**Isobaric-isothermal ensemble (NPT).** Let us now present a first application of the above general derivation. Isobaric-isothermal ensembles are characterized by the fact that the energy and the volume of the system are fixed in average only. Consider for example a periodic system for which the size of the unit cell can vary in one direction, and denote by  $x > 0$  the length of unit cell in this direction (while it is fixed to  $L$  in the two remaining directions). Then,

$$\mathcal{D}_x = [x\mathbb{T} \times (L\mathbb{T})^2]^N, \quad T^*\mathcal{D}_x = [x\mathbb{T} \times (L\mathbb{T})^2]^N \times \mathbb{R}^{3N}.$$

We choose a uniform measure on all possible volumes:

$$\mathcal{X} = (0, +\infty), \quad \lambda(dq dp dx) = 1_{(q,p) \in T^*\mathcal{D}_x} 1_{x>0} dq dp dx.$$

The constraints to be taken into account are  $A_1 = H$  (average energy fixed), and  $A_2(x, q, p) = xL^2$  (average volume fixed).

Applying the results of the general derivation to the NPT case, it is easily seen that the probability measure describing the equilibrium is

$$\rho_{\text{NPT}}(dq dp dx) = Z_{\text{NPT}}^{-1} e^{-\beta PL^2 x} e^{-\beta H(q,p)} 1_{\{q \in [x\mathbb{T} \times (L\mathbb{T})^2]^N\}} dq dp dx,$$

where the Lagrange multiplier associated with the volume constraint is written as  $\beta P$ . The quantity  $P$  can be identified with the pressure.

**Grand canonical ensemble ( $\mu\text{VT}$ ).** We now describe a second application of the above general derivation. Consider a fluid of  $N$  indistinguishable particles. The additional variable describing the microscopic state of the system is the number  $N \in \mathbb{N}^*$  of particles contained in a periodic cubic box of volume  $L^3$ . For a given number  $N$  of particles, the set of admissible configurations is

$$T^*\mathcal{D}_N = (L\mathbb{T})^{3N} \times \mathbb{R}^{3N}.$$

The reference measure for the number  $N$  of particles

$$\sum_{n=1}^{+\infty} \frac{1}{n!} \delta_n(dN)$$

is the uniform measure on the set of positive integers, up to factors  $n!$  which are related to the indistinguishability of the particles. (See for instance Chapter 3 in [Minlos (2000)] for further precision on the construction of the reference measure for the grand-canonical ensemble.) Therefore,

$$\lambda(dq dp dN) = \sum_{n=1}^{+\infty} \frac{1}{n!} 1_{(q,p) \in T^*\mathcal{D}_n} dq dp \delta_n(dN).$$

We denote by  $H_n$  the Hamiltonian function on each space  $T^*\mathcal{D}_n$ , which is a function of the variables  $(q_1, \dots, q_n, p_1, \dots, p_n)$ . The Hamiltonian  $H$  is then defined as  $H(q, p, n) = H_n(q, p)$  for  $(q, p) \in T^*\mathcal{D}_n$ .

The constraints to be taken into account are  $A_1 = H$  (average energy fixed) and  $A_2(x, q, N) = N$  (average number of particles fixed). Applying the results of the general derivation, the grand-canonical equilibrium measure reads:

$$\rho_{\mu VT}(dq dp dN) = Z_{\mu VT}^{-1} \sum_{n=1}^{+\infty} \frac{e^{\beta\mu n}}{n!} e^{-\beta H_n(q,p)} \mathbf{1}_{(q,p) \in T^*\mathcal{D}_n} dq dp \delta_n(dN), \quad (1.44)$$

where  $\beta\mu$  is the Lagrange multiplier associated with the average number constraint.<sup>3</sup> The parameter  $\mu$  can be identified with the chemical potential.

### 1.3 Free energy and its numerical computation

Free energy is a central concept in thermodynamics and in modern studies on biochemical and physical systems. The statistical physics definition of this quantity as the logarithm of the partition function

$$F = -\frac{1}{\beta} \ln \int_{T^*\mathcal{D}} e^{-\beta H(q,p)} dq dp$$

is motivated in Section 1.3.1.

In many applications, the important quantity is actually the *free energy difference* between various macroscopic states of the system, rather than the free energy itself. Free energy differences allow to quantify the relative likelihood of different states. A state should be understood here as either

- (i) the collection of all possible microscopic configurations, distributed according to the canonical measure (1.32), and satisfying a given macroscopic constraint  $\xi(q) = z$ , where  $\xi : \mathcal{D} \rightarrow \mathbb{R}^m$  with  $m$  small. In this case,  $z$  is the index of the state; or
- (ii) the collection of all possible microscopic configurations distributed according to the canonical measure associated with a Hamiltonian depending on some parameter  $\lambda$ . The parameter  $\lambda$  is then the index of the state.

---

<sup>3</sup>The notation  $\mu$  for the chemical potential, standard in the physics and chemistry literature, should not be confused with the notation used for the canonical measure throughout this book.

This is explained in more detail in Section 1.3.2, where two examples are also provided.

Beside these practical motivations to compute free energy differences, a more numerical motivation is to use the free energy to devise algorithms which overcome sampling barriers. Indeed, it is often the case in practice that approximations (1.23) to ensemble averages exhibit a slow convergence. The trajectory generated by the numerical method typically remains trapped for a long time in some region of the phase space, and hops only occasionally to another region, where it also remains trapped for a long time. This occurs as soon as there exist several regions of phase space separated by very low probability areas. Such regions are called *metastable*. The concept of metastability may be formalized in various ways, see Section 2.3.2.2. Chemical and physical intuitions may guide the practitioners of the field toward the identification of some slowly evolving degree of freedom responsible for the metastable behavior of the system. This quantity is a function  $\xi(q)$  of the configuration of the system, where  $\xi : \mathcal{D} \rightarrow \mathbb{R}^m$  with  $m$  small. The framework to consider is therefore the case of transitions indexed by a reaction coordinate. If the function  $\xi$  is well chosen (*i.e.* if the dynamics in the direction orthogonal to  $\xi$  is not too metastable), the free energy can be used as a biasing potential to accelerate the sampling (1.23), see Section 1.3.3.

It is thus important to accurately compute free energy differences in order to assess the relative likelihood of physical states or to build efficient algorithms to overcome sampling barriers. The most important techniques to this end are briefly presented in Section 1.3.4, and then detailed in the following chapters.

### 1.3.1 Absolute free energy

We first define the free energy in Section 1.3.1.1, and then motivate this definition from a macroscopic thermodynamics perspective (Section 1.3.1.2).

#### 1.3.1.1 Definition

We restrict ourselves throughout the book to the canonical ensemble, though most of the concepts and numerical methods considered can be extended to other thermodynamic ensembles (see Section 1.2.3.3). The *absolute free energy* of a system is defined as

$$F = -\frac{1}{\beta} \ln Z_\mu, \quad (1.45)$$

where  $Z_\mu$  is the partition function

$$Z_\mu = \int_{T^* \mathcal{D}} e^{-\beta H(q,p)} dq dp. \quad (1.46)$$

Since the potential energy function  $V$  (hence the Hamiltonian  $H$ ) is defined only up to an additive constant when empirical potential functions are used, so is the absolute free energy. However, this has no consequence on free energy differences, see Section 1.3.2 below.

The free energy (1.45) is called the Helmholtz free energy. Similar free energies can be considered for other thermodynamic ensembles. They are also logarithms of the partition functions multiplied by a factor  $-\beta^{-1}$ . When the isobaric-isothermal ensemble (NPT) is considered, the associated free energy is called the Gibbs free energy.

For separable Hamiltonians (1.2), the partition function can be rewritten as

$$Z_\mu = Z_\nu \left( \frac{2\pi}{\beta} \right)^{3N/2} \prod_{i=1}^N m_i^{3/2}, \quad Z_\nu = \int_{\mathcal{D}} e^{-\beta V(q)} dq,$$

and the only difficulty is the computation of the configurational partition function  $Z_\nu$ . This partition function cannot be computed as such in general. It however has a simple expression for some specific systems, such as the ideal gas, or solids at low temperature (resorting to the phonon spectrum, *i.e.* assuming that the interactions can be approximated by a sum of harmonic interactions), see [Frenkel and Smit (2002); Rickman and LeSar (2002)].

### 1.3.1.2 Relationship with macroscopic thermodynamics

We now motivate the definition (1.45) of the free energy, and also comment on the limits of the theory.

**Analogy with the definition in thermodynamics.** A first motivation for the definition (1.45) relies on an analogy with macroscopic thermodynamics, where the Helmholtz free energy of a system at constant temperature  $T$  is defined as

$$F = U - TS, \quad (1.47)$$

$U$  being the internal energy of the system, and  $S$  its entropy. The microscopic definition of the internal energy is the average energy as given by the laws of statistical physics:

$$\mathbb{E}_\mu(H) = Z_\mu^{-1} \int_{T^* \mathcal{D}} H(q, p) e^{-\beta H(q, p)} dq dp, \quad (1.48)$$

where the normalization constant  $Z_\mu$  is given by (1.46). Besides, the microscopic quantity, proportional to the statistical entropy (1.36) encountered in Section 1.2.3.2,

$$\Sigma = -k_B \int_{T^* \mathcal{D}} \ln \left( \frac{d\mu}{dq dp} \right) d\mu \quad (1.49)$$

is proportional to the mathematical relative entropy of the canonical measure (1.32) with respect to the Lebesgue measure  $dq dp$ . It has many similarities with the (macroscopic) thermodynamic entropy  $S$ , as shown in [Gibbs (1902)]. Replacing  $U$  and  $S$  in (1.47) by their microscopic counterparts defined in (1.48) and (1.49) respectively, we obtain a quantity  $\mathcal{F}$  which should be *similar* to some free energy:

$$\mathcal{F} = \mathbb{E}_\mu(H) - T\Sigma = -\frac{1}{\beta} \ln Z_\mu. \quad (1.50)$$

**Work and heat exchanges.** A second motivation for the definition (1.45) relies on a decomposition of energy exchanges into work and heat for isothermal transformations. This requires however the notion of free energy differences, so that the corresponding discussion is postponed to Section 1.3.2.1.

**Validity and relevance of these motivations.** In spite of the formal analogies highlighted above, the relationships between the microscopic definition of the free energy or entropy, and their counterparts in classical macroscopic thermodynamics are still not completely clear. Quoting [Balian (2007)] (see Section 3.4.6):

Notwithstanding the many interrelations which have been established between the different kinds of entropy, the identification of the thermodynamic entropy and the statistical entropy has not yet been accepted universally. While the former can be measured more or less directly for systems in thermodynamic equilibrium and thus appears to be a property of the system itself, the latter refers to the knowledge of the system by an observer and does have a nature which is partially subjective, or at least anthropocentric and relative. It certainly may appear paradoxical that these two quantities would be equal to one another.

### 1.3.2 Relative free energies

As mentioned above, the quantity of interest in many applications and in this book is not the absolute free energy, but the *free energy differences* between various states. Typical examples studied by computer simulations include the solvation free energy (which is the free energy difference between a molecule *in vacuo* and its counterpart surrounded by solvent molecules), and the binding free energy of two molecules (this free energy difference determines whether a new drug can have an efficient action on a given protein for example). See [Chipot and Pohorille (2007b)] for other relevant examples in chemistry and biophysics.

In this section, we describe more precisely what we mean by *states*, and how a transition between two states can be defined. As already hinted at in the introduction to this section, two cases should be considered: alchemical transitions (Section 1.3.2.1) and transitions indexed by a reaction coordinate (Section 1.3.2.2). In order to fix the ideas, we illustrate each type of transition with a typical example: computation of chemical potential through Widom insertion in the alchemical case (see Section 1.3.2.3), and dimer molecule in a solvent in the reaction coordinate case (see Section 1.3.2.4). These examples will also be our running examples for the numerical illustrations throughout the book.

#### 1.3.2.1 Alchemical transitions

The so-called *alchemical case* considers transitions indexed by an external parameter  $\lambda$ , independent of the microscopic phase space configuration  $(q, p)$ . Typical examples are the intensity of an applied magnetic field for a spin system, or the constants used in the empirical force fields (such as the energy  $\varepsilon$  or the length  $\sigma$  in the Lennard-Jones potential (1.5)). See Section 1.3.2.3 below and Section 2.8 in [Chipot and Pohorille (2007a)] for more examples. The name “alchemical” refers to the fact that the nature of the particles at hand can be modified in the computer simulation by changing the parameters of the potential describing the molecular interactions.

For a given value of  $\lambda$ , the system is described by a Hamiltonian  $H_\lambda$ . A state is then the collection of all possible microscopic configurations  $T^*\mathcal{D}$ , distributed according to the canonical measure

$$\mu_\lambda(dq dp) = \frac{1}{Z_\lambda} e^{-\beta H_\lambda(q,p)} dq dp, \quad Z_\lambda = \int_{T^*\mathcal{D}} e^{-\beta H_\lambda(q,p)} dq dp. \quad (1.51)$$

An alchemical transition transforms the state  $\lambda = 0$  into the state  $\lambda = 1$ .

The corresponding free energy difference is

$$F(1) - F(0) = -\beta^{-1} \ln \left( \frac{\int_{T^* \mathcal{D}} e^{-\beta H_1(q,p)} dq dp}{\int_{T^* \mathcal{D}} e^{-\beta H_0(q,p)} dq dp} \right). \quad (1.52)$$

It is often the case that  $H_\lambda$  depends on  $\lambda$  only through the potential energy  $V_\lambda$ . In this case, the free energy difference simplifies as

$$F(1) - F(0) = -\beta^{-1} \ln \left( \frac{\int_{\mathcal{D}} e^{-\beta V_1(q)} dq}{\int_{\mathcal{D}} e^{-\beta V_0(q)} dq} \right). \quad (1.53)$$

**Work and heat exchanges in a reversible isothermal transformation.** Now that alchemical transitions have been defined, we can come back to the motivation of the definition of free energy in terms of work and heat exchanges, see Section 1.3.1.2. An alchemical transformation can be considered as isothermal since there is a common thermodynamic temperature in the family of measures (1.51), defined through the factor  $\beta$ .

In accordance with (1.48), the energy of the state described by  $H_\lambda$  is

$$U(\lambda) = \mathbb{E}_{\mu_\lambda}(H_\lambda) = Z_\lambda^{-1} \int_{T^* \mathcal{D}} H_\lambda(q,p) e^{-\beta H_\lambda(q,p)} dq dp,$$

while, in view of (1.49), the corresponding microscopic entropy is

$$\mathfrak{S}(\lambda) = -k_B \int_{T^* \mathcal{D}} \ln \left( \frac{d\mu_\lambda}{dq dp} \right) \mu_\lambda(dq dp).$$

A simple computation shows that

$$\frac{d}{d\lambda} \mathbb{E}_{\mu_\lambda}(H_\lambda) = \mathbb{E}_{\mu_\lambda} \left( \frac{\partial H_\lambda}{\partial \lambda} \right) - \beta \left[ \mathbb{E}_{\mu_\lambda} \left( H_\lambda \frac{\partial H_\lambda}{\partial \lambda} \right) - \mathbb{E}_{\mu_\lambda}(H_\lambda) \mathbb{E}_{\mu_\lambda} \left( \frac{\partial H_\lambda}{\partial \lambda} \right) \right].$$

This relation can be rewritten by decomposing the energy variation of the system (as  $\lambda$  changes) as a work contribution, supplemented with some heat exchange:

$$\frac{dU}{d\lambda} = \frac{dW}{d\lambda} + \frac{dQ}{d\lambda}, \quad \frac{dQ}{d\lambda} = T \frac{d\mathfrak{S}}{d\lambda}. \quad (1.54)$$

Here,  $W$  is the so-called reversible work, by definition equal to the variation of the free-energy  $F(\lambda) = -\beta^{-1} \ln Z_\lambda$ :

$$\frac{dW}{d\lambda} = F'(\lambda) = \mathbb{E}_{\mu_\lambda} \left( \frac{\partial H_\lambda}{\partial \lambda} \right).$$

The equality (1.54) has a well-known analogue in standard thermodynamics: During a reversible isothermal transformation  $d\lambda$ , the heat exchanged with the thermostat, defined as the energy variation minus the exerted work  $dU - \delta W = \delta Q$ , is an exact differential  $\delta Q = T d\mathfrak{S}$  involving the entropy  $\mathfrak{S}$  of the system. This interpretation of (1.54) in terms of standard thermodynamics is due to Boltzmann in its original research on the microscopic interpretation of macroscopic thermodynamics (see Section 1.5 in [Gallavotti (1999)] for further precision). Note that in this microscopic framework the heat exchange is related to the variation of the weight  $Z_\lambda^{-1} e^{-\beta H_\lambda(q,p)}$  of the configurations during a reversible transformation.

### 1.3.2.2 Transitions indexed by a reaction coordinate

In the *reaction coordinate case*, the Hamiltonian of the system is kept fixed. A state is a measure on a submanifold of the phase space. These submanifolds are the level sets of some function, the so-called *reaction coordinate*,

$$\xi : \mathcal{D} \rightarrow \mathbb{R}^m,$$

with  $m \leq 3N$ . Examples of such functions are dihedral angles, or distances between two molecular subgroups, as in the example presented in Section 1.3.2.4 below. To  $\xi$  is associated a foliation of the phase space into submanifolds  $\Sigma(z) = \{q \in \mathcal{D} \mid \xi(q) = z\}$ , so that

$$\mathcal{D} = \bigcup_{z \in \mathbb{R}^m} \Sigma(z).$$

We assume in the sequel that the submanifolds  $\Sigma(z)$  are simply connected,<sup>4</sup> and that  $\Sigma(z_1) \neq \Sigma(z_2)$  when  $z_1 \neq z_2$ .

The free energy difference is related to the relative likelihoods of marginal distributions with respect to  $\xi$ . For the canonical measure (1.32), the marginal distribution is by definition

$$\mu^\xi(dz) = \left( \frac{1}{Z_\mu} \int_{\Sigma(z) \times \mathbb{R}^{3N}} e^{-\beta H(q,p)} \delta_{\xi(q)-z}(dq) dp \right) dz.$$

It is the image of the measure  $\mu$  by  $\xi$ . The measure  $\delta_{\xi(q)-z}(dq)$  is defined as in (1.25) and (1.26) through the relation  $\delta_{\xi(q)-z}(dq) dz = dq$ , see also Sections 3.2.1 and 3.3.2 for further precision. In particular, it can be written as

$$\delta_{\xi(q)-z}(dq) = \frac{\sigma_{\Sigma(z)}(dq)}{|\nabla \xi(q)|}, \quad (1.55)$$

<sup>4</sup>This will be important to state ergodicity results for dynamics constrained to remain on  $\Sigma(z)$ .

where  $\sigma_{\Sigma(z)}(dq)$  is the area measure induced by the Lebesgue measure on the manifold  $\Sigma(z)$  when  $\mathcal{D}$  is equipped with the standard Euclidean scalar product. The free energy is then defined as the log-density of the marginal distribution:

$$e^{-\beta F(z)} dz = \mu^\xi(dz).$$

Thus,  $\exp(-\beta[F(1) - F(0)])$  can be interpreted as the relative likelihood of states in  $\Sigma(1)$  compared to states in  $\Sigma(0)$ . More explicitly,

$$F(z) = -\beta^{-1} \ln \left( \frac{1}{Z_\mu} \int_{\Sigma(z) \times \mathbb{R}^{3N}} e^{-\beta H(q,p)} \delta_{\xi(q)-z}(dq) dp \right). \quad (1.56)$$

The free energy can therefore also be seen as some effective potential associated with  $\xi$ . The function  $z \mapsto F(z)$  is called *potential of mean force*. This terminology is motivated by the fact that  $F'(z)$ , called the *mean force*, is some average force exerted on the system when the reaction coordinate is kept constant, see Chapter 3 for further precision.

The free energy difference between the state  $\Sigma(0)$  and the state  $\Sigma(1)$  is finally defined as

$$F(1) - F(0) = -\beta^{-1} \ln \left( \frac{\int_{\Sigma(1) \times \mathbb{R}^{3N}} e^{-\beta H(q,p)} \delta_{\xi(q)-1}(dq) dp}{\int_{\Sigma(0) \times \mathbb{R}^{3N}} e^{-\beta H(q,p)} \delta_{\xi(q)}(dq) dp} \right). \quad (1.57)$$

For separable Hamiltonians, (1.2), the free energy difference can be rewritten as

$$F(1) - F(0) = -\beta^{-1} \ln \left( \frac{\int_{\Sigma(1)} e^{-\beta V(q)} \delta_{\xi(q)-1}(dq)}{\int_{\Sigma(0)} e^{-\beta V(q)} \delta_{\xi(q)}(dq)} \right). \quad (1.58)$$

Notice that when  $|\nabla \xi|$  is constant, the free energy difference  $F(1) - F(0)$  only depends upon  $\xi$  through  $\Sigma(1)$  and  $\Sigma(0)$  in view of (1.55).

**Remark 1.3 (Choice of the reaction coordinate).** *For a given foliation of the configurational space, the free energy difference depends in general on the choice of the reaction coordinate indexing this foliation. Indeed, consider another reaction coordinate  $\tilde{\xi}$ , defining the same level sets, with in particular*

$$\tilde{\Sigma}(0) = \left\{ q \mid \tilde{\xi}(q) = 0 \right\} = \left\{ q \mid \xi(q) = 0 \right\} = \Sigma(0), \quad (1.59)$$

and a similar relation for  $z = 1$ . For instance,  $\tilde{\xi} = f(\xi)$  with any one-to-one increasing function  $f : [0, 1] \rightarrow [0, 1]$  has the same level sets as  $\xi$ , and satisfies (1.59).

In general, the associated free energy differences  $F(1) - F(0)$  and  $\tilde{F}(1) - \tilde{F}(0)$  are different. Indeed, the surface measure  $\sigma_{\Sigma(0)}(dq)$  is somehow intrinsic (it depends only on the submanifold  $\Sigma(0) = \tilde{\Sigma}(0)$  and on the ambient scalar product), while the measure  $\delta_{\xi(q)-z}(dq)$  depends on the gradients of the reaction coordinates through the factors  $|\nabla\xi(q)|^{-1}$ , see the right-hand side of (1.55). It is therefore a modelling choice to decide which reaction coordinate to use, in particular when comparing results of numerical simulations to experimental measurements. Of course, there are no such issues in the alchemical case.

**Remark 1.4 (Relation with the alchemical setting).**

Alchemical transitions can be considered as a special case of transitions indexed by a reaction coordinate, upon introducing the extended variable  $Q = (\lambda, q)$  and the reaction coordinate  $\xi(Q) = \lambda$ . In this case, the geometry of the submanifolds is very simple since  $|\nabla\xi(Q)| = 1$ . The level sets are  $\Sigma(\lambda) = \{\lambda\} \times \mathcal{D}$ , and the measure  $\delta_{\xi(Q)-\lambda}(dQ)$  in the extended space is the Lebesgue measure  $dq$  on  $\mathcal{D}$ .

Besides, the reaction coordinate case is sometimes considered as a limiting case of the alchemical case, using the family of Hamiltonians

$$H_\lambda^\eta(q) = V(q) + \frac{1}{2\eta} \left( \xi(q) - \lambda \right)^2 + \frac{1}{2} p^T M^{-1} p,$$

and letting  $\eta \rightarrow +\infty$ . See Section 5.1.2 for further precision.

1.3.2.3 A typical alchemical transition: Widom insertion

We describe here the running example used to illustrate simulation results for alchemical transitions. We consider a fluid composed of  $N$  particles and enclosed in a domain  $\mathcal{D} = (L\mathbb{T})^d$ , where the physical dimension is  $d = 2$ . The chemical potential is defined as

$$\mu = F(N + 1) - F(N),$$

where  $F(N)$  is the free energy of the system composed of  $N$  particles in the canonical ensemble for the domain  $(L\mathbb{T})^d$  at a given inverse temperature  $\beta$ . In fact, the chemical potential is equal, in the thermodynamic limit, to the Lagrange multiplier  $\mu$  used to define the grand-canonical measure (1.44) (see Section 5.6.3 in [Balian (2007)]).

The chemical potential can be rewritten as

$$\mu = \mu_{\text{id}} + \mu_{\text{ex}},$$

where the so-called “ideal gas contribution”  $\mu_{\text{id}}$  comes from the kinetic part of the partition function, and has an analytic expression (see [Frenkel and Smit (2002)]). The challenge is the computation of the “excess chemical potential”  $\mu_{\text{ex}}$ , which arises from interactions between the fluid particles. Denoting by  $q^N \in \mathcal{D}_N = (L\mathbb{T})^{dN}$  the positions of  $N$  particles,

$$\mu_{\text{ex}} = -\beta^{-1} \ln \left( \frac{\int_{\mathcal{D}_{N+1}} e^{-\beta V_{N+1}(q^{N+1})} dq^{N+1}}{L^d \int_{\mathcal{D}_N} e^{-\beta V_N(q^N)} dq^N} \right), \quad (1.60)$$

where  $V_N(q^N)$  is the potential energy function for a fluid composed of  $N$  particles, and  $\mathcal{D}_N = (L\mathbb{T})^{dN}$  is the associated configuration space. Notice that  $\mu_{\text{ex}} = 0$  when the potential functions are  $V_N = V_{N+1} = 0$  thanks to the factor  $L^d$  in (1.60).

Denoting the canonical measure for a fluid of  $N$  particles by

$$\mu_N(dq^N) = Z_N^{-1} e^{-\beta V_N(q^N)} dq^N, \quad Z_N = \int_{\mathcal{D}_N} e^{-\beta V(q^N)} dq^N,$$

and defining the energy difference between a fluid of  $N$  and  $N + 1$  particles as

$$\Delta_N V(q^N, q) = V_{N+1}(q^{N+1}) - V_N(q^N),$$

when  $q^{N+1} = (q^N, q)$ , the chemical potential (1.60) can be rewritten in the form (1.53) upon defining a potential function on  $\mathcal{D}_{N+1}$ :

$$V_\lambda(q^N, q) = V_N(q^N) + \lambda \Delta_N V(q^N, q). \quad (1.61)$$

Indeed,

$$\mu_{\text{ex}} = -\frac{1}{\beta} \ln \left( \frac{\int_{\mathcal{D}_{N+1}} e^{-\beta V_1(q^{N+1})} dq^{N+1}}{\int_{\mathcal{D}_{N+1}} e^{-\beta V_0(q^{N+1})} dq^{N+1}} \right).$$

Another expression of the excess chemical potential, which will be useful for later purposes, is:

$$\mu_{\text{ex}} = -\beta^{-1} \ln \left( \frac{1}{L^d} \int_{\mathcal{D}_N \times \mathcal{D}_1} e^{-\beta \Delta_N V(q^N, q^1)} \mu_N(dq^N) dq^1 \right). \quad (1.62)$$

The idea of the alchemical transition is therefore to go from a system of  $N + 1$  particles where one of the particles does not have any interaction

with the others, to a system of  $N+1$  fully interacting particles. When there are sufficiently many particles in the simulation box, adding an extra one requires some energy since some space must be created. The alchemical transition consists in progressively switching on the interactions with the  $(N+1)$ -th particle.

The computational results presented in this book have been obtained for a system with pairwise interactions, so that

$$V_N(q^N) = \sum_{1 \leq i < j \leq N} \mathcal{V}(|q_i - q_j|).$$

As in [Hendrix and Jarzynski (2001); Oberhofer *et al.* (2005)], we use a smoothed Lennard-Jones potential (in order to avoid the singularities at the origin). This potential reads

$$\mathcal{V}(r) = \begin{cases} a - br^2, & 0 \leq r \leq 0.8\sigma, \\ \Phi_{\text{LJ}}(r) + c(r - r_{\text{cut}}) - d, & 0.8\sigma \leq r \leq r_{\text{cut}}, \\ 0, & r \geq r_{\text{cut}}, \end{cases} \quad (1.63)$$

where

$$\Phi_{\text{LJ}}(r) = 2\varepsilon \left( \frac{1}{2} \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right),$$

is the Lennard-Jones potential expressed in length units such that the equilibrium position corresponds to  $r = \sigma$ :  $\Phi'_{\text{LJ}}(\sigma) = 0$ . The value  $r_{\text{cut}} = 2.5\sigma$  is a prescribed cut-off radius. The numbers  $a, b, c, d$  ensure that the potential is  $C^1$ .

#### 1.3.2.4 A typical transition indexed by a reaction coordinate:

##### *Dimer in a solvent*

We now describe the running example used to illustrate simulation results for transitions indexed by a reaction coordinate. We consider a system composed of  $N$  particles in a two-dimensional periodic box of side length  $L$ . Among these particles, two particles (numbered 1 and 2 in the following) are designated to form a dimer while the others are solvent particles.

All particles, except the two particles forming the dimer, interact through the purely repulsive WCA pair potential, which is a truncated Lennard-Jones potential [Dellago *et al.* (1999); Straub *et al.* (1988)]:

$$V_{\text{WCA}}(r) = \begin{cases} 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] + \varepsilon & \text{if } r \leq r_0, \\ 0 & \text{if } r > r_0, \end{cases}$$

where  $r$  denotes the distance between two particles,  $\varepsilon$  and  $\sigma$  are two positive parameters and  $r_0 = 2^{1/6}\sigma$ . The interaction potential between the two particles of the dimer is a double-well potential

$$V_S(r) = h \left[ 1 - \frac{(r - r_0 - w)^2}{w^2} \right]^2, \quad (1.64)$$

where  $h$  and  $w$  are two positive parameters. The total energy of the system is therefore, for  $q \in (L\mathbb{T})^{dN}$  with  $d = 2$ ,

$$V(q) = V_S(|q_1 - q_2|) + \sum_{3 \leq i < j \leq N} V_{\text{WCA}}(|q_i - q_j|) + \sum_{i=1,2} \sum_{3 \leq j \leq N} V_{\text{WCA}}(|q_i - q_j|),$$

where  $q_1$  and  $q_2$  are the positions of the two particles forming the dimer.

The potential  $V_S$  has two energy minima. The first one, at  $r = r_0$ , corresponds to the compact state. The second one, at  $r = r_0 + 2w$ , corresponds to the stretched state. The height of the energy barrier separating the two states is  $h$ . Figure 1.7 presents a schematic view of the system.

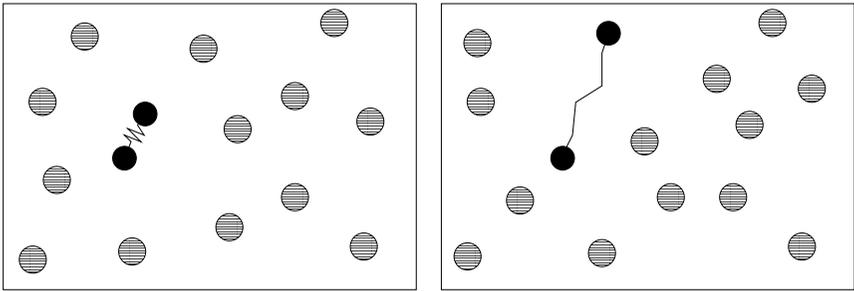


Fig. 1.7 Schematic views of the system, when the dimer is in the compact state (Left), and in the stretched state (Right). The interaction of the particles forming the dimer is described by a double-well potential. All the other interactions are of WCA form.

The reaction coordinate used to describe the transition from the compact to the stretched state is the normalized bond length

$$\xi(q) = \frac{|q_1 - q_2| - r_0}{2w}, \quad (1.65)$$

of the dimer molecule. The compact state (resp. the stretched state) corresponds to the value  $z = 0$  (resp.  $z = 1$ ) of the reaction coordinate.

### 1.3.3 Free energy and metastability

Standard molecular simulation techniques such as those that will be presented in Chapter 2 often experience difficulties in sampling *metastable* potentials. Potentials are called metastable when the corresponding canonical

measure has several regions of high probability separated by low-probability regions. Typical numerical methods spend a lot of time in one given metastable basin, and only rarely escape it to visit another basin. These escapes are rare but fast events. The notion of metastability may be formalized and quantified in several ways, see Section 2.3.2.2 for a more detailed discussion. Some examples of metastable potentials are described in Sections 1.3.3.1 and 1.3.3.2 below.

We motivate in this section the interest of free energy methods for the sampling of metastable potentials. Such methods can be used provided the low- and high-probability regions of the systems are the level sets of some function  $\xi(q)$ , which is still called a reaction coordinate. Alternatively,  $\xi(q)$  can be seen as some slowly evolving degrees of freedom encoding some coarse-grained information on the system. The free energy associated with  $\xi$  may then be used as a biasing potential enforcing transitions from one metastable basin to another. We show an instance of this strategy in Section 1.3.3.3 for the potentials considered in Sections 1.3.3.1 and 1.3.3.2. Of course the reliability of the method crucially depends on the choice of the reaction coordinate. This is a very important problem in practice, unfortunately rather ill-posed.

### 1.3.3.1 A simple example of metastable dynamics

Consider the potential energy

$$V(x, y) = \frac{1}{6} \left[ 4(1 - x^2 - y^2)^2 + 2(x^2 - 2)^2 + ((x + y)^2 - 1)^2 + ((x - y)^2 - 1)^2 \right], \quad (1.66)$$

and a single particle  $q = (x, y)$  evolving according to the overdamped Langevin dynamics:

$$dq_t = -\nabla V(q_t) dt + \sqrt{\frac{2}{\beta}} dW_t.$$

Figure 1.8 presents the level sets of the potential (1.66) and a typical trajectory.

The overdamped Langevin dynamics can be shown to be ergodic for the canonical probability measure  $\nu(dq) = Z^{-1} \exp(-\beta V(q)) dq$  (see Section 2.2.2 for more detail on the overdamped Langevin dynamics and its numerical implementation). The dynamics projected in the  $y$  variable is irrelevant, whereas the time evolution of the  $x$  variable shows that it is a “slow” variable. If the average position  $\mathbb{E}_\nu(x)$  is computed as a time-average along a trajectory, the convergence is very slow (compared to the

convergence of the average  $\mathbb{E}_\nu(y)$  for instance). This suggests to choose  $\xi(x, y) = x$ .

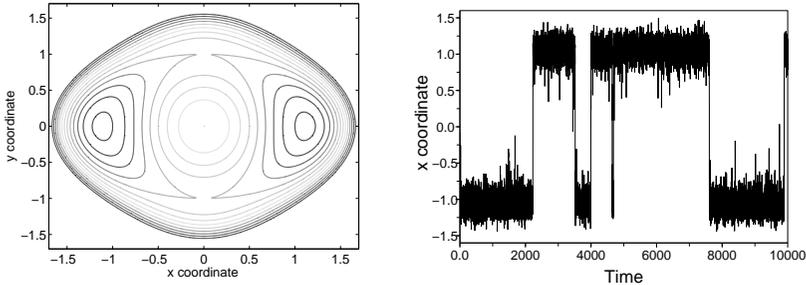


Fig. 1.8 Left: Level sets of the potential (1.66). Right: Projected trajectory in the  $x$  variable for  $\Delta t = 0.01$ ,  $\beta = 6$ .

For later purposes, we compute the free energy profile for the reaction coordinate  $\xi(x, y) = x$ :

$$F(x_2) - F(x_1) = -\beta^{-1} \ln \left( \frac{\psi^\xi(x_2)}{\psi^\xi(x_1)} \right), \quad (1.67)$$

where the marginals  $\psi^\xi$  of the equilibrium canonical distribution are

$$\psi^\xi(x) = \int_{\mathbb{R}} e^{-\beta V(x, y)} dy.$$

This profile is illustrated in Figure 1.9, together with

$$F'(x) = \frac{\int_{\mathbb{R}} \partial_x V(x, y) e^{-\beta V(x, y)} dy}{\int_{\mathbb{R}} e^{-\beta V(x, y)} dy}.$$

Notice that  $F'$  is the opposite of the averaged force experienced in the direction of the reaction coordinate (the so-called mean force). There is a high free energy barrier at  $x = 0$ , which corresponds to a small value of  $\psi^\xi(x)$ . This barrier is at the origin of the metastable behavior since it separates two regions of high probability.

### 1.3.3.2 Entropic and energetic barriers

Free energy barriers can have two origins, related to either energetic or entropic bottlenecks. We give below two toy examples of purely energetic and purely entropic barriers. Of course, in general, both components are

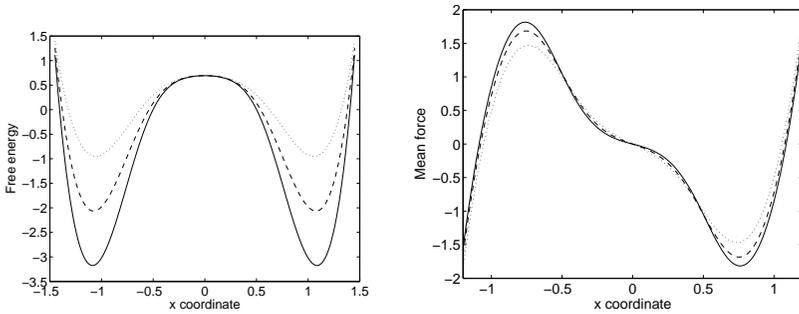


Fig. 1.9 Left: Potential of mean force for the potential plotted in Figure 1.8, using the  $x$  coordinate as reaction coordinate. From top to bottom:  $\beta = 2$  (dotted line),  $\beta = 3$  (dashed line),  $\beta = 4$  (solid line). Right: Associated mean forces.

mixed, and it is not so obvious to decide whether the metastability of the dynamics rather has an energetic or an entropic origin (except in some limiting temperature regime, see the discussion at the end of this section).

**Purely energetic barrier.** Consider  $q = (q_1, \dots, q_N) \in \mathbb{R}^N$ ,  $p \in \mathbb{R}^N$ , and

$$H(q, p) = W(q_1) + V(q_2, \dots, q_N) + \frac{1}{2}p^T M^{-1}p, \quad (1.68)$$

where  $W$  is a one-dimensional double-well potential  $W(q_1) = h(q_1^2 - 1)^2$  with  $h$  large enough. Then, choosing the first coordinate  $q_1$  as a reaction coordinate:  $\xi(q) = q_1$ , it holds (up to a multiplicative constant which does not depend on  $z$ ):

$$e^{-\beta F(z)} = \int_{\mathbb{R}^{2N-1}} e^{-\beta H(z, q_2, \dots, q_N, p_1, \dots, p_N)} dq_2 \dots dq_N dp_1 \dots dp_N,$$

so that

$$F(z_2) - F(z_1) = W(z_2) - W(z_1).$$

In this case, it is clear that free energy barriers are purely of energetic origin.

**Purely entropic barrier.** Entropic barriers are often encountered in complex systems with many degrees of freedom. In this case, the system typically has enough energy to overcome the energetic barriers it can encounter, but has not, somehow, got its energy concentrated in the right modes or directions. It is expected that entropic barriers increase with the

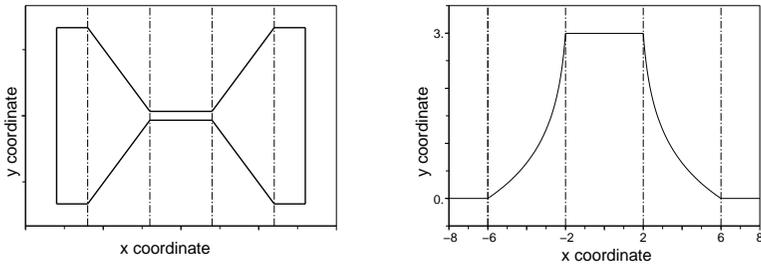


Fig. 1.10 Left: Potential for which entropic barriers have to be overcome, in the case  $L_1 = 2$ ,  $L_2 = 4$  and  $L_3 = 2$ . The potential is 0 in the region enclosed by the curve, and  $+\infty$  outside. Right: Associated free energy profile when the  $x$  coordinate is the reaction coordinate ( $\beta = 1$ ).

dimensionality of the system (think of a random walk in a high-dimensional space).

A toy model of an entropic barrier is the potential presented in Figure 1.10. The potential is zero inside the curve, and  $+\infty$  outside, so that the system is confined in the bone-shaped region. Here,  $q = (x, y) \in \mathcal{D} = \{q \in \mathbb{R}^2 \mid V(q) = 0\}$ . Denote by  $d$  the width of the tunnel between the two metastable regions, by  $2L_1$  its length, by  $L_2$  the length of the transition region, and by  $L_3$  the length of the initial and final rectangular domains, which are of heights  $\Delta$ . We choose  $\xi(q) = x$  as the reaction coordinate. Then,

$$F(x) = \begin{cases} -\beta^{-1} \ln d & \text{when } |x| \leq L_1, \\ -\beta^{-1} \ln \left( d + \frac{\Delta - \delta}{L_2} (|x| - L_1) \right) & \text{when } L_1 \leq |x| \leq L_1 + L_2, \\ -\beta^{-1} \ln \Delta & \text{when } L_1 + L_2 \leq |x| \leq L_1 + L_2 + L_3. \end{cases} \quad (1.69)$$

There is a free energy barrier in the tunnel region, arising from the contraction of the phase space volume: Less configurations are accessible, although the energy has not changed. This barrier has no energy component in it since the average energy for a fixed value of the reaction coordinate is zero.

Figure 1.11 presents a typical trajectory in the case  $L_1 = L_3 = 2$ ,  $L_2 = \Delta = 4$ ,  $\delta = 0.2$ , for a Metropolis random walk with isotropic Gaussian moves of variance  $2\tau/\beta$  (see Section 2.1.2 for further precision on the Metropolis algorithm). Here, this amounts to proposing a new position  $\tilde{q}^{n+1}$  as

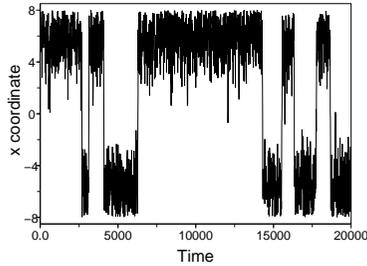


Fig. 1.11 Typical trajectory of the variable  $x$  for the potential presented in Figure 1.10, when a Metropolis dynamics is used, for the parameters  $\tau = 0.1$  and  $\beta = 1$ . The time variable is defined as the number of iterations times the typical time  $\tau$ .

$$\tilde{q}^{n+1} = q^n + \sqrt{\frac{2\tau}{\beta}} G^n,$$

where  $(G^n)_{n \geq 0}$  are independent and identically distributed centered Gaussian random variables of identity covariance; and setting  $q^{n+1} = \tilde{q}^{n+1}$  when  $\tilde{q}^{n+1} \in \mathcal{D}$ , and  $q^{n+1} = q^n$  otherwise. The simulation results show that the  $x$  coordinate only significantly varies on long timescales, which is a typical signature of metastability.

**Temperature dependence of the free energy barrier.** The temperature dependence of the free energy barrier is a good indicator of the nature of the bottleneck. Indeed, in the case of a purely energetic barrier (1.68), the ratio of the marginal distributions

$$e^{-\beta(F(z_1) - F(z_0))} = e^{-\beta(W(z_1) - W(z_0))}$$

varies exponentially as a function of  $\beta$ , whereas, for the example (1.69) of purely entropic barrier, this ratio does not depend on  $\beta$ . In general, it is expected that free energy barriers at low temperatures (*i.e.* in the limit  $\beta \rightarrow +\infty$ ) are mostly of energetic nature, in accordance with large deviation principles [Freidlin and Wentzell (1998)]. On the other hand, at high temperatures (in the limit  $\beta \rightarrow 0$ ),

$$\begin{aligned} F(z_1) - F(z_0) &= -\frac{1}{\beta} \ln \left( \frac{\int_{\mathcal{D}} e^{-\beta V(q)} \delta_{\xi(q) - z_1}(dq)}{\int_{\mathcal{D}} e^{-\beta V(q)} \delta_{\xi(q) - z_0}(dq)} \right) \\ &\simeq -\frac{1}{\beta} \ln \left( \frac{\int_{\mathcal{D}} \delta_{\xi(q) - z_1}(dq)}{\int_{\mathcal{D}} \delta_{\xi(q) - z_0}(dq)} \right), \end{aligned}$$

provided the integrals

$$I(z) = \int_{\mathcal{D}} \delta_{\xi(q)-z}(dq)$$

are finite for  $z_0$  and  $z_1$ . In this case the free energy difference is controlled at first order by the entropic contribution. Indeed,  $I(z)$  measures the accessible phase space for the constraint  $\xi(q) = z$ , and some entropy can be defined from this volume according to Boltzmann's definition of the entropy as the logarithm of a density of states.

### 1.3.3.3 Free energy biased sampling

In the simple examples considered in Sections 1.3.3.1 and 1.3.3.2, the slowly evolving variable is known. There is a clear free energy barrier when using  $\xi(x, y) = x$  as a reaction coordinate for (1.67) and (1.69). It is then possible to bias the dynamics in the  $x$  variable in order to remove the free energy barrier. More precisely, we now sample the modified potential

$$V(q) - F(\xi(q)).$$

Notice that the free energy associated to the reaction coordinate  $\xi$  for this modified potential is constant:

$$-\beta^{-1} \ln \int_{\Sigma(z)} e^{-\beta(V-F\circ\xi)(q)} \delta_{\xi(q)-z}(dq) = \beta^{-1} \ln Z_\mu.$$

The above formula is a consequence of the definition (1.56) of the free energy  $F(z)$ , using also the equality  $F(\xi(q)) = F(z)$  on  $\Sigma(z)$ . The marginal law of  $Z^{-1} e^{-\beta(V-F\circ\xi)(q)} dq$  along  $\xi$  is therefore the uniform law.

If  $\xi$  completely describes the metastability of the potential  $V$  as in the previous examples, the modified potential  $V - F\circ\xi$  is no longer metastable. An efficient importance sampling method can then be obtained, especially when  $F$  does not vary too much (see Section 2.4.1.4 for further precision on importance sampling). We now numerically illustrate this strategy.

**Application to the two-dimensional double-well potential.** Consider the system described by the potential (1.66). Figure 1.12 presents the new potential  $V - F\circ\xi$  (where the free energy bias, computed with standard quadrature rules, has been applied for  $|x| \leq 1.7$ ) and a typical trajectory of the overdamped Langevin dynamics for the potential  $V - F\circ\xi$ , projected on the  $x$  coordinate. The comparison with Figure 1.8 shows that the transitions from the region  $x < 0$  to the region  $x > 0$  are now sufficiently frequent in order to attain good sampling accuracies.

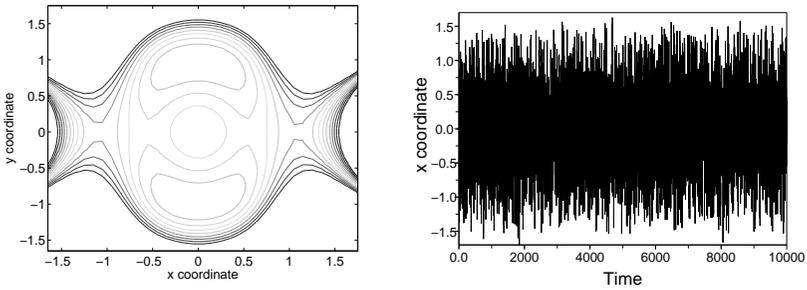


Fig. 1.12 Left: Modified potential  $V - F \circ \xi$ . Right: Projected trajectory in the  $x$  variable for  $\Delta t = 0.01$ ,  $\beta = 6$  for the dynamics associated with the modified potential.

**Application to the entropic barrier problem.** Figure 1.13 presents the results for a Metropolis random-walk dynamics biased by the free energy (1.69) in the case of the potential presented in Figure 1.10 (see Section 1.3.3.2 for a brief description of the dynamics). As in the previous case, the metastability is removed, and many transitions are observed from one well to the other (compare with Figure 1.11). The effect of the free energy bias is to increase the likelihood of regions close to the transition zone, so that many more crossings are attempted.

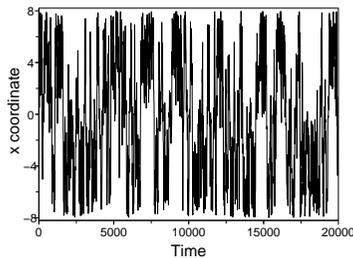


Fig. 1.13 Typical trajectory for the potential exhibiting an entropic barrier when the dynamics is biased by the analytically-known free energy. The numerical parameters are the same as for Figure 1.11.

### 1.3.4 Computational techniques for free energy differences

We present in this section the key ideas behind the methods currently available to compute free energy differences. Some of these techniques are

suiting both for alchemical transitions and transitions indexed by a reaction coordinate, but not all of them. In our opinion, the currently available techniques fall within the following four classes:

- (i) The first technique, dating back to [Kirkwood (1935)], is *thermodynamic integration*, which mimics the quasi-static evolution of a system as a succession of equilibrium samplings (this amounts to an infinitely slow switching between the initial and final states). In practice, it allows to compute free energy differences by integrating the derivative of the free energy, which happens to be a canonical average for a fixed value of the reaction coordinate or alchemical parameter. This technique can be used both for alchemical transitions and transitions indexed by reaction coordinates, see Chapter 3;
- (ii) The second one is based on straightforward sampling methods. In the alchemical case, the *free energy perturbation method*, introduced in [Zwanzig (1954)], recasts free energy differences as usual canonical averages (see Section 2.4.1). In the reaction coordinate case, usual sampling methods can also be employed, relying on *histogram methods* (see Section 2.5);
- (iii) A more recent class of methods relies on dynamics with an imposed schedule for the reaction coordinate or the alchemical parameter. These techniques therefore use *nonequilibrium dynamics*. Equilibrium properties can however be recovered from the nonequilibrium trajectories with a suitable exponential reweighting, see [Jarzynski (1997b, a)]. This technique can handle both alchemical transitions and transitions indexed by reaction coordinates, see Chapter 4. It also has many similarities with free-energy perturbation since the corresponding free-energy estimators have the same mathematical structure (exponential averages);
- (iv) Finally, *adaptive biasing dynamics* may be used in the reaction coordinate case. The switching schedule is not imposed *a priori*, but a biasing term in the dynamics forces the transition by penalizing the regions which have already been visited. This biasing term can be a biasing force as for the *Adaptive Biasing Force* technique of [Darve and Pohorille (2001)], or a biasing potential as for the Wang-Landau method [Wang and Landau (2001b, a)], *nonequilibrium metadynamics* [Iannuzzi *et al.* (2003)] or Self-Healing Umbrella Sampling [Marsili *et al.* (2006)].

We refer to Figure 1.14 for a schematic comparison of the computational methods in the reaction coordinate case.

We now give a flavor of these approaches, preferentially in the alchemical setting for simplicity. The remainder of the book is devoted to a thorough presentation of these techniques. Recall that the free energy is defined, up to an additive constant (unimportant as long as free energy differences are concerned), by (1.52) or (1.53) in the alchemical case, and by (1.57) or (1.58) in the reaction coordinate case.

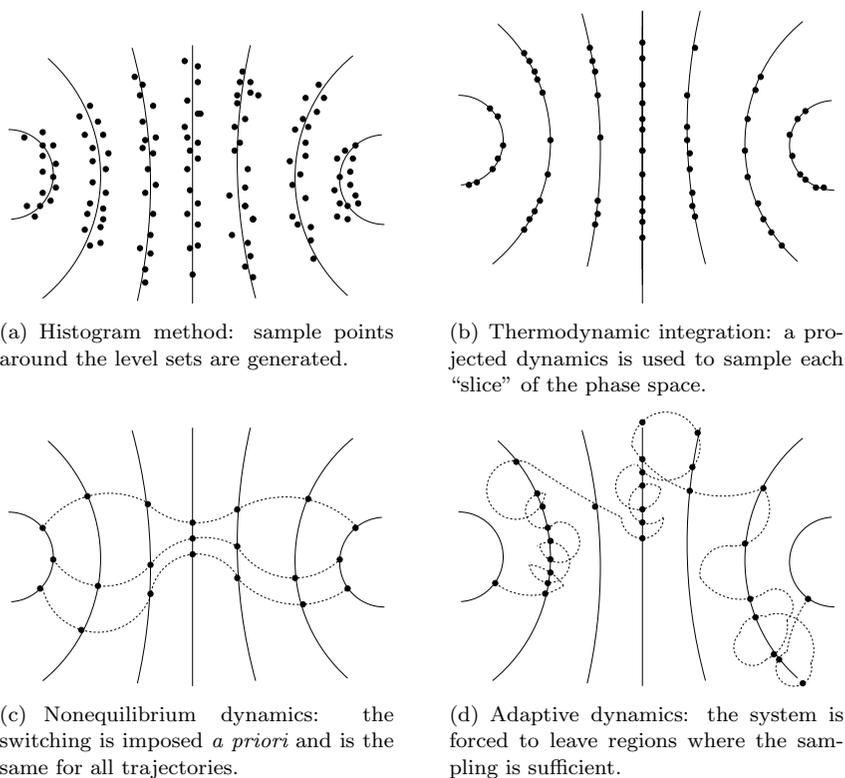


Fig. 1.14 Cartoon comparison of the different techniques to compute free energy differences in the reaction coordinate case.

### 1.3.4.1 Thermodynamic integration

Thermodynamic integration consists in remarking that

$$F(\lambda) - F(0) = \int_0^\lambda F'(s) ds, \quad (1.70)$$

and that the derivative

$$F'(\lambda) = \frac{\int_{T^* \mathcal{D}} \frac{\partial H_\lambda}{\partial \lambda}(q, p) e^{-\beta H_\lambda(q, p)} dq dp}{\int_{T^* \mathcal{D}} e^{-\beta H_\lambda(q, p)} dq dp}$$

is the canonical average of  $\partial_\lambda H_\lambda$  with respect to the canonical measure

$$\mu_\lambda(dq dp) = Z_\lambda^{-1} e^{-\beta H_\lambda(q, p)} dq dp.$$

In practice,  $F'(\lambda_i)$  is computed using classical sampling techniques for a sequence of values  $\lambda_i \in [0, 1]$ . The integral on the right-hand side of (1.70) is then integrated numerically to obtain the free energy difference profile. The extension to transitions indexed by a reaction coordinate is presented in Chapter 3 (for dynamics in position space in Section 3.2 and phase space dynamics in Section 3.3).

### 1.3.4.2 Methods based on straightforward sampling

**Free energy perturbation.** Free energy perturbation is a technique which is restricted to the computation of free energy differences in the alchemical case (see however Remark 1.4 for an extension of the alchemical setting to the reaction coordinate case). It consists in rewriting the free energy difference as

$$\Delta F = -\beta^{-1} \ln \int_{T^* \mathcal{D}} e^{-\beta(H_1 - H_0)} d\mu_0.$$

An approximation of  $\Delta F$  is then obtained by generating configurations  $(q^n, p^n)$  distributed according to  $\mu_0$  and computing the empirical average

$$\frac{1}{N} \sum_{n=1}^N e^{-\beta(H_1 - H_0)(q^n, p^n)}.$$

However, the initial and the final distributions  $\mu_0$  and  $\mu_1$  often hardly overlap. Intermediate steps should then be considered, or some importance sampling strategy should be used to improve the numerical accuracy, see Section 2.4.1.

It is also possible to resort to bridge sampling. In this case, the free energy difference  $\Delta F$  is estimated using sample points from  $\mu_0$  and  $\mu_1$ , see Section 2.4.2.

**Histogram methods.** In the reaction coordinate case, a naive algorithm to compute approximate free energy differences would be to sample configurations using a simple dynamics ergodic with respect to the canonical measure (see Chapter 2 for further precision), and to compute approximations of the marginal law in the reaction coordinate. More precisely, this can be done in practice by discretizing the values of the reaction coordinate into small intervals, and approximating the free energy by computing the canonical average of the indicator function of these intervals in the limit when the interval width  $\Delta z$  goes to 0. Defining

$$\chi_{z,\Delta z}(q) = \frac{1}{\Delta z} \mathbf{1}_{|\xi(q)-z|\leq\Delta z/2},$$

it holds

$$\begin{aligned} -\frac{1}{\beta} \ln \mathbb{E}_\mu(\chi_{z,\Delta z}) &= -\frac{1}{\beta} \ln \left( \frac{1}{Z_\mu} \int_{T^*\mathcal{D}} \frac{\mathbf{1}_{|\xi(q)-z|\leq\Delta z/2}}{\Delta z} e^{-\beta H(q,p)} dq dp \right) \\ &\rightarrow F(z) = -\frac{1}{\beta} \ln \left( \frac{1}{Z_\mu} \int_{T^*\mathcal{D}} e^{-\beta H(q,p)} \delta_{\xi(q)-z}(dq) dp \right) \end{aligned} \quad (1.71)$$

when  $\Delta z \rightarrow 0$ . However, the metastable features of the dynamics used for sampling usually prevent such a simple strategy from being efficient, see Section 1.3.3. The idea of histogram methods is to sample configurations centered on some level set  $\Sigma(z)$ , typically by sampling canonical measures associated with modified potentials

$$V(q) + \frac{1}{2\eta} (\xi(q) - z)^2,$$

where  $\eta > 0$  is a small parameter, and to construct a global sample for the canonical measure  $\mu(dq dp)$  by concatenating the sample points (with some appropriate weighting factor), see Section 2.5 for further precision. Once this global sample is obtained, an approximation of the free energy is obtained with (1.71) (for  $\Delta z$  small enough).

### 1.3.4.3 Nonequilibrium dynamics

Free energy differences can be expressed as a nonlinear average over nonequilibrium trajectories, using the so-called Jarzynski equality, see (1.74) below. This equality can easily be obtained for a system governed by Hamiltonian dynamics, with *initial conditions at equilibrium*, canonically distributed according to  $\mu_0$ , and subjected to a switching schedule  $\Lambda : [0, T] \rightarrow \mathbb{R}$  with  $\Lambda(0) = 0$  and  $\Lambda(T) = 1$ . More precisely, we consider initial conditions  $(q(0), p(0)) \sim \mu_0$ , which are evolved according to

the following non-autonomous ordinary differential equation for  $0 \leq t \leq T$  (compare with (1.8)):

$$\begin{cases} \frac{dq}{dt}(t) = \nabla_p H_{\Lambda(t)}(q(t), p(t)), \\ \frac{dp}{dt}(t) = -\nabla_q H_{\Lambda(t)}(q(t), p(t)). \end{cases} \quad (1.72)$$

Defining by  $\phi^\Lambda$  the associated flow, the work performed on the system starting from some initial conditions  $(q, p)$  is

$$\mathcal{W}(q, p) = \int_0^T \frac{\partial H_{\Lambda(t)}}{\partial \lambda}(\phi_t^\Lambda(q, p)) \Lambda'(t) dt = H_1(\phi_T^\Lambda(q, p)) - H_0(q, p). \quad (1.73)$$

The last equality is obtained by noticing that

$$\begin{aligned} \frac{d}{dt} \left( H_{\Lambda(t)}(\phi_t^\Lambda(q, p)) \right) = \\ \frac{\partial H_{\Lambda(t)}}{\partial \lambda}(\phi_t^\Lambda(q, p)) \Lambda'(t) + \left( \begin{array}{c} \nabla_q H_{\Lambda(t)}(\phi_t^\Lambda(q, p)) \\ \nabla_p H_{\Lambda(t)}(\phi_t^\Lambda(q, p)) \end{array} \right) \cdot \partial_t \phi_t^\Lambda(q, p), \end{aligned}$$

and the second term on the right-hand side vanishes in view of (1.72). Then,

$$\int_{T^* \mathcal{D}} e^{-\beta \mathcal{W}(q, p)} d\mu_0(q, p) = Z_0^{-1} \int_{T^* \mathcal{D}} e^{-\beta H_1(\phi_T^\Lambda(q, p))} dq dp.$$

Since  $\phi_T^\Lambda$  defines a change of variables of Jacobian 1, the above equality can be restated as

$$\mathbb{E}_{\mu_0}(e^{-\beta \mathcal{W}}) = \frac{Z_1}{Z_0} = e^{-\beta(F(1) - F(0))}, \quad (1.74)$$

where the expectation is taken with respect to initial conditions distributed according to  $\mu_0$ . The extension to stochastic dynamics, for transitions indexed by a reaction coordinate or an alchemical parameter, is presented in Chapter 4.

In view of the equality (1.74), it is already clear that the lowest values of the work dominate the nonlinear average (1.74), and the distribution of weights  $e^{-\beta \mathcal{W}(q, p)}$  is often degenerate in practice. This prevents in general an accurate numerical computation of the (1.74), and raises issues very similar to the ones encountered with free-energy perturbation. Refined strategies are therefore needed to use nonequilibrium methods in practice (see Chapters 4 and 6).

### 1.3.4.4 Adaptive dynamics

Adaptive dynamics may be seen as some adaptive importance sampling strategy, with a biasing potential at time  $t$  function of the reaction coordinate. The biasing potential converges in the longtime limit to the free energy by construction of the dynamics.

To illustrate this strategy, we consider the case of the Adaptive Biasing Force (ABF) method [Darve and Porohille (2001); Hénin and Chipot (2004)] in the simple example when the reaction coordinate  $\xi(q) = q_1$  has values in  $\mathbb{T}$ , while the remaining coordinates  $q_{2\dots N}$  belong to  $\mathbb{R}^{N-1}$ . Recall that, when  $\xi$  adequately describes the metastabilities of the system, the dynamics biased by the free energy is less metastable than the original dynamics (see Section 1.3.3 for two typical examples).

Let us assume that we know the free energy  $F$ . Denoting by  $q_t = (q_{1,t}, q_{2\dots N,t})$  the current configuration of the system, the overdamped Langevin dynamics associated with the modified potential  $V - F \circ \xi$  reads

$$\left\{ \begin{array}{l} dq_t = -\left(\nabla V(q_t) - F'(q_{1,t}) e_1\right) dt + \sqrt{\frac{2}{\beta}} dW_t, \\ F'(z) = \mathbb{E}_\nu\left(\partial_{q_1} V(q) \mid \xi(q) = z\right) = \frac{\int_{\mathbb{R}^{N-1}} \partial_{q_1} V(z, q_{2\dots N}) e^{-\beta V(z, q_{2\dots N})} dq_{2\dots N}}{\int_{\mathbb{R}^{N-1}} e^{-\beta V(z, q_{2\dots N})} dq_{2\dots N}}, \end{array} \right. \quad (1.75)$$

where  $e_1 = (1, 0, \dots, 0)^T$  is the unit vector in the  $q_1$  direction. Denote by

$$\tilde{\nu}(dq) = \tilde{Z}^{-1} \exp\left(-\beta(V(q) - F(q_1))\right) dq$$

the stationary measure of the process (1.75). The equilibrium mean force  $F'(z)$  can actually be rewritten as a canonical average with respect to  $\tilde{\nu}$ , conditionally on  $q_1 = z$ :

$$F'(z) = \mathbb{E}_\nu\left(\partial_{q_1} V(q) \mid \xi(q) = z\right) = \mathbb{E}_{\tilde{\nu}}\left(\partial_{q_1} V(q) \mid \xi(q) = z\right). \quad (1.76)$$

Indeed, the bias  $F(\xi(q))$  is constant when  $\xi(q)$  is kept constant. Therefore, conditional averages with respect to  $\tilde{\nu}$  for  $\xi(q) = z$  fixed are equal to conditional averages with respect to the canonical measure (1.33) since the factor  $e^{-\beta F(\xi(q))}$  cancels out in the numerator and denominator of the conditional average.

Now, of course,  $F$  is not known in practice. In view of (1.75)-(1.76), it seems natural to replace, in the dynamics (1.75), the conditional expectation with respect to the stationary measure in the expression of the

equilibrium mean force, by the conditional expectation with respect to the current law of  $q_t$ :

$$\begin{cases} dq_t = -\left(\nabla V(q_t) - F'_t(q_{1,t}) e_1\right) dt + \sqrt{\frac{2}{\beta}} dW_t, \\ F'_t(z) = \mathbb{E}\left(\partial_{q_1} V(q_t) \mid \xi(q_t) = z\right). \end{cases} \quad (1.77)$$

Notice that the biasing potential  $F_t$  now explicitly depends on the time variable. Denoting by  $\psi_t(q) dq$  the law of  $q_t$  at time  $t$  (intuitively, the distribution of configurations obtained by simulating an infinite number of replicas interacting only through the common bias they are constructing), the biasing force  $F'(z)$  can be rewritten in a form closer to the expression in (1.75):

$$F'_t(z) = \frac{\int_{\mathbb{R}^{N-1}} \partial_{q_1} V(z, q_{2\dots N}) \psi_t(z, q_{2\dots N}) dq_{2\dots N}}{\int_{\mathbb{R}^{N-1}} \psi_t(z, q_{2\dots N}) dq_{2\dots N}}.$$

We now motivate why the adaptive dynamics (1.77) may be relevant. The distribution of the variable  $\xi(q_t) = q_{1,t}$  is given by the marginal law with density

$$\psi_t^\xi(z) = \int_{\mathbb{R}^{N-1}} \psi_t(z, q_{2\dots N}) dq_{2\dots N}.$$

A simple computation (see Section 5.2.3.1) shows that

$$\partial_t \psi_t^\xi(z) = \frac{1}{\beta} \partial_z^2 \psi_t^\xi(z).$$

The above diffusion equation implies that  $\psi_t^\xi$  converges (exponentially fast) to the uniform distribution on  $\mathbb{T}$ . Therefore, the metastable features associated with  $\xi$  are suppressed. Heuristically, the simple diffusion equation in the direction  $q_1$  is not too surprising since the biasing force  $F'_t$  aims precisely at counteracting in average the force experienced by the system in the direction  $q_1$ .

Besides, the dynamics (1.77) in the  $q_{2\dots N}$  variable (at fixed  $z$ ) is an overdamped Langevin dynamics associated with the potential  $V(z, q_{2\dots N})$ . Assuming that the dynamics is at equilibrium conditionally on the  $z$  variable, the distribution of the variable  $q_{2\dots N}$  at fixed  $z$  is equal to the canonical conditional distribution:

$$\frac{\psi_t(z, q_{2\dots N})}{\psi_t^\xi(z)} dq_{2\dots N} = Z_z^{-1} e^{-\beta V(z, q_{2\dots N})} dq_{2\dots N}.$$

Recall also that the marginal law  $\psi_t^\xi$  converges to the uniform law. On the other hand,  $\tilde{\nu}(dq)$  is the unique probability measure whose marginal distribution in the  $\xi$  variable is the uniform law, while the conditional distributions at fixed values of  $\xi$  are equal to the canonical conditional distributions. This motivates the convergence of  $\psi_t(q) dq$  towards  $\tilde{\nu}(dq)$ , and therefore the convergence of  $F_t$  towards  $F$ .

The above presentation naturally suggests a parallel implementation of the dynamics through many replicas constructing a shared biasing potential. This plain parallel implementation can be enhanced through some selection process on the replicas (see Section 6.2). There exist also adaptive dynamics where the biasing potential  $F_t$  is updated, in contrast to the method presented here where the derivative of the biasing potential is updated. See Section 5.1 for further precision.

## 1.4 Summary of the mathematical tools and structure of the book

Table 1.2 presents in a synthetic manner the techniques used from a mathematical viewpoint for each of the methods presented in Section 1.3.4. This explains the construction of the book: we present the methods in what we consider to be the increasing order of mathematical complexity.

Table 1.2 Mathematical theories used for each free energy technique (MCs = Markov chains, SDEs = Stochastic differential equations).

Free energy perturbation	Time homogeneous MCs and SDEs	Chapter 2
Histogram methods	Time homogeneous MCs and SDEs	Chapter 2
Thermodynamic integration	Projected SDEs and MCs	Chapter 3
Nonequilibrium dynamics	Nonhomogenous MCs and SDEs	Chapter 4
Adaptive dynamics	Nonlinear SDEs and MCs	Chapter 5
Selection procedures	Particle systems and jump processes	Chapter 6

For the reader's convenience, see the dependency diagram in Figure 1.15, which highlights the prerequisites for each chapter. In particular, Sections 2.1 and 2.2 cover some material which will be of constant use for the remainder of the book.

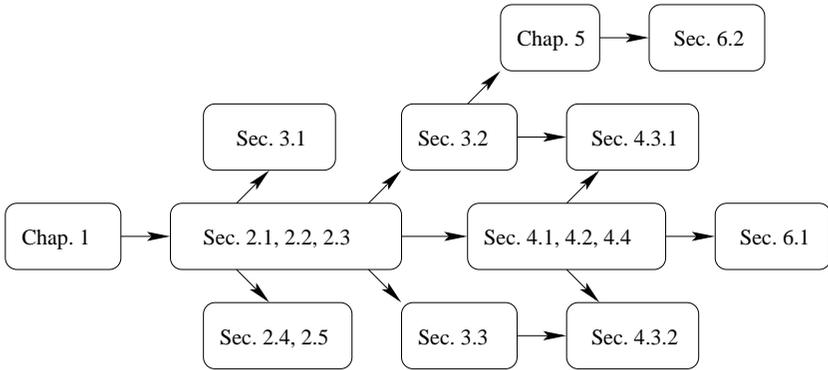


Fig. 1.15 Interdependence of the chapters and sections.