Part I

A Modeling Hierarchy for Simulations

1

Introduction

1.1 What is this book about?

1.1.1 Simulation of real systems

Computer simulations of real systems require a *model* of that reality. A model consists of both *a representation of the system* and *a set of rules* that describe the behavior of the system. For dynamical descriptions one needs in addition a specification of the *initial state* of the system, and if the response to external influences is required, a specification ofthe *external influences*.

Both the model and the method of solution depend on the *purpose* of the simulation: they should be *accurate* and *efficient*. The model should be chosen accordingly. For example, an accurate quantum-mechanical description of the behavior of a many-particle system is not efficient for studying the flow of air around a moving wing; on the other hand, the Navier–Stokes equations – efficient for fluid motion – cannot give an accurate description of the chemical reaction in an explosion motor. Accurate means that the simulation will reliably (within a required accuracy) predict the real behavior of the real system, and efficient means "feasible with the available technical means." This combination of requirements rules out a number of questions; whether a question is answerable by simulation depends on:

- the state of theoretical development (models and methods of solution);
- the computational capabilities;
- the possibilities to implement the methods of solution in algorithms;
- the possibilities to validate the model.

Validation means the assessment of the accuracy of the model (compared to physical reality) by critical experimental tests. Validation is a crucial part of modeling.

3

1.1.2 System limitation

We limit ourselves to models of the real world around us. This is the realm
of chemistry, biology and material sciences, and includes all industrial and
practical applications. We do not include the formation of stars and galax-
ies (*stellar dynamics*) or the physical processes in hot plasma on the sun's
surface (*astrophysics*); neither do we include the properties and interactions
of elementary particles (*quantum chromodynamics*) or processes in atomic
nuclei or neutron stars. And, except for the purposes of validation and
demonstration, we shall not consider unrealistic models that are only meant
to test a theory. To summarize: we shall look at literally "down-to-earth"
systems consisting of atoms and molecules under non-extreme conditions of
pressure and temperature.

This limits our discussion in practice to systems that are made up of
interacting *atomic nuclei*, which are specified by their mass, charge and spin,
electrons, and *photons* that carry the electromagnetic interactions between
the nuclei and electrons. Occasionally we may wish to add gravitational
interactions to the electromagnetic ones. The internal structure of atomic
nuclei is of no consequence for the behavior of atoms and molecules (if we
disregard radioactive decay): nuclei are so small with respect to the spatial
spread of electrons that only their *monopole* properties as total charge and
total mass are important. Nuclear excited states are so high in energy
that they are not populated at reasonable temperatures. Only the spin
degeneracy of the nuclear ground state plays a role when nuclear magnetic
resonance is considered; in that case the nuclear magnetic dipole and electric
quadrupole moment are important as well.

In the normal range of temperatures this limitation implies a practical
division between electrons on the one hand and nuclei on the other: while
all particles obey the rules of quantum mechanics, the quantum character
of electrons is essential but the behavior of nuclei approaches the classical
limit. This distinction has far-reaching consequences, but it is rough and
inaccurate. For example, protons are light enough to violate the classical
rules. The validity of the classical limit will be discussed in detail in this
book.

1.1.3 Sophistication versus brute force

Our interest in *real* systems rather than simplified model systems is con-
sequential for the kind of methods that can be used. Most real systems
concern some kind of condensed phase: they (almost) never consist of iso-
lated molecules and can (almost) never be simplified because of inherent

symmetry. Interactions between particles can (almost) never be described by mathematically simple forms and often require numerical or tabulated descriptions. Realistic systems usually consist of a very large number of interacting particles, embedded in some kind of environment. Their behavior is (almost) always determined by statistical averages over ensembles consisting of elements with random character, as the random distribution of thermal kinetic energy over the available degrees of freedom. That is why statistical mechanics plays a crucial role in this book.

The complexity of real systems prescribes the use of methods that are easily extendable to large systems with many degrees of freedom. Physical theories that apply to simple models only, will (almost) always be useless. Good examples are the very sophisticated statistical-mechanical theories for atomic and molecular fluids, relating fluid structural and dynamic behavior to interatomic interactions. Such theories work for atomic fluids with simplified interactions, but become inaccurate and intractable for fluids of polyatomic molecules or for interactions that have a complex form. While such theories thrived in the 1950s to 1970s, they have been superseded by accurate simulation methods, which are faster and easier to understand, while they predict liquid properties from interatomic interactions much more accurately. Thus sophistication has been superseded by brute force, much to the dismay of the sincere basic scientist.

Many mathematical tricks that employ the simplicity of a toy model system cannot be used for large systems with realistic properties. In the example below the brute-force approach is applied to a problem that has a simple and elegant solution. To apply such a brute-force method to a simple problem seems outrageous and intellectually very dissatisfying. Nevertheless, the elegant solution cannot be readily extended to many particles or complicated interactions, while the brute-force method can. Thus not only sophistication in physics, but also in mathematics, is often replaced by brute force methods. There is an understandable resistance against this trend among well-trained mathematicians and physicists, while scientists with a less elaborate training in mathematics and physics welcome the opportunity to study complex systems in their field of application. The field of simulation has made theory much more widely applicable and has become accessible to a much wider range of scientists than before the "computer age." Simulation has become a "third way" of doing science, not instead of, but in addition to theory and experimentation.

There is a danger, however, that applied scientists will use "standard" simulation methods, or even worse use "black-box" software, without realizing on what assumptions the methods rest and what approximations are
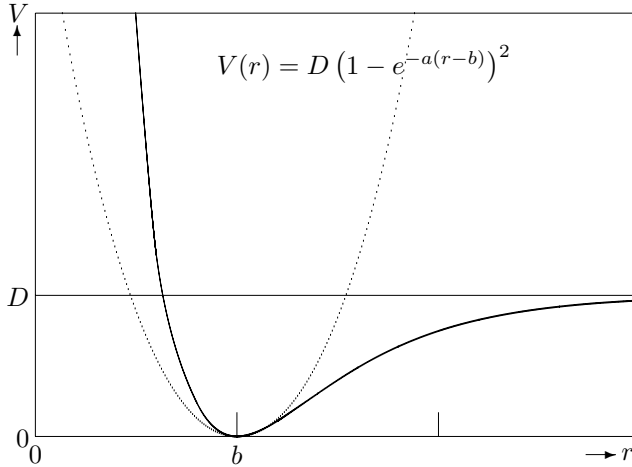
Figure 1.1 Morse curve with $a = 2/b$ (solid curve). Dotted curve: parabola with same curvature as Morse curve at $r = b$: $V = Da^2(r - b)^2$.

implied. This book is meant to provide the necessary scientific background and to promote awareness for the limitations and inaccuracies of simulating the "real world".

Example: An oscillating bond
In this example we use brute-force simulation to attack a problem that could be approached analytically, albeit with great difficulty. Consider the classical bond length oscillation of a simple diatomic molecule, using the molecule hydrogen fluoride (HF) as an example. In the simplest approximation the potential function is a parabola:

$$V(r) = \tfrac{1}{2}k(r - b)^2, \tag{1.1}$$

with r the H–F distance, k the force constant and b the equilibrium distance. A better description of the potential function is the *Morse function* (see Fig. 1.1)

$$V(r) = D\left(1 - e^{-a(r-b)}\right)^2, \tag{1.2}$$

where D is the dissociation energy and a is a constant related to the steepness of the potential. The Morse curve is approximated near the minimum at $r = b$ by a parabola with force constant $k = 2Da^2$.

The Morse curve (Morse, 1929) is only a convenient analytical expression that has some essential features of a diatomic potential, including a fairly good agreement with vibration spectra of diatomic molecules, but there is no theoretical justification for this particular form. In many occasions we may not even have an analytical form for the potential, but know the potential at a number of discrete points, e.g., from quantum-chemical calculations. In that case the best way to proceed is to construct the potential function from *cubic spline interpolation* of the computed points. Be-

Table 1.1 *Data for hydrogen fluoride*

mass H	m_{H}	1.0079	u
mass F	m_{F}	18.9984	u
dissocation constant	D	569.87	kJ/mol
equilibrium bond length	b	0.09169	nm
force constant	k	5.82×10^5	$\mathrm{kJ\,mol^{-1}\,nm^{-2}}$

cause cubic splines (see Chapter 19) have continuous second derivatives, the forces
will behave smoothly as they will have continuous first derivatives everywhere.

A little elementary mechanics shows that we can split off the translational motion
of the molecule as a whole, and that – in the absence of rotational motion – the
bond will vibrate according to the equation of motion:

$$\mu \ddot{r} = -\frac{dV}{dr}, \tag{1.3}$$

where $\mu = m_{\mathrm{H}} m_{\mathrm{F}} / (m_{\mathrm{H}} + m_{\mathrm{F}})$ is the reduced mass of the two particles. When we
start at time $t = 0$ with a displacement Δr and zero velocity, the solution for the
harmonic oscillator is

$$r(t) = b + \Delta r \cos \omega t, \tag{1.4}$$

with $\omega = \sqrt{k/\mu}$. So the analytical solution is simple, and we do not need any nu-
merical simulation to derive the frequency of the oscillator. For the Morse oscillator
the solution is not as straightforward, although we can predict that it should look
much like the harmonic oscillator with $k = 2Da^2$ for small-amplitude vibrations.
But we may expect anharmonic behavior for larger amplitudes. Now numerical sim-
ulation is the easiest way to derive the dynamics of the oscillator. For a spline-fitted
potential we *must* resort to numerical solutions. The extension to more complex
problems, like the vibrations of a molecule consisting of several interconnected har-
monic oscillators, is quite straightforward in a simulation program, while analytical
solutions require sophisticated mathematical techniques.

The reader is invited to write a simple molecular dynamics program that uses
the following very general routine `mdstep` to perform one dynamics step with the
velocity-Verlet algorithm (see Chapter 6, (6.83) on page 191). Define a function
`force(r)` that provides an array of forces $\boldsymbol{F}$, as well as the total potential energy
V, given the coordinates $\boldsymbol{r}$, both for the harmonic and the Morse potential. You
may start with a one-dimensional version. Try out a few initial conditions and
time steps and look for energy conservation and stability in long runs. As a rule
of thumb: start with a time step such that the fastest oscillation period contains
50 steps (first compute what the oscillation period will be). You may generate
curves like those in Fig. 1.2. See what happens if you give the molecule a rotational
velocity! In this case you of course need a two- or three-dimensional version. Keep
to "molecular units": *mass*: u, *length*: nm, *time*: ps, *energy*: kJ/mol. Use the data
for hydrogen fluoride from Table 1.1.

The following function performs one velocity-Verlet time step of MD on a system
of n particles, in m (one or more) dimensions. Given initial positions r, velocities v
and forces F (at position r), each as arrays of shape (n, m), it returns r, v, F and

8   *Introduction*

time (fs)
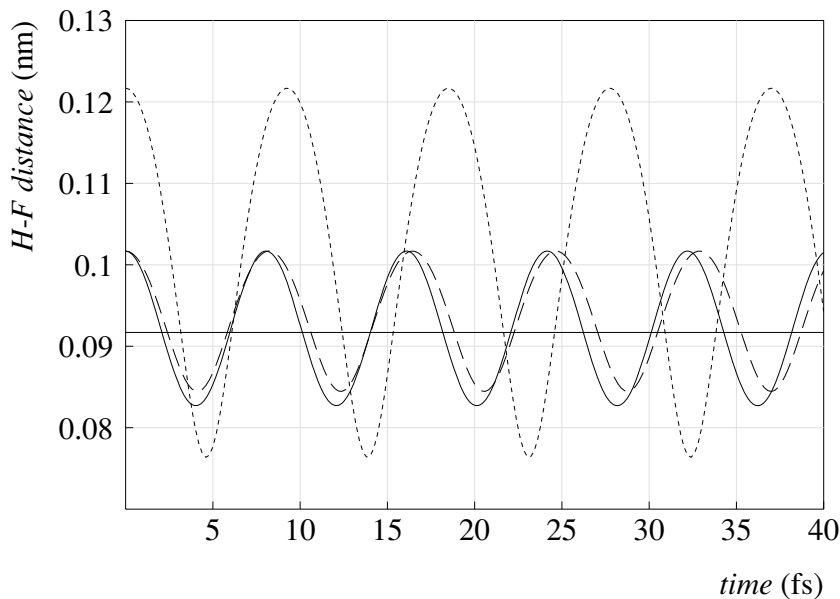
Figure 1.2 Oscillation of the HF bond length, simulated with the harmonic oscil-
lator (solid curve) and the Morse curve (long dash), both with initial deviation
from the equilibrium bond length of 0.01 nm, Dotted curve: Morse oscillator with
initial deviation of 0.03 nm, showing increased anharmonic behavior. Note that the
frequency of the Morse oscillator is lower than that of the harmonic oscillator. A
time step of 0.2 fs was used; the harmonic oscillator simulation is indistinguishable
from the analytical solution.

the potential energy V one time step later. For convenience in programming, the
inverse mass should be given as an array of the same shape (n, m) with repeats of
the same mass for all m dimensions. In Python this $n \times m$ array `invmass` is easily
generated from a one-dimensional array `mass` of arbitrary length n:

`invmass=reshape(repeat(1./mass,m),(alen(mass),m)),`

or equivalently

`invmass=reshape((1./mass).repeat(m),(alen(mass),m))`

An external function `force(r)` must be provided that returns $[F, V]$, given r. V is
not actually used in the time step; it may contain any property for further analysis,
even as a list.

PYTHON PROGRAM 1.1 **mdstep(invmass,r,v,F,force,delt)**
General velocity-Verlet Molecular Dynamics time step

```
01  def mdstep(invmass,r,v,F,force,delt):
02  # invmass: inverse masses [array (n,m)] repeated over spatial dim. m
03  # r,v,F: initial coordinates, velocities, forces [array (n,m)]
04  # force(r): external routine returning [F,V]
05  # delt: timestep
```

```
06   # returns [r,v,F,V] after step
07     v=v+0.5*delt*invmass*F
08     r=r+v*delt
09     FV=force(r)
10     v=v+0.5*delt*invmass*FV[0]
11     return [r,v,FV[0],FV[1]]
```

Comments
As mentioned in the *Preface* (page xiii), it is assumed that *scipy* has been imported. The initial
values of r, v, F, V are valid at the time before the step, and normally available from the output of
the previous step. To start the run, the routine `force(r)` must have been called once to initiate
F. The returned values are valid at the end of the step. The arguments are not modified in place.

1.2 A modeling hierarchy

The behavior of a system of particles is in principle described by the rules of
relativistic quantum mechanics. This is – within the limitation of our system
choices – the highest level of description. We shall call this *level 1*. All other
levels of description, such as considering atoms and molecules instead of
nuclei and electrons, classical dynamics instead of quantum dynamics, or
continuous media instead of systems of particles, represent approximations
to level 1. These approximations can be ordered in a hierarchical sense from
fine atomic detail to coarse macroscopic behavior. Every lower level loses
detail and loses applicability or accuracy for a certain class of systems and
questions, but gains applicability or efficiency for another class of systems
and questions. The following scheme lists several levels in this hierarchy.

LEVEL 1 RELATIVISTIC QUANTUM DYNAMICS	
System	**Rules**
Atomic nuclei (mass, charge, spin), electrons (mass, charge, spin), photons (frequency)	Relativistic time-dependent quantum mechanics; Dirac's equation; (quantum) electrodynamics

Approximation	**No Go**
Particle velocities small compared to velocity of light	Electrons close to heavy nuclei; hot plasmas

LEVEL 2 QUANTUM DYNAMICS	
System	**Rules**
Atomic nuclei, electrons, photons	Non-relativistic time-dependent Schrödinger equation; time-independent Schrödinger equation; Maxwell equations

| **Approximation** Born–Oppenheimer approx.: electrons move much faster than nuclei | | **No Go** Electron dynamics (e.g., in semiconductors); fast electron transfer processes; dynamic behavior of excited states | |

LEVEL 3 ATOMIC QUANTUM DYNAMICS

System	**Rules**
Atoms, ions, molecules, (photons)	Atoms move in effective potential due to electrons; atoms may behave according to time-dependent Schrödinger equation

---

| **Approximation** Atomic motion is classical | | **No Go** Proton transfer; hydrogen and helium at low temperatures; fast reactions and high-frequency motions | |

LEVEL 4 MOLECULAR DYNAMICS

System	**Rules**
Condensed matter: (macro)molecules, fluids, solutions, liquid crystals, fast reactions	Classical mechanics (Newton's equations); statistical mechanics; molecular dynamics

---

| **Approximation** Reduce number of degrees of freedom | | **No Go** Details of fast dynamics, transport properties | |

LEVEL 5 GENERALIZED LANGEVIN DYNAMICS ON REDUCED SYSTEM

System	**Rules**
Condensed matter: large molecular aggregates, polymers, defects in solids, slow reactions	Superatoms, reaction coordinates; averaging over local equilibrium, constraint dynamics, free energies and potentials of mean force.

| **Approximation** Neglect time correlation and/or spatial correlation in fluctuations | | **No Go** Correlations in motion, short-time accuracy | |

LEVEL 6 SIMPLE LANGEVIN DYNAMICS	
System	**Rules**
"Slow" dynamic (non-equilibrium) processes and reactions	Accelerations given by systematic force, friction, and noise; Fokker–Planck equations

Approximation	**No Go**
Neglect inertial terms: coarse graining in time	Dynamic details

LEVEL 7 BROWNIAN DYNAMICS	
System	**Rules**
Coarse-grained non-equilibrium processes; colloidal systems; polymer systems	Velocities given by force and friction, plus noise; Brownian (diffusive) dynamics; Onsager flux/force relations

Approximation	**No Go**
Reduce description to continuous densities of constituent species	Details of particles

LEVEL 8 MESOSCOPIC DYNAMICS	
System	**Rules**
As for level 7: self-organizing systems; reactive non-equilibrium systems	Density description: mass conservation plus dynamic flux equation, with noise.

Approximation	**No Go**
Average over "infinite" number of particles	Spontaneous structure formation driven by fluctuations

LEVEL 9 REACTIVE FLUID DYNAMICS	
System	**Rules**
Non-equilibrium macroscopic mixture of different species (as the atmosphere for weather forecasting	Energy, momentum and mass conservation; reactive fluxes