

# 7 Molecular Dynamics simulations

<sup>1</sup> The aim of molecular dynamics (MD) simulations is to model a system in microscopic detail, over a physical length of time relevant for certain properties of interest. The systems that are studied may consist of, for example, individual atoms, molecules, macromolecules, proteins in a solvent, etc.

An important point is that one studies the molecules through *classical equations of motions*, based on an approximate representation of the inter-molecule and/or intra-molecular forces. The corresponding force laws may be empirically derived or in some cases can be motivated by quantum mechanical calculations.

Molecular dynamics simulations are intimately connected to statistical mechanics, and in fact, for interfering macroscopic properties, concepts of statistical mechanics need to be used. This in particular means that the precise *microstate* (given for example in terms of the positions and velocities of all atoms) of an MD simulation is unimportant, instead we are interested in *ensemble averages of macroscopic variables*, such as temperature, pressure, diffusion coefficient, etc.

In principle, carrying out a proper ensemble average would mean to average over many different simulations of the system. This is usually impossible. One can however resort to the *ergodic hypothesis* which postulates that the ensemble average is equal to the time average of a specific system. We can hence hope to accurately measure the macroscopic thermodynamic properties of a system by looking at a single realization for a long enough time and average our measurements over this time span.

## 7.1 Interaction potentials

Within the Born-Oppenheimer approximation, the total wave function  $\psi_{\text{tot}}$  is separated into the nuclear  $\psi_{\text{n}}$  and electronic wave function  $\psi_{\text{e}}$ ,

$$\psi_{\text{tot}}(R, r) = \psi_{\text{n}}(R)\psi_{\text{e}}(R; r), \quad (7.1)$$

where  $R$  and  $r$  are the coordinates and momenta of the nuclei and electrons, respectively. Thus, the electronic wave function  $\psi_{\text{e}}$  only parametrically depends on the position, not on the dynamics, of the nuclei. As a result of this approximation, Eq. 7.1 separates into two equations, a time-dependent Schrödinger equation for the motion of the nuclei, and a time-independent Schrödinger equation for the electronic dynamics.

---

<sup>1</sup>Lecture notes by Prof. Dr. Volker Springel, extended by Prof. Dr. Frauke Gräter, slightly modified by Prof. Dr. Friedrich Röpke

## 7 Molecular Dynamics simulations

The Born-Oppenheimer approximation allows to treat the electronic wave function of a system as a function of only the nuclear coordinates. Within this framework, the calculation of electronic energies requires the solution of the time-independent Schrödinger equation for the electrons. This, however, is prohibitive due to the large number of electrons in real-world molecular systems.

As a second approximation, therefore, a classical force field for the calculation of the potential energy of the system as a function of nuclear coordinates is derived, from fitting to the quantum-mechanical ground state energy and/or from experimental observables.

Within these approximations, in which electronic degrees of freedom of each atom are neglected, the potential energy for a system containing  $N$  atoms can be written down as

$$V(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{j>i}^N v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{j>i}^N \sum_{k>j}^N v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (7.2)$$

where  $v_2$  are pair-wise interaction potentials,  $v_3$  are triplet potentials, etc. While purely atomistic systems are typically described by only pair potentials,  $V(\mathbf{r}_1, \dots, \mathbf{r}_N)$  of molecular systems includes most commonly also 3-body and 4-body terms, see below. A pair potential only depends on the pair separation,  $v_2(\mathbf{r}_i, \mathbf{r}_j) = v_2(|\mathbf{r}_i - \mathbf{r}_j|)$ .

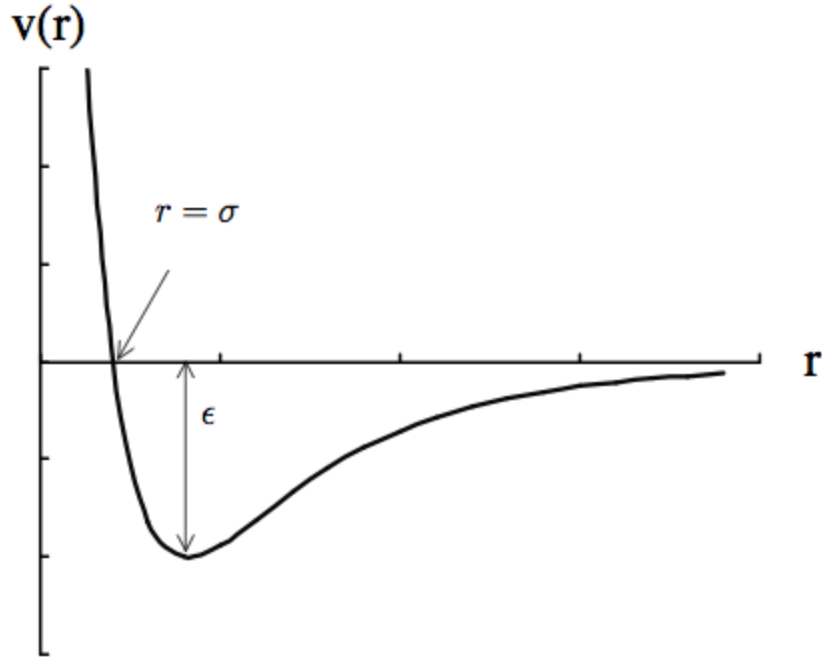
With the kinetic energy of  $\sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i}$ , we arrive at a conservative Hamiltonian system,

$$H = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (7.3)$$

for which we can readily derive equations of motion in the standard way. Once the equations of motion are written down, we can integrate them as an ordinary N-body system, using, for example, the Leapfrog or Verlet integration schemes. While these integration schemes are of low-order, recall that they are symplectic, hence they have particularly good stability properties for long-term integrations of conservative systems.

### 7.1.1 Non-bonded interactions

Let us first consider the simplest case of a monatomic liquid such as argon. It features attractive forces at large distances due to London dispersion or 'van der Waals interactions', and strong repulsive forces at short distances due to the Pauli repulsion. London dispersion can be considered as an attraction between fluctuating dipoles, and scales with  $r_{ij}^{-6}$ , while the Pauli repulsion scales with  $\frac{e^{-r_{ij}}}{r_{ij}}$ .



A potential often used to approximately describe the interaction between pairs of atoms is the 12-6 Lennard-Jones potential:

$$v(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (7.4)$$

Here  $\epsilon$  characterizes the interaction strength, and  $\sigma$  the range. There is a minimum of the potential at  $r_0 = 2^{1/6}\sigma \simeq 1.12\sigma$ , with  $v(r_0) = -\epsilon$ . For distances below  $r_0$ , the force is (strongly) repulsive, mimicking the exponential dependency of the Pauli repulsion. For larger distances, it is attractive and approaches zero quite quickly.

For example, for argon suitable parameters to describe the potential are  $\epsilon/k_B = 120$  K and  $\sigma = 3.4 \times 10^{-8}$  cm.

One should bear in mind that MD simulations of monatomic systems typically neglect multi-body terms such as  $v_3$  in Eq. 7.1. Up to 10 % of the lattice energy of solid state argon has been estimated to arise from non-additive non-pairwise terms. Despite of this significant contribution, these terms are rarely included in computer simulations, first because the pairwise approximation is reasonable, and secondly because triplet terms would render computer simulations in most cases too expensive. Hence, parameters for the Lennard-Jones potential are *effective* parameter taking any higher-order interactions into account.

We note that highly simplified pair potentials might be of interest for general investigations of liquids and comparison to theory. Examples are the hard-sphere potential

$$v^{\text{HS}}(r_{ij}) = \begin{cases} \infty & \text{if } r_{ij} < \sigma \\ 0 & \text{if } r_{ij} \geq \sigma \end{cases} \quad (7.5)$$

## 7 Molecular Dynamics simulations

or the soft-sphere potential

$$v^{\text{SS}}(r_{ij}) = \epsilon(\sigma/r_{ij})^\nu, \quad (7.6)$$

with 'harder' potentials for higher  $\nu$ .

For the long-range electrostatic attraction or repulsion of ions, Coulomb interactions between the charges is added to the Lennard-Jones potential,

$$v^{\text{Coul}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (7.7)$$

where  $q_i$  and  $q_j$  are the atomic charges, and  $\epsilon_0$  is the electrical permittivity of space.

### 7.1.2 Bonded interactions

When it comes to molecules, interactions within each molecule through chemical bonds need to be considered. The simplest way to describe a chemical bond is a harmonic potential. The two parameters, the spring constant and the equilibrium length, can be derived from spectroscopy and X-ray structures, respectively. This applies analogously to angles within a set of three atoms connected by bonds.

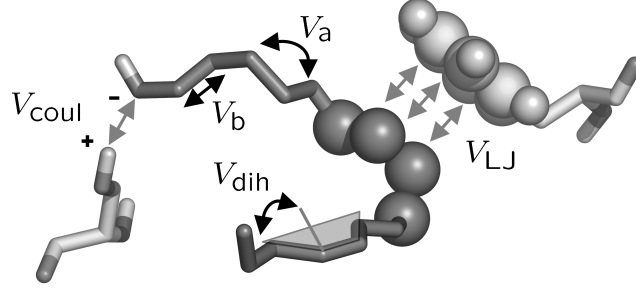
Torsional rotations around molecular bonds are very critical as they involve barriers in the range of  $k_B T$ , with  $k_B$  the Boltzmann constant, and  $T$  the (laboratory or body) temperature, and thus are heavily sampled. They can be described among others by a functional form put forward by Ryckaert and Bellemans ?

$$v_{\text{dih}}(\phi_{ijkl}) = \sum_{n=0}^5 C_n (\cos(\psi))^n, \quad (7.8)$$

where  $\psi = \phi - 180^\circ$ , and  $i, j, k$  and  $l$  are the atoms involved in the dihedral, and  $C_n$  are constants in  $\text{kJ mol}^{-1}$ .

### 7.1.3 Example for a molecular force field

Fig. 7.1 illustrates the components of a force field comprising bonded interactions, namely energy terms for bonds,  $v_b$ , angles,  $v_a$ , proper and improper dihedrals,  $v_{\text{dih}}$ , and additional non-bonded interactions, specifically non-polar Lennard-Jones interactions,  $v_{\text{LJ}}$ , and electrostatic interactions between partial charges of the atoms,



**Figure 7.1:** Scheme to illustrate the components of a typical mechanical force field as used for the simulations of macromolecules such as proteins. Black arrows: bonded interactions:  $v_b$ , bond-stretching potential,  $v_a$ , angle-bending potential,  $v_{dih}$ , dihedral (out-of-plane) potential. Grey arrows: non-bonded interactions:  $v_{Coul}$ , Coulomb potential,  $v_{LJ}$ , Lennard-Jones potential.

$v_{Coul}$ . It is given by

$$\begin{aligned}
 v(R) &= v_b + v_a + v_{dih} + v_{Coul} + v_{LJ} \\
 &= \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,0})^2 \\
 &+ \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \\
 &+ \sum_{\text{dihedrals}} \sum_{n=0}^5 C_n (\cos(\psi))^n \\
 &+ \sum_{\text{atoms } i} \sum_{\text{atoms } j > i} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}.
 \end{aligned} \tag{7.9}$$

The force field parameters for bonded interactions comprise the equilibrium bond length  $l_{i,0}$  and angle  $\theta_{i,0}$ , the respective force constants  $k_i$ , and the Ryckaert-Bellemans constants  $C_n$  for the dihedrals. Non-bonded interactions are parametrized in terms of partial charges  $q_i$  for Coulombic interactions, and the parameters  $\epsilon_{ij}$  and  $\sigma_{ij}$ , defining the depth and width of the Lennard-Jones potential, respectively. Parameters are derived by fitting to experimental thermodynamic or structural quantities, and/or to higher (quantum) level calculations.

## 7.2 Statistical mechanics aspects

Usually, MD simulations are first evolved to reach a certain dynamical equilibrium state in which any memory of the initial conditions has been completely erased. One is then often interested in time averages  $\bar{A}$  of some macroscopic quantity  $A(\Gamma)$ , which itself can be calculated in terms of the microstate  $\Gamma$  of the system. Formally,

## 7 Molecular Dynamics simulations

we are interested in the quantity

$$\bar{A} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t_0}^{t_0 + \tau} A(\Gamma(t)) dt. \quad (7.10)$$

In a typical MD simulation, this average is simply carried out as an average over a finite number of time steps:

$$\bar{A} \simeq \frac{1}{N_{\text{steps}}} \sum_{k=1}^{N_{\text{steps}}} A(\Gamma_k) \quad (7.11)$$

Alternatively, one might be interested in simulating the system (far) away from equilibrium. In such non-equilibrium simulations, a perturbation is added to the Hamiltonian,

$$H^{\text{NE}} = H + A(\mathbf{r}, \mathbf{p}) \cdot F(t), \quad (7.12)$$

with  $F(t)$  as the time-dependent applied field, and  $A(\mathbf{r}, \mathbf{p})$  as a function of the positions and momenta of all particles. A simple example is the generation of shear flow to measure the shear viscosity of a fluid. In this case, at each MD step, an external force is applied along the x-coordinate to each atom, the magnitude of which depends on the y-position of the atom.

The nature of the thermodynamic ensemble decides on some of the macroscopic quantities that can be calculated. The ensemble in turn is defined by the choice of fixed macroscopic parameters ( $NpT$ ,  $NVT$  etc). The next sections describe the techniques to fix the temperature and the pressure, the two quantities also very often fixed in a corresponding laboratory experiment.

### 7.2.1 Temperature and pressure adjustment

The nature of the simulated statistical ensemble decides on the macroscopic thermodynamic quantities that can be calculated from the MD simulation. In turn, the ensemble is fixed by the choice we make on the macroscopic parameters that we keep constant.

In basic MD simulations, we follow the classical equations of motion for a system with given particle number  $N$ , volume  $V$ , and total energy  $E$ . This corresponds to the *microcanonical ensemble* of statistical mechanics and is referred to in the context of MD as “ $NVE$  ensemble”. Consequently, the average results obtained from such simulations are microcanonical ensemble averages. Every microstate (with the given same energy, by definition) is occurs with the same probability.

However, experimentally much more accessible are systems in which the temperature is constant, not the energy. These  $NVT$  ensembles represent the *canonical ensemble*. In the canonical ensemble, each microstate of energy  $E$  has the probability

$$p = \frac{1}{Z} e^{-\frac{E}{k_B T}}, \quad (7.13)$$

of occurring, and the expectation value of a macroscopic variable  $A$  is given by

$$\langle A \rangle = \frac{1}{Z} \int d\Gamma A(\Gamma) \exp \left( -\frac{E(\Gamma)}{k_B T} \right), \quad (7.14)$$

which is the classical Boltzmann-Gibbs average. The partition function

$$Z = \int d\Gamma \exp \left( -\frac{E}{k_B T} \right) \quad (7.15)$$

acts essentially as a normalization factor.

How do we realize a constant temperature of our MD system? This is of central importance as the temperature is often a key thermodynamic control variable. Obviously, we first have to define the temperature of the system. This is achieved by introducing an instantaneous kinetic temperature  $\mathcal{T}$  as

$$\frac{3}{2} k_B \mathcal{T} = \left\langle \frac{m \mathbf{v}_i^2}{2} \right\rangle = \frac{1}{N} \sum_{i=1}^N \frac{m \mathbf{v}_i^2}{2}. \quad (7.16)$$

Then,  $\mathcal{T}$  is averaged over many microstates produced by an MD simulation. The result corresponds to the temperature of the system, i.e.

$$T = \frac{1}{N_{\text{steps}}} \sum_k \mathcal{T}(\Gamma_k) \quad (7.17)$$

Now, how can we control this temperature? If the current (instantaneous) temperature  $\mathcal{T}$  is very different from the desired target temperature  $T$ , we may, for example, re-scale all velocities according to

$$\mathbf{v}'_i = \sqrt{\frac{T}{\mathcal{T}}} \mathbf{v}_i. \quad (7.18)$$

Note that this will not automatically lead to a thermodynamic equilibrium distribution of the velocities, where for each Cartesian component we expect a Gaussian, i.e.

$$p(v_x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{v_x^2}{2\sigma^2} \right) \quad (7.19)$$

with  $\sigma^2 = k_B T / m$ .

To implement a so-called “isokinetic simulation” one re-scales  $\mathcal{T}$  to  $T$  in each time step. Such a simple re-scaling of velocities, however, leads to some residual distortions in the distribution function, which do not exactly reflect thermodynamic equilibrium when this procedure is applied. The equations of motion then differ from Newtonian mechanics.

This can be mitigated by the use of the Nosé-Hoover ‘thermostat’. Here one adds a thermostat variable as an additional degree of freedom, and uses it to drive the

## 7 Molecular Dynamics simulations

velocities to the desired temperature through a suitable friction/antifriction term in the equations of motion. For example, we may use:

$$m_i \dot{\mathbf{v}}_i = \mathbf{F}_i - 2\alpha \mathbf{v}_i, \quad \text{with} \quad (7.20)$$

$$\frac{d\alpha}{dt} = \frac{1}{\tau_E} \left( \sum_i \frac{1}{2} m_i \mathbf{v}_i^2 - \frac{3}{2} N k_B T \right) \quad (7.21)$$

where  $\tau_E$  is a parameter that controls the timescale over which the temperature is regulated and  $\alpha$  is the thermostat variable.

Also important is the  $NpT$ -ensemble, in which the volume of the system is adjusted such that the pressure of the system is kept constant. Again, a direct re-scaling the coordinates does not result in a thermodynamically defined ensemble as it violates Newtonian mechanics. A solution is the Parrinello-Rahman 'barostat', which analogously to the Nose-Hoover thermostat is an extended ensemble method. The box vectors are subject to an equation of motion, and the equation of motion is coupled to this.

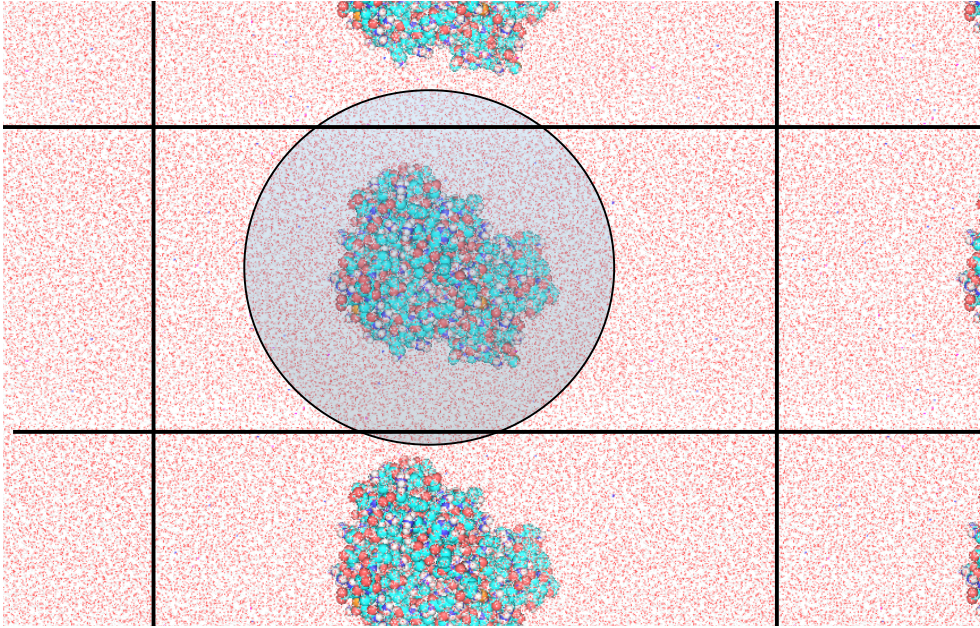
Molecular dynamics simulations can also be used to approximate other types of ensembles. Of particular importance is the *grand canonical ensemble* ( $\mu VT$ -Ensemble) in which the chemical potential  $\mu$  instead of the particle number is held constant.

## 7.3 Practical aspects

### 7.3.1 Boundary conditions

As simulated MD systems only contain particle numbers that are very much smaller than those in any macroscopic sample, surface effects would easily spoil any attempt to approximate the continuum limit if they cannot be efficiently suppressed. For this reason, one usually adopts periodic boundary conditions, because this largely eliminates surface effects in a simple way. The system is periodically replicated in all three Cartesian coordinates. Particles with coordinates beyond the limits of the simulation box will be 'moved' into the simulation system on the other side.





For a fluid of Lennard Jones particles, the box length should be a multiple of  $\sigma$ , such that the interaction of a particle with its periodic image is negligibly small. If the potential is long-ranged, e.g.  $v(r) \sim r^{-1}$ , periodic boundary conditions can induce errors in particular of processes/observables with large long-range contributions. These include phonons in solids, light scattering factors, or phase transitions.

### 7.3.2 Initial conditions

In case of a liquid or gas, the initial conditions might not require particular care as their details should anyway be quickly forgotten in a proper MD simulation as it transits to equilibrium. One then usually sets up the initial particle positions on a simple regular grid. However, particles in solids typically have relaxation time scales much longer than the simulation time scale, and are typically placed according to an experimental X-ray structure. Also, care must be taken to start with reasonable molecular bond lengths within each molecule. Otherwise, forces at the first MD steps are extraordinarily high, eventually causing instabilities.

To speed up the settling to thermodynamic equilibrium, it is helpful to draw random initial velocities from an appropriate Maxwellian distribution at the temperature of choice.

### 7.3.3 Finite range interactions

Formally, calculating all the pairwise forces for  $N$  molecules is a  $\mathcal{O}(N^2)$  problem, which would rather seriously limit the system sizes that can be studied. However, at  $r = 3\sigma$ , the Lennard-Jones potential has dropped already to about  $v(r) \simeq -0.005\epsilon$ , and this low binding energy indicates the weak forces that act at distances this large and beyond. It is hence clear that forces at large distances become negligible, and

## 7 Molecular Dynamics simulations

the full  $N^2$ -interactions do not all have to be calculated. One therefore commonly introduces a cut-off radius into the potential, e.g. in the form of a hard cut-off:

$$v_c(r) = \begin{cases} v(r) & \text{for } r \leq r_c, \\ 0 & \text{otherwise.} \end{cases} \quad (7.22)$$

As an alternative to a hard cut-off at  $r_c$  one may also modify the potential such that it smoothly drop to zero, for example in the form

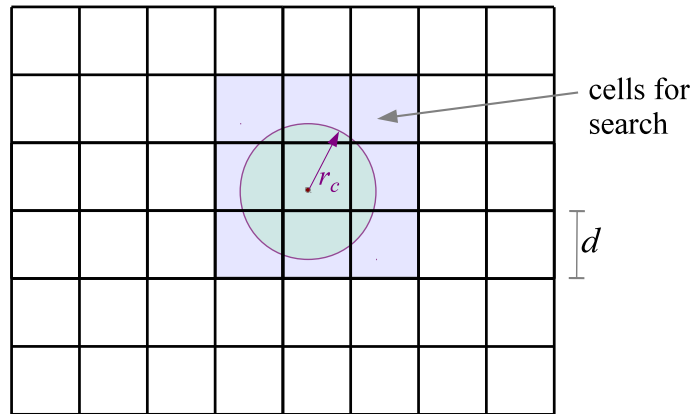
$$v_c(r) = \begin{cases} v(r) - v(r_c) - v'(r_c)(r - r_c) & \text{for } r \leq r_c, \\ 0 & \text{otherwise.} \end{cases} \quad (7.23)$$

Note that the latter form modifies the potential everywhere, which can introduce a small bias in the results.

But independently of how exactly it is carried out, the most important feature of the cut-off is that it reduces the force calculation problem from an unwieldy  $\mathcal{O}(N^2)$  to one of order  $\mathcal{O}(N)$ , provided one has a fast (ideally zeroth-order) method of finding the neighbors that are actually participating in the interactions. Two methods can be used for this:

### Search grids

The simplest approach is to use a Cartesian search grid with cell size  $d \geq r_c$  overlaid over the system. In a first step, one bins all the particles onto this grid, using for example link-lists as a book-keeping device to organize lists of particles contained in every single cell.



Then, in the calculation of the forces for a given particle, one only considers the particles in the same search grid cell as the target particle, as well as the 26 surrounding cells (in three dimensions). Since we have  $d \geq r_c$ , this search will already guarantee that we definitely find all interacting neighbors with distance up to  $r_c$ , independent of where our target point lies with respect to the target cell.

We note that a search grid with  $d = r_c$  still leads to many in principle superfluous look-ups of potential interaction neighbors. This is because one effectively checks all

particles in a volume of  $(3d)^3 = (3r_c)^3$ , while contributing to the interactions is only the spherical volume  $(4\pi/3)r_c^3$  around the target particle. This sphere covers hence only 15.5% of the region that is actually searched. Assuming a roughly uniform distribution of the particles, one then needs to carry out  $\sim 6$  times more distance computations than really used later for the interactions. By using a somewhat finer search grid, one can reduce this overhead somewhat and increase the efficiency of the neighbor search further.

### Range search with a tree

A generalization of the search grid idea is to use a *search tree*. Here one considers a hierarchical grid in which the cell size from level to level differs by a factor of 2. This is simply the oct-tree we considered in the gravity calculation with a ‘tree-code’, except that we do not need the multipole moments here.

The interaction neighbors out to a distance  $r_c$  can then simply be found by performing a special tree walk: A tree node is opened when it has a geometric overlap with the search region (which is a sphere or box of size  $r_c$  around the target position), otherwise the walk across the branch is terminated. An advantage of this method is that it works well also for variable  $r_c$ ; in fact, the search region is allowed to be widely different from particle to particle, but the method always works with nearly constant efficiency. In most MD simulations, the particle density is fairly constant, and  $r_c$  has a global value, too, hence search trees are not needed. However, in smoothed particle hydrodynamics (SPH), this is different. Here one needs, just like in MD, interaction neighbors out to certain distance, but these distances may vary widely in space and time.

### Neighbour lists

For simulations with a fairly homogeneous distribution of particles, neighbour list searches are very efficient, and work equally well for MD and Monte Carlo simulations (see next Chapter). Let us consider a sphere with a cut-off radius  $r_c$  around a particle  $i$ , e.g. within a Lennard Jones fluid. We now define a larger sphere with radius  $r_l$  around the same particle  $i$ . A list is constructed that contains for each particle  $i$  all ‘neighbours’, i.e. all particles within the larger sphere, i.e. within  $r_l$ . Now, at the first MD step, the neighbours list is constructed, and used in a number of subsequent MD steps. After some  $\tau_l$ , the neighbour list is newly constructed. Parameters for this Verlet type of neighbour list are  $r_c$  and  $\tau$ , which are chosen such that particles within the short layer between  $r_c$  and  $r_l$  do only very rarely move out of the sphere with  $r_c$  within the time scale  $\tau$ . The longer  $\tau$  the larger should be the buffer. The speed-up for a homogeneous system with as little as 500 particles can be already in the range of a factor of 2-3. Typical values for molecular systems are  $r_c = 1$  nm,  $r_l = 1.5$  nm, with a  $\tau$  of for example 10 integration steps.

### 7.3.4 Long-range interactions

Lennard-Jones interactions decay with  $r^{-6}$ , i.e. relatively rapidly (see above). Electrostatic interactions, instead, scale with  $1/r$ , and neglecting long-range forces gives rise to serious errors. Thus, beyond a cut-off  $r_c$ , long-range electrostatic forces are commonly still considered, though through computationally cheaper approximative means.

One option is the description by solutions of the Poisson equation. For obtaining the forces in this case, the same techniques as in the gravitational dynamics can be applied. They are discussed in Chapter 5ff.

### 7.3.5 Time integration

A very popular scheme for MD simulations is the so-called Verlet time integrator <sup>?</sup>. This is given by

$$\mathbf{r}_i^{(n+1)} = 2\mathbf{r}_i^{(n)} - \mathbf{r}_i^{(n-1)} + \mathbf{a}_i^{(n)}(\Delta t)^2. \quad (7.24)$$

Note that this can in principle work without storing the velocities explicitly, but one needs two copies of the old positions. If one also wants to know the velocities, these are computed in this scheme from

$$\mathbf{v}_i^{(n)} = \frac{1}{2\Delta t} \left( \mathbf{r}_i^{(n+1)} - \mathbf{r}_i^{(n-1)} \right). \quad (7.25)$$

The Verlet scheme may look a bit strange at first, but actually it is really *identical* to the ordinary Leapfrog. The latter can be written as

$$\mathbf{v}_i^{(n+1/2)} = \mathbf{v}_i^{(n)} + \mathbf{a}_i^{(n)} \frac{\Delta t}{2}, \quad (7.26)$$

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + \mathbf{v}_i^{(n+1/2)} \Delta t, \quad (7.27)$$

$$\mathbf{v}_i^{(n+1)} = \mathbf{v}_i^{(n+1/2)} + \mathbf{a}_i^{(n+1)} \frac{\Delta t}{2}. \quad (7.28)$$

If we now plug in equation (7.26) into equation (7.27), consider the same equation a second time for index  $n$  instead of  $n+1$ , and subtract the two from each other, we obtain the Verlet scheme (7.24). This also implies that the Verlet scheme, just like leapfrog, is a second-order method, time-centered, time-reversible, and symplectic (see Chapter 2). The Euler scheme comes with the same computational costs but smaller numerical stability for Newton's equations of motion. When using Euler but not Verlet, velocities diverge quickly for a set of non-interacting particles in a box with repulsive walls, even if the time steps is chosen smaller than the residence time of particles within the repulsive potential.

Time steps of MD simulations are chosen such that they are smaller than the fastest motions of the simulated molecular system. Vibrational frequencies

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (7.29)$$

where  $k$  is the spring constant and  $\mu$  the reduced mass of a bond, are highest for bonds involving the lightest atoms, that is, hydrogen. Accordingly, MD simulations with chemical bonds to hydrogen, as is the case for water and virtually all organic matter, can only be stably run with a time step of 1 fs or lower. For many applications, however, the intramolecular dynamics of atoms relative to one another is not of interest (and anyways classically not well described). A pragmatic solution is fixing the high frequency interactions such as hydrogen bonds, allowing time steps of  $\sim 2$  fs. To further reduce the time step, other degrees of freedom can be neglected. A first obvious candidate is the rotation around bonds with three identical atoms, such as methyl ( $\text{CH}_3$ ) groups, or the modelling of more than one atom as one particle, which is called coarse-graining and very problem-dependent.

## 7.4 Integrating other equations of motion: Langevin and Brownian Dynamics

A stochastic alternative to Newton's equations of motion are Langevin dynamics. In Langevin dynamics, friction and random forces are added to the systematic forces due to the potential energy between particles. One can consider it as a heat bath (analogous to the temperature coupling schemes discussed above), but now here with *random* collisions with the bath.

$$m\ddot{\mathbf{x}}(t) = \nabla V(\mathbf{x}(t)) - \gamma m\dot{\mathbf{x}}(t) + \mathbf{R}(t) \quad (7.30)$$

with  $\gamma$  as the damping constant or collision parameter, and  $\mathbf{R}(t)$  a random force vector, which averages to zero over time and follows a stationary Gaussian process. The higher  $\gamma$ , the less the dynamics is governed by the inertia and the more diffusive (Brownian) it is. The stochastic forces typically mimic collisions with solvent molecules (e.g. water around a biomolecule), which are not explicitly considered as particles.

A generalized Verlet scheme can be used to integrate Langevin dynamics.

$$\mathbf{v}_i^{(n+1/2)} = \mathbf{v}_i^{(n)} + m^{-1} \frac{\Delta t}{2} [-\nabla V(\mathbf{r}_i^{(n)}) - \gamma m \mathbf{v}_i^{(n)} + \mathbf{R}^{(n)}], \quad (7.31)$$

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + \mathbf{v}_i^{(n+1/2)} \Delta t, \quad (7.32)$$

$$\mathbf{v}_i^{(n+1)} = \mathbf{v}_i^{(n+1/2)} + m^{-1} \frac{\Delta t}{2} [-\nabla V(\mathbf{r}_i^{(n+1)}) - \gamma m \mathbf{v}_i^{(n+1)} + \mathbf{R}^{(n+1)}]. \quad (7.33)$$

It reduces to velocity-Verlet (see above) when  $\gamma$  is zero. It is, however, only applicable in the regime of small  $\gamma$ . The superscript of  $\mathbf{R}$  simply implies that it is chosen randomly at each integration step.

At high friction, i.e the overdamped regime, inertia can be ignored,  $m\ddot{\mathbf{x}}(t) = 0$ , resulting in Brownian dynamics of the form

$$\dot{\mathbf{x}}(t) = \frac{\mathbf{D}}{k_B T} \nabla V(\mathbf{x}(t)) + \mathbf{R}(t) \quad (7.34)$$

## 7 Molecular Dynamics simulations

$\mathbf{D}$  is a diffusion tensor, or  $k_B T \mathbf{Z}^{-1}$ , with  $\mathbf{Z}$  as the friction tensor. An integration scheme by Ermak and McCammon is

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + \frac{\Delta t}{k_B T} \mathbf{D}^n \nabla V(\mathbf{r}_i^{(n)}) + \mathbf{R}^{(n)}. \quad (7.35)$$

A few comments with regard to  $\mathbf{R}^{(n)}$ : It averages to zero over time,  $\langle \mathbf{R}^{(n)} \rangle = 0$ . and has a variance linearly proportional to  $\mathbf{D}^n$ ,  $\langle (\mathbf{R}^{(n)})(\mathbf{R}^{(m)})^T \rangle = 2\mathbf{D}^n \Delta t$ . Let's consider a simple example, a particle moving in one dimension in absence of a systematic force, and random numbers  $r_n$  being drawn from a uniform distribution in the interval  $[-0.5, 0.5]$ . Then the integration algorithm simplifies to

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + (24D\Delta t)^{\frac{1}{2}} r_n \quad (7.36)$$

Note that in BD, for that the variance of the random walk scales linearly in time, the random displacements scale with the square-root of the time-step. The time step for integrating BD is typically much larger than in MD or LD. The definition of the diffusion tensor defines the hydrodynamics of the system.