Force fields, thermo- and barostats

Berk Hess

What is a force field?

A force field usually consists of three parts:

- a set of functional forms

- parameters for the functional forms that, usually, depend on the atom types

- a set of building blocks, e.g. for amino acid residues

Gromacs interaction function support

Often one can split a force field into two main contributions:

Bonded interactions work only within molecules:

$$V_b(\boldsymbol{r}_i, \ldots, \boldsymbol{r}_N) = \sum_{m \in molecules} \left(\sum_{i,j \in bonds/pairs(m)} V_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j) + \right.$$

$$\sum_{i,j,k \in angles(m)} V_{ijk}(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k) +$$

$$\left. \sum_{i,j,k,l \in dihedrals(m)} V_{ijkl}(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k, \boldsymbol{r}_l) \right)$$

Non-bonded interactions work between and within molecules:

$$V_{nb}(\boldsymbol{r}_i, \ldots, \boldsymbol{r}_N) = \sum_{i<j} V_{ij}(\| \boldsymbol{r}_i - \boldsymbol{r}_j \|) = \sum_{i<j} V_{ij}(r_{ij})$$

Non-bonded interactions

Non-bonded interactions in GROMACS are pair-additive and centro-symmetric:

Electrostatic interactions

- Coulomb $\quad V_{ij} = \dfrac{q_i q_j}{4\pi\epsilon_0\epsilon}\dfrac{1}{r_{ij}}$

- tabulated $\quad V_{ij} = \dfrac{q_i q_j}{4\pi\epsilon_0\epsilon}\, f_C(r_{ij})$

Van der Waals interactions

- Lennard Jones $\quad V_{ij} = C_{ij}^{(12)}\dfrac{1}{r_{ij}^{12}} - C_{ij}^{(6)}\dfrac{1}{r_{ij}^{6}}$

- tabulated $\quad V_{ij} = C_{ij}^{(12)} f_R(r_{ij}) - C_{ij}^{(6)} f_D(r_{ij})$

- Buckingham $\quad V_{ij} = A_{ij}\exp(-B_{ij}r_{ij}) - \dfrac{C_{ij}}{r_{ij}^{6}}$

Bonded interactions

Bonds (two-body):

- harmonic bond stretching: $V_b\left(r_{ij}\right) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})^2$

- 4th power bond bond stretching: $V_b\left(r_{ij}\right) = \frac{1}{4}k_{ij}^b\left(r_{ij}^2 - b_{ij}^2\right)^2$

- ...

Angles (three-body):

- harmonic angle potential: $V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta(\theta_{ijk} - \theta_{ijk}^0)^2$

- cosine based potential: $V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta\left(\cos(\theta_{ijk}) - \cos(\theta_{ijk}^0)\right)^2$

- ...

Constraints

In GROMACS bonds are usually replaced by constraints,
i.e. the distances between bonded atoms are fixed to their
equilibrium values

Constraints are actually a more faithful representation of reality,
since the dynamics of bonds is of quantum nature

Removing the bond vibrations with constraints allows one to increase

the timestep from 0.5 or 1 fs to 2 or 4 fs

Bonded interactions (continued)

Improper dihedrals (four-body), for planarity or stereochemistry:

- harmonic

Improper dihedrals (four-body):

- single cosine: $V_d(\phi_{ijkl}) = k_\phi(1 + \cos(n\phi - \phi_s))$

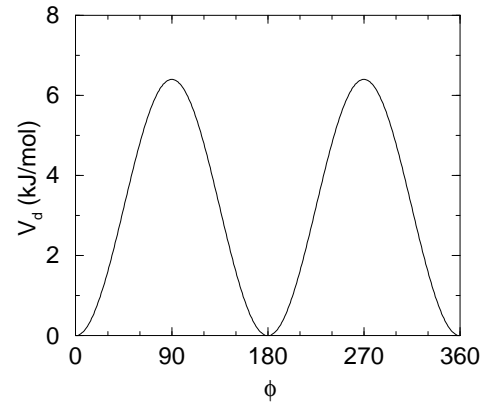- Ryckaert-Bellemans: $V_{rb}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n(\cos(\psi))^n$
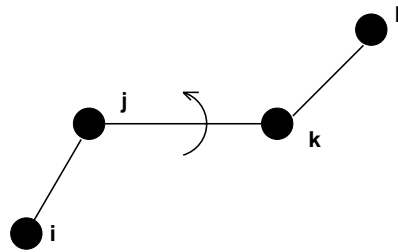
- ...

1-4 interactions (two-body):

- (scaled) Lennard-Jones interactions

- (scaled) Coulomb interactions

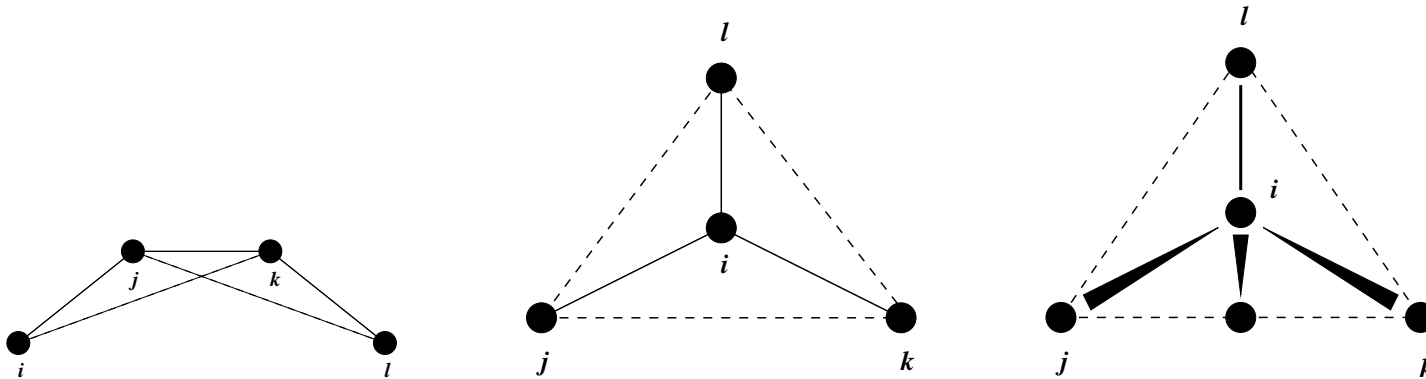1-4 interactions go together with dihedrals, more on that later

Dihedrals

Proper dihedral

Also 1-4 interactions between atoms i and j contribute to the torsional potential around ϕ!

Improper dihedral

The choice of force field

The choice of force field is one of the most important factors in an MD study

A good force field requires years of work by many people
All parameters for all molecules should be consistent!

The first question is which of the force field have parameters for the molecules present in your system

When some parameters are missing:

- search the literature
- if no success, you need to do force field parametrization :-(

Biomolecular force fields in GROMACS

GROMOS (6 versions)	OPLS-AA	AMBER (7 versions)
united atom aliphatic CH_n	all-atom	all-atom
proteins hydrocarbons	proteins hydrocarbons many liquids	proteins DNA/RNA

An AMBER port for Gromacs is available at:
http://folding.stanford.edu/ffamber/

Force field parametrization
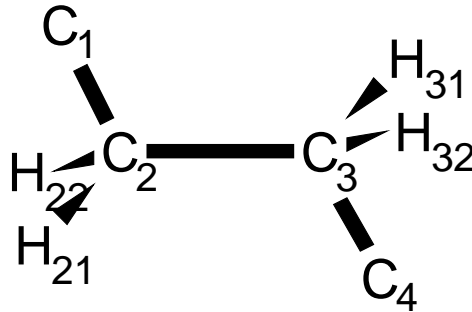
Charges and Lennard-Jones parameters:

Very tricky!
Try to do it the same way as the force field has been parametrized

Bonded terms:

Relatively easy, from quantum calculations But beware of the 1-4

terms for dihedrals!

Torsional potential setup

dihedrals:

- 9: C_1-C_2-C_3-C_4, H_{21}-C_2-C_3-C_4, H_{22}-C_2-C_3-C_4, ...
- 1 (GROMOS96): only C_1-C_2-C_3-C_4

1-4 LJ/Coulomb pair interactions:

- 9: C_1-C_2-C_3-C_4, H_{21}-C_2-C_3-C_4, H_{22}-C_2-C_3-C_4, ...

The original idea behind this is:
the dihedral potentials give the 2-3 dependencies only

the 1-4 pair potentials give the 1-4 dependencies only

Torsional potential parametrization

What to do when parameters for a certain dihedral are not present in the force field?

1) Scan the dihedral in steps of e.g. 15 degrees and optimize the structures using quantum calculations (e.g. Gaussian)

2) Determine the force-field energies for the optimized quantum structures using `mdrun -rerun`,
including all interations, except for the dihedral to parametrize

3) Determine the difference between the quantum and force-field potential energies
and fit one of the dihedral functional forms to this

GROMACS topology file setup

Your topology is contained in the `.top` file

The `.top` file "includes" the force field `.itp` files, which are stored in the directory `$GMXDIR/shared/top`

For proteins your molecular topology, i.e. the set of bonded interactions for your molecule, can be generated from the force field builing blocks with the program `pdb2gmx`

Topology file

```
; Include forcefield parameters
#include "ffG53a6.itp"


[ moleculetype ]
; Name               nrexcl
Methanol               3


... molecule definition stuff, next slide ...


; Include water topology
#include "spc.itp"


[ system ]
; Name
Methanol in water


[ molecules ]
; Compound        #mols
Methanol               1
SOL                  640
```

Molecule definition

```
[ moleculetype ]
; Name                nrexcl
Methanol                 3


[ atoms ]
;   nr        type   resnr residue   atom   cgnr       charge        mass
     1         CH3       1     SER      CB      1        0.266      15.035
     2          OA       1     SER      OG      1       -0.674     15.9994
     3           H       1     SER      HG      1        0.408       1.008


[ bonds ]
;  ai     aj funct         c0         c1
    1      2      2      gb_18      gb_18
    2      3      2       gb_1       gb_1


[ angles ]
;  ai     aj     ak funct         c0         c1
    1      2      3      2      ga_12      ga_12
```

Force field files

ffG53a6.itp:

```
#define _FF_GROMOS96
#define _FF_GROMOS53A6


[ defaults ]
; nbfunc          comb-rule       gen-pairs       fudgeLJ fudgeQQ
  1               1               no              1.0     1.0


#include "ffG53a6nb.itp"
#include "ffG53a6bon.itp"
```

ffoplsaa.itp:

```
#define _FF_OPLS
#define _FF_OPLSAA


[ defaults ]
; nbfunc          comb-rule       gen-pairs       fudgeLJ fudgeQQ
1                 3               yes             0.5     0.5


#include "ffoplsaanb.itp"
#include "ffoplsaabon.itp"
```

OPLS-AA files

ffoplsaanb.itp:

```
[ atomtypes ]
; full atom descriptions are available in ffoplsaa.atp
; name    bond_type    mass      charge   ptype      sigma        epsilon
 opls_001   C   6    12.01100    0.500     A     3.75000e-01   4.39320e-01 ; S
 opls_002   O   8    15.99940   -0.500     A     2.96000e-01   8.78640e-01 ; S
 opls_003   N   7    14.00670   -0.570     A     3.25000e-01   7.11280e-01 ; S
...
```

ffoplsaabon.itp:

```
[ bondtypes ]
; i      j   func         b0              kb
  OW     HW      1     0.09572     502080.0    ; For TIP4F Water - wlj 1/98
  OW     LP      1     0.01750     753120.0    ;            -idem-
  C*     HC      1     0.10800     284512.0    ;
...


[ angletypes ]
; i      j      k   func       th0          cth
  HW     OW     HW     1    109.500      627.600    ; For TIP4F Water - wl
  HW     OW     LP     1     54.750      418.400    ; For TIP4F Water - wl
```

GROMOS96 files

ffG53a6nb.itp:

```
[ atomtypes ]
;name    at.num    mass         charge  ptype        c6              c12
    O     8   0.000        0.000      A   0.0022619536        1e-06
   OM     8   0.000        0.000      A   0.0022619536   7.4149321e-07
   OA     8       0.000        0.000      A   0.0022619536   1.505529e-06
...


[ nonbond_params ]
; i      j func          c6              c12
     OM          O  1   0.0022619536    8.611e-07
     OA          O  1   0.0022619536    1.38651e-06
     OA         OM  1   0.0022619536    2.258907e-06
...


[ pairtypes ]
; i      j func          c6              c12
      O          O  1   0.0022619536   7.4149321e-07
     OM          O  1   0.0022619536   7.4149321e-07
...
```

GROMOS96 files

ffG53a6bon.itp:

```
; ICB(H)[N]    CB[N] B0[N]
;
#define gb_1        0.1000  1.5700e+07
; H  -  OA      750
;
#define gb_2        0.1000  1.8700e+07
; H  -  N (all) 895
...


;  ICT(H)[N]  CT[N]  (T0[N])
;
#define ga_1        90.00      380.00
; NR(heme)  -  FE  -  C          90
;
#define ga_2        90.00      420.00
; NR(heme)  -  FE  -  NR(heme)  100
...
```

Making topologies

Proteins:

- use `pdb2gmx`

Small molecules:

- write a topology by hand, but be very careful about the details of the force field used

(hetero-)polymers:

- write building blocks (an .rtp file)
- then use `pdb2gmx`

Thermostats and barostats

Ensembles

"Plain" molecular dynamics is Hamiltonian dynamics and this conserves energy, leading to an constant NVE or microcanonical ensemble.

However, in most cases a constant NVT (canonical) or NPT ensemble will be a more suitable representation of the real system.

To achieve this, the equation of motions need to be modified to include a thermostat and/or barostat.

There are two types of thermostats:

- global thermostats
- local thermostats

The Berendsen thermostat

A global thermostat, weak coupling to a bath:

$$\frac{\mathsf{d}T}{\mathsf{d}t} = \frac{T_0 - T}{\tau}$$

To achieve this first order decay of temperature deviations, all velocities are scaled with the factor:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left\{ \frac{T_0}{T(t - \frac{\Delta t}{2})} - 1 \right\} \right]^{1/2}$$

τ_T is set with the parameter `tau_t` in the mdp file

Different T_0 and τ_T can be used for subgroups of the system.

The Nosé-Hoover thermostat

A global thermostat, using the extended-ensemble approach.

An additional degree of freedom ξ is added to the Hamiltonian:

$$\frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i} - \xi \frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} \quad , \qquad \frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{1}{Q}\left(T - T_0\right)$$

The dynamics for the extended Hamiltonian is canonical, but leads to an NVT ensemble for the original system.

In the mdp file not the mass Q is set, but the period of the oscillations.

Global thermostats

Berendsen thermostat
advantage: little influence on kinetics
disadvantage: does not generate an ensemble (but error is small)

Nosé-Hoover thermostat
advantage: produces an NVT ensemble
disadvantage: can induce oscillations in the dynamics

Temperature jump
from 300 to 330 K
for 1000 H_2O

Legend:
— Berendsen, τ=0.2 ps
- - - exp($-t/\tau$), τ=0.5 ps
— Nose–Hoover, period=1 ps

Axes: T (K) vs Time (ps)
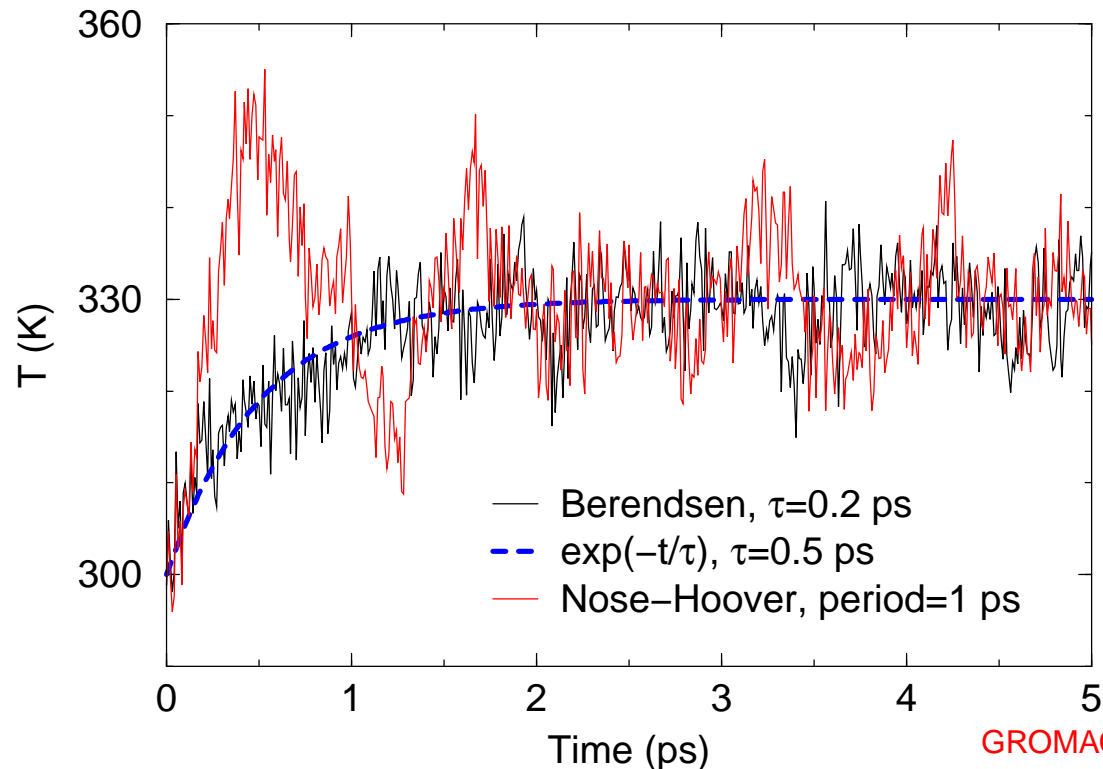
Ergodicity

Global thermostats only work well when all degrees of freedom in the system are coupled strong enough to one another.

Consider the following system:
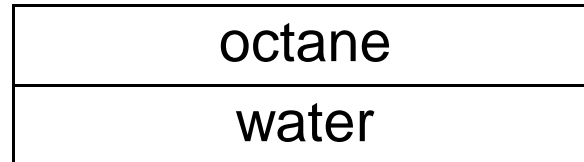
octane
water

Here the octane and water slabs are very weakly coupled.
Using one global thermostat can lead to a temperature difference between the octane and water slabs.

A solution is: using separate thermostats for octane and water
But what about the 3 degrees of freedom coupling the octane and water slabs?
GROMACS removes the center of mass motion for octane and water.

Local thermostats

One can also use Stochastic Dynamics (SD):

$$m_i \frac{\mathsf{d}^2 \boldsymbol{r}_i}{\mathsf{d}t^2} = -m_i \xi \frac{\mathsf{d}\boldsymbol{r}_i}{\mathsf{d}t} + \boldsymbol{F}_i(\boldsymbol{r}) + \overset{\circ}{\boldsymbol{\eta}}_i$$

advantages: NVT ensemble, always ergodic, no oscillations
disadvantage: all velocity correlations decay with $\exp(\xi\, t)$

In the overdamped limit SD reduces to Brownian Dynamics:

$$\frac{\mathsf{d}\boldsymbol{r}_i}{\mathsf{d}t} = \frac{1}{\gamma_i} \boldsymbol{F}_i(\boldsymbol{r}) + \overset{\circ}{\boldsymbol{\eta}}_i$$

BD is not useful for dynamics,
but can be very useful for equilibrating bad starting structures.

Barostats

Berendsen barostat, or weak coupling:

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_p}$$

The box/coordinate scaling matrix $\boldsymbol{\mu}$ is given by

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\,\tau_p}\beta_{ij}\{P_{0ij} - P_{ij}(t)\}$$

Parrinello-Rahman barostat:

$$\frac{d\boldsymbol{b}^2}{dt^2} = V\boldsymbol{W}^{-1}\boldsymbol{b}'^{-1}\left(\boldsymbol{P} - \boldsymbol{P}_0\right)$$

$$\frac{d^2\boldsymbol{r}_i}{dt^2} = \frac{\boldsymbol{F}_i}{m_i} - \boldsymbol{M}\frac{d\boldsymbol{r}_i}{dt} \quad , \qquad \boldsymbol{M} = \boldsymbol{b}^{-1}\left[\boldsymbol{b}\frac{d\boldsymbol{b}'}{dt} + \frac{d\boldsymbol{b}}{dt}\boldsymbol{b}'\right]\boldsymbol{b}'^{-1}$$

Calculating non-bonded interactions

Neighbor searching

In principle for a periodic system of N atoms we would need to determine $N(N+1)/2$ pair interactions, plus the interactions will all periodic images.

When interactions are shorted ranged we can use cut-off's.

For efficiency we used a fixed neighbor list for n_s steps:

- construct neighbor list for all charge groups pairs within r_c

- n_s times:
    - evaluate V and F for all pairs in the list
    - integrate the equations of motion

- construct neighbor list for all charge groups pairs within r_c

- ...

Charge groups

We use charge groups because of two reasons:

- Neighbor searching efficiency
  Example: For a pair of 3-site water molecules we need 9 atom pair distances, but only 1 charge group distance

- Avoiding electrostatic artifacts due to cut-off's
  Example: For water without charge groups partial charges enter the cut-off, with charge groups dipoles enter the cut-off
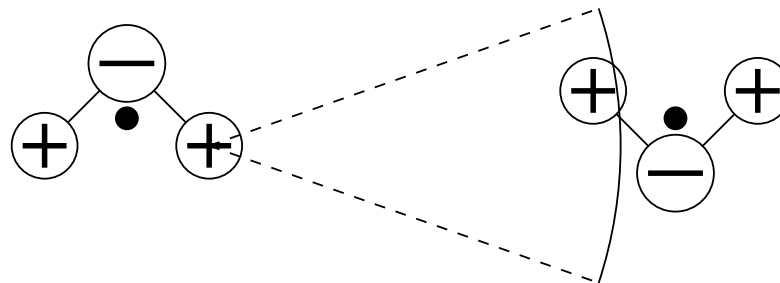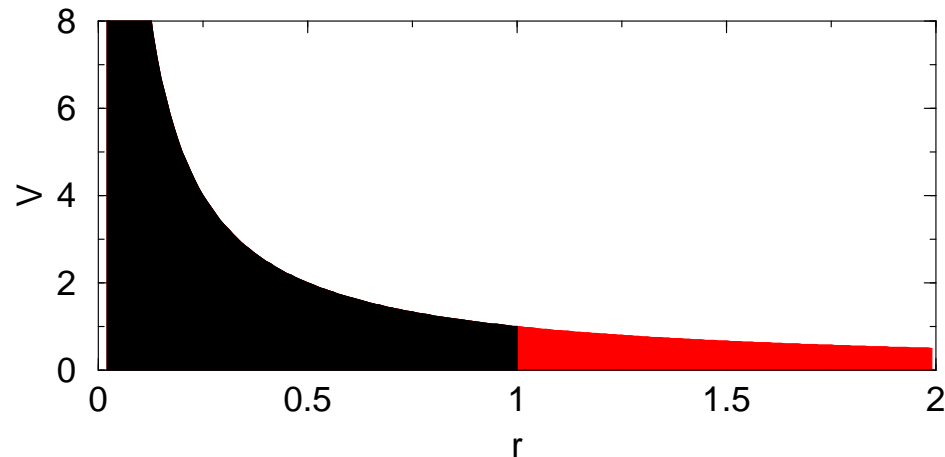
Cut-off, charge groups and diffusion
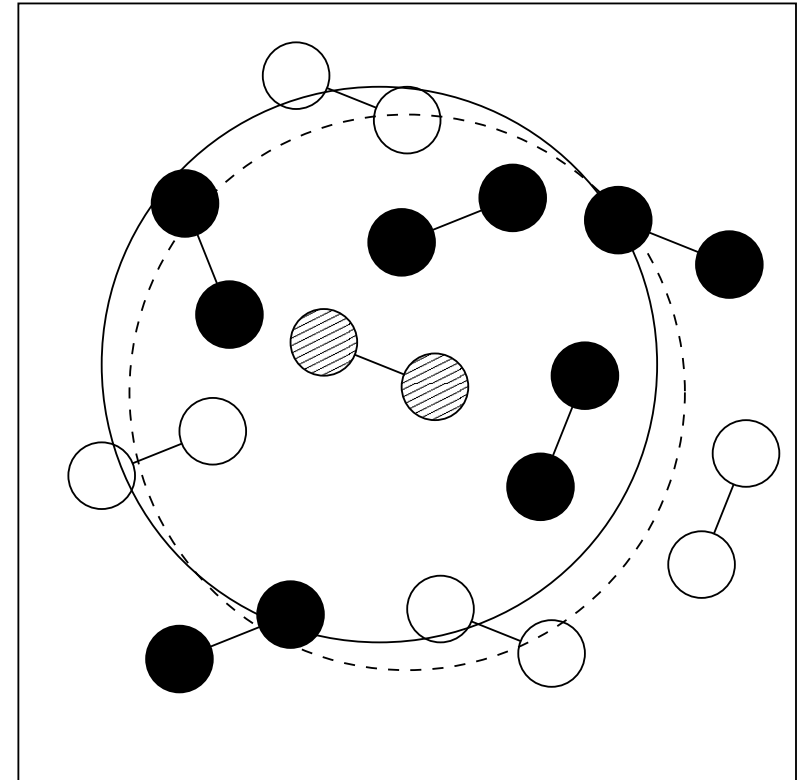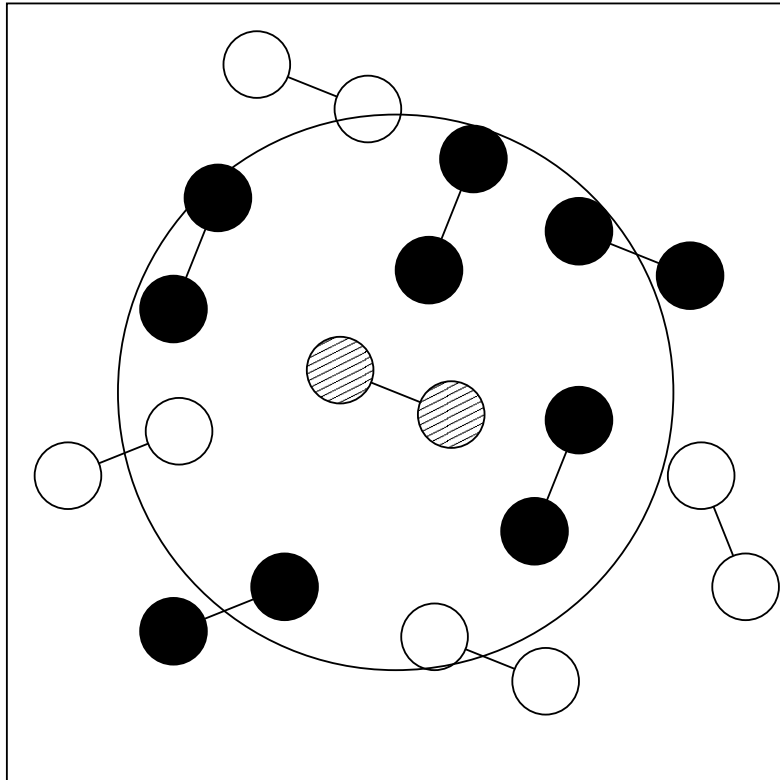
Cut-off, charge groups and diffusion
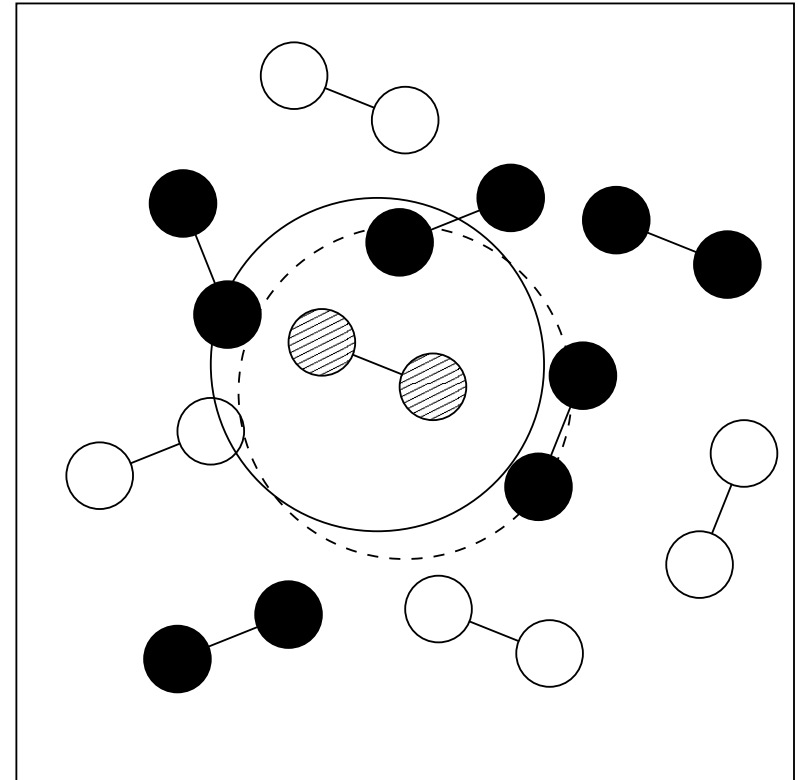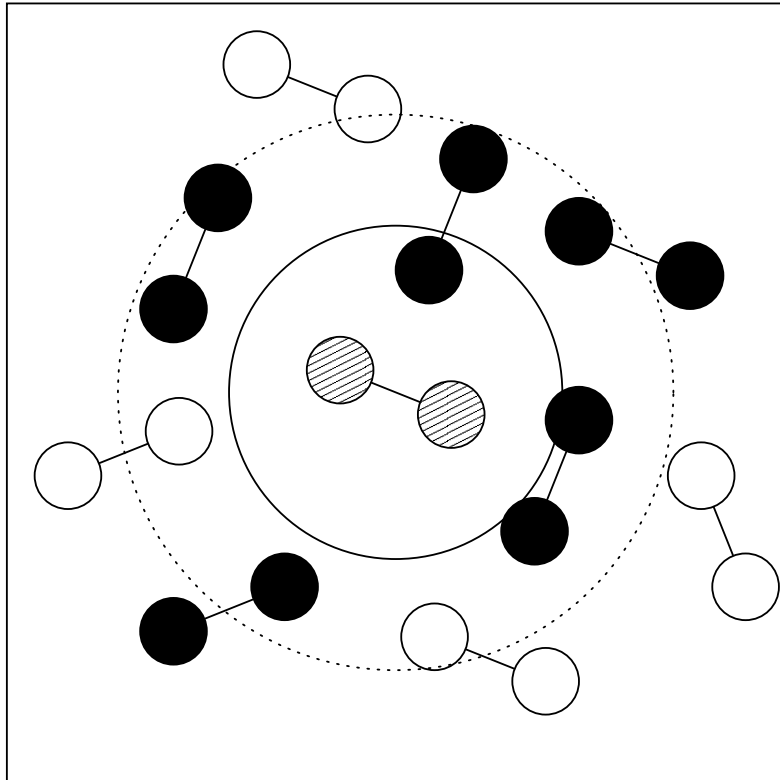
Electrostatics with cut-off

Cutting off interactions abruptly causes artefacts:

- In water: over-orientation of dipoles: anti-parallel before the cut-off, parallel beyond the cut-off

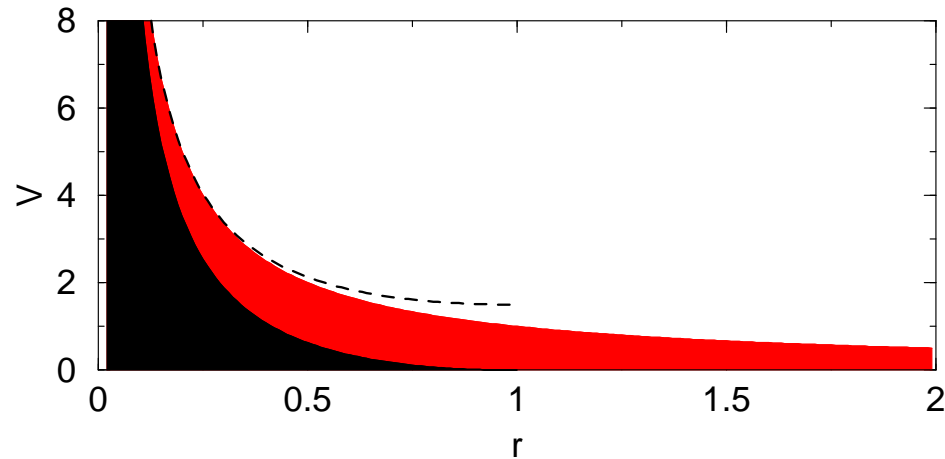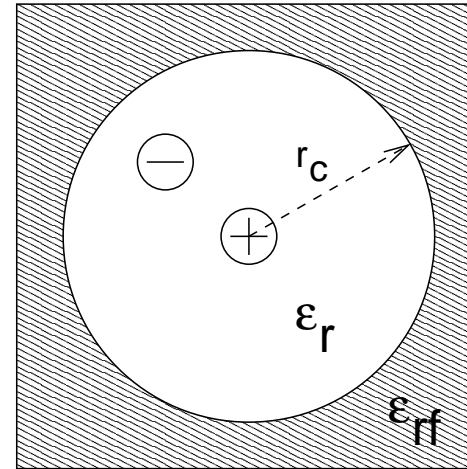- Like charged ions will accumulate just beyond the cut-off

Some artefacts get worse with increasing cut-off distance!!!

This is because Coulomb goes as $1/r$, but the surface of the cut-off

sphere goes as r^2

Reaction-field electrostatics

$$\varepsilon_{rf} = \infty \quad \Rightarrow$$

$$V = f \frac{q_i q_j}{\varepsilon_r} \left[\frac{1}{r_{ij}} + \frac{1}{2} \frac{r_{ij}^2}{r_c^3} - \frac{3}{2} \frac{1}{r_c} \right]$$
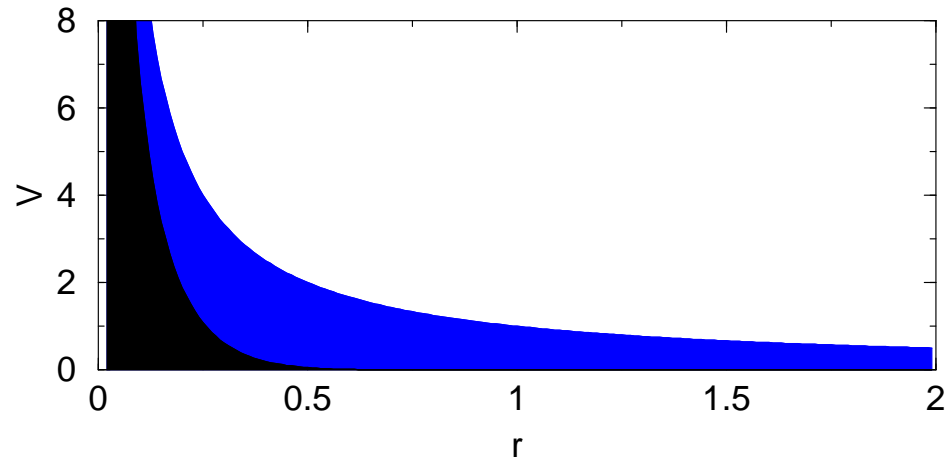

Advantage: much less cut-off artefacts

Disadvantage: $V = 0$ beyond r_c

Particle Mesh Ewald (PME) electrostatics

Split the electrostatics into a short- and long-range part
Do the short-range part with particle-particle interactions
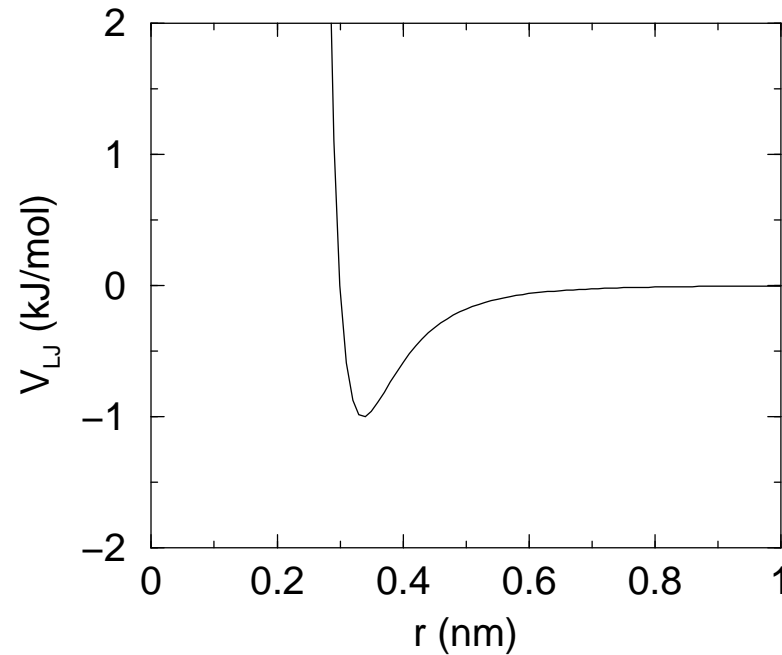Solve the long-range part on a grid using FFT's

Advantages: no cut-off artefacts, infinite interaction range
Disadvantages: computationally more expensive, periodicity (but
this is due to the boundary conditions)

Can be used with systems with net charge:
a uniform background charge is applied
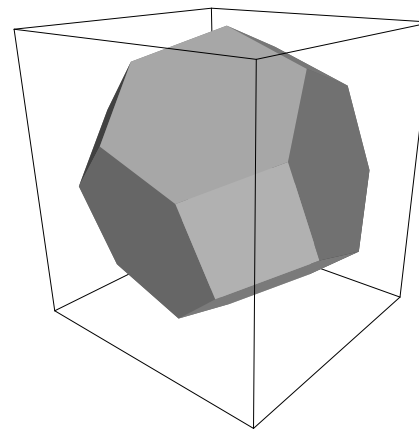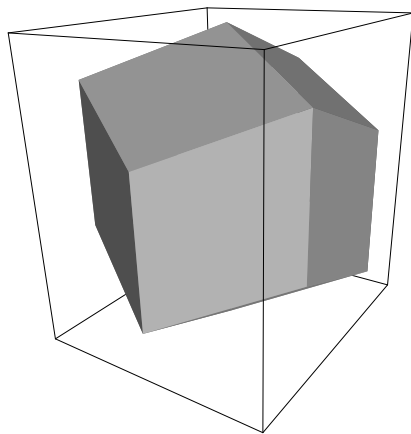
Lennard-Jones potential

LJ decays fast: $-r^{-6}$

but the dispersion contributes up to relatively long distances:

$$V_{disp} = \int_{r_c}^{\infty} \rho_n \langle C^{(6)} \rangle \, 4\pi \, r^2 \, r^{-6} \mathrm{d}r = -\rho_n \langle C^{(6)} \rangle \frac{4\pi}{3} \, r_c^{-3}$$

Some useful box shapes

box type	image distance	box volume	box vectors			box vector angles		
			a	**b**	**c**	$\angle\mathbf{bc}$	$\angle\mathbf{ac}$	$\angle\mathbf{ab}$
cubic	d	d^3	d 0 0	0 d 0	0 0 d	$90°$	$90°$	$90°$
rhombic dodecahedron	d	$\frac{1}{2}\sqrt{2}\,d^3$ $0.707\,d^3$	d 0 0	0 d 0	$\frac{1}{2}d$ $\frac{1}{2}d$ $\frac{1}{2}\sqrt{2}\,d$	$60°$	$60°$	$90°$
truncated octahedron	d	$\frac{4}{9}\sqrt{3}\,d^3$ $0.770\,d^3$	d 0 0	$\frac{1}{3}d$ $\frac{2}{3}\sqrt{2}\,d$ 0	$-\frac{1}{3}d$ $\frac{1}{3}\sqrt{2}\,d$ $\frac{1}{3}\sqrt{6}\,d$	$71.53°$	$109.47°$	$71.53°$

Neighbor search with a grid
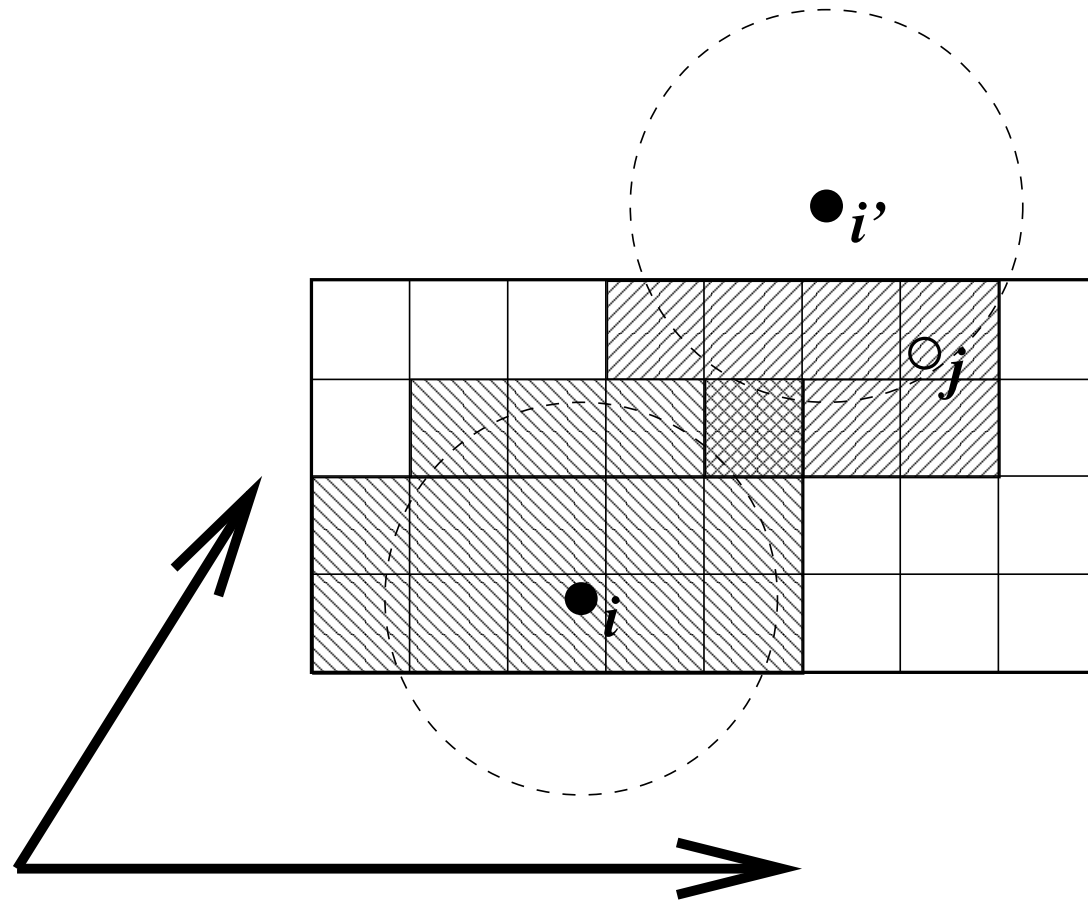
grid search in two dimensions, the arrows are the box vectors

Neighbor search/interaction setting

Neighbor search is C-code ⇒ slow
Force loops are SSE/Altivec/... assembly ⇒ fast

For efficiency you want to use nstlist = 10

To avoid electrostatic articfacts use PME

force field	neighb. list	Elec. cut-off	PME grid	VdW type	VdW cut-off	disp. corr.
GROMOS united atom	1.0	1.0	0.135	shift cut-off	0.9 1.4	yes no
OPLS-AA all atom	0.9	0.9	0.125	shift cut-off	0.8 1.4	yes no