

Chapter 3

Algorithms

3.1 Introduction

In this chapter we first give describe some general concepts used in GROMACS: *periodic boundary conditions* (sec. 3.2) and the *group concept* (sec. 3.3). The MD algorithm is described in sec. 3.4: first a global form of the algorithm is given, which is refined in subsequent subsections. The (simple) EM (Energy Minimization) algorithm is described in sec. 3.10. Some other algorithms for special purpose dynamics are described after this.

A few issues are of general interest. In all cases the *system* must be defined, consisting of molecules. Molecules again consist of particles with defined interaction functions. The detailed description of the *topology* of the molecules and of the *force field* and the calculation of forces is given in chapter 4. In the present chapter we describe other aspects of the algorithm, such as pair list generation, update of velocities and positions, coupling to external temperature and pressure, conservation of constraints. The *analysis* of the data generated by an MD simulation is treated in chapter 8.

3.2 Periodic boundary conditions

The classical way to minimize edge effects in a finite system is to apply *periodic boundary conditions*. The atoms of the system to be simulated are put into a space-filling box, which is surrounded by translated copies of itself (Fig. 3.1). Thus there are no boundaries of the system; the artifact caused by unwanted boundaries in an isolated cluster is now replaced by the artifact of periodic conditions. If the system is crystalline, such boundary conditions are desired (although motions are naturally restricted to periodic motions with wavelengths fitting into the box). If one wishes to simulate non-periodic systems, such as liquids or solutions, the periodicity by itself causes errors. The errors can be evaluated by comparing various system sizes; they are expected to be less severe than the errors resulting from an unnatural boundary with vacuum.

There are several possible shapes for space-filling unit cells. Some, like the *rhombic dodecahedron* and the *truncated octahedron* [18] are closer to being a sphere than a cube is, and are therefore

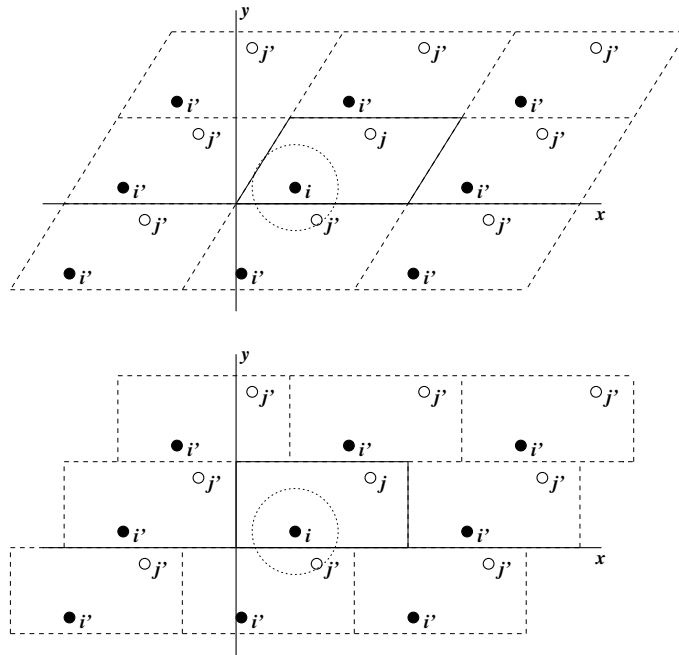


Figure 3.1: Periodic boundary conditions in two dimensions.

better suited to the study of an approximately spherical macromolecule in solution, since fewer solvent molecules are required to fill the box given a minimum distance between macromolecular images. At the same time, rhombic dodecahedra and truncated octahedra are special cases of *triclinic* unit cells; the most general space-filling unit cells that comprise all possible space-filling shapes [19]. For this reason, GROMACS is based on the triclinic unit cell.

GROMACS uses periodic boundary conditions, combined with the *minimum image convention*: only one – the nearest – image of each particle is considered for short-range non-bonded interaction terms. For long-range electrostatic interactions this is not always accurate enough, and GROMACS therefore also incorporates lattice sum methods such as Ewald Sum, PME and PPPM.

GROMACS supports triclinic boxes of any shape. The simulation box (unit cell) is defined by the 3 box vectors \mathbf{a} , \mathbf{b} and \mathbf{c} . The box vectors must satisfy the following conditions:

$$a_y = a_z = b_z = 0 \quad (3.1)$$

$$a_x > 0, \quad b_y > 0, \quad c_z > 0 \quad (3.2)$$

$$|b_x| \leq \frac{1}{2} a_x, \quad |c_x| \leq \frac{1}{2} a_x, \quad |c_y| \leq \frac{1}{2} b_y \quad (3.3)$$

Equations 3.1 can always be satisfied by rotating the box. Inequalities (3.2) and (3.3) can always be satisfied by adding and subtracting box vectors.

Even when simulating using a triclinic box, GROMACS always keeps the particles in a brick-shaped volume for efficiency, as illustrated in Fig. 3.1 for a 2-dimensional system. Therefore, from the output trajectory it might seem that the simulation was done in a rectangular box. The program `trjconv` can be used to convert the trajectory to a different unit-cell representation.

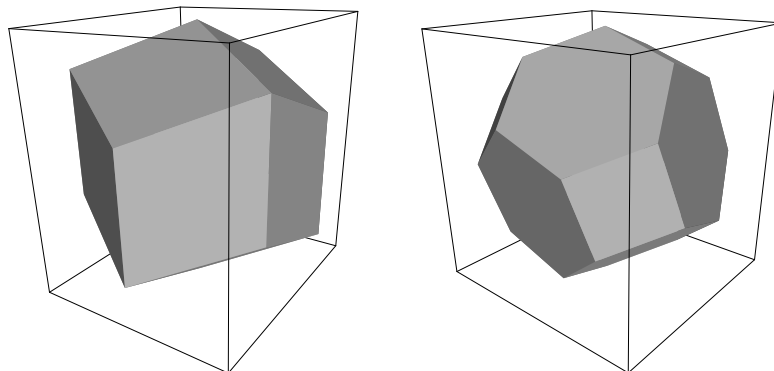


Figure 3.2: A rhombic dodecahedron and truncated octahedron (arbitrary orientations).

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

Table 3.1: The cubic box, the rhombic dodecahedron and the truncated octahedron.

It is also possible to simulate without periodic boundary conditions, but it is usually more efficient to simulate an isolated cluster of molecules in a large periodic box, since fast grid searching can only be used in a periodic system.

3.2.1 Some useful box types

The three most useful box types for simulations of solvated systems are described in Table 3.1. The rhombic dodecahedron (Fig. 3.2) is the smallest and most regular space-filling unit cell. Each of the 12 image cells is at the same distance. The volume is 71% of the volume of a cube having the same image distance. This saves about 29% of CPU-time when simulating a spherical or flexible molecule in solvent. There are two different orientations of a rhombic dodecahedron that satisfy equations 3.1, 3.2 and 3.3. The program `editconf` produces the orientation which has a square intersection with the xy-plane. This orientation was chosen because the first two box vectors coincide with the x and y-axis, which is easier to comprehend. The other orientation can

be useful for simulations of membrane proteins. In this case the cross-section with the xy-plane is a hexagon, which has an area which is 14% smaller than the area of a square with the same image distance. The height of the box (c_z) should be changed to obtain an optimal spacing. This box shape not only saves CPU time, it also results in a more uniform arrangement of the proteins.

3.2.2 Cut-off restrictions

The minimum image convention implies that the cut-off radius used to truncate non-bonded interactions may not exceed half the shortest box vector:

$$R_c < \frac{1}{2} \min(\|\mathbf{a}\|, \|\mathbf{b}\|, \|\mathbf{c}\|), \quad (3.4)$$

because otherwise more than one image would be within the cut-off distance of the force. When a macromolecule, such as a protein, is studied in solution, this restriction alone is not sufficient: in principle, a single solvent molecule should not be able to ‘see’ both sides of the macromolecule. This means that the length of each box vector must exceed the length of the macromolecule in the direction of that edge *plus* two times the cut-off radius R_c . It is, however, common to compromise in this respect, and make the solvent layer somewhat smaller in order to reduce the computational cost. For efficiency reasons the cut-off with triclinic boxes is more restricted. For grid search the extra restriction is weak:

$$R_c < \min(a_x, b_y, c_z) \quad (3.5)$$

For simple search the extra restriction is stronger:

$$R_c < \frac{1}{2} \min(a_x, b_y, c_z) \quad (3.6)$$

Each unit cell (cubic, rectangular or triclinic) is surrounded by 26 translated images. A particular image can therefore always be identified by an index pointing to one of 27 *translation vectors* and constructed by applying a translation with the indexed vector (see 3.4.3). Restriction (3.5) ensures that only 26 images need to be considered.

3.3 The group concept

The GROMACS MD and analysis programs use user-defined *groups* of atoms to perform certain actions on. The maximum number of groups is 256, but each atom can only belong to six different groups, one each of the following:

temperature-coupling group The temperature coupling parameters (reference temperature, time constant, number of degrees of freedom, see 3.4.4) can be defined for each T-coupling group separately. For example, in a solvated macromolecule the solvent (that tends to generate more heating by force and integration errors) can be coupled with a shorter time constant to a bath than is a macromolecule, or a surface can be kept cooler than an adsorbing molecule. Many different T-coupling groups may be defined. See also center of mass groups below.

freeze group Atoms that belong to a freeze group are kept stationary in the dynamics. This is useful during equilibration, *e.g.* to avoid badly placed solvent molecules giving unreasonable kicks to protein atoms, although the same effect can also be obtained by putting a restraining potential on the atoms that must be protected. The freeze option can be used, if desired, on just one or two coordinates of an atom, thereby freezing the atoms in a plane or on a line. When an atom is partially frozen, constraints will still be able to move it, even in a frozen direction. A fully frozen atom can not be moved by constraints. Many freeze groups can be defined. Frozen coordinates are unaffected by pressure scaling; in some cases this can produce unwanted results, particularly when constraints are also used (in this case you will get very large pressures). Accordingly, it is recommended to avoid combining freeze groups with constraints and pressure coupling. For the sake of equilibration it could suffice to start with freezing in a constant volume simulation, and afterward use position restraints in conjunction with constant pressure.

accelerate group On each atom in an “accelerate group” an acceleration a^g is imposed. This is equivalent to an external force. This feature makes it possible to drive the system into a non-equilibrium state and enables the performance of non-equilibrium MD and hence to obtain transport properties.

energy-monitor group Mutual interactions between all energy-monitor groups are compiled during the simulation. This is done separately for Lennard-Jones and Coulomb terms. In principle up to 256 groups could be defined, but that would lead to 256×256 items! Better use this concept sparingly.

All non-bonded interactions between pairs of energy-monitor groups can be excluded (see sec. 7.3). Pairs of particles from excluded pairs of energy-monitor groups are not put into the pair list. This can result in a significant speedup for simulations where interactions within or between parts of the system are not required.

center of mass group In GROMACS the center of mass (COM) motion can be removed, for either the complete system or for groups of atoms. The latter is useful, *e.g.* for systems where there is limited friction (*e.g.* gas systems) to prevent center of mass motion to occur. It makes sense to use the same groups for temperature coupling and center of mass motion removal.

Compressed position output group In order to further reduce the size of the compressed trajectory file (`.xtc` or `.trj`), it is possible to store only a subset of all particles. All x-compression groups that are specified are saved, the rest are not. If no such groups are specified, then all atoms are saved to the compressed trajectory file.

The use of groups in GROMACS tools is described in sec. 8.1.