Biomolecular modeling II

Marcus Elstner and Tomáš Kubař

2019, December 4

Molecular dynamics simulation

how to get things moving

Equations of motion

$$m \cdot \ddot{\vec{r}} = \vec{F}$$

ordinary differential equations of second order

- have to be solved numerically
- solution proceeds in discreet steps of length Δt
- numerical integration starts at time t_0,
  where the initial conditions are specified
  – the positions $\vec{r}_0$ and the velocities $\vec{v}_0$
- calculations of forces at $\vec{r}_0$ to get accelerations $\ddot{\vec{r}}_0$
- then, an integrator calculates $\vec{r}$ and $\vec{v}$ at time $t_0 + \Delta t$
- accelerations → step → accelerations → step → . . .

Verlet integration method

for the development of the method:

take a virtual step in positive time and in ‘negative’ time, and apply Taylor expansion up to second order:

$$\begin{aligned} r(t+\Delta t) &= r(t) + \dot{r}(t)\cdot\Delta t + \tfrac{1}{2}\ddot{r}(t)\cdot\Delta t^2 \\ r(t-\Delta t) &= r(t) - \dot{r}(t)\cdot\Delta t + \tfrac{1}{2}\ddot{r}(t)\cdot\Delta t^2 \end{aligned}$$

add both equations – eliminate the velocity $\dot{r}$:

$$\begin{aligned} r(t+\Delta t) &= 2\cdot r(t) - r(t-\Delta t) + \ddot{r}(t)\Delta t^2 \\ \ddot{r}(t) &= a(t) = \frac{F(t)}{m} = -\frac{1}{m}\frac{\partial V}{\partial r}(t) \end{aligned}$$

Verlet integration method

another, equivalent formulation – velocity Verlet

$$\begin{aligned} r(t+\Delta t) &= r(t) + v(t)\cdot\Delta t + \tfrac{1}{2}a(t)\cdot\Delta t^2 \\ v(t+\Delta t) &= v(t) + \tfrac{1}{2}\left(a(t) + a(t+\Delta t)\right)\cdot\Delta t \end{aligned}$$

yet another – Leap-frog

$$\begin{aligned} v(t+\tfrac{1}{2}\Delta t) &= v(t-\tfrac{1}{2}\Delta t) + a(t)\cdot\Delta t \\ r(t+\Delta t) &= r(t) + v(t+\tfrac{1}{2}\Delta t)\cdot\Delta t \end{aligned}$$

both: better numerical precision than Verlet normal form

Δt – crucial parameter

Let us say: we want to obtain a trajectory over a time interval T
- – we perform M steps
- – we have to evaluate the forces on atoms $M = T/\Delta t$ times

Computational cost of the calculation of forces
- – major computational effort
- – determines how many steps we can afford to make

Δt – crucial parameter

- we neglect contributions in Δt^3 and higher orders
  → error per step in the order of Δt^3
- keep the step short → make the error small
  but need too many steps to simulate certain time T
- make the step long → cut computational cost
  but increase the error and decrease stability
- compromise needed

Δt – crucial parameter

- fastest motion – hydrogen atoms, period around 10 fs
- rule of thumb – stable integration with $\Delta t \leq \frac{1}{10}$ fastest period
- practically, Δt of 1 fs is used,
  increase to 2 fs possible with a special treatment of bonds
- 1M calculations of forces needed for a trajectory of 1 ns
- large systems – multi-ns simulations routinely, μs possible

Total energy with a force field

$$V(R^N) =$$

$$\begin{aligned}
= \quad & \frac{1}{2}\sum_i k_i(r_i - r_i^0)^2 + \frac{1}{2}\sum_j k_j^\vartheta(\vartheta_j - \vartheta_j^0)^2 + \frac{1}{2}\sum_n V_n \cdot \cos\left[n\omega - \gamma_n\right] \\
+ \quad & \sum_i^N \sum_{j=i+1}^N \left\{ 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right) + \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \right\}
\end{aligned}$$

and get forces as derivatives with respect to atomic coordinates:

$$F_i^x = -\frac{\partial V}{\partial x_i}$$

Verlet integration method

$$\begin{aligned} r(t+\Delta t) &= 2\cdot r(t) - r(t-\Delta t) + \ddot{r}(t)\Delta t^2 \\ \ddot{r}(t) &= a(t) = \frac{F(t)}{m} = -\frac{1}{m}\frac{\partial V}{\partial r}(t) \end{aligned}$$

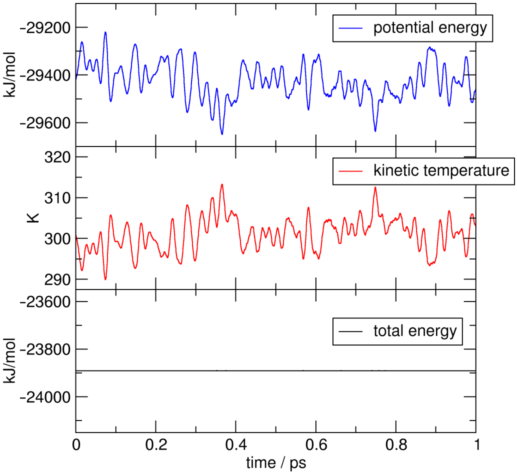
and choose an appropriate time step Δt

Temperature and pressure

what you simulate is what you would measure

Energies and temperature

Solution of equations of motion – conserves total / internal energy



Energies and temperature

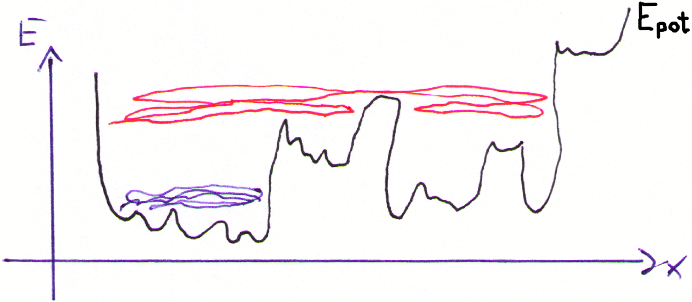
Solution of equations of motion – conserves total / internal energy

what we need – to control basic simulation parameters
– temperature and possibly pressure

significance of temperature
- – determines which structures of the system are accessible
- – different dynamics at high and at low temperatures

Energies and temperature

high E – multiple different structural ‘classes’ are reached
low E – restricted available structures



difference $E - E_{\mathrm{pot}}$ corresponds to E_{kin} and temperature

Isolated system

- exchanges with surroundings neither energy (heat / work)
  nor matter (particles)
- total energy of system: $E = E_{\mathrm{kin}} + E_{\mathrm{pot}} = \mathrm{const}$
- individually, E_{kin} and E_{pot} fluctuate in the course of time
  as they are being transformed into each other
- is what we get when using the Verlet method for a molecule

Isolated system

kinetic theory of gases → relation of E_{kin} and temperature:

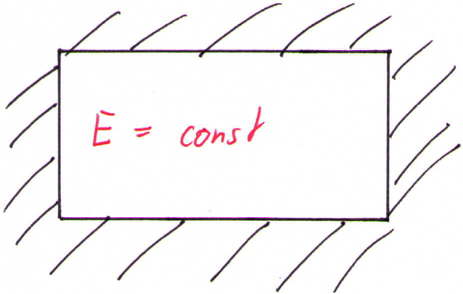
$$\begin{aligned}\left\langle E_{\text{kin}}\right\rangle &= \frac{3}{2}NkT \\ \text{where}\ \left\langle E_{\text{kin}}\right\rangle &= \frac{1}{2}\sum_i m_i \left\langle v_i^2\right\rangle\end{aligned}$$

‘local’ T – fluctuates in time; may differ between parts of system

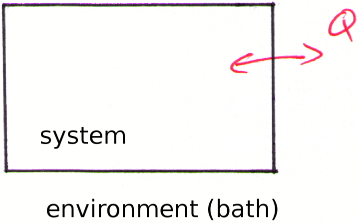
Isolated and closed system

experimental setup (a test tube with a sample)
- usually in thermodynamic equilibrium with the surroundings
- temperature of system = temperature of suroundings

isolated system



closed system



system

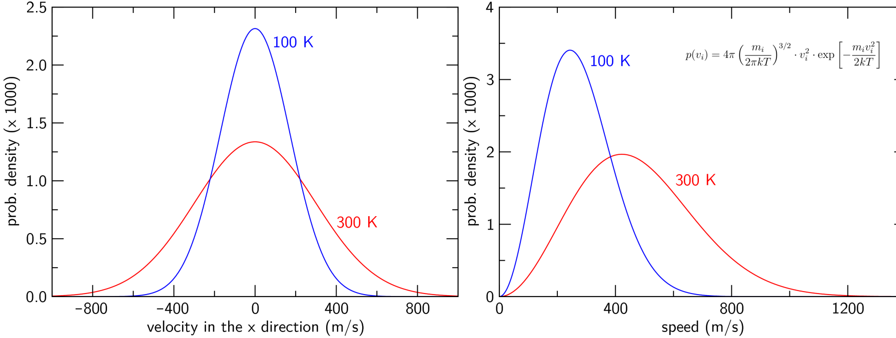
environment (bath)

Closed system

- thermal contact of system with surroundings
- exchange of energy in the form of heat
  until the temperature of surroundings is reached
- canonical ensemble
- velocity / speed of atoms – Maxwell–Boltzmann distribution

Canonical ensemble

Maxwell–Boltzmann distribution of velocity / speed (N_2, IG)



Idea for constant temperature – scale the velocities

'instantaneous temp.' T in a simulation deviates from target T_ref:

$$T(t) = \frac{2}{3}\frac{E_\text{kin}(t)}{N k_\text{B}} \neq T_\text{ref} \qquad\qquad E_\text{kin}(t) = \frac{1}{2}\sum_i m_i \cdot v_i(t)^2$$

how to control T? scale the velocities by a certain factor λ,
so that temperature is modified and T_ref is reached:

$$T_\text{ref} = \frac{1}{\frac{3}{2} N k_\text{B}} \cdot \frac{1}{2}\sum_i m_i \left(\lambda \cdot v_i\right)^2 = \lambda^2 \cdot \frac{1}{\frac{3}{2} N k_\text{B}} \cdot \frac{1}{2}\sum_i m_i v_i^2 = \lambda^2 \cdot T$$

thus: scale all velocities by $\lambda = \sqrt{T_\text{ref} / T} \;\rightarrow\; T_\text{ref}$ reached exactly

Problem: rescaling of velocities affects the 'natural' dynamics
of the molecular system too strongly → unstable simulation

Berendsen thermostat

How to avoid the drastic changes to the dynamics?
adjust velocities more smoothly, in the direction of T_{ref}

- temperature changes between two time steps according to

$$\Delta T = \frac{\Delta t}{\tau} \left(T_{\text{ref}} - T \right)$$

- rate of change of T (due to the change of velocities) is proportional to the deviation of actual T from T_{ref}
- constant of proportionality – relaxation time τ

Berendsen thermostat

- velocities are scaled by λ:

$$\begin{aligned} T_{\text{new}} &= T + \Delta T = T + \frac{\Delta t}{\tau}\left(T_{\text{ref}} - T\right) \\ \lambda &= \sqrt{\frac{T_{\text{new}}}{T}} = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{\text{ref}}}{T} - 1\right)} \end{aligned}$$

- usually: $\tau = 0.1 - 10$ ps
- T will fluctuate around the desired value T_{ref}

Berendsen thermostat

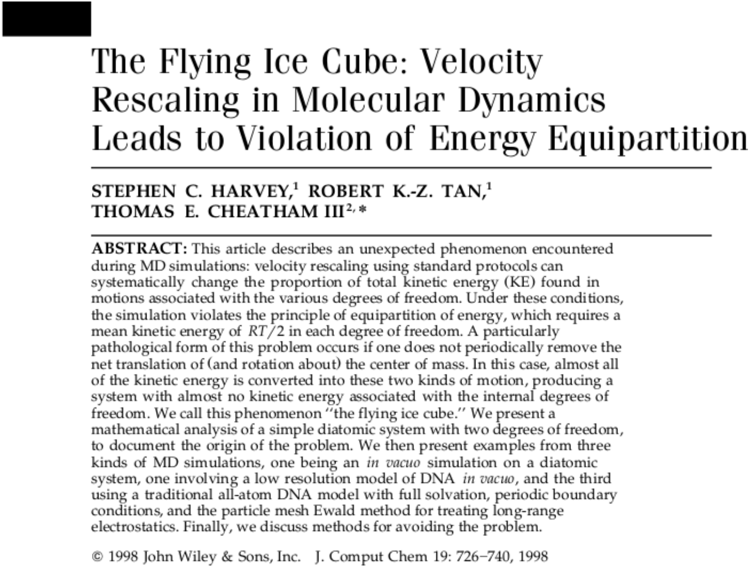
Question – is the statistical thermodynamics correct?

- in other words – is distribution of velocities correct (M–B)?
- it is important to make sure that canonical ensemble is sampled because every thermodynamic quantity A is obtained as an ensemble average:

$$\langle A \rangle = \frac{1}{Z} \int \rho(\vec{r}) \cdot A(\vec{r}) \, \mathrm{d}\vec{r}$$

- if sampling is wrong → wrong density ρ → wrong averages
- answer: it is not correct – may even lead to problems like . . .

Berendsen thermostat



[redacted]

The Flying Ice Cube: Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition

STEPHEN C. HARVEY,[1] ROBERT K.-Z. TAN,[1]
THOMAS E. CHEATHAM III[2,*]

ABSTRACT: This article describes an unexpected phenomenon encountered during MD simulations: velocity rescaling using standard protocols can systematically change the proportion of total kinetic energy (KE) found in motions associated with the various degrees of freedom. Under these conditions, the simulation violates the principle of equipartition of energy, which requires a mean kinetic energy of $RT/2$ in each degree of freedom. A particularly pathological form of this problem occurs if one does not periodically remove the net translation of (and rotation about) the center of mass. In this case, almost all of the kinetic energy is converted into these two kinds of motion, producing a system with almost no kinetic energy associated with the internal degrees of freedom. We call this phenomenon "the flying ice cube." We present a mathematical analysis of a simple diatomic system with two degrees of freedom, to document the origin of the problem. We then present examples from three kinds of MD simulations, one being an *in vacuo* simulation on a diatomic system, one involving a low resolution model of DNA *in vacuo*, and the third using a traditional all-atom DNA model with full solvation, periodic boundary conditions, and the particle mesh Ewald method for treating long-range electrostatics. Finally, we discuss methods for avoiding the problem.

© 1998 John Wiley & Sons, Inc. J. Comput Chem 19: 726–740, 1998

Nosé–Hoover thermostat

- generates the correct canonical ensemble → good choice
- conceptionally and mathematically > difficult to understand
- heat bath is treated not as an external element
  rather as an integral part of the system;
  is assigned an additional degree of freedom
  s with fictitious mass Q
- eqns of motion for this extended system ($3N + 1$ DOF):

$$\begin{aligned}\ddot{r}_i &= \frac{F_i}{m_i} - s \cdot \dot{r}_i \\ \dot{s} &= \frac{1}{Q}\left(T - T_{\mathrm{ref}}\right)\end{aligned}$$

Introducing pressure

chemical reality – constant pressure rather than constant volume
goal – implement such conditions in simulations, too

How to calculate pressure?
– first, calculate virial of force

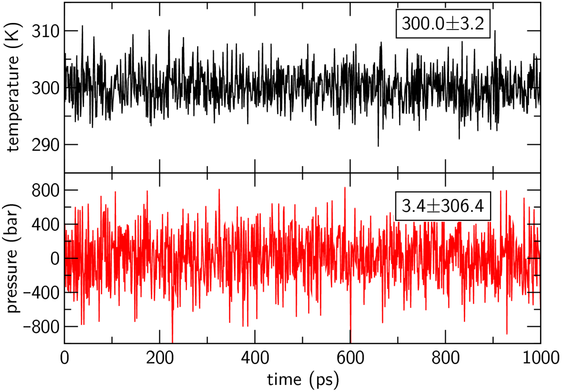
$$\Xi = -\frac{1}{2}\sum_{i<j} \vec{r}_{ij} \cdot \vec{F}_{ij}$$

($\vec{r}_{ij}$ distance of atoms i and j, $\vec{F}_{ij}$ – force between them)

$$P = \frac{2}{3V} \cdot (E_{\mathrm{kin}} - \Xi) = \frac{2}{3V} \cdot \left(\frac{1}{2}\sum_{i} m_i \cdot |\vec{v}_i|^2 + \frac{1}{2}\sum_{i<j} \vec{r}_{ij} \cdot \vec{F}_{ij} \right)$$

Measuring pressure

T and P in an NPT simulation of a DNA oligomer in water
($T_{\text{ref}} = 300$ K, $P_{\text{ref}} = 1.0$ bar)



Controlling pressure

we can calculate the pressure
  – so how do we maintain it at a desired value?

barostat – algorithm that is equivalent of a thermostat,
  just that it varies volume of the box instead of velocities

alternatives are available:

- Berendsen barostat
    - – direct rescaling of box volume
    - – system coupled to a 'force / pressure bath' – piston
- Parrinello–Rahman barostat
    - – extended-ensemble simulation
    - – additional DOF for the piston

System boundary and the solvent

Biomolecule in solution

typical MD simulations – molecular system in aqueous solution
preferably – make the system as small as possible (reduce cost)

straightforward solution – single molecule of solute (protein, DNA)
with a smallest possible number of H_2O molecules
typical – several thousand H_2O molecules in a box $n \times n \times n$ nm

issue – everything is close to the surface,
while we are interested in a molecule in bulk solvent

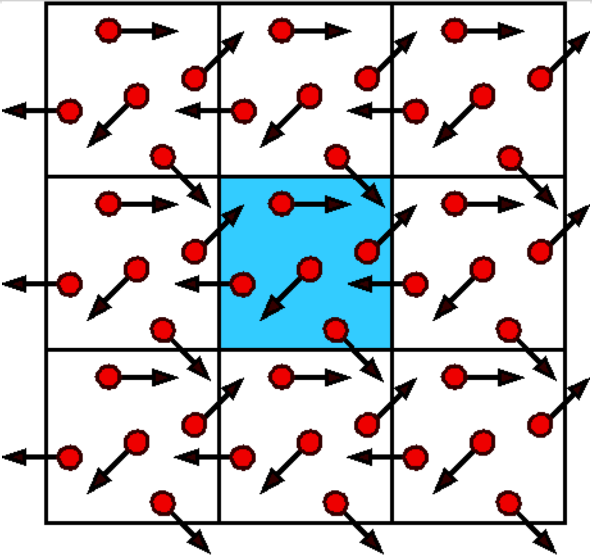
Periodic boundary conditions

- elegant way to avoid these problems
- molecular system placed in a regular-shaped box
- the box is virtually replicated in all spatial directions

Periodic boundary conditions

- elegant way to avoid these problems
- molecular system placed in a regular-shaped box
- the box is virtually replicated in all spatial directions
- positions (and velocities) of all particles are identical in all replicas, so that we can keep only one copy in the memory
- this way, the system is infinite – no surface!
- the atoms near the wall of the simulation cell interact with the atoms in the neighboring replica

Periodic boundary conditions

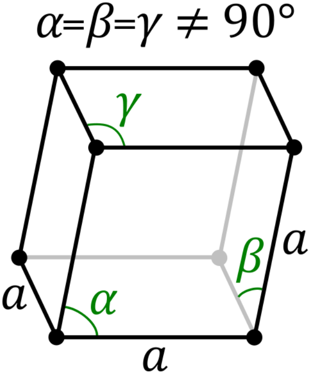


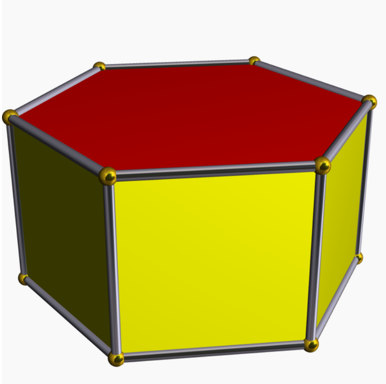
PBC – features

- only coordinates of the unit cell are kept in memory
- atom that leaves the box enters it on the other side
- carefull accounting of the interactions of atoms necessary!
  simplest way – minimum image convention:
  an atom interacts with the nearest copy of every other
  – interaction with two different images of another atom,
  or even with another image of itself is avoided

PBC – box shape

may be simple – cubic or orthorhombic, parallelepiped (specially, rhombohedron), or hexagonal prism

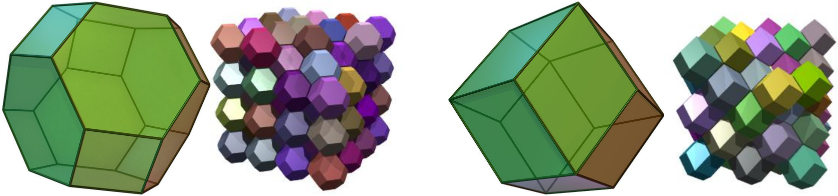




PBC – box shape

. . . but also more complicated

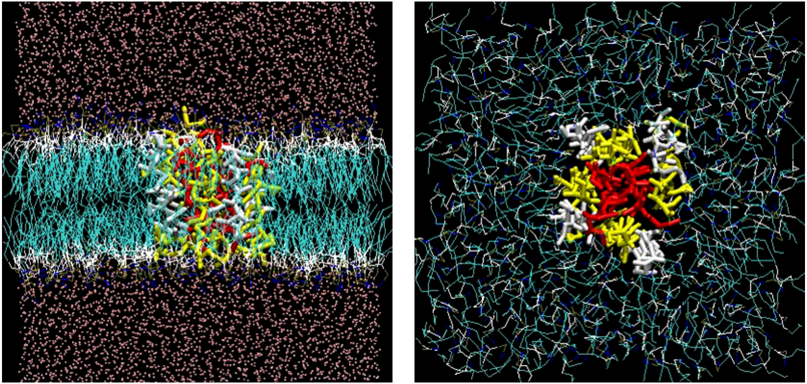
- truncated octahedral or rhombic dodecahedral
- quite complex equations for interactions & eqns of motion



advantage for simulation of spherical objects (globular proteins)
– no corners far from the molecule filled with unnecessary H_2O

PBC – box shape

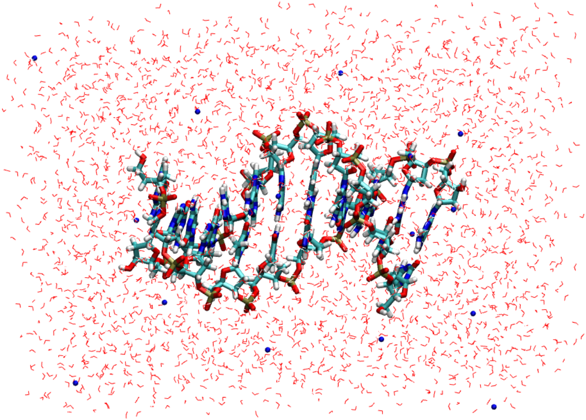
2D objects – phase interfaces, membrane systems
  – usually treated in a slab geometry



Water in biomolecular simulations

most simulations – something in aqueous solutions

H_2O – usually (many) thousands of molecules



Water in biomolecular simulations

most simulations – something in aqueous solutions
H_2O – usually (many) thousands of molecules

example – simulation of DNA decanucleotide:

- PBC box 3.9 × 4.1 × 5.6 nm (smallest meaningful)
- 630 atoms in DNA, 8346 atoms in water and 18 Na^+
- concentration of DNA: 18 mmol/L – very high!
- of all pair interactions: 86 % are water–water,
  most of the others involve water

Water models

most interactions involve H_2O

→ necessary to pay attention to its description

model of water must be simple enough (computational cost)

and accurate enough, at the same time

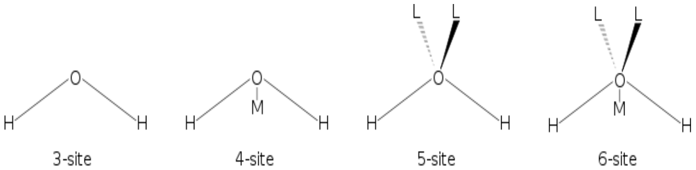
water models – usually rigid

– bond lengths and angles do not vary – constraints

molecule with three sites (atoms in this case), or up to six sites

– three atoms and virtual sites corresponding

to a ‘center’ of electron density or lone electron pairs



Water models

TIP3P (or SPC)

- most frequently used
- 3 atoms with 3 rigid bonds, charge on every atom ($-0.834/+0.417$)
- only the O atom has non-zero LJ parameters (optimization)

TIP4P

- negative charge placed on virtual site M rather than on the O
- electric field around the molecule described better

TIP5P

- 2 virtual sites L with negative charges near the O – lone pairs
- better description of directionality of H-bonding etc.
  (radial distribution function, temperature of highest density)