Introduction to Classical Molecular Dynamics

Giovanni Chillemi  g.chillemi@cineca.it
HPC department, CINECA
MD ingredients

- Coordinates
- Velocities
- Force field
- Topology
- Input parameters
- Trajectories
- Analysis
The equations that describe the temporal evolution of a physical system is called **equation of motion**. There are several different equations of motions, which characterize the motion in different types of physical systems:

- Time-dependent Schrödinger's Equation
  - for quantum-mechanical system
- **Newton’s Equation**
  - for classical-mechanical system
- Langevin's Equation
  - for stochastic system
Molecules are quantum-mechanical systems whose motion should be described by Schrödinger's Equation. However, technical difficulties make solving Schrödinger's Equation for large systems impractical.

Therefore the motion of a molecule is usually approximated by the laws of classical mechanics and by Newton's equation of Motion. In its most simplistic form Newton's second law of motion states:

$$ f_i = m_i \cdot a_i $$

where $m_i$ is the mass of particle $i$, $a_i$ is its acceleration. The force $f_i$ is given as the derivative of the potential energy function $V$:

$$ f_i = - \frac{\partial V}{\partial r_i} $$

where $r_i$ is the position of particle $i$. 
Potential energy function

\[
V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) = \sum_{\text{bond}} \frac{1}{2} k_{b_n} (b_n - b_{0_n})^2 + \sum_{\text{angle}} \frac{1}{2} k_{\theta_n} (\theta_n - \theta_{0_n})^2 + \\
\sum_{\text{improper dihedral}} \frac{1}{2} k_{\xi_n} (\xi_n - \xi_{0_n})^2 + \sum_{\text{dihedral}} k_{\phi_n} [1 + \cos (m_n \phi_n - \delta_n)] + \\
\sum_{\text{nonbonded pairs (ij)}} \left( \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{\varepsilon_r r_{ij}}
\]

non bonded interactions

bonded interactions
<table>
<thead>
<tr>
<th>Model</th>
<th>Degree of freedom</th>
<th>Example of predicted properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Considered</td>
<td>Removed</td>
</tr>
<tr>
<td>Quantum mechanic</td>
<td>Nucleus, electrons</td>
<td>Nucleons</td>
</tr>
<tr>
<td>Polarizable atoms</td>
<td>Atoms, dipoles</td>
<td>Electrons</td>
</tr>
<tr>
<td>Non polarizable atoms</td>
<td>Solute atoms,</td>
<td>Dipoles</td>
</tr>
<tr>
<td></td>
<td>solvent atoms</td>
<td></td>
</tr>
<tr>
<td>Implicit solvent</td>
<td>Solute atoms</td>
<td>Solvent atoms</td>
</tr>
</tbody>
</table>
## A Brief History

<table>
<thead>
<tr>
<th>Year</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1964</td>
<td>Liquid Argon (Rahman Phys Rev)</td>
</tr>
<tr>
<td>1977</td>
<td>Small protein in vacuo (Mc Cammon Karplus Nature)</td>
</tr>
<tr>
<td>1988</td>
<td>First Protein in explicit water (Levitt PNAS)</td>
</tr>
<tr>
<td>From 1995</td>
<td>Protein-DNA Complexes – Membrane Proteins- Complex Systems</td>
</tr>
</tbody>
</table>
Bond Stretching Energy

\[ V = \sum_{\text{bond}} \frac{1}{2} k_b (b_n - b_{0n})^2 + \ldots \]

- \( k_b \) is the spring constant of the bond
- \( b_0 \) is the bond length at equilibrium

Unique \( k_b \) and \( b_0 \) assigned for each bond pair, i.e. C-C, O-H

Principle of bond stretching (left), and the bond stretching potential (right).
Bond Stretching Force

\[
f_i = -\frac{\partial V^{\text{bond}}}{\partial r_i} = -\frac{\partial V^{\text{bond}}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial r_i} = k_b \left( r_{ij} - b_0 \right) \frac{r_{ij}}{r_{ij}}
\]

\[
f_j = -f_i
\]

If atom \(i\) and \(j\) are closer than \(b_0\), the bond force separates them.

If atom \(i\) and \(j\) are farther than \(b_0\), the bond force draws them nearer.
Bending Energy

\[ + \sum_{\text{angle}} \frac{1}{2} k_{\theta_n} \left( \theta_n - \theta_{0_n} \right)^2 + \]

\( k_{\theta} \) is the spring constant of the bending.
\( \theta_0 \) is the angle bending at equilibrium.

Unique parameters for angle bending are assigned to each bonded triplet of atoms based on their types (e.g. C-C-C, C-O-C, C-C-H, etc.)

Principle of angle vibration (left) and the bond angle potential (right).
Torsional o Dihedral Energy

\[ + \sum_{\text{dihedral}} k_{\phi_n} [1 + \cos (m_n \phi_n - \delta_n)] + \]

\( \phi = \text{angle} \)
\( \delta = \text{phase} \)
\( m = \text{number of peaks in a full rotation} \)

Principle of proper dihedral angle (left, in trans form) and the dihedral angle potential (right).
Improper Dihedral Energy

The energy required to deform a group of atoms from its equilibrium angle, $x_0$. Used for tetrahedral or planar groups

\[ + \sum_{\text{improper dihedral}} \frac{1}{2} k \xi_n \left( \xi_n - \xi_{0n} \right)^2 \]

Again this system can be modeled by a spring, and the energy is given by the Hookean potential with respect to the planar angle.
The “Hookean” potential

$k_\theta (\theta - \theta_o)^2$ or $k_b (r - r_o)^2$

$k_b$ and $k_\theta$ broaden or steepen the slope of the parabola
The larger the value of $k$, the more energy is required to deform an angle (or bond) from its equilibrium value.
Lennard Jones (Van der Waals) interactions

\[ V(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right], \]

\[ \sum_{\text{nonbonded pairs (ij)}} \left( \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{\varepsilon r_{ij}} \]

non bonded interactions
Electrostatic interactions

The $q_i$ and $q_j$ are the partial atomic charges for atoms $i$ and $j$, separated by a distance $r_{ij}$. $\varepsilon_r$ is the relative dielectric constant. For gas phase calculations $\varepsilon_r$ is normally set to 1. Larger values of $\varepsilon_r$ are used to approximate the dielectric effect of intervening solute ($\varepsilon \sim 60$-80) or solvent atoms in solution.

\[
V_{\text{Coulomb}}(r) = \sum_{\text{nonbonded pairs } (ij)} \left( \frac{C^{(12)}_{ij}}{r_{ij}^{12}} - \frac{C^{(6)}_{ij}}{r_{ij}^6} \right) + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{\varepsilon_r r_{ij}}
\]
To simulate our finite system in liquid conditions, we apply the pbc: i.e. the system box is virtually surrounded in all directions by copy of itself. An atom close to a box border interacts with the atoms in another pbc image. The non-bonded interactions are only calculated between atom pairs closer than a spherical cut-off.
BOX dimension

The edge of cubic box must be large enough to avoid interactions of the solute with itself. Its minimal dimension therefore depends by the chosen cut-off for the non bonded interactions.

Edge of the box = 3.2 nm
Electrostatic interactions: Particle Mesh Ewald (PME)

The cut-off radius method for electrostatic interactions is particularly inaccurate for charged molecules such as DNA or dipolar groups such as alpha helices.

PME corrects these errors and helps maintain a short cut-off in the real space: i.e., the number of atom pairs is reduced.

A charge group is a neutral charge group composed by several partially charged atoms of a chemical group. Electrostatics can be calculated between charge groups instead that atom pairs.

Charge groups were first introduced to reduce artifacts in the electrostatics calculation but they can also speed up the calculations; given a pair of water molecules for instance, we only need to determine one atom distance instead of nine (or sixteen for a four-site water model).

Note that an atom type is not a physical feature. O8 is defined with a different atom type than O9. In fact, their bond constants with C7 and atomic charges are different.
The potential energy function, together with the parameters required to describe the behavior of different kinds of atoms and bonds \((k_b, k_{\theta}, k_{\phi}, C_{ij}, \ldots)\), is called a **force field**. Several force fields are currently used and the choice depends from the studied system. Some force field are better suited for nucleic acids, for example, while others for membrane proteins.
Available forcefield in Gromacs (4.6.5)

1. AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
2. AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
3. AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
4. AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
5. AMBER99SB protein, nucleic AMBER94 (Hornak et al., Proteins 65, 712-725, 2006)
6. AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)
7. AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
8. CHARMM27 all-atom force field (with CMAP) - version 2.0
9. GROMOS96 43a1 force field
10. GROMOS96 43a2 force field (improved alkane dihedrals)
11. GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
12. GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
13. GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
15. OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
16. [DEPRECATED] Encad all-atom force field, using full solvent charges
17. [DEPRECATED] Encad all-atom force field, using scaled-down vacuum charges
18. [DEPRECATED] Gromacs force field (see manual)
19. [DEPRECATED] Gromacs force field with hydrogens for NMR
Numeric integration of Newton's equation of motion is typically done step by step using methods that are called **Finite Difference** methods.

These methods use the information available at time $t$ to predict the system's coordinates and velocities at a time $t + \delta t$, where $\delta t$ is a short time interval and are based on a Taylor expansion of the position at time $t + \delta t$

\[
r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2} a(t)\delta t^2 + ...
\]
Integration of the equation of motion

\[ r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \]

- **Verlet integrator**

\[ r(t + \delta t) = r(t) - v(t + \frac{1}{2}\delta t)\delta t \]

\[ v(t + \frac{1}{2}\delta t) = v(t - \frac{1}{2}\delta t) + a(t)\delta t \]

- **Leap-frog integrator**

\[ r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \]

\[ v(t + \delta t) = v(t) + [a(t) + a(t + \delta t)]\frac{\delta t}{2} \]

- **Velocity Verlet**
Choice of the timestep

The length of the timestep must be small compared to the period of the highest frequency motions being simulated.

<table>
<thead>
<tr>
<th>Force characteristics</th>
<th>Relaxation time (fs)</th>
<th>Time step (fs)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High frequency motion</strong>: bond stretching vibrations</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Medium frequency motion</strong>: angle bending, proper and improper dihedral angle deformation, LJ and short range Coulombian interactions</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td><strong>Low frequency motion</strong>: long range coulombian interactions</td>
<td>1000</td>
<td>20</td>
</tr>
</tbody>
</table>

The bond stretching vibrations are generally of minimal interest in the study of biomolecular structure and function. Therefore this degree of freedom is usually kept frozen with constraint algorithms such as Shake, Settle, Lincs and a typical timestep is 2 fs (2 x 10^{-15} s).
L’applicazione di vincoli geometrici per mantenere la lunghezza dei legami covalenti durante la simulazione permette di usare un time step fino a 2 fs

In SHAKE (storicamente il primo metodo per l’applicazione dei vincoli in un codice di dinamica molecolare) l’equazione del moto è integrata soddisfacendo contemporaneamente i vincoli sulla distanza degli atomi legati, usando il metodo dei *moltiplicatori di Lagrange*

Metodo iterativo
L'algoritmo LINCS (proposto nel 1997, venti anni dopo SHAKE) risolve in maniera non iterativa vincoli tra atomi legati ed angoli isolati.

Figure 3.9: The three position updates needed for one time step. The dashed line is the old bond of length \(d\), the solid lines are the new bonds. \(l = d \cos \theta\) and \(p = (2d^2 - l^2)^{\frac{1}{2}}\).

SETTLE è un algoritmo non iterativo per applicare i vincoli ad una molecola d'acqua.
Constraints

constraints:

none
No constraints except for those defined explicitly in the topology, i.e. bonds are represented by a harmonic (or other) potential or a Morse potential (depending on the setting of `morse`) and angles by a harmonic (or other) potential.

h-bonds
Convert the bonds with H-atoms to constraints.

all-bonds
Convert all bonds to constraints.

h-angles
Convert all bonds and additionally the angles that involve H-atoms to bond-constraints.

all-angles
Convert all bonds and angles to bond-constraints.
Timescale

- **Protein Folding** - milliseconds/seconds (10^{-3}-1s)
- **Ligand Binding** - micro/milliseconds (10^{-6}-10^{-3} s)
- **Enzyme catalysis** - micro/milliseconds (10^{-6}-10^{-3} s)
- **Conformational transitions** - pico/nanoseconds (10^{-12}-10^{-9} s)
- **Collective vibrations** - 1 picosecond (10^{-12} s)
- **Bond vibrations** - 1 femtosecond (10^{-15} s)
Topology

The topology file describes the atoms composing a molecule and their bond connections.

Es: flexspc.itp in gromacs

```
[ moleculetype ]
; molname    nrexcl
SOL         2

[ atoms ]
; id    at type    res nr    res name    at name    cg nr    charge    mass
 1    OW_sp 1       SOL       OW        1       -0.82     15.99940
 2    HW_sp 1       SOL       HW1       1       0.41      1.00800
 3    HW_sp 1       SOL       HW2       1       0.41      1.00800

[ bonds ]
; i    j    funct    length    force.c.
 1    2    1       0.1       345000  0.1     345000
 1    3    1       0.1       345000  0.1     345000

[ angles ]
; i    j    k    funct    angle    force.c.
 2    1    3    1       109.47   383     109.47   383
```
Constraints in Topology

Only in case of water, the constraint algorithm can be selected in the topology file

[ bonds ]
: i     j       funct   length    force.c. 
1      2          1         0.1       345000
1      3        0.1       345000
flexspc.itp

[ angles ]
: i     j       k       funct   angle   force.c. 
2     1        3         1        109       383

[ settles ]
: OW funct doh dhh spc.itp
1              1       0.1    0.16333
Fig. 4.2.70  Molecular topology building block definition

Solute building block: Benzoic acid (neutral)
Name: BA

a. Atoms

Topology
Initial velocities

The initial velocity of each atom is random assigned through a Maxwell-Boltzmann distribution that is function of the temperature.
To recapitulate..
THE GLOBAL MD ALGORITHM

1. Input initial conditions

Potential interaction $V$ as a function of atom positions
Positions $r$ of all atoms in the system
Velocities $v$ of all atoms in the system

\[ \downarrow \]

\textbf{repeat 2, 3, 4} for the required number of steps:

2. Compute forces

The force on any atom

\[ F_i = -\frac{\partial V}{\partial r_i} \]

is computed by calculating the force between non-bonded atom pairs:

\[ F_i = \sum_j F_{ij} \]

plus the forces due to bonded interactions (which may depend on 1, 2, 3, or 4 atoms), plus restraining and/or external forces.

The potential and kinetic energies and the pressure tensor are computed.

\[ \downarrow \]

3. Update configuration

The movement of the atoms is simulated by numerically solving

Newton's equations of motion

\[ \frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \]

or

\[ \frac{dr_i}{dt} = v_i; \quad \frac{dv_i}{dt} = \frac{F_i}{m_i} \]

\[ \downarrow \]

4. if required: Output step

write positions, velocities, energies, temperature, pressure, etc.
The method discussed above is appropriate for the micro-canonical ensemble: constant $N$ (number of particles), $V$ (volume) and $E_T$ (total energy = $E + E_{\text{kin}}$).

Note that if time step is short enough, the system loses/gains no net energy (potential + kinetic) when running MD in the NVE ensemble.

When simulating biological macromolecules, it might be more appropriate to simulate under constant Temperature ($T$) or constant Pressure ($P$):

- Canonical ensemble: NVT
- Isothermal-isobaric: NPT
Simulating at constant $T$: the Berendsen scheme

Bath supplies or removes heat from the system as appropriate

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T}$$

where $\tau$ determines how strong the bath influences the system

Exponentially scale the velocities at each time step by the factor $\lambda$:

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

$T$: “kinetic” temperature

Simulating at constant T: the Berendsen scheme

A small $\tau$, close to the timestep (strong thermostat), is useful in the equilibration phase, when the quick decreasing of the potential energy could increase too much the kinetic energy of the protein.

A bigger $\tau$, e.g. equal to ten times the timestep (weak thermostat), is useful in the production phase, when we want to keep at minimum the perturbation to the conformational sampling.

Simulating at constant $P$: the Berendsen scheme

Couple the system to a pressure bath

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_p}$$

A change in pressure $P$ is related to a change in volume $V$

To regulate pressure: exponentially scale the volume of the simulation box at each time step by a factor $\mu$

$$\mu(t) = \left[1 - k_T \frac{\Delta t}{\tau_p} (P_o - P(t))\right]^{\frac{1}{3}}$$

where $k_T$: isothermal compressibility
$\tau_p$: coupling constant

Sample input file of gromacs

http://manual.gromacs.org/current/online/mdp.html

title = Yo
cpp = /lib/cpp
include = -I../top
define =
integrator = md
dt = 0.002
nsteps = 500000
nstxout = 5000
nstvout = 5000
nstlog = 5000
nstenergy = 250
nstxout-compressed = 250
compressed-x-grps = Protein
energygrps = Protein SOL
nstlist = 10
ns-type = grid
rlist = 0.8
coulombtype = cut-off
rcoulomb = 1.4
rvdw = 0.8
tcoupl = Berendsen
tc-grps = Protein SOL
tau-t = 0.1 0.1
ref-t = 300 300
Pcoupl = Berendsen
tau-p = 1.0
compressibility = 4.5e-5
ref-p = 1.0
gen-vel = yes
gen-temp = 300
gen-seed = 173529
constraints = all-bonds

tau-t = 0.1 0.1
ref-t = 300 300
Pcoupl = Berendsen
tau-p = 1.0
compressibility = 4.5e-5
ref-p = 1.0
gen-vel = yes
gen-temp = 300
gen-seed = 173529
constraints = all-bonds
Atoms covalently bound are defined as first neighbours second neighbours and so on. ...
LJ and electrostatic interactions **are not calculated among** first and second neighbours since they are considered in the stretching (first) or in the bending potential (second)
The standard non-bonding interactions are too strong for the third neighbours and are reduced (interactions 1-4; list 1-4)
Conformational sampling

Initial coordinates have bad contacts, causing high energies and forces.

Minimization finds a nearby local minimum.

Equilibration escapes local minima with low energy barriers.

Energy

Conformation

Basic simulation samples thermally accessible states.
La superficie potenziale di una molecola è definita da un gran numero di minimi locali (configurazioni stabili dove tutte le derivate prime della funzione energia potenziale rispetto le coordinate sono nulle e tutte le derivate seconde sono non negative) o punti di sella (stati di transizione).
Gli algoritmi di minimizzazione trovano il minimo **locale** della funzione di energia potenziale. Cioè quello che viene raggiunto procedendo lungo il gradiente negativo della superficie di energia potenziale.

In generale NON trovano il minimo globale.

Due algoritmi molto diffusi sono

- *steepest descent*
- *conjugate gradient*

Entrambi del **primo ordine** utilizzano cioè la derivata prima della funzione potenziale rispetto alle coordinate.
Il metodo *steepest descent* utilizza solo il gradiente del potenziale nel punto per calcolare lo spostamento.

È più rapido nella singola iterazione ma meno preciso nel raggiungere il minimo locale quindi è più adatto ad una prima minimizzazione poco accurata.

Il metodo *conjugate gradient* utilizza, oltre al gradiente istantaneo del potenziale, anche quello dello step precedente.

È più accurato nel trovare il minimo locale ma è più lento nei primi passi di minimizzazione quindi è spesso usato dopo un ciclo di *steepest descent*.
Partendo da A, lo *steepest descents* percorre A-B-C (o A-D-F) mentre il *conjugate gradient*, pesando il gradiente A-B con quello B-C riesce a percorrere A-B-O.
steep
A steepest descent algorithm for energy minimization. The maximum step size is `emstep [nm]`, the tolerance is `emtol [kJ mol\(^{-1}\) nm\(^{-1}\)]`.

cg
A conjugate gradient algorithm for energy minimization, the tolerance is `emtol [kJ mol\(^{-1}\) nm\(^{-1}\)]`. CG is more efficient when a steepest descent step is done every once in a while, this is determined by `nstcgsteep`.

`emtol: (100.0) [kJ mol\(^{-1}\) nm\(^{-1}\)]`
the minimization is converged when the maximum force is smaller than this value

`emstep: (0.01) [nm]`
initial step-size

`nstcgsteep: (1000) [steps]`
frequency of performing 1 steepest descent step while doing conjugate gradient energy minimization.