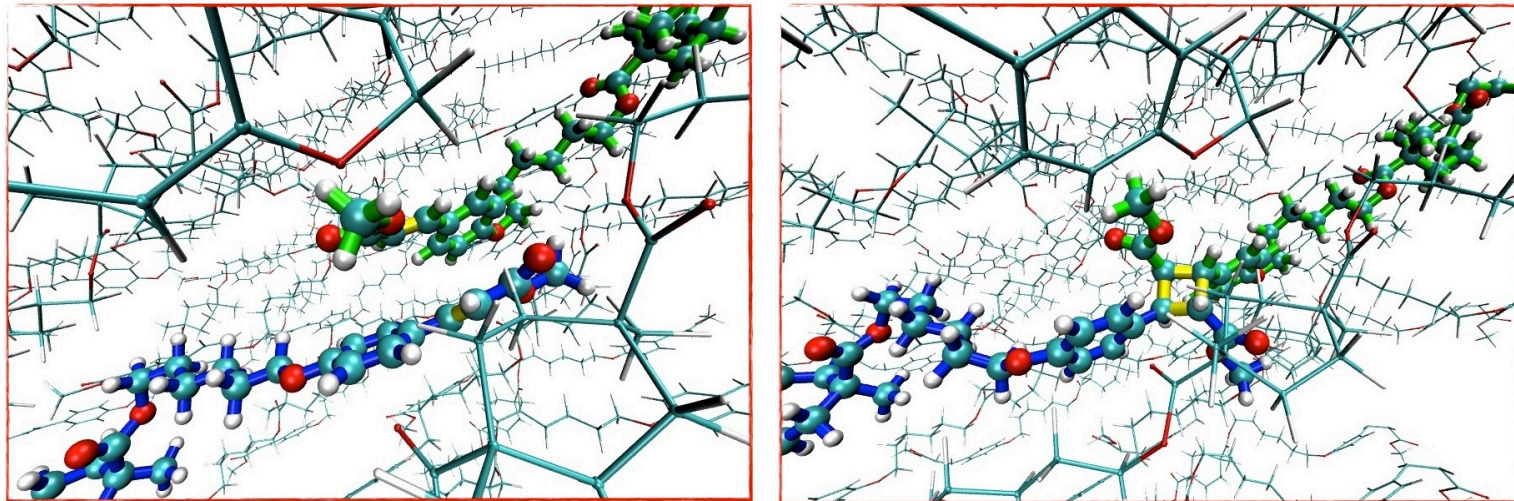# Something on force fields + Molecular Dynamics

Molecular and Materials Modelling (MMM)
HCI D4

**Daniele Passerone**

Empa, Swiss Federal Laboratories for Materials Science and Technology
Ueberlandstrasse 129, 8600 Dübendorf
daniele.passerone@empa.ch

# Classical MD: Force fields

# Molecular dynamics and its evolution

## Empirical classical interaction functions for molecular simulation

**P.H. Hünenberger and W.F. van Gunsteren**
*Laboratory of Physical Chemistry, ETH-Zentrum, CH-8092 Zürich, Switzerland*

### 1. Introduction

With the continuing increase of the power of computers, the past decades have seen a rapid increase in the number, performance and accuracy of theoretical computational methods in chemistry [1,2]. One can distinguish three major classes of methods for the theoretical study of molecular properties, listed in order of decreasing computational expenses: (i) *ab initio* molecular orbital methods [3]; (ii) semiempirical molecular orbital methods [4,5]; and (iii) empirical classical force-field methods. The computational expenses of *ab initio* methods are of order $O(N_f^4)$ (Hartree–Fock level) or higher (configuration interaction, many-body perturbation theory), $N_f$ being the number of basis functions used. Density functional approaches and semiempirical methods scale as $O(N_f^3)$ or lower. The costs of empirical methods scale as $O(N_a^2)$ down to nearly $O(N_a)$, where $N_a$ stands for the number of elementary particles (atoms or groups of atoms). Independently of the scaling with the system size, the evaluation of an empirical interaction function remains usually much cheaper than any other method (size of the prefactor to the scaling) and currently allows for the simulation of systems typically up to $10^5$–$10^6$ atoms

1997: $10^5$-$10^6$ atoms
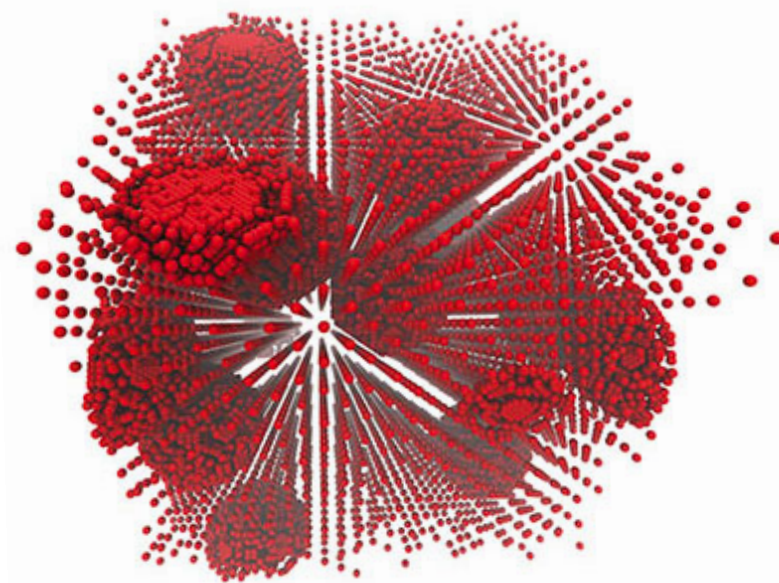


Germany 2013: ($4.125*10^{12}$) particles on Garching's SuperMUC, 146,016 cores used to reach an actual processing power of 591.2 teraFLOPS!!!!

*In liquid form, 4.125 trillion molecules of the noble gas krypton would occupy the volume of a cube whose edges are 6.3 micrometers long. Thus the simulation computation pushes forward into a domain in which it should soon be possible to directly compare the results of simulations with the results of measurements – an important advance on the way to reliable insights into properties of matter.*

# Why simulation? Which choices?
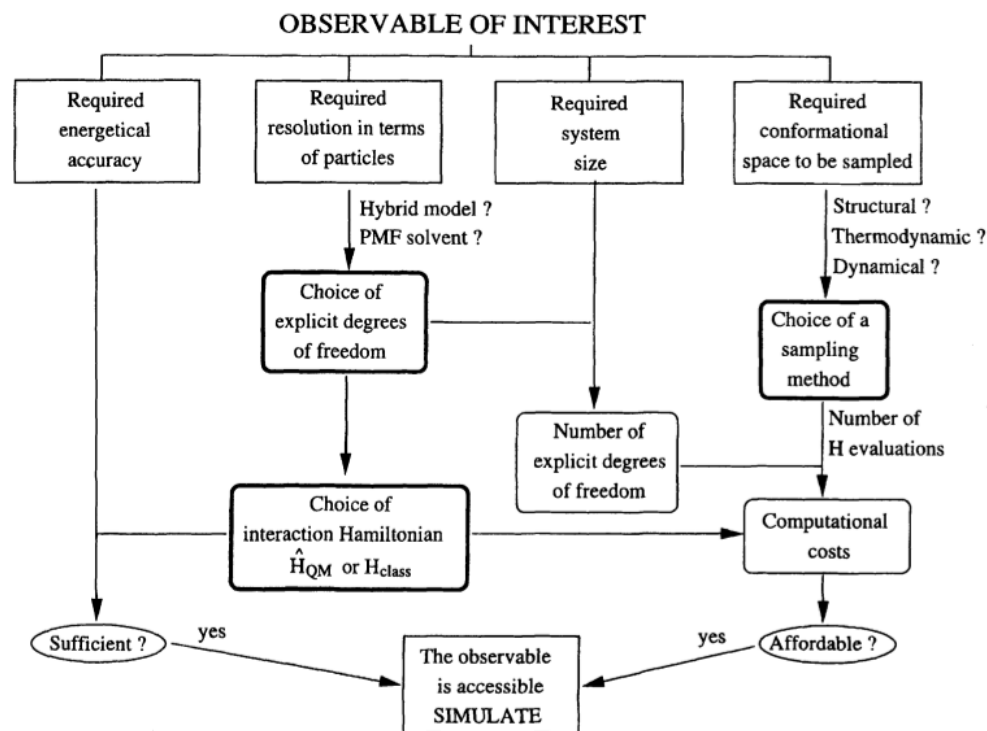
- Hühnenberger + van Gunsteren 1997:     HvG97



Fig. 1. Schematic representation of the basic choices made while building a model of the molecular system in order to simulate an observable of interest. The thick-line boxes represent the three essential choices and the global scheme of the present text.

# Choice of the explicit degrees of freedom of the model

The choice of the elementary unit is the first step in the design of an empirical force field. This choice affects:

• the computational effort

• the extent of conformational space that can be searched (timescale)

• maximum resolution in terms of particles and processes

• the energetical accuracy in the interactions

• the type of observables that can be accessible
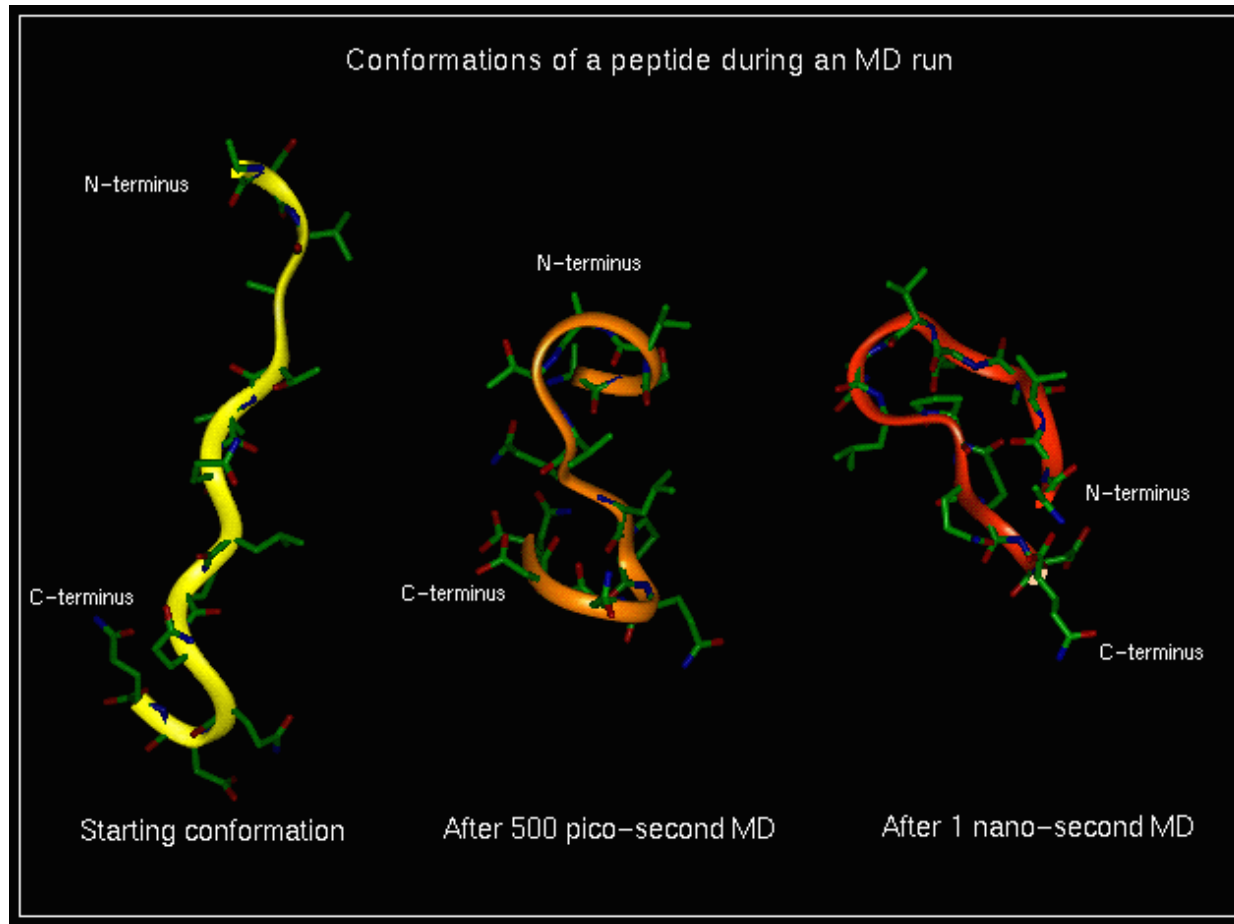
# Hierarchy of explicit degrees of freedom

Table 1 *Hierarchy of explicit degrees of freedom included in the model*

| Elementary unit | Phase | Type of interaction (operator/function) | Degrees of freedom averaged out | Reference |
|---|---|---|---|---|
| **Electrons and nuclei** | Gas phase | *Ab initio*, density functional: First-principles quantum mechanical Hamiltonian, Born–Oppenheimer surface | None | [3] |
| | | Semiempirical: Approximated Hamiltonian | None | [4,5] |
| | Explicit solvent | *Idem*, supermolecule methods | None | [14] |
| | Implicit solvent | *Idem*, additional reaction field contribution | Solvent | [15–19] |
| **United atoms** | Gas phase | | | |
| All atoms | | Classical empirical interaction function | Electronic | [64, 77] |
| United atom (aliphatic groups only) | | *Idem* | Aliphatic H | [64, 77] |
| United atom (all $CH_n$ groups) | | *Idem* | All H bound to C | [64, 77] |
| United atoms (all) | | *Idem* | All H | [64, 77] |
| *Idem* | Explicit solvent | *Idem*, including explicit solvent terms | *Idem* | [12] |
| *Idem* | Implicit solvent | *Idem*, possible corrections in the functional form, parameters, by additional terms or in the equations of motion | Solvent | [12] |
| **Atom groups as 'bead(s)'** E.g. amino acids in proteins represented by one or a few beads | Implicit solvent (or crystal) | Statistics-based interaction function | Side chain | [25] |
| **Molecules** Represented by a sphere, a rod or a disk | Liquid phase (or crystal) | Average intermolecular interaction function | Intramolecular | [257] |

# Kind of force fields
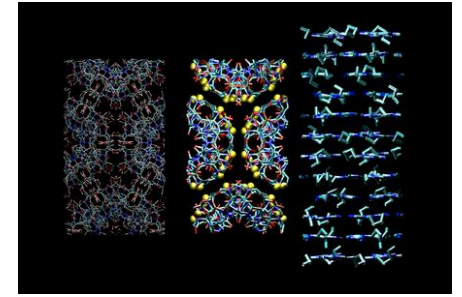
- **Gas-phase force fields**

- **Condensed-phase force fields:**

  A. **Molecules**

  B. Polymers

  C. **Solids**

- Mean-solvent force fields

- Low-resolution force fields

- Hybrid force fields

# Need of a classical approximation



Conformations of a peptide during an MD run

Starting conformation | After 500 pico-second MD | After 1 nano-second MD

# Length and time scale problems

# Simulating nuclei+electrons



- Solving Schrödinger equation for several atoms (see next lectures for DFT)

- Dynamics: Discretizing the equation of motion with timesteps of the order of fractions of femtosecond; expensive wavefunction optimizations at each timestep

- Explicit solvent effects

# More simply... classical approach

- Electrons are treated implicitly: only nuclei are considered

- Classical equations of motion govern the dynamics (e.g., Newton's equation)

- No solution of the quantum electronic problem at each dynamics step

- Electron effect is embedded in the interaction: example, a C-C bond

- This is valid only under some assumptions

# Possible interaction terms (from HvG97, p. 37)

Table 2 *n-Body interaction terms found in common force fields*

| (n) | Subset | Type | Term |
|---|---|---|---|
| 1 | All atoms | P | Kinetic energy |
| | Charged atoms | P | Interaction with an external electric field |
| | Surface atoms | P | Stochastic/frictional force on a macromolecule |
| | Listed or all atoms | U | Atomic positional restraining |
| 2 | All-atom pairs | P | Pairwise nonbonded interaction (point charges, point charge/point dipole etc., van der Waals, solvent accessible surface area interaction) |
| | Bonded atoms | P | Covalent bond |
| | H-bonded atoms | P | H-bonding interaction (acceptor–donor) |
| | Listed atom pairs | U | Distance restraining |
| 3 | All-atom triples | P | Triple nonbonded interactions (expensive, seldom used) |
| | Atoms in bond angle | P | Covalent bond-angle bending |
| | Pairs of bond | P | Bond–bond cross-term |
| | Bond in angle | P | Bond–angle cross-term |
| 4 | Atoms in dihedrals | P | Torsional interaction, improper dihedral interaction |
| | H-bonded atoms | P | H-bonding (acceptor-antecedent, acceptor, hydrogen, donor) |
| | Pairs of angle | P | Angle–angle cross-term (around one centre) |
| | Atoms in dihedral | P | Bond–dihedral cross-term (central bond), angle–angle–torsion cross-term |
| | Atoms in dihedral | U | J-value restraining, local elevation |
| $\geq 5$ | Covalent neighbours | P | Other cross-terms among bonds, angles and dihedrals |
| N | All atoms | P | Point polarizability |
| | All atoms | U | Radius of gyration unfolding force |

(n): order of the term, i.e. the number of particles involved in the interaction term, N indicates all atoms; Subset: subset of atoms for which the term is calculated, either from a list or all atoms (pairs, triples, respectively); Type: physical (P) or 'unphysical' (U) term.

# Assumptions underlying empirical interaction functions

*"The only justification of empirical force fields resides in their ability to reproduce and predict a vast amount of experimental results... It is useful to try to understand the reason of the agreement (or the cause of discrepancies) by considering the relationship between the energy terms of the force field and the underlying QM reality"*Important assumptions are:

- Implicit degrees of freedom and the assumption of weak correlation
  *example: fluctuations of implicit electronic degrees of freedom can be neglected at fixed nuclei*

- Energy terms and the assumption of transferability
  *Functionally simple energy terms valid in several "physical situations" and not around a given stable configuration of a molecule*

- Coordinate redundancy and the assumption of transferability
  *There are more than 3N-6 energy terms...*

- Choices made in the averaging process
  *A certain energy term is resulting on averaging different molecular situation (all, for example, with a C-C single bond)*
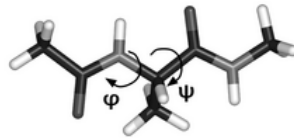
# Preparing the model for a molecule

**1.** Getting the coordinates from some database (e.g., pdb file)
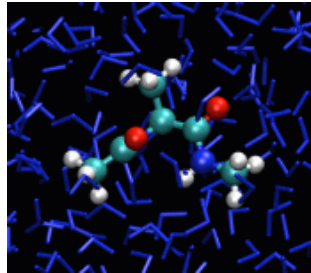
| ATOM | 1 HC1 | 6.274 | 2.310 | −2.248 | 0.00 | 0.00 | H |
|---|---|---|---|---|---|---|---|
| ATOM | 2 CT1 | 5.273 | 1.875 | −2.280 | 0.00 | 0.00 | C |
| ATOM | 3 HC1 | 5.330 | 0.863 | −1.872 | 0.00 | 0.00 | H |
| ATOM | 4 HC1 | 4.958 | 1.814 | −3.324 | 0.00 | 0.00 | H |
| ATOM | 5 C | 4.257 | 2.702 | −1.481 | 0.00 | 0.00 | C |
| ATOM | 6 O | 3.131 | 2.238 | −1.290 | 0.00 | 0.00 | O |
| ATOM | 7 N | 4.520 | 3.006 | 1.025 | 0.00 | 0.00 | N |

**2.** Defining the topology



alanine dipeptide

φ=61.21   ψ=−53.27
initial configuration

φ=−73.30   ψ=65.71
final configuration

```
BONDS
!
!V(bond) = Kb(b − b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb        b0
!
CT1   HC1   340.000000000   1.090000000
CT3   HC2   340.000000000   1.090000000
CT2   H11   340.000000000   1.090000000
N     H     434.000000000   1.010000000
CT4   H12   340.000000000   1.090000000
C     O     570.000000000   1.229000000
C     N     490.000000000   1.335000000
CT1   C     317.000000000   1.522000000
CT2   CT3   310.000000000   1.526000000
CT2   C     317.000000000   1.522000000
N     CT2   337.000000000   1.449000000
N     CT4   337.000000000   1.449000000
```

**3.** Defining the interaction parameters

```
21 !NBOND
   2      3      2      4      1      2     11     12
  11     13     11     14      9     10      7      8
  19     20     19     21     19     22     17     18
   5      6      5      7      2      5     15     16
  15     17      9     11      9     15      7      9
  17     19

36 !NTHETA
   5      7      8      4      2      5      3      2      4
   3      2      5      1      2      3      1      2      4
   1      2      5     15     17     18     13     11     14
  12     11     13     12     11     14     10      9     11
  10      9     15      9     11     12      9     11     13
   9     11     14      8      7      9      7      9     10
  21     19     22     20     19     21     20     19     22
  18     17     19     17     19     20     17     19     21
  17     19     22      6      5      7      5      7      9
   2      5      6      2      5      7     16     15     17
  15     17     19     11      9     15      9     15     16
   9     15     17      7      9     11      7      9     15
```
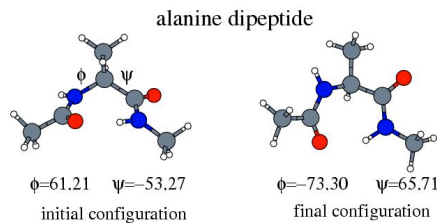
```
45 !NPHI
   6      5      7      8      5      7      9     10
   4      2      5      6      4      2      5      7
   3      2      5      6      3      2      5      7
   2      5      7      8      1      2      5      6
   1      2      5      7     16     15     17     18
  15     17     19     20     15     17     19     21
  15     17     19     22     14     11      9     15
  13     11      9     15     12     11      9     15
  10      9     11     12     10      9     11     13
  10      9     11     14     10      9     15     16
  10      9     15     17      9     15     17     18
   8      7      9     10      8      7      9     11
   8      7      9     15      7      9     11     12
   7      9     11     13      7      9     11     14
  18     17     19     20     18     17     19     21
  18     17     19     22      5      9      7      8
  15     19     17     18      6      5      7      9
   5      7      9     11      5      7      9     15
   2      5      7      9     16     15     17     19
  11      9     15     16     11      9     15     17
   9     15     17     19      7      9     15     16
   7      9     15     17      2      7      5      6
   9     17     15     16
```

# The first step is always the atom type

| No. | Atom type | Description | No. | Atom type | Description |
|---|---|---|---|---|---|
| 1 | c | $sp^2$ carbon in C=O, C=S | 2 | c1 | $sp^1$ carbon |
| 3 | c2 | $sp^2$ carbon, aliphatic | 4 | c3 | $sp^3$ carbon |
| 5 | ca | $sp^2$ carbon, aromatic | 6 | n | $sp^2$ nitrogen in amides |
| 7 | n1 | $sp^1$ nitrogen | 8 | n2 | $sp^2$ nitrogen with 2 subst., real double bonds |
| 9 | n3 | $sp^3$ nitrogen with 3 subst. | 10 | n4 | $sp^3$ nitrogen with 4 subst. |
| 11 | na | $sp^2$ nitrogen with 3 subst. | 12 | nh | amine nitrogen connected to aromatic rings |
| 13 | no | Nitrogen in nitro groups | 14 | o | $sp^2$ oxygen in C=O, COO$^-$ |
| 15 | oh | $sp^3$ oxygen in hydroxyl groups | 16 | os | $sp^3$ oxygen in ethers and esters |
| 17 | s2 | $sp^2$ sulfur (p=S, C=S, etc.) | 18 | sh | $sp^3$ sulfur in thiol groups |
| 19 | ss | $sp^3$ sulfur in —SR and S—S | 20 | s4 | hypervalent sulfur, 3 subst. |
| 21 | s6 | hypervalent sulfur, 4 subst. | 22 | p2 | $sp^2$ phosphorus (C=P, etc.) |
| 23 | p3 | $sp^3$ phosphorus, 3 subst. | 24 | p4 | hypervalent phosphorus, 3 subst. |
| 25 | p5 | hypervalent phosphorus, 4 subst. | 26 | hc | hydrogen on aliphatic carbon |
| 27 | ha | hydrogen on aromatic carbon | 28 | hn | hydrogen on nitrogen |
| 29 | ho | hydrogen on oxygen | 30 | hs | hydrogen on sulfur |
| 31 | hp | hydrogen on phosphorus | 32 | f | any fluorine |
| 33 | cl | any chlorine | 34 | br | any bromine |
| 35 | i | any iodine | | | |
| 36 | cc(cd) | inner $sp^2$ carbon in conjugated ring systems | 37 | ce(cf) | inner $sp^2$ carbon in conjugated chain systems |
| 38 | cp(cq) | bridge aromatic carbon in biphenyl systems | 39 | cu | $sp^2$ carbon in three-membered rings |
| 40 | cv | $sp^2$ carbon in four-membered rings | 41 | cx | $sp^3$ carbon in three-membered rings |
| 42 | cy | $sp^3$ carbon in four-membered rings | 43 | nb | aromatic nitrogen |
| 44 | nc(nd) | inner $sp^2$ nitrogen in conjugated ring systems, 2 subst. | 45 | ne(nf) | inner $sp^2$ nitrogen in conjugated chain systems, 2 subst. |
| 46 | pb | aromatic phosphorus | 47 | pc(pd) | inner $sp^2$ phosphorus in conjugated ring systems, 2 subst. |
| 48 | pe(pf) | inner $sp^2$ phosphorus in conjugated chain systems, 2 subst. | 49 | px | conjugated phosphorus, 3 subst. |
| 50 | py | conjugated phosphorus, 4 subst. | 51 | sx | conjugated sulfur, 3 subst. |
| 52 | sy | conjugated sulfur, 4 subst. | 53 | h1 | hydrogen on aliphatic carbon with 1 electron-withdrawal group |
| 54 | h2 | hydrogen on aliphatic carbon with 2 electron-withdrawal groups | 55 | h3 | hydrogen on aliphatic carbon with 3 electron-withdrawal groups |
| 56 | h4 | hydrogen on aromatic carbon with 1 electron-withdrawal group | 57 | h5 | hydrogen on aromatic carbon with 2 electron-withdrawal groups |

Nos. 1–35 are the basic atom types and Nos. 36–55 are the special atom types.

# Hybridization
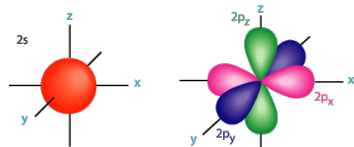
| | Linear | Trigonal planar | Tetrahedral | Trigonal bipyramidal | Octahedral |
|---|---|---|---|---|---|
| Atomic orbitals mixed | one s one p | one s two p | one s three p | one s three p one d | one s three p two d |
| Hybrid orbitals formed | two sp | three sp² | four sp³ | five sp³d | six sp³d² |
| Unhybridized orbitals remaining | two p | one p | none | four d | three d |
| Orientation | | | | | |
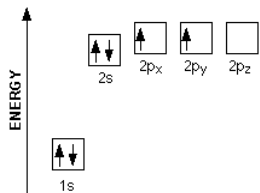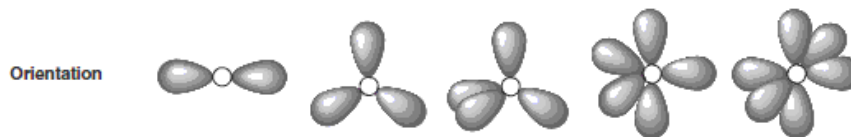
Example: carbon: $1s^2$ **2s² 2p²**

four valence electrons per atom
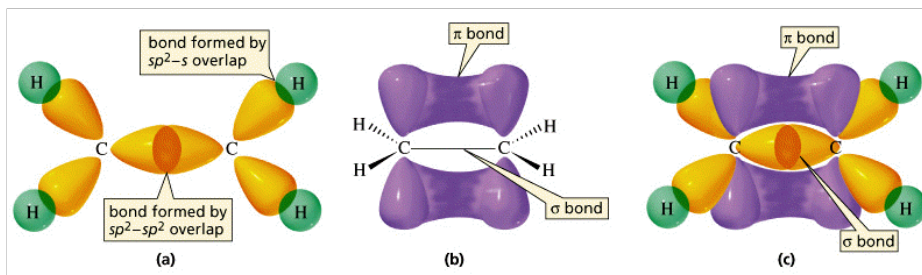
ethylene: double bond
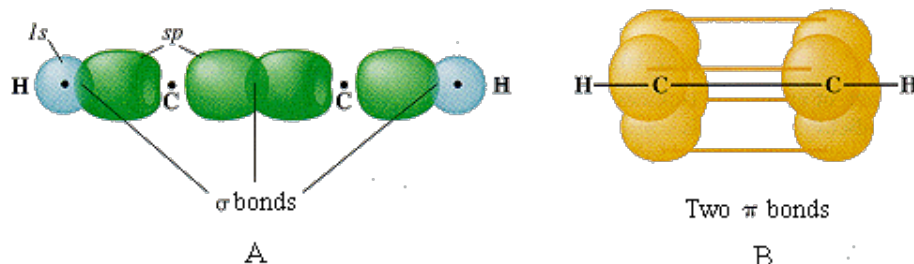
acetylene: triple bond

# Strength of interactions (1 Ha=627.5 kcal/mol ; 1eV= 23 kcal/

### NON-COVALENT

- Hydrophobic: < 10 kcal/mol

- Electrostatic: 1-20 kcal/mol

- Hydrogen bond: 2-30 kcal/mol

- Stacking of aromatic molecules (pi-pi): 0-10 kcal/mol

- van der Waals: 0.1-1 kcal/mol

### COVALENT

- C-O bond:  81 kcal/mol 1.43 Angstrom

- C-C bond: 86 kcal/mol 1.54 Angstrom

- C-H bond: 103 kcal/mol 1.11 Angstrom

- C=C bond: 143 kcal/mol 1.33 Angstrom

# The classical Hamiltonian: AMBER force field

$$V(r) = \boxed{E_{bonded}} + \boxed{E_{nonbonded}}$$

$$= \underbrace{\sum_{bonds} K_b(b - b_0)^2}_{} + \underbrace{\sum_{angles} K_\theta(\theta - \theta_0)^2}_{}$$

$$+ \sum_{dihedrals} (V_n/2)(1 + \cos[n\phi - \delta])$$

$$+ \sum_{nonbij} (A_{ij}/r_{ij}^{12}) - (B_{ij}/r_{ij}^6) + (q_i q_j/r_{ij})$$

# A visual summary from a commercial software (from Accelrys)

**Table 24.** Common potential terms in major forcefields supported by Accelrys

| name | illustrated | form of the term | forcefield[1] |
|---|---|---|---|
| quadratic bond-stretching | | $k(r - r_0)^2$ | AMBER, CHARMm, UFF |
| quartic bond-stretching | | $k_2(r - r_0)^2 + k_3(r - r_0)^3 + k_4(r - r_0)^4$ | CFF |
| Morse bond-stretching | | $k\left[1 - e^{(-\alpha(r_b - r_3^0))}\right]^2$ | CVFF, ESFF |
| quadratic angle-bending | | $k(\theta - \theta_0)^2$ | AMBER, CHARMm, CVFF |
| quartic angle-bending | | $k_2(\theta - \theta_0)^2 + k_3(\theta - \theta_0)^3 + k_4(\theta - \theta_0)^4$ | CFF |
| cosine angle-bending | | various | ESFF, UFF |
| single-cosine torsion | | $k(1 + \cos(n\phi - \phi_0))$ or similar | AMBER, CHARMm, CVFF |
| three-term cosine torsion | | $k_1[1 - \cos(\phi - \phi_{01})] + k_2[1 - \cos(2\phi - \phi_{02})] + k_3[1 - \cos(3\phi - \phi_{03})]$ | CFF |
| cosine-Fourier torsion | | $k(1 \pm \cos n\phi)$ | UFF |
| sin-cos torsion | | $k_\phi\left(\dfrac{\sin^2\theta_1 \sin^2\theta_2}{\sin^2\theta_1^0 \sin^2\theta_2^0} + \text{sign}\dfrac{\sin^\alpha\theta_1 \sin^2\theta_2}{\sin^\alpha\theta_1^0 \sin^\alpha\theta_2^0}\cos[n\phi]\right)$ | ESFF |
| improper cosine out-of-plane | | $k[1 + \cos(n\chi - \chi_0)]$ or similar | AMBER, CVFF, UFF |
| improper quadratic out-of-plane | | $k(\chi - \chi_0)^2$ | CHARMm |
| improper square out-of-plane, imprope | | $k\chi^2$ | CVFF |
| Wilson (or umbrella) out-of-plane | | $k\chi^2$ | CFF, ESFF, UFF |
| | | $k(\cos\chi - \cos\chi_0)^2$ | UFF |
| pyrimid-height out-of-plane | | not used | none |

| Type | Diagram | Equation | Force fields |
|---|---|---|---|
| 6-9 van der Waals | | $\left[\dfrac{*_{ij}}{r_{ij}^{9}} - \dfrac{*_{ij}}{r_{ij}^{6}}\right]$ or $\varepsilon[2(r*/r)^{9} - 3(r*/r)^{6}]$ | CFF, ESFF, UFF |
| 6-12 van der Waals | | $\left[\dfrac{A_{ij}}{r_{ij}^{12}} - \dfrac{B_{ij}}{r_{ij}^{6}}\right]$ or $\varepsilon[(r*/r)^{12} - 2(r*/r)^{6}]$ | AMBER, CHARMm, CVFF |
| electrostatic | | $\dfrac{q_i q_j}{\varepsilon r_{ij}}$ or similar | AMBER, CFF, CHARMm, CVFF, ESFF, UFF |
| quadratic bond-bond | | $k(r - r_0)(r' - r'_0)$ | CFF, CVFF[2] |
| quadratic bond-angle | | $k(r - r_0)(\theta - \theta_0)$ | CFF, CVFF |
| angle-angle | | $k(\theta - \theta_0)(\theta' - \theta'_0)$ | CFF, CVFF |
| end bond-torsion | | $(b - b_0)[k_1\cos\phi + k_2\cos2\phi + k_3\cos3\phi]$ | CFF |
| center bond-torsion | | $(b' - b'_0)[k_1\cos\phi + k_2\cos2\phi + k_3\cos3\phi]$ | CFF |
| angle-torsion | | $(\theta - \theta_0)[k_1\cos\phi + k_2\cos2\phi + k_3\cos3\phi]$ | CFF |
| angle-angle-torsion | | $k\cos\phi(\theta - \theta_0)(\theta' - \theta'_0)$ | CFF, CVFF |
| improper out-of-plane--out-of-plane, improper | | $k[1 - \cos2\chi]^{1/2}[1 - \cos2\chi']^{1/2}$ | CVFF |

# Bond energy terms
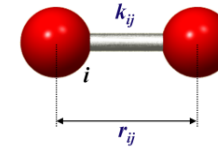
- Assuming that the system will not deviate "too much" from equilibrium, a quadratic expansion is valid

$$U_b(r) = K \left(r - r_{eq}\right)^2$$

- although a Morse expansion would be more faithful to the exact result.

$$U(d) = D_e \left[1 - e^{-a(d-d_0)}\right]^2$$

Representation of bond energy terms

C-H bond
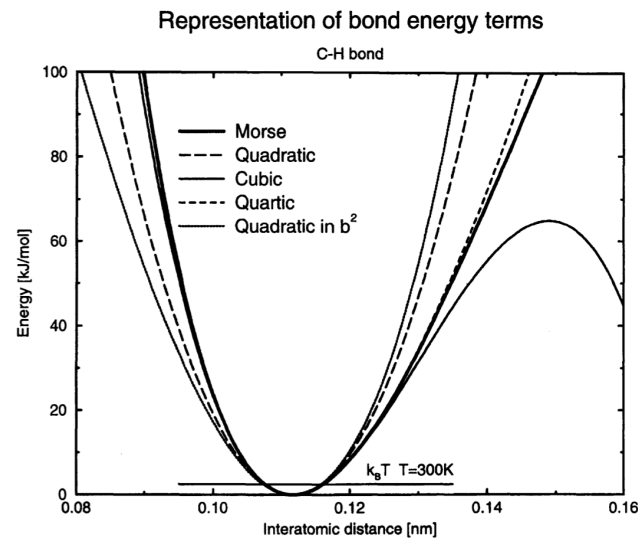
# Typical bond parameter values

| Atom pair | r_eq in Å | K_r in kcal/ (molÅ$^2$) |
|:---:|:---:|:---:|
| C = O | 1.229 | 570 |
| C - C2 | 1.522 | 317 |
| C - N | 1.335 | 490 |
| C2 - N | 1.449 | 337 |
| N - H | 1.01 | 434 |

# Molecular dynamics

Molecular and Materials Modelling (MMM)
HCI D4

**Daniele Passerone**

Empa, Swiss Federal Laboratories for Materials Science and Technology
Ueberlandstrasse 129, 8600 Dübendorf
daniele.passerone@empa.ch

# Acknowledgments for today

- Marcella Iannuzzi (UZH) *Some plots for MD*

- Furio Ercolessi (Udine, Italy) *MD primer*

- Goodstein - Goldstein - Allen/Tildesley books
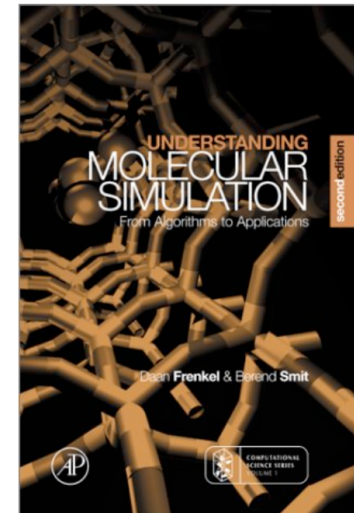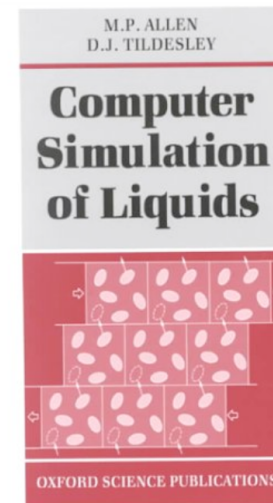
# Hamilton's equations

- In Hamiltionian mechanics, a classical physical system is described by a set of **canonical coordinates** (q,p) and a function (Hamiltonian) that represents the total energy of the system. All possible pairs of **(q,p)** represent the **phase space.**

- In a simple onedimensional case,

$$p = mv \; ; \quad H(p,q) = T(p) + U(q) = \frac{p^2}{2m} + U(q)$$

Hamilton's equations are a system of two first-order differential equations governing motion.

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m}$$

$$\dot{p} = -\frac{\partial H}{\partial q} = -\frac{\partial U}{\partial q} = f(q)$$

(a)

(b)
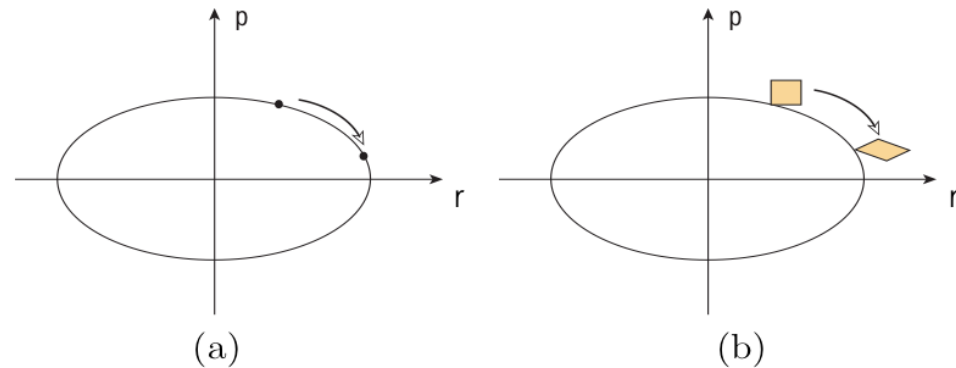
The motion of a point in phase space is confined to a shell with constant energy.

The area of a phase space element is conserved.

# Ensembles

- Microcanonical ensemble *(NVE)*

$$P(p,q)dpdq \propto \delta(H(p,q) - H_0)dpdq$$

- Canonical ensemble *(NVT)*

$$P(p,q)dpdq \propto \exp\left(-\frac{H(p,q)}{k_B T}\right) dpdq$$

- Isothermal-isobaric *(NPT)*

$$P(p,q,V)dpdqdV \propto \exp\left(-\frac{H(p,q) + P_{ext}V}{k_B T}\right) dpdqdV$$

Goal of MD: obtaining ensemble averages from trajectory

$$\langle A \rangle = \int dpdq P(p,t) A(p,q) = \int dt A(p(t), q(t))$$

# Hamilton's principle

- The starting point is "Hamilton's principle" in classical mechanics. Such principle considers the entire motion of a system between times $t_1$ and $t_2$.

- **The motion of the system from time $t_1$ to time $t_2$ is such that the line integral (called the action or the action integral)**

$$I = \int_{t_1}^{t_2} \mathcal{L}\,\mathrm{d}t,$$

**where** *L=T-V,* **has a stationary value for the actual path of the motion.**
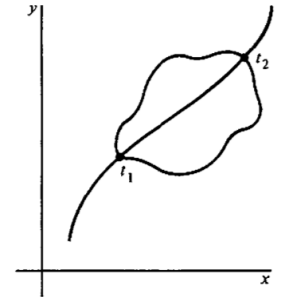
- This corresponds to say that the variation of the line integral *I* for fixed t1 and t2 is zero:

$$\delta I = \delta \int_{t_1}^{t_2} \mathcal{L}(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n, t)\,\mathrm{d}t = 0.$$

# Derivation of Lagrange's equations

Let us consider a function $f$ of many independent variables $yi$ and their derivatives $\dot{y}i$, all function of the parametric variable $x$. Then a variation of the integral $J$,

$$\delta J = \delta \int_1^2 f(y_1(x); y_2(x), \ldots, \dot{y}_1(x); \dot{y}_2(x), \ldots, x) dx$$

is obtained by considering $J$ as a function of parameter $\alpha$ that labels a possible set of curves $y_1(x, \alpha)$. Thus, we may introduce $\alpha$ by setting:

$$
\begin{aligned}
y_1(x, \alpha) &= y_1(x, 0) + \alpha \eta_1(x), \\
y_2(x, \alpha) &= y_2(x, 0) + \alpha \eta_2(x), \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots
\end{aligned}
$$

the variation functions vanishing at the endpoints. The variation becomes:

$$\delta J = \delta \int_1^2 \sum_i \left( \frac{\partial f}{\partial y_i} - \frac{d}{dx} \frac{\partial f}{\partial \dot{y}_i} \right) \delta y_i \, dx$$

and since the $y$ variables are independent, the variations are also independent. Hence, the only possibility to set to zero the variation of $J$ is that the coefficients of $dy_i$ separately vanish.
In the language of Hamilton's principle, this leads to Lagrange's equations:

$$\frac{d}{dx} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \qquad i = 1, \ldots, n$$

and it can be shown that if $V$ does not depend on $t$, the total energy of a system **$T+V$** is conserved.

# Molecular dynamics

- Lagrange equation of motion can be written under some conditions in the more familiar (Newton's law)

$$-\vec{\nabla}_i V = \vec{F}_i = m_i \vec{a}_i$$

- In classical **molecular dynamics** we thus follow the laws of classical mechanics.

- Given an initial set of positions and velocities, the subsequent time evolution is *in principle* completely determined.

- The trajectory is a way to obtain a set of configurations distributed according to some statistical distribution function, or statistical ensemble.

- Example **microcanonical**: in the **HAMILTONIAN formulation** (trivially, *H=E*) we sample the ensemble where the total energy is a constant $H_0$.

$$\mathcal{H} = \sum_i \frac{p_i^2}{2m_i} + V(q_1, \ldots, q_N), \text{ with } p_i = \frac{\partial L}{\partial \dot{q}_i}$$

$$P(p, q)dpdq \propto \delta(H(p, q) - H_0)dpdq$$

# Lagrangian vs. Hamiltonian

- Summarizing and bringing everything together: **Lagrangian equations:**

$$\frac{d}{dx}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \qquad i = 1, \ldots, n$$

- With the usual definitions of $T$ and $V$, these equations lead to

$$-\vec{\nabla}_i V = \vec{f}_i = m_i \ddot{\vec{r}}_i$$

- we can then transform the problem into a system of equations of first order, introducing canonical momenta and the Hamiltonian:

$$p_k = \frac{\partial L}{\partial \dot{q}_k} \qquad H = \sum_k p_k \dot{q}_k - L(\vec{q}, \dot{\vec{q}}) \qquad \dot{p}_k = -\frac{\partial H}{\partial q_k}$$

$$\dot{q}_k = \frac{\partial H}{\partial p_k}$$

- and in cartesian coordinates and simple cases this means that velocities are the momenta divided by the mass and the time derivative of the momentum is the force.

# Conserved quantities (1)

- In certain circumstances, a particular generalized momentum $p$ can be conserved, i.e. its time derivative is 0. The requirement is that the Hamiltonian is independent of the corresponding (generalized) coordinate $q$.

- If potential and $T$ do not depend explicitly on time, $H$ is conserved (**total energy is conserved**)

- For any set of particle, it is possible to choose **six** generalised coordinates that correspond to translations of the center of mass and rotations about the center of mass for the system as the whole. CHANGES IN THE REMAINING *3N-6* coordinates involve relative motion of the particles.

- If the potential depends only on the MAGNITUDE of the distances and without external field, then $H$ and $L$ are independent of these six coordinates. The corresponding **conserved** conjugate momenta are:

$$\textbf{TOTAL LINEAR MOMENTUM} \quad \vec{P} = \sum_i \vec{p_i}$$

$$\textbf{TOTAL ANGULAR MOMENTUM} \quad \vec{L} = \sum_i \vec{r_i} \times \vec{p_i} = \sum_i m_i \vec{r_i} \times \dot{\vec{r_i}}$$

# Time integration algorithm

- The engine of a molecular dynamics program is its time integration algorithm, required to integrate the equation of motion of interacting particles and follow their trajectory.

- Time integration algorithms are based on *finite difference methods*, where time is discretized on a finite grid, the *time step $\Delta$* beign the distance between consecutive points on the grid.

- I know positions, velocities... at time *t,* and the scheme gives the same quantities at time t+$\Delta$.

- All that is connected with *truncation errors* (Taylor's expansion, etc.) and *round-off* errors (precision of the machine, implementation).

- Decreasing $\Delta$ improves *truncation errors,* and *round-off* errors dominate for small $\Delta$.

- One of the most popular algorithms: **Verlet algorithm**

# Derivation of Verlet from discretization of least action principle

We start from the discretization of the action integral:

$$I = \int_{t_1}^{t_2} \mathcal{L}\, dt \simeq \Delta \sum_{n=0}^{N-1} T(q^{(n)}, \dot{q}^{(n)}) - V(q^{(n)})$$

$$\simeq \Delta \sum_{n=0}^{N} \frac{1}{2} m \left( \frac{q^{(n+1)} - q^{(n)}}{\Delta} \right)^2 - V(q^{(n)})$$

and we note that the variation becomes a standard derivation in this case.

$$0 = \frac{\partial I}{\partial q^{(k)}} = \frac{m}{\Delta^2}(q^{(k)} - q^{(k-1)}) - \frac{m}{\Delta^2}(q^{(k+1)} - q^{(k)}) - \frac{\partial V(q^k)}{\partial q^{(k)}}$$

$$q^{(k+1)} = 2q^{(k)} - q^{(k-1)} + F(q^{(k)}) \frac{\Delta^2}{m}$$

This algorithm has an error of $O(\Delta^4)$ and does not include explicit calculation of the velocities.

# Velocity Verlet algorithm

An alternative implementation is the so-called velocity Verlet. Positions, velocities and accelerations at time $t+1$ are obtained from the same quantities at time $t$ in the following way:

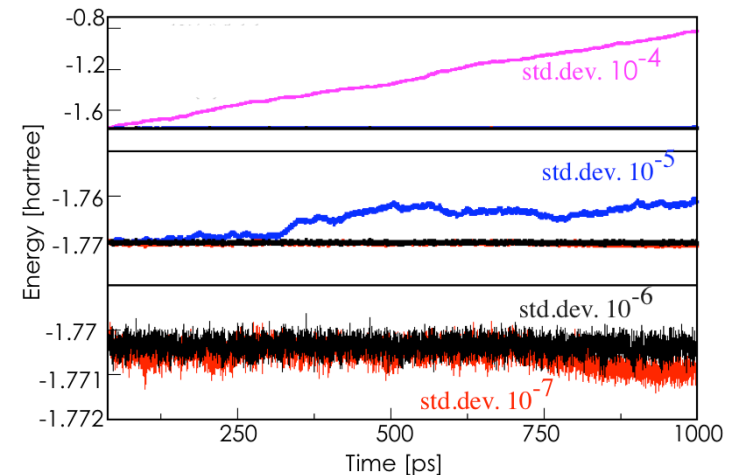$$q^{(k+1)} = q^{(k)} + v^{(k)}\Delta + \frac{1}{2m}F(q^{(k)})\Delta^2$$

$$v^{(k+\frac{1}{2})} = v^{(k)} + \frac{1}{2m}F(q^{(k)})\Delta$$

$$v^{(k+1)} = v^{(k+\frac{1}{2})} + \frac{1}{2m}F(q^{(k+1)})\Delta$$

*Update forces*

Both Verlet form have nice properties:

- Conserve the total energy

- Conserve phase space area

- Are time reversible

- Satisfy symplectic condition



(Marcella Iannuzzi)

4 trajectories of 64 $H_2O$ molecules at 330K, with TIP3P, different deviation of the forces (in Hartree/Bohr) from the correct values.

# Predictor-corrector algorithms

- Another important class of methods to integrate the equations of motion

- The most common of these algorithms are due to Gear, and consist of three steps:

- *Predictor.* From the positions and their time derivatives up to a certain order $q$, all known at time $t$, one "predicts" the same quantities at time $t+\Delta t$, by means of a Taylor expansion. Among these quantities are, of course accelerations **a**.

$$\mathbf{r}^P(t+\delta t) = \mathbf{r}(t) + \delta t\,\mathbf{v}(t) + \tfrac{1}{2}\delta t^2\mathbf{a}(t) + \tfrac{1}{6}\delta t^3\mathbf{b}(t) + \ldots$$
$$\mathbf{v}^P(t+\delta t) = \mathbf{v}(t) + \delta t\,\mathbf{a}(t) + \tfrac{1}{2}\delta t^2\mathbf{b}(t) + \ldots$$
$$\mathbf{a}^P(t+\delta t) = \mathbf{a}(t) + \delta t\,\mathbf{b}(t) + \ldots$$
$$\mathbf{b}^P(t+\delta t) = \mathbf{b}(t) + \ldots$$

- *Force evaluation.* The **force** is computed taking the gradient of the potential at the predicted positions. The resulting acceleration = **f**/m will be in general different from the "predicted acceleration". The difference is an "error signal"

$$\Delta\mathbf{a}(t+\delta t) = \mathbf{a}^c(t+\delta t) - \mathbf{a}^P(t+\delta t)$$

- *Corrector.* This error signal is used to "correct" positions and their derivatives. All corrections are proportional to the error signal, with a coefficient of proportionality being a "magic number" determined to maximize the stability of the algorithm.

$$\mathbf{r}^c(t+\delta t) = \mathbf{r}^P(t+\delta t) + c_0\,\Delta\mathbf{a}(t+\delta t)$$
$$\mathbf{v}^c(t+\delta t) = \mathbf{v}^P(t+\delta t) + c_1\,\Delta\mathbf{a}(t+\delta t)$$
$$\mathbf{a}^c(t+\delta t) = \mathbf{a}^P(t+\delta t) + c_2\,\Delta\mathbf{a}(t+\delta t)$$
$$\mathbf{b}^c(t+\delta t) = \mathbf{b}^P(t+\delta t) + c_3\,\Delta\mathbf{a}(t+\delta t)$$

# (Almost) no algorithm will save us from divergence of trajectories

- A tiny error in the initial coordinate will soon lead to mutual divergence of two trajectories, which in turn will probably both differ from the exact solution of Newton's equation

- This should not disturb since the goal of MD is to have exact solutions for times comparable with the correlation times of interest (time correlation functions) and to generate states sampled from a certain statistical ensemble (example: microcanonical). So energy conservation is important.

- Energy conservation experiences degradation as the time step is increased

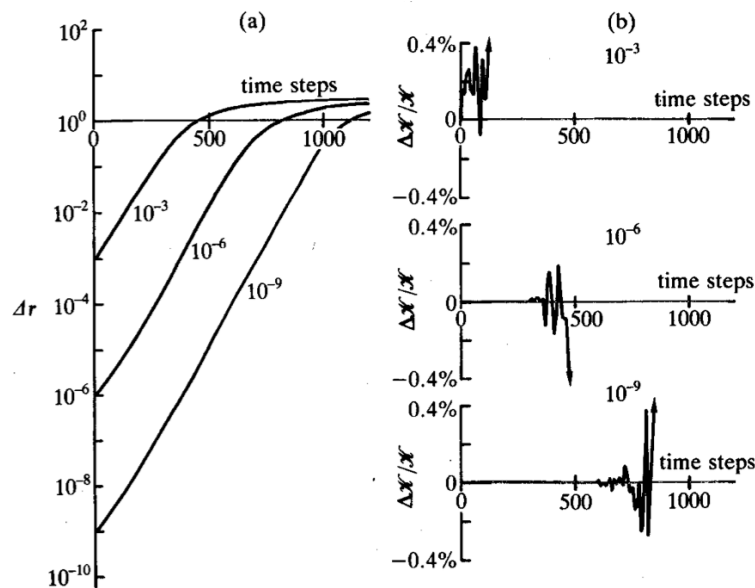- Shorter time-steps at high temperatures and for rugged PES.



Fig. 3.1 The divergence of trajectories in molecular dynamics. Atoms interacting through the potential $v^{RLJ}$ $(r)$, eqn (1.10a), were used, and a dense fluid state was simulated ($\rho^* = 0.6$, $T^* = 1.05, \delta t^* = 0.005$). The curves are labelled with the initial displacement in units of $\sigma$. (a) $\Delta r$ is the phase space separation between perturbed and reference trajectories. (b) $\Delta \mathcal{K}/\mathcal{K}$ is the percentage difference in kinetic energies.

# Shifted potentials

- The truncation of the intermolecular potential at a cutoff introduces some difficulties in defining a consistent potential and force in the MD method

- The function $V(r_{ij})$ used in a simulation contains a discontinuity at $r_{ij}=r_c$ : whenever a pair of atoms/molecules cross this boundary, the total energy will not be conserved. One can shift the potential function by an amount $v_c=V(r_c)$

$$v^S(r_{ij}) = \begin{cases} v(r_{ij})-v_c & r_{ij} \leqslant r_c \\ 0 & r_{ij} > r_c \end{cases}$$

- Still, the force is discontinuous at $r_{ij}=r_c$ : one can modify this shift for having also the force continuous. In this second modification, the form of the potential is modified, and the averages have to be corrected at the end of simulation (Stoddard and Ford 1973)

$$v^{SF}(r_{ij}) = \begin{cases} v(r_{ij})-v_c-\left(\dfrac{dv(r_{ij})}{dr_{ij}}\right)_{r_{ij}=r_c} (r_{ij}-r_c) & r_{ij} \leqslant r_c \\ 0 & r_{ij} > r_c \end{cases}$$
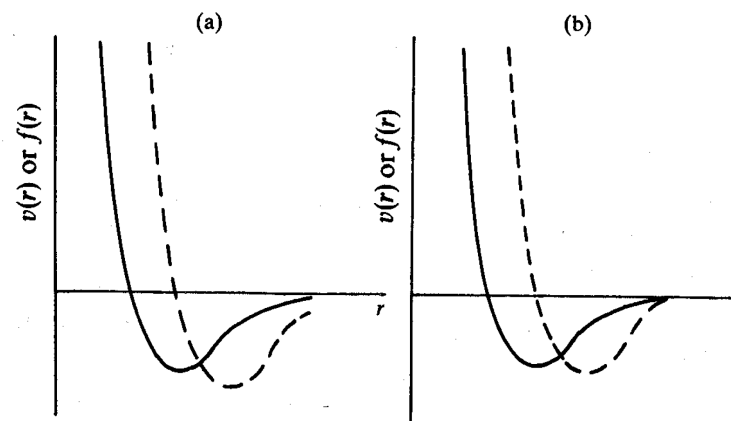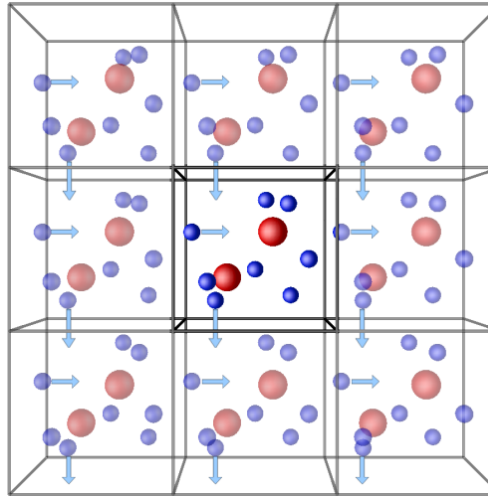


Fig. 5.3 Magnitude of the pair potential (solid line) and force (dashed line) for (a) the Lennard-Jones potential and (b) its shifted-force modification.

# Periodic boundary conditions



- A box is replicated throughout space to form an infinite lattice

- As a molecule moves in the original box, its periodic image in each of the neighbouring boxes moves in exactly the same way

- Thus, as a molecule leaves the central box, one of its images will enter through the opposite face

- There are no surface molecules in the periodic directions

- Only the coordinates of the central box must be stored

# Minimum image convention

●In principle, one must include interactions between all images in the periodic boxes (periodic boundary conditions).

●In case of short range forces, the **minimum image convention (m.i.c.) can be used:**

●Consider molecule 1 to rest at the centre of a region which has the same shape/size of the simulation box

● m.i.c. assumes that molecule 1 interacts only with all the molecules whose centres lie within this region: 1/2 *N(N-1)* interaction terms

●If now we suppose that we use a potential with a finite range $R_c$ : when separated by a distance $d > R_c$, molecules do not interact

●To be consistent with the m.i.c., the box must have size larger than $2R_c$ along each direction

●This gives a reduction by a factor $4\pi R_c/3L^3$ in the number of neighbours.

●Then, particle **1** in the box can interact with images of particle **4,** or with particle **4** itself

●We choose the closest image of particle **4** to particle **1 since *at most one* 1-4** pair (formed by 1 and all possible images of **4**) will interact

●**Only the closest image is candidate to interact: the others certainly do not.**
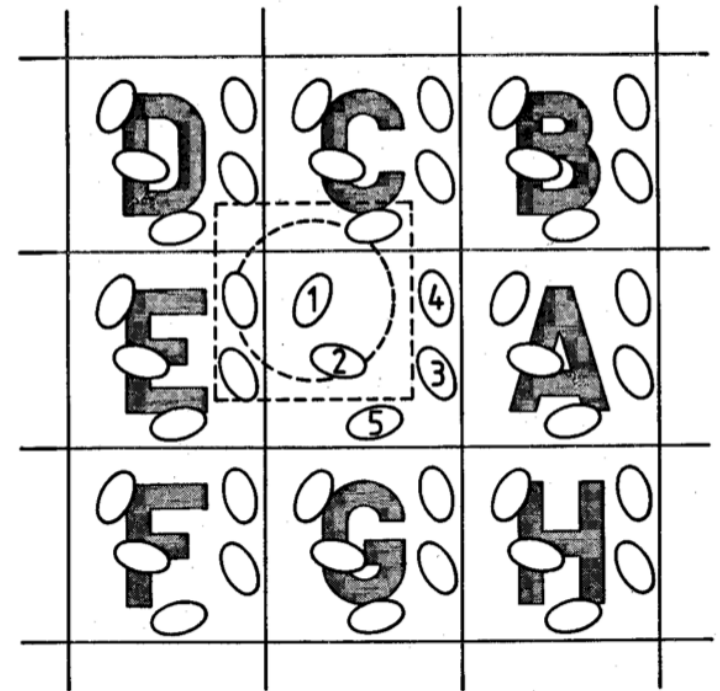
# Applying periodic boundary conditions

- We check whether the **coordinates** are within the box $(-L_x/2, L_x/2)$ ,$(-L_y/2, L_y/2)$ ,$(-L_z/2, L_z/2)$ and we "refold" them (**ANINT:nearest integer**):

```
IF (X(I)>LX/2) X(I)=X(I)-LX
IF (X(I)<-LX/2) X(I)=X(I)+LX
IF (Y(I)>LY/2) Y(I)=Y(I)-LY
IF (Y(I)<-LY/2) Y(I)=Y(I)+LY
IF (Z(I)>LZ/2) Z(I)=Z(I)-LZ
IF (Z(I)<-LZ/2) Z(I)=Z(I)+LZ
```

*OR*

```
X(I)=X(I)-LX*ANINT(X(I)/LX)
Y(I)=Y(I)-LY*ANINT(Y(I)/LY)
Z(I)=Z(I)-LZ*ANINT(Z(I)/LZ)
```

- Similar procedure for the **distance** between two particles, using the minimum image convention:

```
XIJ=X(I)-X(J)
YIJ=Y(I)-Y(J)
ZIJ=Z(I)-Z(J)
IF (XIJ>LX/2) XIJ=XIJ-LX
IF (XIJ<-LX/2) XIJ=XIJ+LX
IF (YIJ>LY/2) YIJ=YIJ-LY
IF (YIJ<-LY/2) YIJ=YIJ+LY
IF (ZIJ>LZ/2) ZIJ=ZIJ-LZ
IF (ZIJ<-LZ/2) ZIJ=ZIJ+LZ
```

*OR*

```
XIJ=X(I)-X(J)-LX*ANINT((X(I)-X(J))/LX)
YIJ=Y(I)-Y(J)-LY*ANINT((Y(I)-Y(J))/LY)
ZIJ=Z(I)-Z(J)-LZ*ANINT((Z(I)-Z(J))/LZ)
```

- Then the square of the distance is computed and compared with the square of cutoff of the potential:

```
RIJSQ=XIJ*XIJ+YIJ*YIJ+ZIJ*ZIJ
IF (RIJSQ < RCUTSQ)THEN
         !COMPUTE FORCES
END IF
```

# Initialization of a simulation

- In a crystal, small random displacements are added to the lattice position, a few percent of the lattice spacing is adequate

- Initial velocities are assigned taking them from a Maxwell distribution at a certain temperature $T$. The linear momentum (and angular as well for molecules) is usually subtracted to have a conserved zero total momentum in the simulation.

- For the rest of the simulation, it depends on the kind of example we want to simulate.

- A trivial way is through a "modified Velocity Verlet algorithm", a simple method that is only used to reach equilibration:

$$v^{(k+\frac{1}{2})} = \sqrt{\frac{T_0}{T^{(k)}}} v^{(k)} + \frac{1}{2} a^{(i)} \Delta$$

- Any physical quantity will approach its equilibrium value exponentially with time with a certain relaxation time.

# How to find the neighbours of an atom

**CELL LIST**: scales like *C\*O(N)*
*only about 16% of the atoms interact*
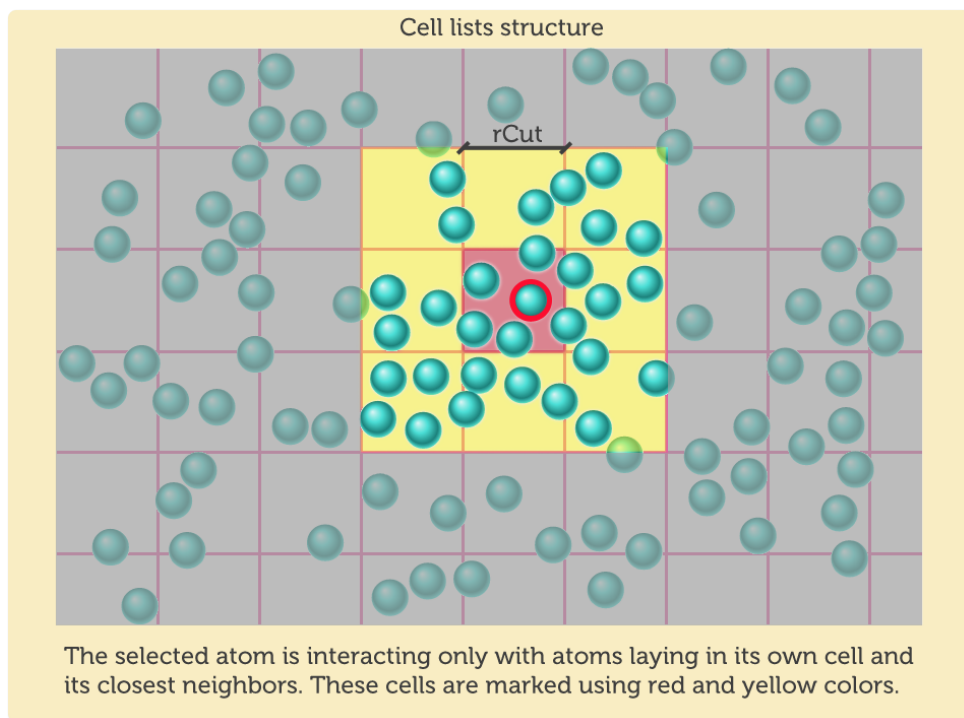
**VERLET LIST**: scales like *O(N²)*
*combined with cell list: O(N)*
*careful skin value: update every several steps*



Cell lists structure

rCut

The selected atom is interacting only with atoms laying in its own cell and its closest neighbors. These cells are marked using red and yellow colors.



Verlet (neighbor) list structure

d    rCut

The selected atom is interacting only with atoms laying close enough to it. Atoms included in its Verlet list lay on the red and yellow areas.
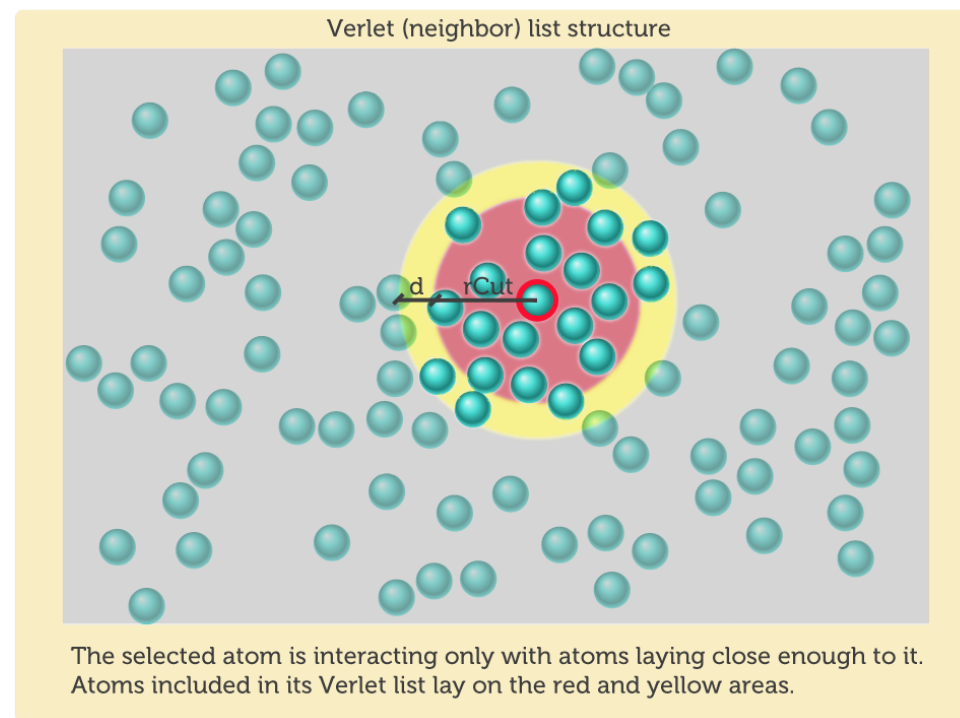
(From: Piotr Janik, Next-Generation Molecular Workbench, http://blog.concord.org/optimizing-short-range-and-long-range-atomic-interactions)

# Verlet neighbour lists
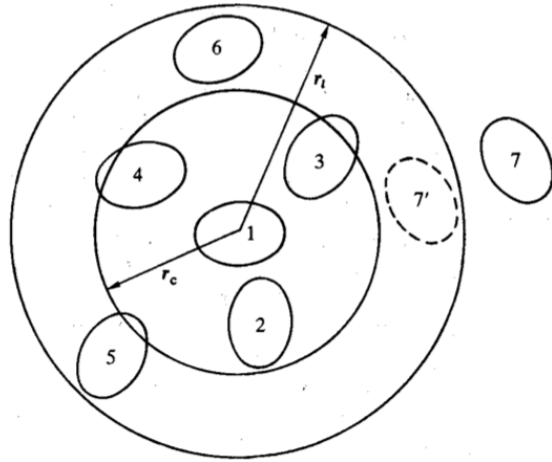


**Table 5.1** *Time saving using a Verlet neighbour list* [*Thompson* 1983].

| List Radius | Update[a] interval | Time[b] N = 256 | N = 500 |
|---|---|---|---|
| no list | — | 3.33 | 10.00 |
| 2.60 | 5.78 | 2.24 | 4.93 |
| 2.70 | 12.50 | 2.17 | 4.55 |
| 2.90 | 26.32 | 2.28 | 4.51 |
| 3.10 | 43.48 | 2.47 | 4.79 |
| 3.43 | 83.33 | 2.89 | — |
| 3.50 | 100.00 | — | 5.86 |

[a]Update interval is the average number of steps between updates. It is essentially independent of system size.
[b]Time is CPU time per step, in seconds. The runs were performed on a PDP 11/70.

- In the inner loops of the MD programs, we consider a molecule $i$ and loop over all molecules $j$ to calculate the minimum image separations.

- If the separation is greater than cutoff, the neighbor is not considered. This method has a scaling of $O(N^2)$

- Verlet (1967) suggested a technique to have list of neighbors updated only every several steps. Namely:

- Choose a sphere with radius $r_1 > r_c$ **(skin)**;

- Build an array **LIST** with all neighbors of all atoms, and a indexing array **POINT**

- Use the list in the force/energy evaluation routine

- The list is reconstructed when the atom with the maximum velocity can in principle cross the whole skin

- Compromise between thick skin (less updates, more neighbors) and thin skin (more updates, less neighbors)

# MDBENCH

● A benchmark for MD simulations developed by Furio Ercolessi in the 80's and maintained until 2000.

```
        MACHINE, COMPILER, COMPILATION OPTIONS          TIME    DATE
--------------------------------------------------------- ------  -------

DEC DS20 (EV6,4MB), OSF1 4.0, f90 V5.2-705 [^] ...............  4.31s  26May99
DEC DS20 (EV6,4MB), OSF1 4.0, f77 V5.2-705 [^] ...............  4.35s  26May99
DEC DS20 6/500(EV6,4MB),f90 V5.2-705, -O5 -tune host -arch host 4.56s+ 16Apr99
[^]  -O5 -arch host -tune host -fast


SGI Octane R12000 (300MHz/2MB SC), f77 7.2.1.3 [^] ...........  4.91s+ 14Apr99
[^] -64 -O3 -mips4 -r10000 -LNO:fis=0:pref=0:cs2=2M:ou=2
    -TENV:X=4 -OPT:O1=0:IEEE_a=3:ro=3:pad=on
    -INLINE:=ON:preempt -TARG:pl=ip30


DEC 21164(EV56)/600MHz/4MB,Unix V4.0 878, f77 V5.0-138-3678F [^] 5.52s  27Apr99
[^] -O5 -fast -unroll 0 -assume noaccuracy_sensitive
    -speculate all -tune host


SGI Origin 2000 R10000 (250MHz/4MB SC), f77 7.1 [^] ..........  5.82s+ 19Feb98
[^] -64 -O3 -mips4 -r10000 -LNO:fis=0:pref=0:cs2=4M:fus=2
    -TENV:X=3 -OPT:O1=0:IEEE_a=3:ro=3:pad=on
    -INLINE:=ON:preempt -TARG:pl=ip27


DEC 21164(EV56)/533MHz, Unix V4.0 878, f77 V5.0-138-3678F [^] .  6.06s  27Apr99
[^] -O5 -fast -unroll 0 -assume noaccuracy_sensitive
    -speculate all -tune host

DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492, [1] ...........  6.38s  11Feb98
DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492, as [1] w/o -g3 . 6.44s  11Feb98
DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492, [2] ...........  6.52s  11Feb98
DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492, as [2] w/o -g3 . 6.66s  11Feb98
DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492, [3] ...........  7.17s  11Feb98
DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492, -O5 ..........  8.28s  11Feb98
DEC 500au (EV56/2MB), Unix 4.0C, f90 V5.0-492 ...............  9.01s  11Feb98
```

```
FPS164, apftn64, opt=3 ........................................  668 s    Jul89
FPS164, apftn64, opt=2 ........................................  688 s    Jul89

Vax 6410, VMS, Vax Fortran 5.4-79, opt .......................  648 s   06Feb91

Vax 8800, VMS, Vax Fortran 5.4, opt ..........................  854 s   14Mar90

Vax 6310, VMS, Vax Fortran 5.4, opt .......................... 1326 s   13Mar90

VaxStation 3100/38, VMS, Vax Fortran 5.4, opt ............... 1378 s   13Mar90

Gould 32/97, Unix, EPCL fort, opt ........................... 1418 s   21Dec88

Vax 6210, VMS, Vax Fortran 4.7, opt ......................... 1873 s

VaxStation 3500, VMS, Vax Fortran 4.7, opt................... 2017 s
VaxStation 3500, Ultrix, Berkeley f77 1.0, opt ............. 2275 s   13Mar90

Microvax II, VMS, Vax Fortran 5.4, opt ...................... 6941 s   14Mar90

Vax 750,  VMS, Vax Fortran 4.7, opt ......................... 6973 s

VaxStation 2000, VMS, Vax Fortran 5.4, opt .................. 7255 s   14Mar90
VaxStation 2000, Ultrix, f77, opt .......................... 8306 s   23Dec88

MacIntosh ClassicII, 16Mhz 68030/25Mhz 68882, Absoft 3.2, -O  8590 s  04Jun96
MacIntosh ClassicII, 16Mhz 68030/25Mhz 68882, Absoft 3.2 ... 10234 s  04Jun96

Apple PowerBook 520, 68LC040, Absoft f77 2.4, -M -H ........  8664 s  12Jun96
Apple PowerBook 520, 68LC040, Absoft f77 2.4, -M -C -B -Q -H  8981 s  12Jun96
```

● On my Macbook pro (without any tweaking) about 0.8 s, on a Nehalem core 0.5 s
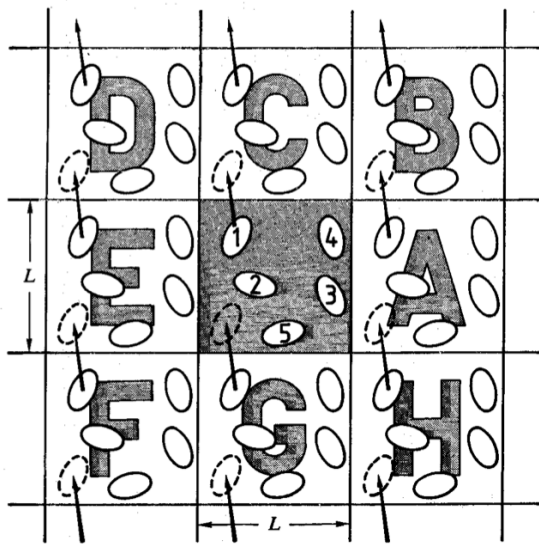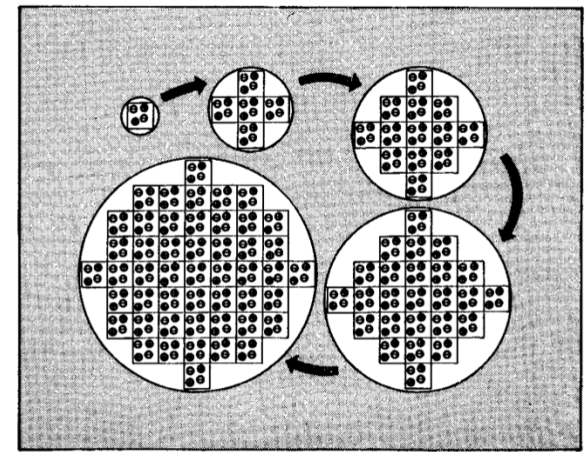
# Long range forces



Fig. 5.7 Building up the sphere of simulation boxes. We illustrate a very small system of two ion pairs for simplicity. The shaded region represents the external dielectric continuum of relative permittivity $\varepsilon_s$.

We need a technique for efficiently summing the interaction between a charged ion and all its periodic images.

Ion 1 in the figure interacts with ion 2, ion 2A, 2B, 2C... and all other images.

The potential energy can be written as

$$V_{\text{Coulomb}} = \frac{1}{2} \sum_{\vec{n}}{}' \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{z_i z_j}{|\vec{r}_{ij} + \vec{n}|} \right),$$

where the sum runs over all lattice points $\vec{n} = (n_x L, n_y L, n_z L)$, the prime indicate that we omit $i=j$ in the same cell for $n=0$.

For long-range potentials, this sum is **conditionally convergent: it depends on the order by which we perform the sum over images. Next lecture: Ewald.**

# Simple statistical quantities to measure

We can consider the instantaneous value of a physical property $A$:

$$A(t) = f(\vec{r}_1(t), \ldots, \vec{r}_n(t), \vec{v}_1(t), \ldots, \vec{v}_n(t))$$

and obtain its average

$$\langle A \rangle = \frac{1}{N_T} \sum_{t=1}^{N_T} A(t)$$

Trivial properties to be averaged are **potential energy, kinetic energy, total energy, temperature, caloric curve (see next lecture about melting)**.

$$\frac{N_f}{2} k_B T = \langle K \rangle \qquad\qquad K = \sum_i \frac{p_i^2}{2m_i}$$

Next time...

# Mean square displacement

- The mean square displacement of atoms in a simulation can be easily computed by its definition:

$$\text{MSD} = \left\langle |\vec{r}(t) - \vec{r}(0)|^2 \right\rangle$$

- with average over all atoms; be careful about refolding.

- In particular this can be used to compute the diffusion coefficient $D$ (in 3D)

$$D = \lim_{t \to \infty} \frac{1}{6t} \left\langle |\vec{r}(t) - \vec{r}(0)|^2 \right\rangle$$

# Pressure

- Measurement of pressure in an MD simulation is based on the Clausius virial function

$$W(\vec{r}_1, \ldots, \vec{r}_N) = \sum_{i=1}^{N} \vec{r}_i \cdot \vec{F}_i^{\text{TOT}}$$

- and an external force (from the container) and an internal one (from interatomic interactions) can be calculated.

- The final formula allows to estimate pressure from MD *(virial equation)*:

$$PV = Nk_BT + \frac{1}{D}\left\langle \sum_{i=1}^{N} \vec{r}_i \cdot \vec{F}_i \right\rangle$$