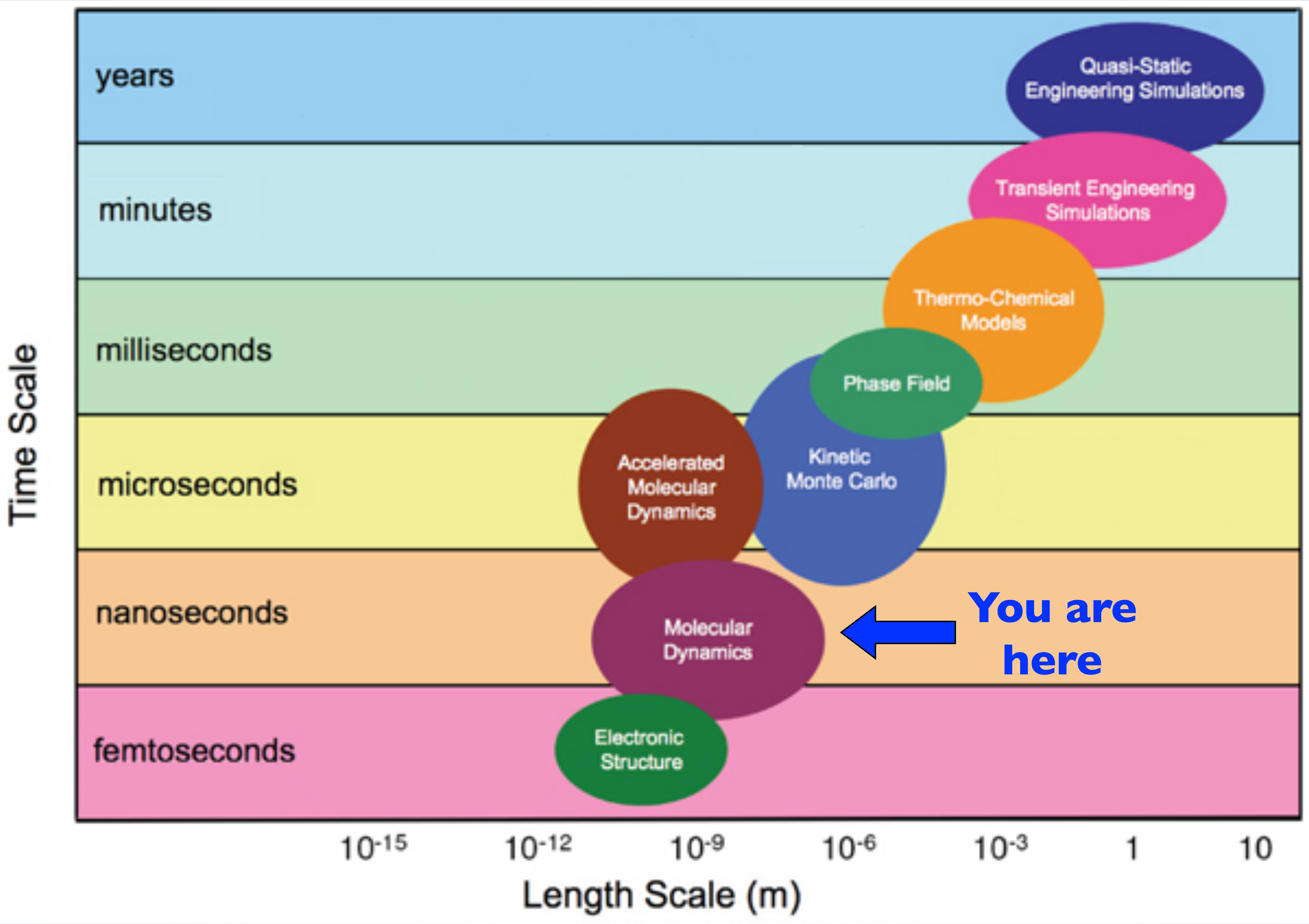
MODULE 2: MOLECULAR DYNAMICS

Principles and Theory

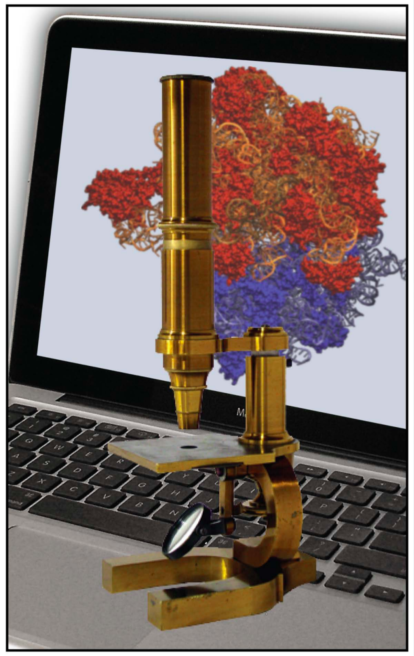
I. Introduction



http://www.lbl.gov/CS/html/exascale4energy/nuclear.html

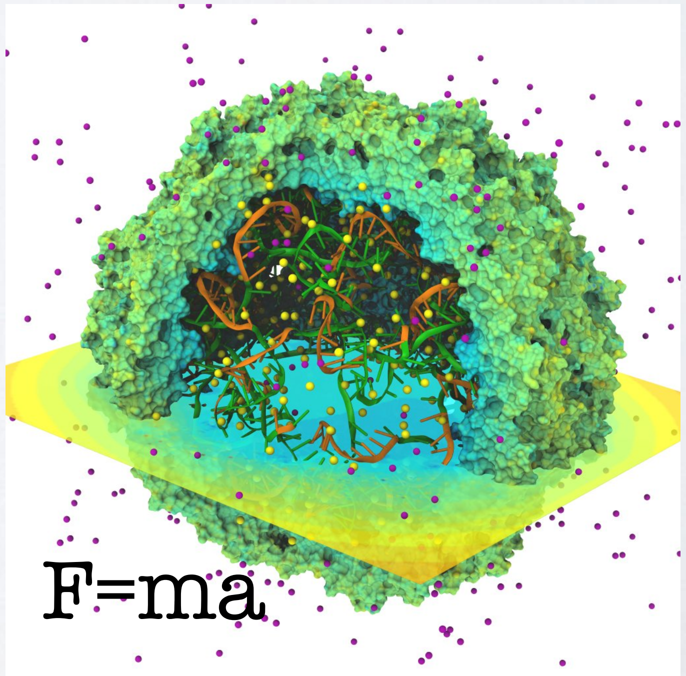
What is molecular dynamics?

- A computational microscope



- An experiment on a computer

- A simulation of the classical mechanics of atoms



http://www.ks.uiuc.edu http://www.123rf.com

Why is it useful?

- By simulating atomic and molecular motions, we can gain **atomistic insight** into molecular structure and kinetics

- Powerful experimental techniques (X-ray diffraction, NMR) can resolve atomic structure, but not dynamics

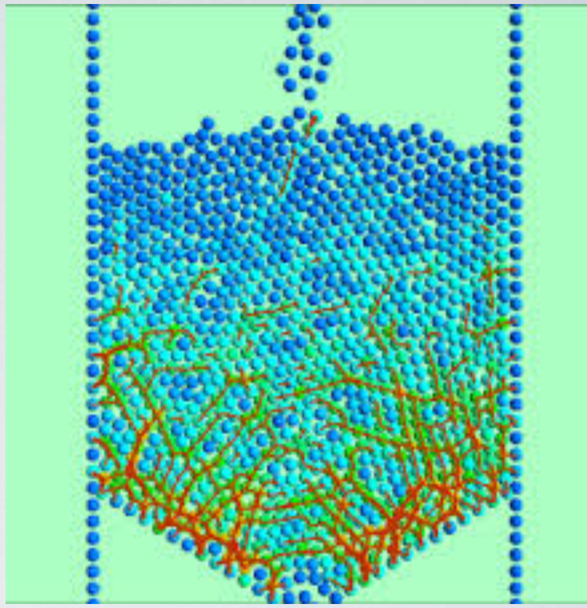
- We can **predict and understand** molecular behavior and compare / interpret experimental observations

- Total control of molecular forces, structure, and conditions

- In principle, it can furnish **all** classical thermodynamics about **any** molecular system*

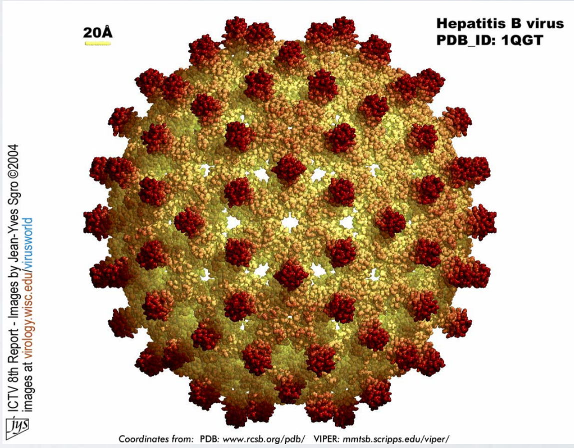
* subject to available force fields and sufficient computational power!

What is it used for?



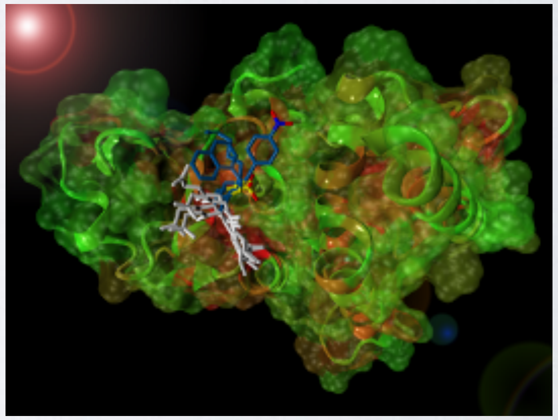
- Materials property prediction

  \- bulk modulus, surface tension, shear viscosity, thermal conductivity, flow, gelation



- Biomolecular modeling

  \- protein folding, viral capsids, cell membranes, ion transport



- Ligand and drug design

  \- docking, interaction, sterics

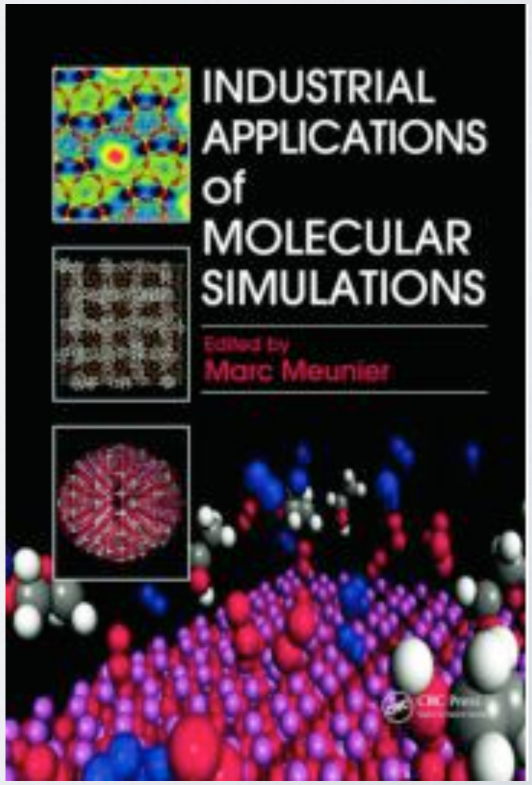
- High-throughput molecular screening

  \- drugs, surfactants, self-assembling materials

http://cssb.biology.gatech.edu/sites/default/files/flhm1.png http://www.virology.wisc.edu/virusworld/viruslist.php?virus=hpb http://www.ph.biu.ac.il/~rapaport/visint/05_examp_grandyn.html

Is it used in industry?

- **YES!**
- Computer power (just) continues to follow Moore's Law, computation gets cheaper every year
- Reliable and validated computational exploration and testing is **much** cheaper and quicker than an R&D lab!
- MD is now a standard tool in pharma, nuclear, chemical, oil, aerospace, electronics, and plastics
- MD is maturing into an "off-the-shelf" tool similar to the emergence of CFD in the 90's

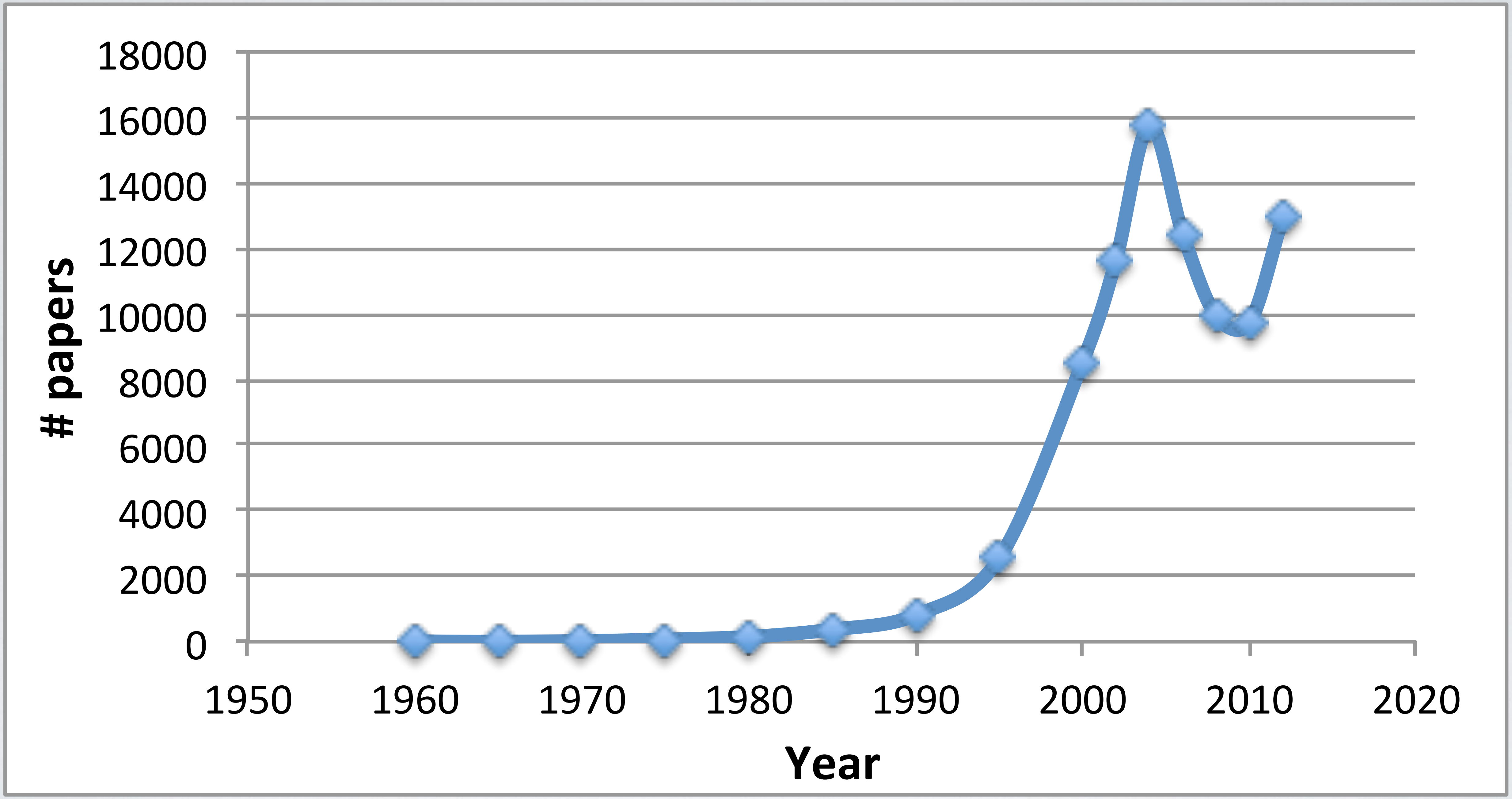


INDUSTRIAL APPLICATIONS of MOLECULAR SIMULATIONS

Edited by Marc Meunier

Academic publishing trends

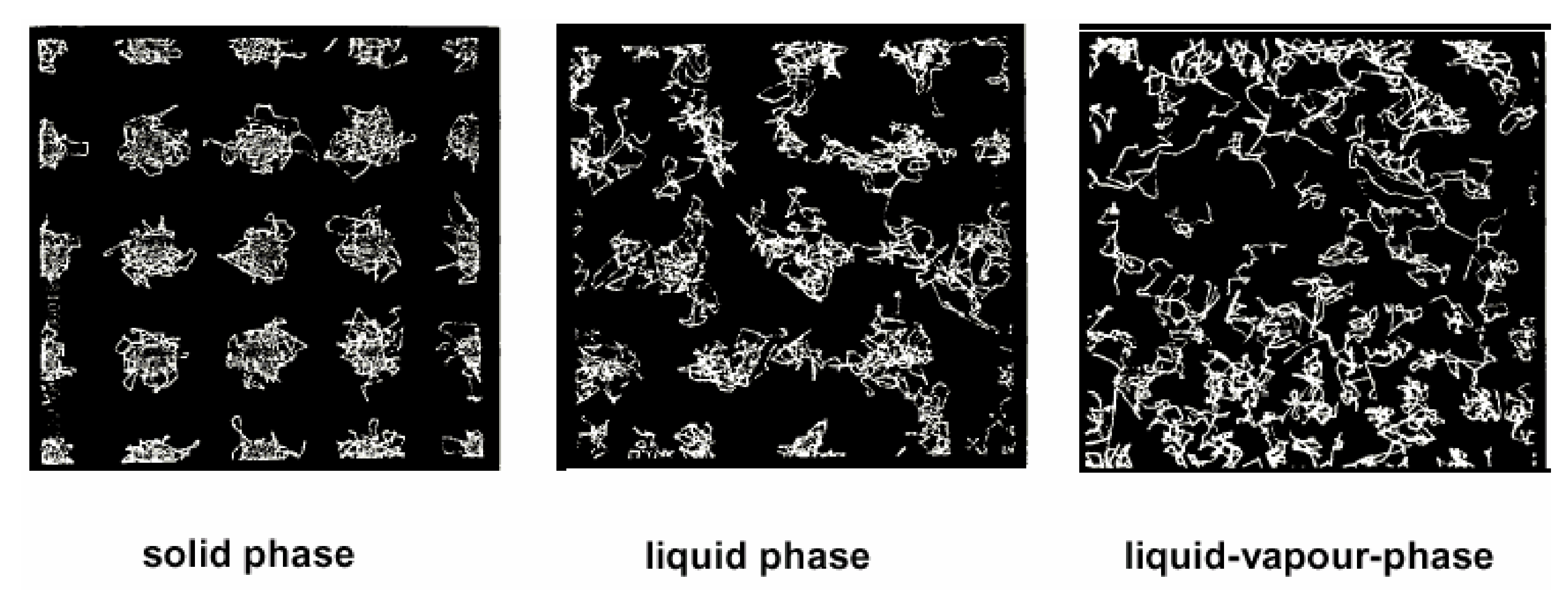
- Scopus abstract/title/keyword search "molecular dynamics"



II. History

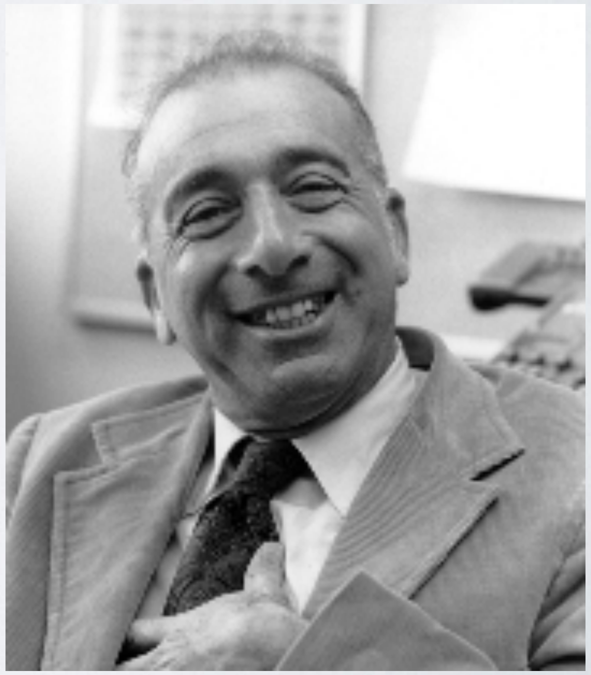
First MD simulation

- Alder & Wainwright (1957) invent molecular dynamics and perform first simulations of the hard sphere fluid


solid phase liquid phase liquid-vapour-phase

- Berni Alder receives Boltzmann Medal (2001) and National Medal of Science (2009) for this work

- Currently Professor Emeritus at UC Davis



Alder, B. J. and Wainwright, T. E. J. Chem. Phys. 27, 1208 (1957)

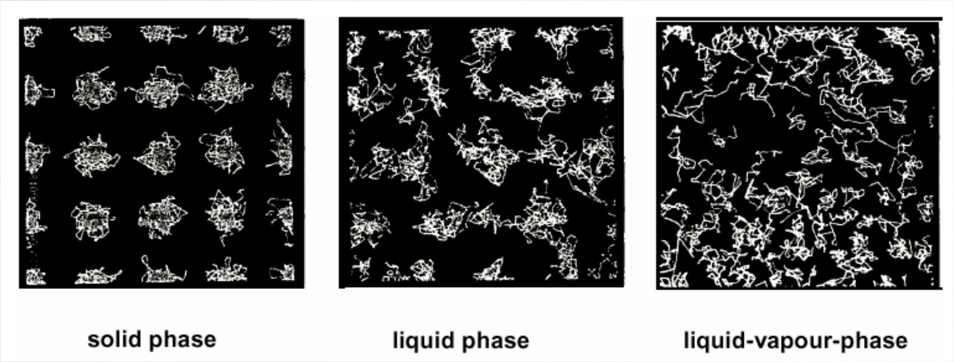
Milestones in MD

1957

Alder & Wainwright

First MD simulation of hard sphere fluid

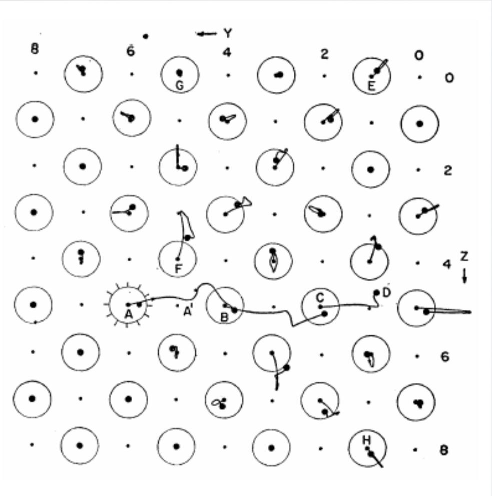
Alder, B.J. and Wainwright, T.E. J. Chem. Phys. 27 1208 (1957)



solid phase

liquid phase

liquid-vapour-phase



1960

Gibson *et al.*

Simulation of Cu radiation damage

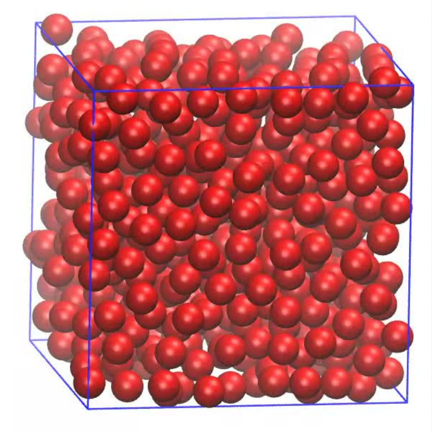
Gibson, J.B., Goland, A.N., Milgram, M., and Vineyard, G.H. Phys. Rev. 120 1229 (1960)

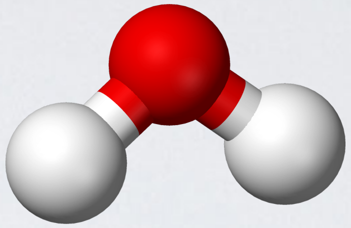
1964

Rahman

First simulation of liquid Ar using realistic potential

Rahman, A. Phys. Rev. A136 405 (1964)





1974

Rahman & Stillinger

First simulation of liquid water

Stillinger, F.H. and Rahman, A. J. Chem. Phys. 60 1545 (1974)

1977

McCammon *et al.*

First protein simulation (BPTI) [8.8ps]

McCammon, J.A., Gelin, B.R., and Karplus, M. Nature 267 585 (1977)

1994

York *et al.*

BPTI hydrated xtal [1ns]

York, D.M., Wlodawer, A., Pedersen, L.G. and Darden, T.A. PNAS 91 18 8715 (1994)

1998

Duan & Kollman

Villin headpiece in water [1μs]

Duan, Y., and Kollman, P.A. Science 282 5389 740 (1998)

2010

Shaw *et al.*

BPTI in water [1ms]

Shaw, D.E. et al. Science 330 341 (2010)

III. Basic Principles

The fundamental idea

- MD simulates atomic motions using classical mechanics
- Running a simulation is like cooking - just follow the recipe!
- Three ingredients:

    1. An initial system configuration $[\vec{r}(t=0), \vec{v}(t=0)]$
    2. Interaction potentials for system $V(\vec{r})$
    3. A way to integrate F=ma

The fundamental idea

- Laplace's Demon / "The Clockwork Universe"

"Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the beings which compose it, if moreover this intelligence were vast enough to submit these data to analysis, it would embrace in the same formula both the movements of the largest bodies in the universe and those of the lightest atom; to it **nothing would be uncertain, and the future as the past would be present to its eyes**."

- Pierre Simon de Laplace (1749-1827)

This is basically molecular dynamics!

But what about quantum effects?

- Classical MD treats atoms* as point particles that move deterministically via Newton's equations of motion

- Is this a valid description of atomic dynamics? **YES.**

(1) Born-Oppenheimer allows us to treat electrons implicitly. Their effect is "baked in" to nuclear interaction potential.

$\tau_{elec} \sim 10^{-18}$ s

$\tau_{nuc} \sim 10^{-15}$ s

Separation of time scales argues for pseudo-equilibrium of electrons with respect to nuclei

---

* or coarse-grained groups of atoms called "united atoms"

But what about quantum effects?

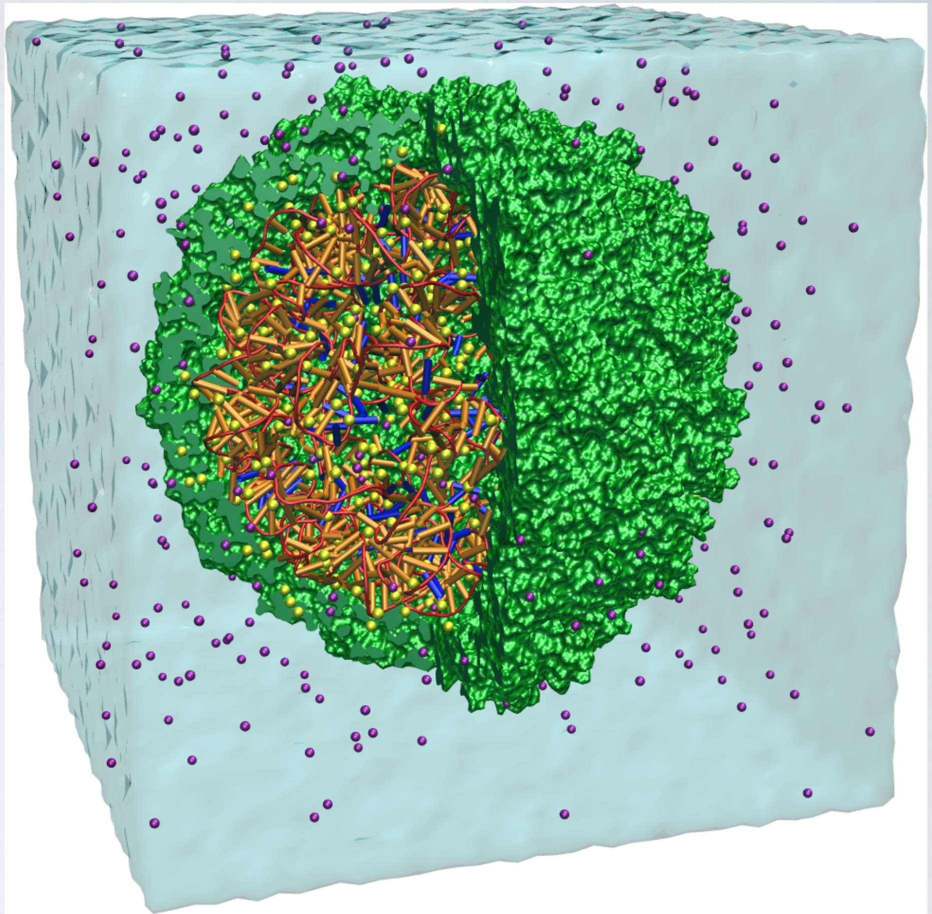
(2) The Schrödinger equation for nuclei replaced by F=ma

de Broglie wavelength: $\Lambda_H \sim 1$Å, $\Lambda_C \sim 0.3$Å
characteristic atomic separation: d ~3Å

For all but lightest atoms d >> Λ, allowing us to treat atoms as point particles and use classical mechanics*

**The quantum behavior of light elements (e.g., H, He, Ne) requires special treatment by fixing bond lengths or lumping light atoms into united atoms*

Ingredient 1: Initial configuration



Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign

- Specification of initial atomic coordinates and velocities
- Classical mechanics is deterministic: **initial state and interaction rules fully specify the system's future***
- Wind up Laplace's clockwork universe and — in principle — a "vast intelligence" could compute the future of the system
- Our intelligence is insufficiently vast — the equations are hard! — and thus **we resort to numerical simulation**

* neglecting numerical integration errors and finite precision (i.e., uncertainty)

www.ks.uiuc.edu

Initializing coordinates

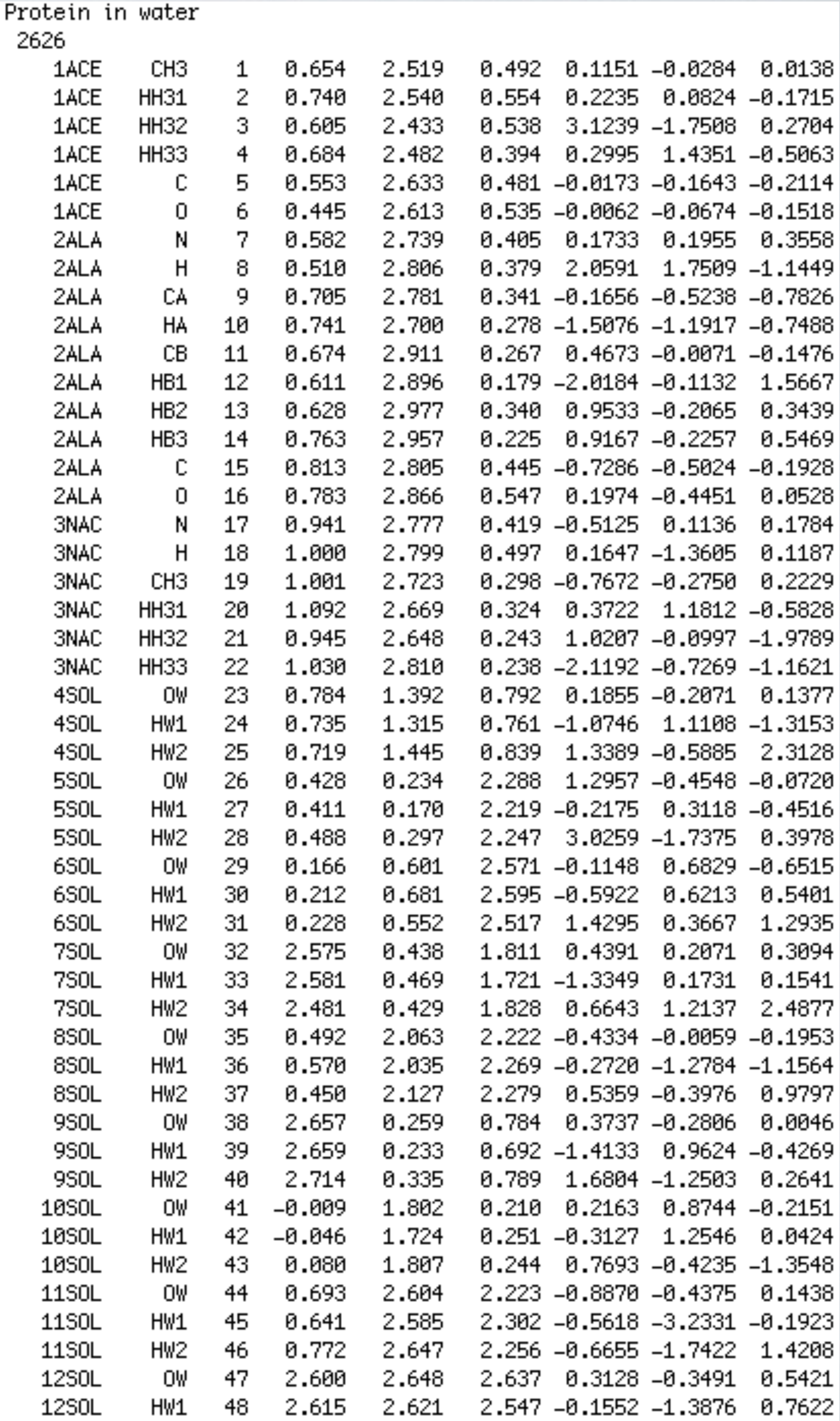
- Initial configurations can be generated "by hand" or short scripts for simple systems (e.g., liquid Ar, bulk Al)

- Software tools for complex systems (e.g., proteins, complex defect structures)

PRODRG (http://davapc1.bioch.dundee.ac.uk/prodrg/)
ATP (http://compbio.biosci.uq.edu.au/atb/)
PyMOl (http://www.pymol.org/)
Chimera (http://www.cgl.ucsf.edu/chimera/)

- Common protein structures are in Protein Data Bank

PDB (www.rcsb.org/pdb)



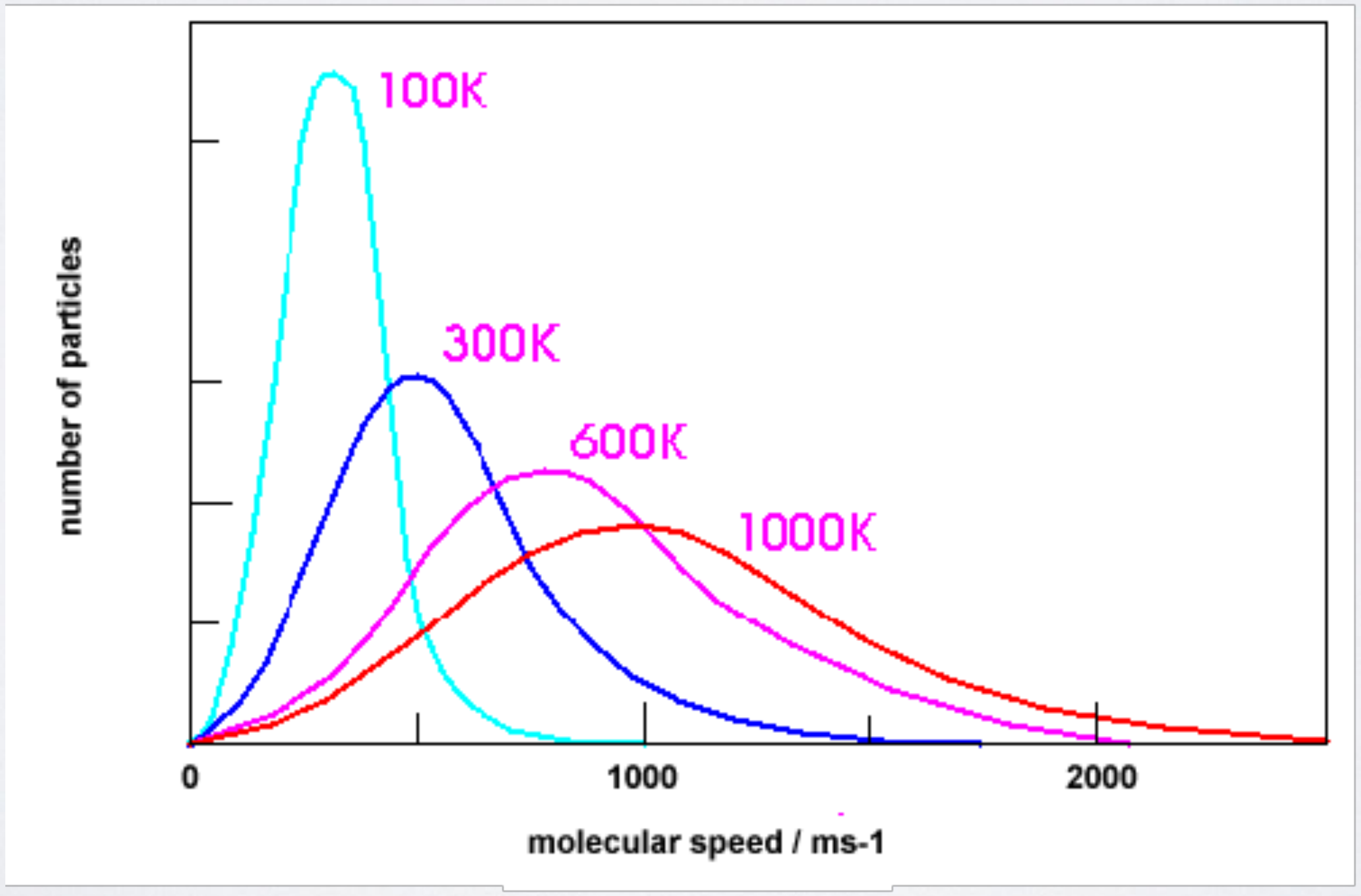
```
Protein in water
 2626
    1ACE    CH3    1   0.654   2.519   0.492  0.1151 -0.0284  0.0138
    1ACE   HH31    2   0.740   2.540   0.554  0.2235  0.0824 -0.1715
    1ACE   HH32    3   0.605   2.433   0.538  3.1239 -1.7508  0.2704
    1ACE   HH33    4   0.684   2.482   0.394  0.2995  1.4351 -0.5063
    1ACE      C    5   0.553   2.633   0.481 -0.0173 -0.1643 -0.2114
    1ACE      O    6   0.445   2.613   0.535 -0.0062 -0.0674 -0.1518
    2ALA      N    7   0.582   2.739   0.405  0.1733  0.1955  0.3558
    2ALA      H    8   0.510   2.806   0.379  2.0591  1.7509 -1.1449
    2ALA     CA    9   0.705   2.781   0.341 -0.1656 -0.5238 -0.7826
    2ALA     HA   10   0.741   2.700   0.278 -1.5076 -1.1917 -0.7488
    2ALA     CB   11   0.674   2.911   0.267  0.4673 -0.0071 -0.1476
    2ALA    HB1   12   0.611   2.896   0.179 -2.0184 -0.1132  1.5667
    2ALA    HB2   13   0.628   2.977   0.340  0.9533 -0.2065  0.3439
    2ALA    HB3   14   0.763   2.957   0.225  0.9167 -0.2257  0.5469
    2ALA      C   15   0.813   2.805   0.445 -0.7286 -0.5024 -0.1928
    2ALA      O   16   0.783   2.866   0.547  0.1974 -0.4451  0.0528
    3NAC      N   17   0.941   2.777   0.419 -0.5125  0.1136  0.1784
    3NAC      H   18   1.000   2.799   0.497  0.1647 -1.3605  0.1187
    3NAC    CH3   19   1.001   2.723   0.298 -0.7672 -0.2750  0.2229
    3NAC   HH31   20   1.092   2.669   0.324  0.3722  1.1812 -0.5828
    3NAC   HH32   21   0.945   2.648   0.243  1.0207 -0.0997 -1.9789
    3NAC   HH33   22   1.030   2.810   0.238 -2.1192 -0.7269 -1.1621
    4SOL     OW   23   0.784   1.392   0.792  0.1855 -0.2071  0.1377
    4SOL    HW1   24   0.735   1.315   0.761 -1.0746  1.1108 -1.3153
    4SOL    HW2   25   0.719   1.445   0.839  1.3389 -0.5885  2.3128
    5SOL     OW   26   0.428   0.234   2.288  1.2957 -0.4548 -0.0720
    5SOL    HW1   27   0.411   0.170   2.219 -0.2175  0.3118 -0.4516
    5SOL    HW2   28   0.488   0.297   2.247  3.0259 -1.7375  0.3978
    6SOL     OW   29   0.166   0.601   2.571 -0.1148  0.6829 -0.6515
    6SOL    HW1   30   0.212   0.681   2.595 -0.5922  0.6213  0.5401
    6SOL    HW2   31   0.228   0.552   2.517  1.4295  0.3667  1.2935
    7SOL     OW   32   2.575   0.438   1.811  0.4391  0.2071  0.3094
    7SOL    HW1   33   2.581   0.469   1.721 -1.3349  0.1731  0.1541
    7SOL    HW2   34   2.481   0.429   1.828  0.6643  1.2137  2.4877
    8SOL     OW   35   0.492   2.063   2.222 -0.4334 -0.0059 -0.1953
    8SOL    HW1   36   0.570   2.035   2.269 -0.2720 -1.2784 -1.1564
    8SOL    HW2   37   0.450   2.127   2.279  0.5359 -0.3976  0.9797
    9SOL     OW   38   2.657   0.259   0.784  0.3737 -0.2806  0.0046
    9SOL    HW1   39   2.659   0.233   0.692 -1.4133  0.9624 -0.4269
    9SOL    HW2   40   2.714   0.335   0.789  1.6804 -1.2503  0.2641
   10SOL     OW   41  -0.009   1.802   0.210  0.2163  0.8744 -0.2151
   10SOL    HW1   42  -0.046   1.724   0.251 -0.3127  1.2546  0.0424
   10SOL    HW2   43   0.080   1.807   0.244  0.7693 -0.4235 -1.3548
   11SOL     OW   44   0.693   2.604   2.223 -0.8870 -0.4375  0.1438
   11SOL    HW1   45   0.641   2.585   2.302 -0.5618 -3.2331 -0.1923
   11SOL    HW2   46   0.772   2.647   2.256 -0.6655 -1.7422  1.4208
   12SOL     OW   47   2.600   2.648   2.637  0.3128 -0.3491  0.5421
   12SOL    HW1   48   2.615   2.621   2.547 -0.1552 -1.3876  0.7622
```

Initializing velocities

- Bad idea to start atoms from rest (absolute zero = 0 K) due to thermal shock upon starting simulation

- Standard approach is to draw velocities randomly from a Maxwell-Boltzmann distribution at the temperature, T

$$f_{\mathbf{v}}(v_x, v_y, v_z) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{m(v_x^2 + v_y^2 + v_z^2)}{2kT}\right]$$



http://ibchem.com/IB/ibfiles/states/sta_img/MB2.gif

Ingredient 2: Interaction potentials

- The net force acting on each atom in the system is a result of its interactions with all other atoms
- These interaction amount to a set of rules known as a **force field** or **interaction potential**
- Accurate, robust, and transferable force fields are critical to perform physically realistic molecular simulations
- Force field development is an academic industry

metals:	EAM (Daw & Baskes), MEAM (Baskes)
biomolecules:	Amber (Kollman, UCSF), GROMOS (U. Groningen), CHARMM (Karplus, Harvard), OPLS (Jorgensen, Yale), MARTINI [coarse grained] (Marrink, U. Groningen)
n-alkanes:	TraPPE (Siepmann, U. Minnesota), MM2 (Allinger, UGA)
water:	SPC (Berendsen), SPC/E (Berendsen), TIPnP(Jorgensen), ST2 (Stillinger & Rahman)
general:	DREIDING (Mayo et al.), DISCOVER(Rappe et al.), UFF (Hagler et al.)

Energy, force, and acceleration

- The potential energy of the system is a complicated function of atomic coordinates (this is why we have to *simulate numerically* rather than *calculate analytically*)

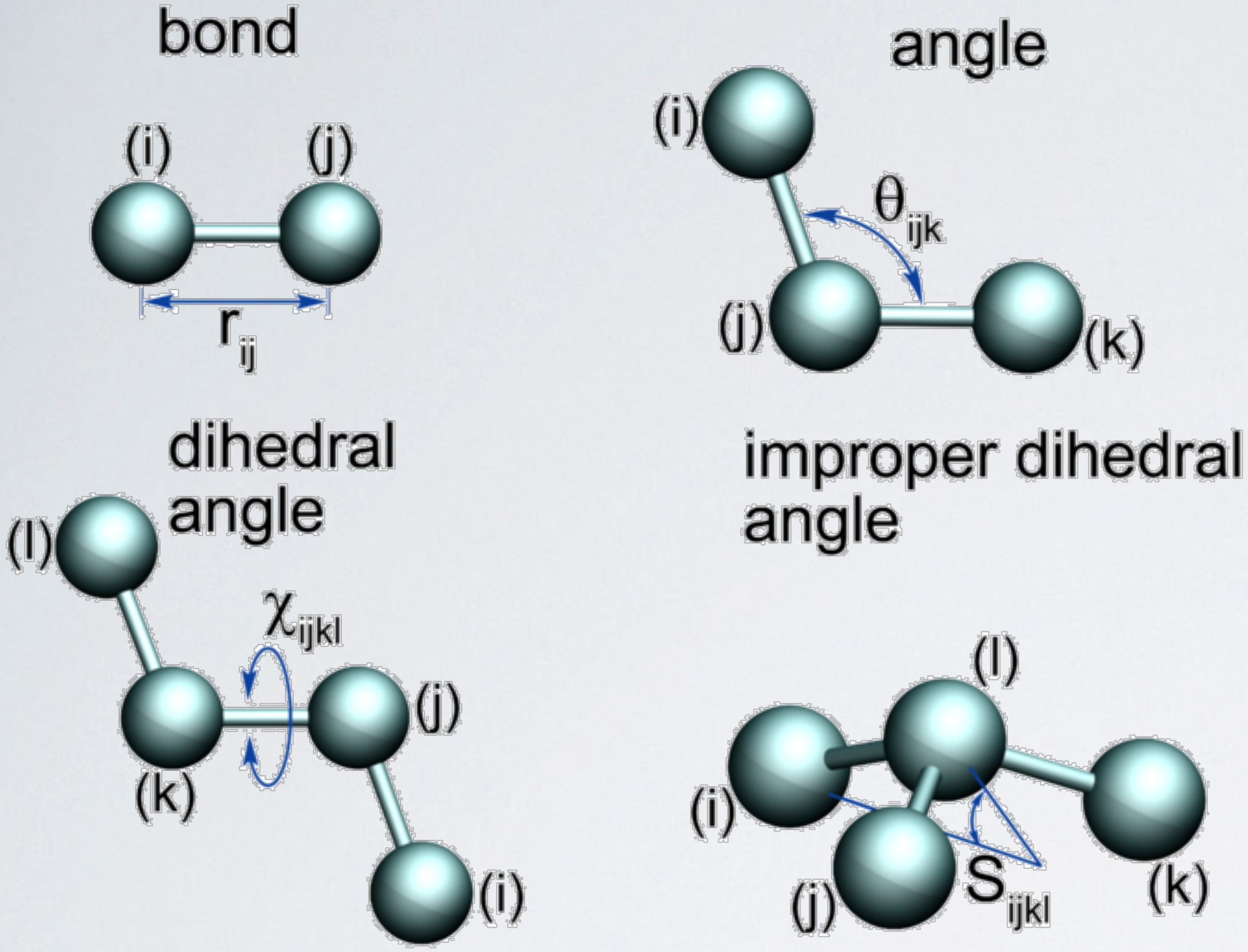
- The net force on atom i is the negative gradient of the potential energy wrt the atomic coordinates

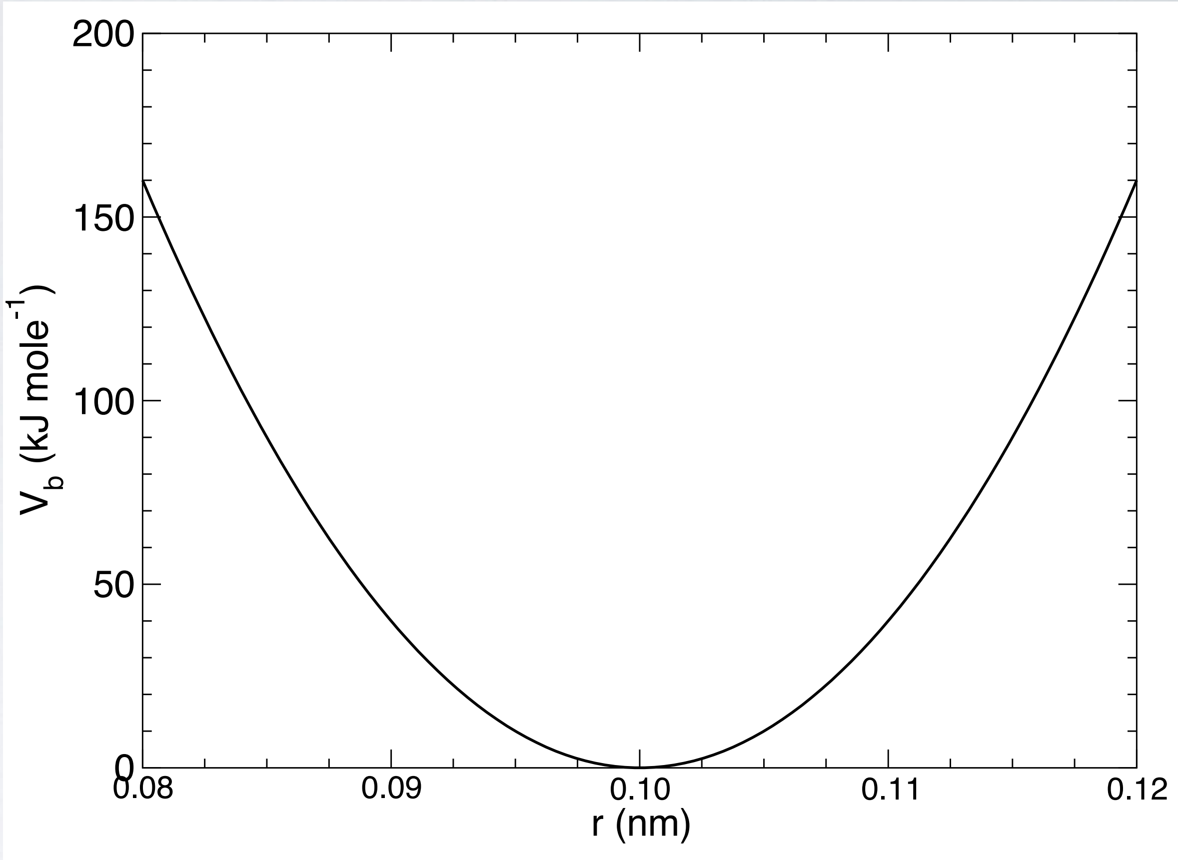
$$F_i = -\nabla_i [V(r_1, r_2, ..., r_N)]$$

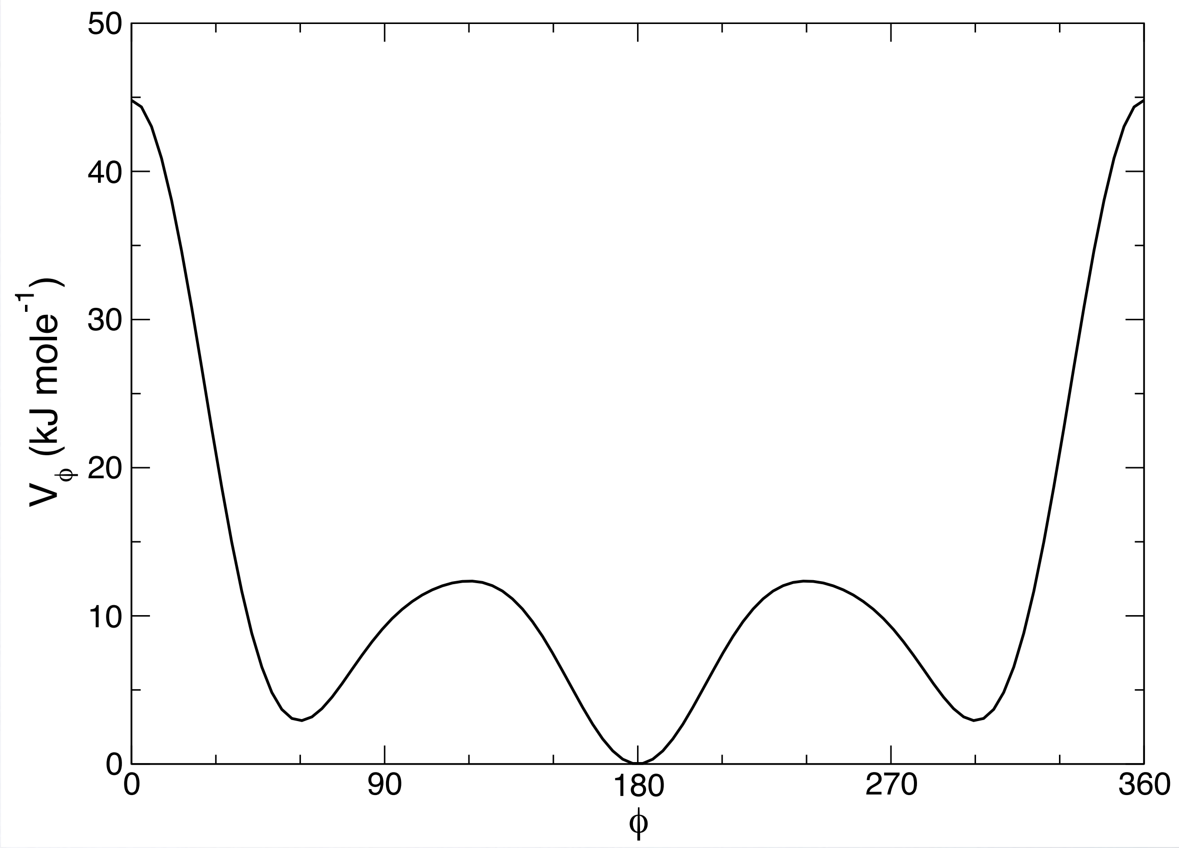
- The potential energy is typically broken into four parts:

$$V(\vec{r}) = V_{bonded} + V_{non-bonded} + V_{restraints} + V_{field}$$

Bonded







$$V_b\left(r_{ij}\right) = \frac{1}{2}k^b_{ij}(r_{ij} - b_{ij})^2$$

$$V_a(\theta_{ijk}) = \frac{1}{2}k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2$$

$$V_{rb}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n(\cos(\psi))^n$$

$$V_{id}(\xi_{ijkl}) = \frac{1}{2}k_\xi(\xi_{ijkl} - \xi_0)^2$$

http://www.mbnexplorer.com/users_guide/users_guide743x.png

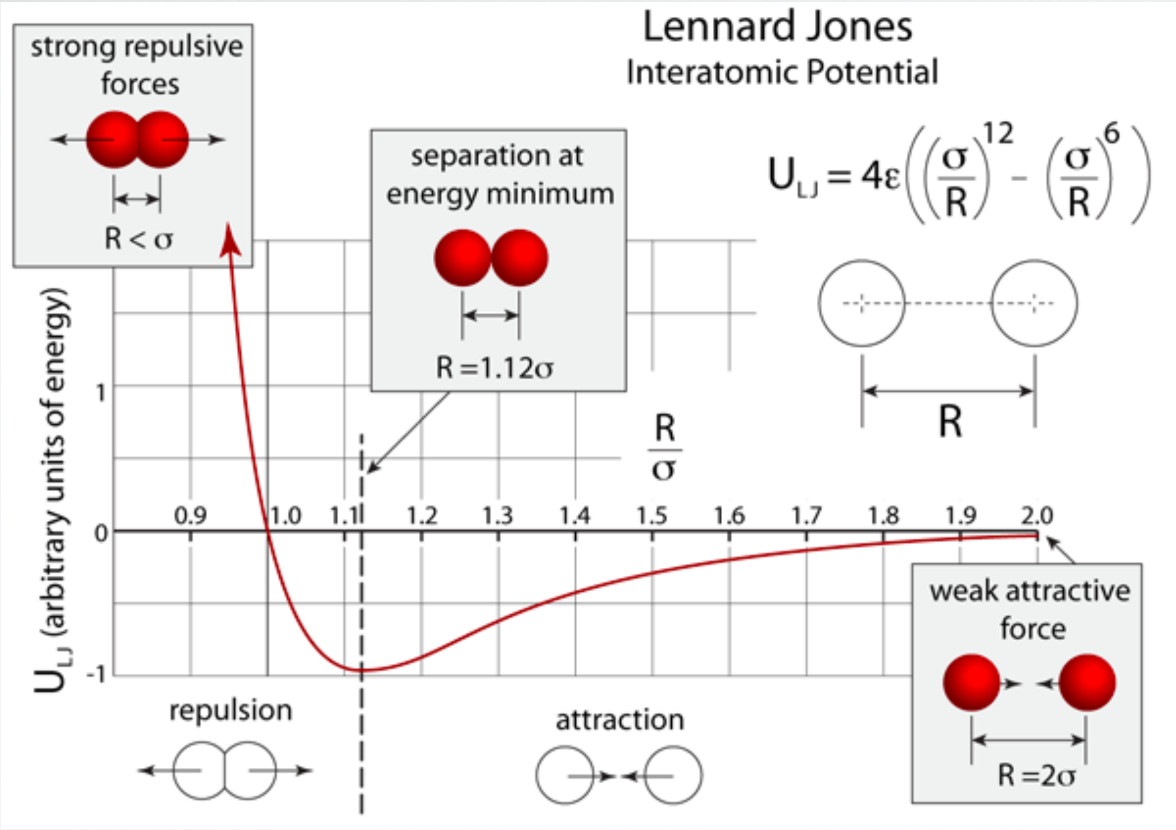
Hess et al. GROMACS Manual v. 4.5.4 (2010)

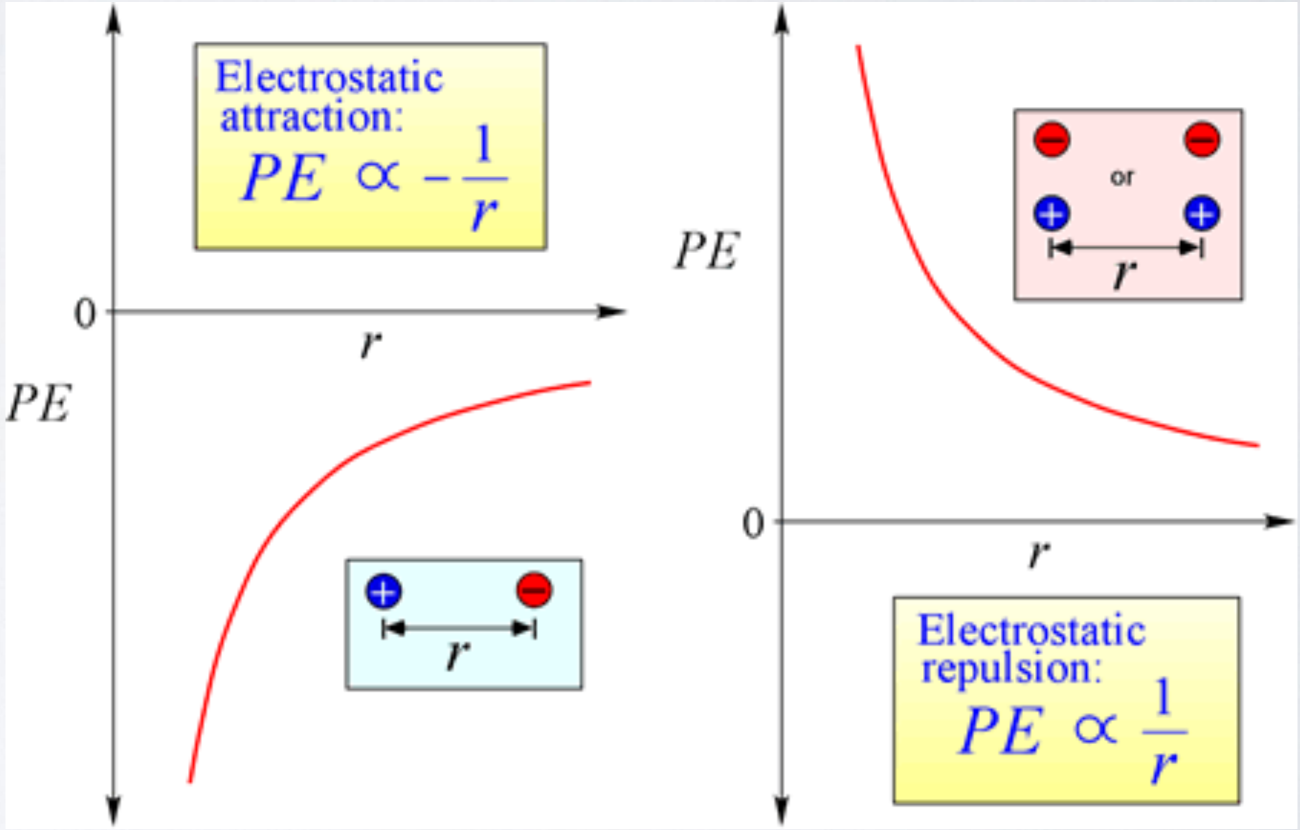
Non-bonded

- Approximate full *n*-body interactions as pairwise additive for simplicity and computational efficiency (cf. (M)EAM)

- van der Waals

- Coulomb





$$V_{LJ}(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]$$

$$V_{Coul}(r_{ij}) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}$$

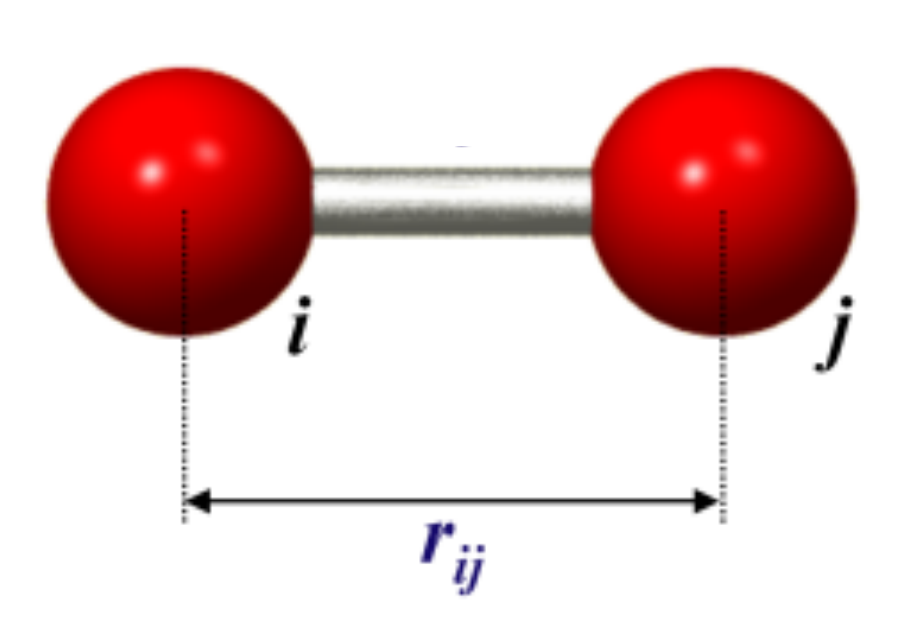
http://atomsinmotion.com/book/chapter5/md

http://guweb2.gonzaga.edu/faculty/cronk/chemistry/images/graph-electrostatic-PE-alt.gif

Restraints

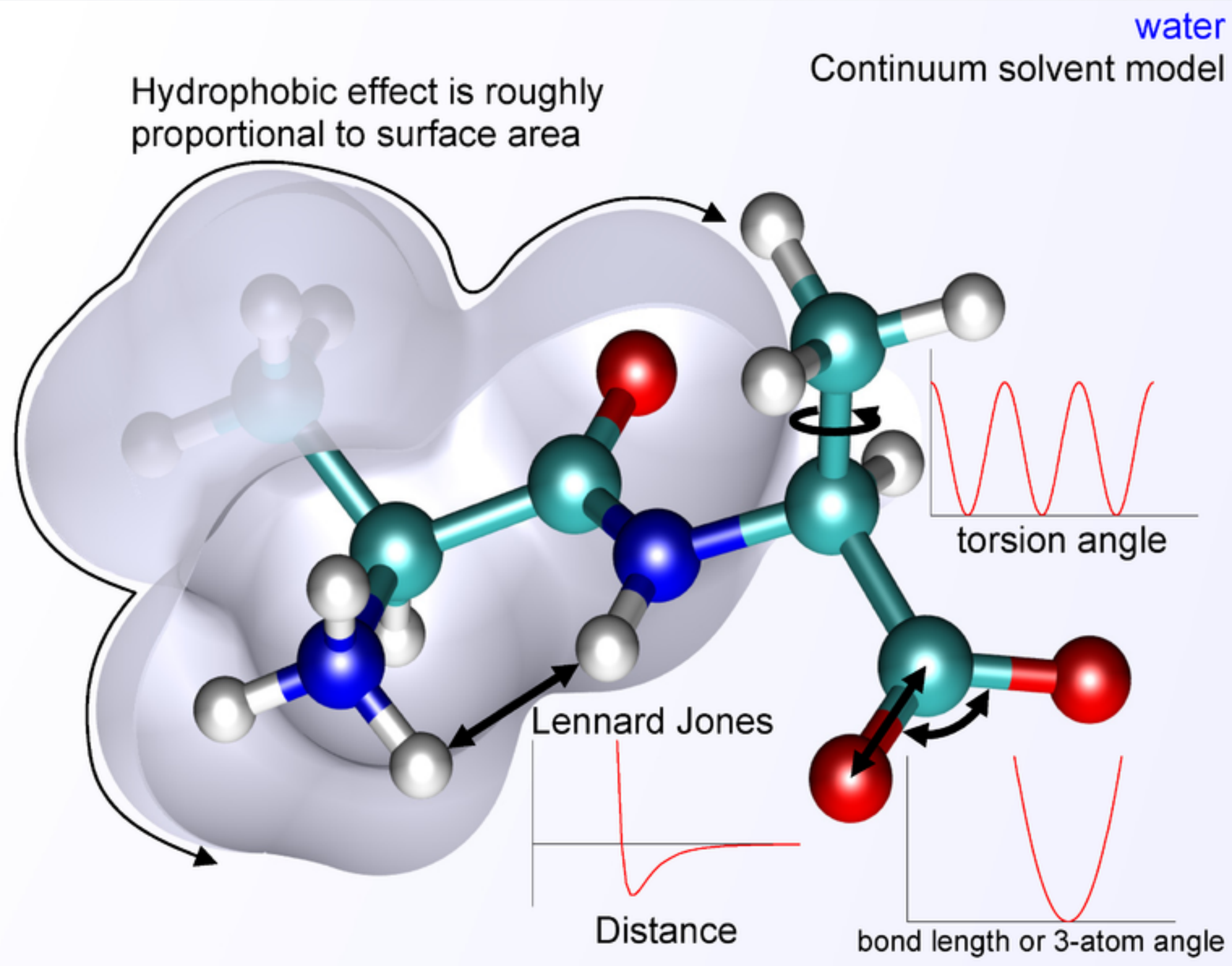
- Restraints can be part of, or supplemental, to a force field

- Many applications, common uses include:
  - fixed bond lengths and angles (esp. for light atoms)
  - artificially immobilize part of the system (e.g., rigid walls or boundary condition)



Fields

- Fields are commonly used to model:
  1. external potentials (e.g., electric, magnetic, flow)
  2. continuum solvation (no explicit solvent molecules)



http://en.wikipedia.org/wiki/File:MM_PEF.png

EAM / MEAM

- Multi-body potential widely used for metallic solids

  EAM - **E**mbedded **A**tom **M**odel
  MEAM - **M**odified **E**mbedded **A**tom **M**odel

- Inherently many-body ⇒ slower than pairwise additive FF
  (2x - EAM, 3-5x - MEAM)

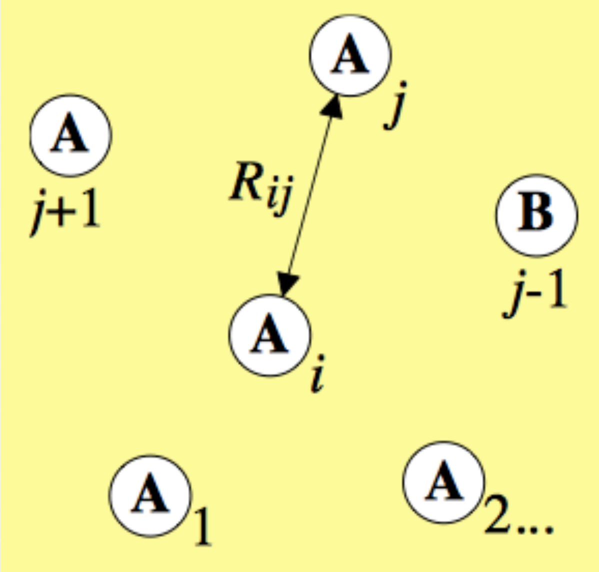
$$E_{\text{total}} = \frac{1}{2}\sum_{i,j}^{N} \Phi_{ij}\left(r_{ij}\right) + \sum_{i}^{N} F_i\left(n_i\right)$$

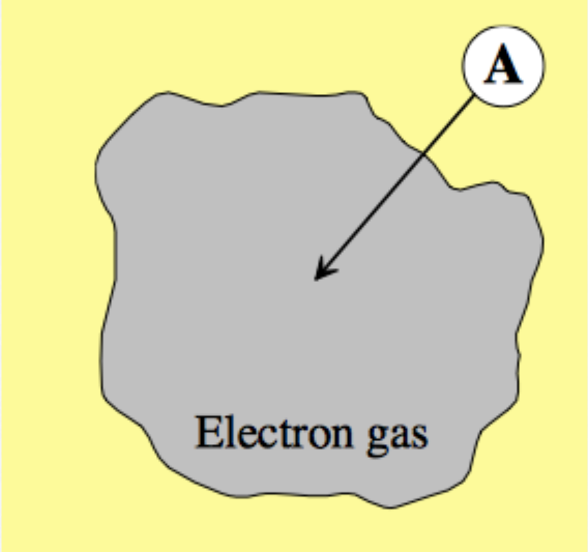
pairwise potential

interatomic separation

local e⁻ density

embedding function





Daw & Baskes PRB 29 12 6443 (1984)   Baskes PRL 59 2666 (1987)   http://potfit.sourceforge.net/wiki/doku.php   http://www.ide.titech.ac.jp/~takahak/pub/ORAN/EAMlecture.pdf

EAM / MEAM

- Local e⁻ density functions

EAM

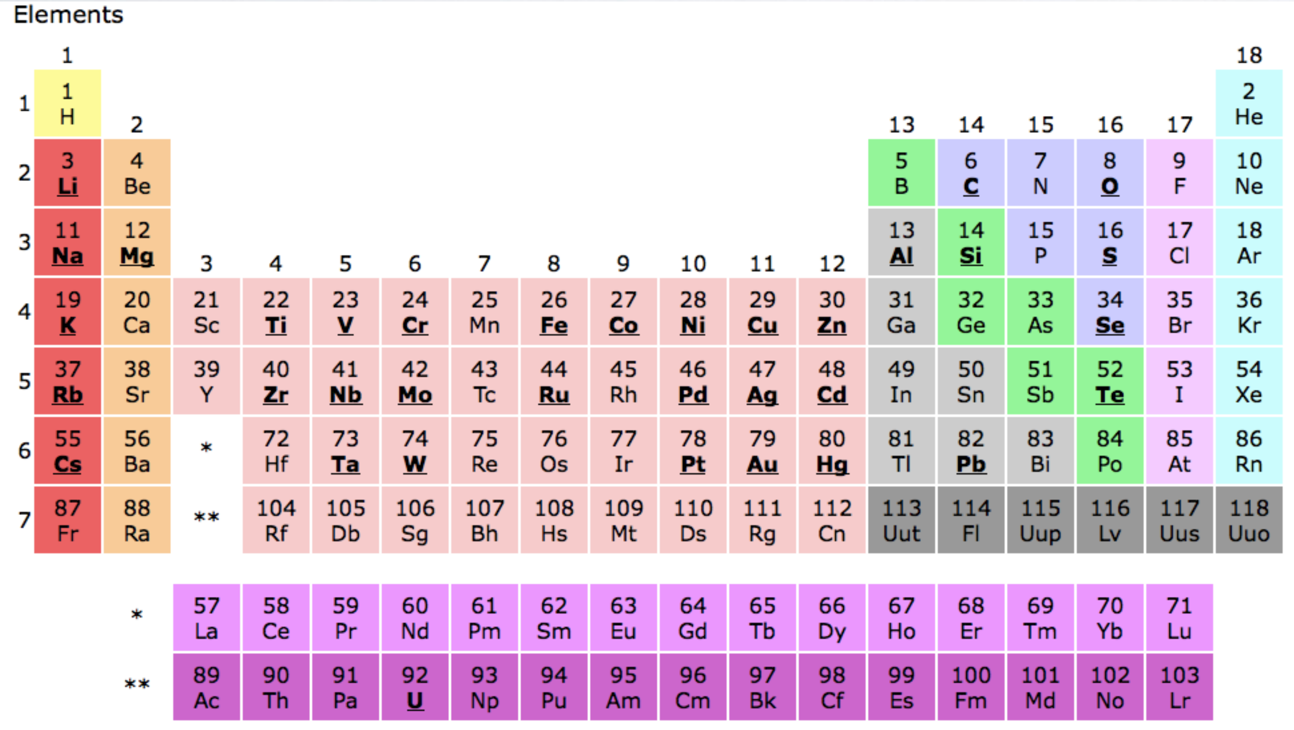
$$n_i = \sum_{j \neq i} \rho_j(r_{ij})$$

MEAM

$$n_i = \sum_{j \neq i}^{N} \rho_j\left(r_{ij}\right) + \frac{1}{2} \sum_{j,k \neq i}^{N} f_{ij}\left(r_{ij}\right) \underset{\substack{\uparrow \\ \text{3-body radial}}}{f_{ik}} \left(r_{ik}\right) \underset{\substack{\uparrow \\ \text{3-body angular}}}{g_i} \left(\cos\theta_{jik}\right)$$

➔ 3-body term in MEAM improves agreement for directional bonding (bcc, hcp, diamond)

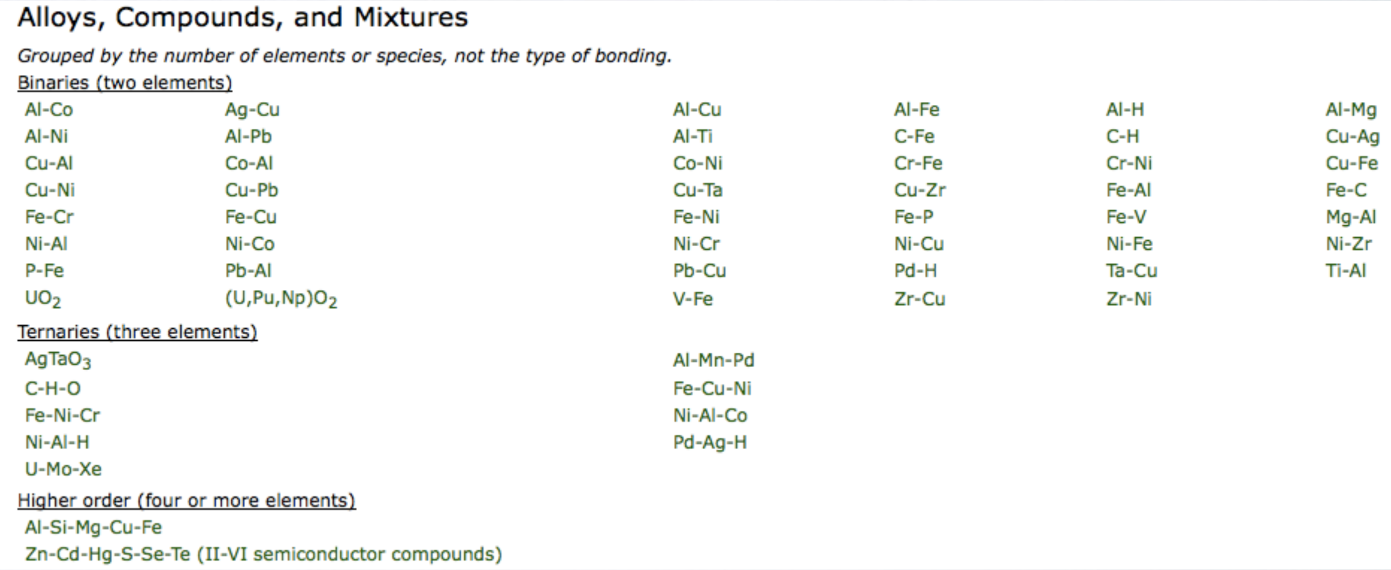
- NIST Interatomic Potentials Repository



Elements

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 **Li**	4 Be											5 B	6 **C**	7 N	8 **O**	9 F	10 Ne
3	11 **Na**	12 **Mg**											13 **Al**	14 **Si**	15 P	16 **S**	17 Cl	18 Ar
4	19 **K**	20 Ca	21 Sc	22 **Ti**	23 **V**	24 **Cr**	25 Mn	26 **Fe**	27 **Co**	28 **Ni**	29 **Cu**	30 **Zn**	31 Ga	32 Ge	33 As	34 **Se**	35 Br	36 Kr
5	37 **Rb**	38 Sr	39 Y	40 **Zr**	41 **Nb**	42 **Mo**	43 Tc	44 **Ru**	45 Rh	46 **Pd**	47 **Ag**	48 **Cd**	49 In	50 Sn	51 Sb	52 **Te**	53 I	54 Xe
6	55 **Cs**	56 Ba	*	72 Hf	73 **Ta**	74 **W**	75 Re	76 Os	77 Ir	78 **Pt**	79 **Au**	80 **Hg**	81 Tl	82 **Pb**	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	**	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
		*	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
		**	89 Ac	90 Th	91 Pa	92 **U**	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

http://www.ctcms.nist.gov/potentials/



Alloys, Compounds, and Mixtures

Grouped by the number of elements or species, not the type of bonding.

Binaries (two elements)

Al-Co
Al-Ni
Cu-Al
Cu-Ni
Fe-Cr
Ni-Al
P-Fe
UO_2
Ag-Cu
Al-Pb
Co-Al
Cu-Pb
Fe-Cu
Ni-Co
Pb-Al
$(U,Pu,Np)O_2$
Al-Cu
Al-Ti
Co-Ni
Cu-Ta
Fe-Ni
Ni-Cr
Pb-Cu
V-Fe
Al-Fe
C-Fe
Cr-Fe
Cu-Zr
Fe-P
Ni-Cu
Pd-H
Zr-Cu
Al-H
C-H
Cr-Ni
Fe-Al
Fe-V
Ni-Fe
Ta-Cu
Zr-Ni
Al-Mg
Cu-Ag
Cu-Fe
Fe-C
Mg-Al
Ni-Zr
Ti-Al

Ternaries (three elements)

$AgTaO_3$
C-H-O
Fe-Ni-Cr
Ni-Al-H
U-Mo-Xe
Al-Mn-Pd
Fe-Cu-Ni
Ni-Al-Co
Pd-Ag-H

Higher order (four or more elements)

Al-Si-Mg-Cu-Fe
Zn-Cd-Hg-S-Se-Te (II-VI semiconductor compounds)

Ingredient 3: Integrators

- [initial atomic coordinates and velocities] + [force field]
  ⇒ entire future (and past!) modeled by **F=ma**
- Analytical solutions for the dynamical evolution cannot be computed for all but the simplest systems (>2 body)
- Solve Newton's equations by numerical integration
  ⇒ computers ideally suited to rapid, repetitive calculations
- Solving by hand would require thousands of years!

Verlet algorithm

- Many possible integration algorithms exist
  (e.g., explicit/implicit Euler, Gear predictor-corrector, nth order Runge-Kutta, Beeman, Newmark-beta)

- The method of choice is the **Verlet algorithm**

  ✓ **fast**
  ✓ **simple**
  ✓ **low-memory**
  ✓ **stable**
  ✓ **time-reversible**
  ✓ **symplectic (phase space volume & E conserving)**

  ✗ **poor accuracy for large time steps (Δt must be small)**

- First recorded use by Delambre in 1791
  Popularized in MD by Loup Verlet in 1967

Verlet algorithm

- Derived from Taylor series:

$$r(t+\delta t) = r(t) + \dot{r}(t)\delta t + \frac{1}{2}\ddot{r}(t)\delta t^2 + \ldots$$

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

$$r(t-\delta t) = r(t) - \dot{r}(t)\delta t + \frac{1}{2}\ddot{r}(t)\delta t^2 + \ldots$$

$$= r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots$$

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 + \mathcal{O}\left(\delta t^4\right)$$

$$v(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} + \mathcal{O}\left(\delta t^2\right)$$

Velocity & leapfrog Verlet

- Original Verlet slightly inconvenient:
  1) need to store two sets of positions
  2) position error is O(dt^4), whereas velocity is O(dt^2)

- Velocity-Verlet

$$x(t+\delta t) = x(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \mathcal{O}\left(\delta t^3\right)$$

$$v(t+\delta t) = v(t) + \frac{a(t)+a(t+\delta t)}{2}\delta t + \mathcal{O}\left(\delta t^3\right)$$

- Leapfrog-Verlet

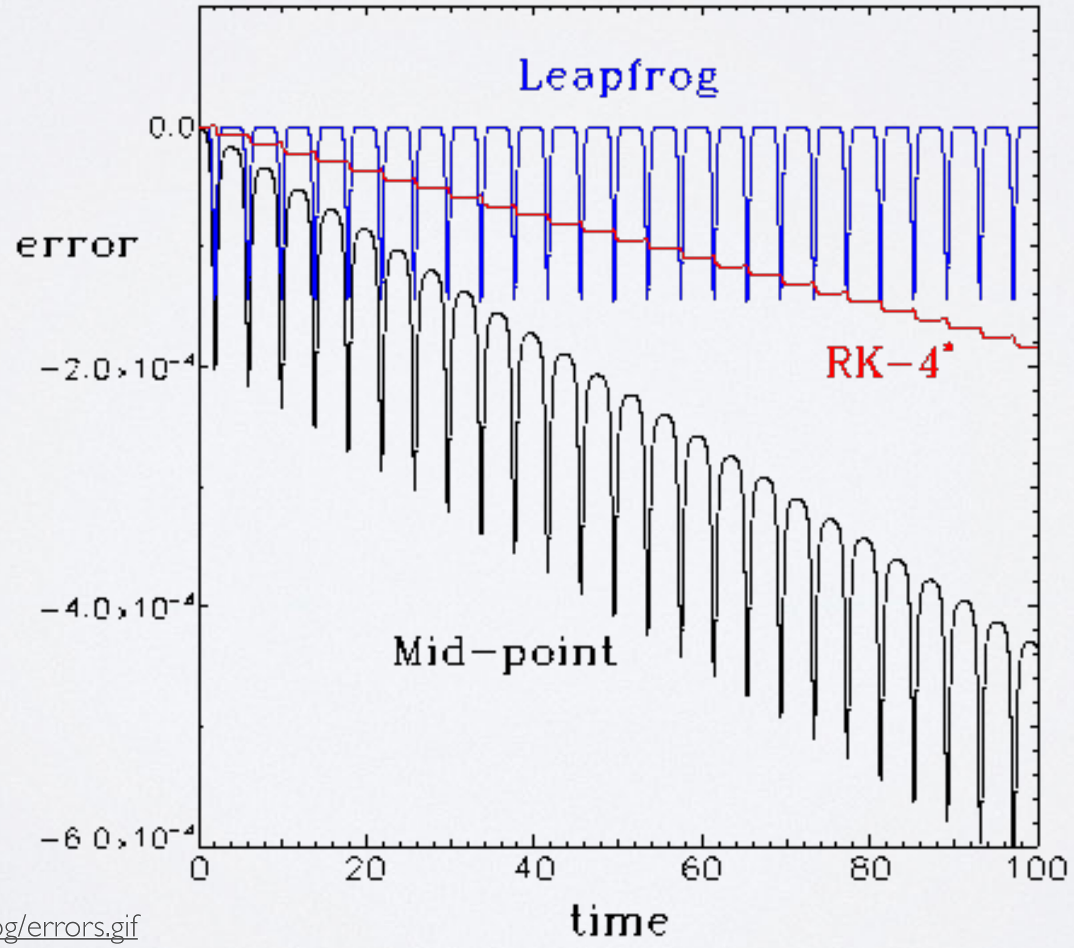
$$x(t+\delta t) = x(t) + v\left(t+\frac{1}{2}\delta t\right)\delta t + \mathcal{O}\left(\delta t^3\right)$$

$$v\left(t+\frac{1}{2}\delta t\right) = v\left(t-\frac{1}{2}\delta t\right) + a(t)\delta t + \mathcal{O}\left(\delta t^3\right)$$

Time-reversibility

- Higher order integration algorithms have higher per step accuracy, enabling longer time steps and faster simulations (e.g., Runge-Kutta, Gear predictor-corrector)

- **But**, do not respect time reversibility of Newton's equations causing energy drift and error accumulation



http://einstein.drexel.edu/courses/Comp_Phys/Integrators/leapfrog/errors.gif

Chaos! Disaster!

- No matter what integrator we use, we introduce numerical errors due to truncation and round-off
- Trajectories are intrinsically chaotic, in the strict math sense of "sensitive dependence on initial conditions"
  (i.e., positive Lyupanov exponents)
- So the simulation trajectories produced by two different machines diverge exponentially in time!
- **How can we possibly trust MD simulation?!?**

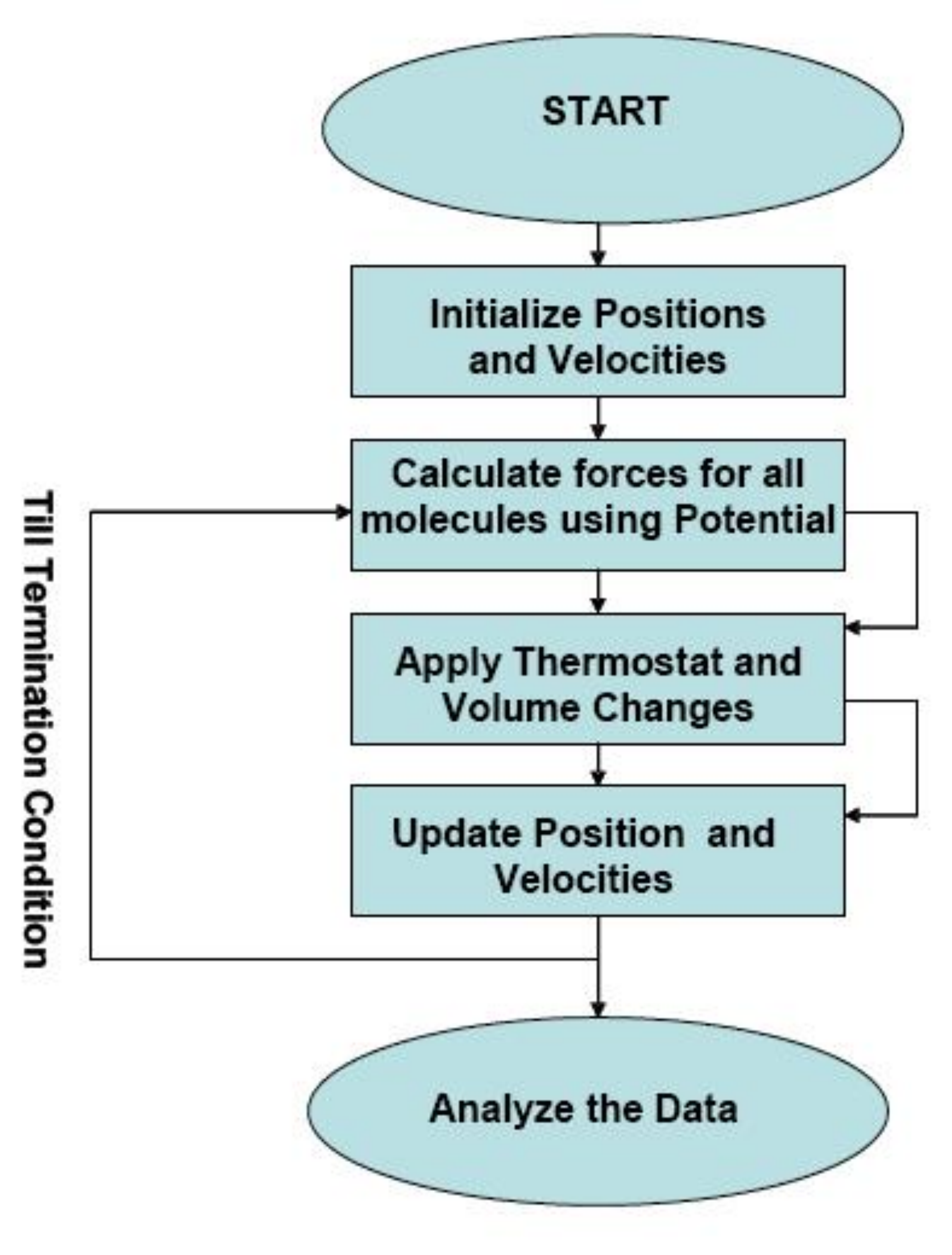
Statistics and shadow orbits

- Two answers:

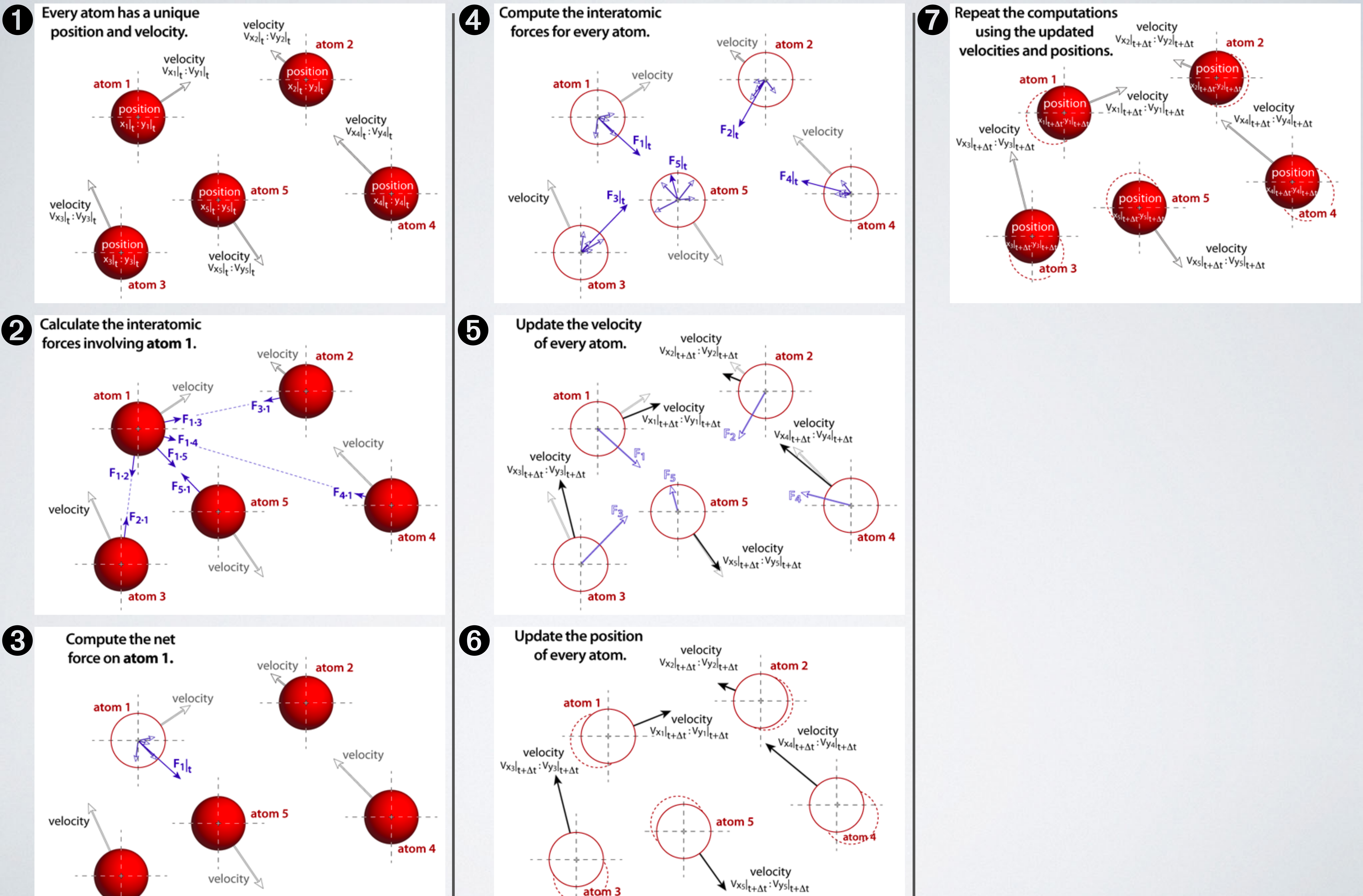
- **1. Shadow orbit.** Symplectic integrators do not conserve the true Hamiltonian, but do conserve a slightly perturbed "shadow Hamiltonian" so simulation trajectories do not diverge "too far" from the true trajectory.

- **2. Statistics.** Often we do not care about reproducing the exact long time trajectory, just generating a sequence of states from the equilibrium distribution from which we can compute statistical thermodynamic properties.

Simulation Overview



Simulation Overview



1. Every atom has a unique position and velocity.
2. Calculate the interatomic forces involving **atom 1**.
3. Compute the net force on **atom 1**.
4. Compute the interatomic forces for every atom.
5. Update the velocity of every atom.
6. Update the position of every atom.
7. Repeat the computations using the updated velocities and positions.

http://atomsinmotion.com/book/chapter5/md

IV. Advanced Topics

Ensembles

- Naturally MD ensemble is microcanonical (NVE):

  N - fixed # atoms
  V - fixed volume
  E - fixed energy

- What if we want to simulate in other thermodynamic ensembles that are closer to experimental systems?

  Canonical (isothermal-isochoric) - **NVT**
  Isothermal-isobaric - **NPT**
  Isenthalpic-isobaric - **NPH**

- MD is typically restricted to fixed **N**

Thermostats

- The temperature of a classical system is defined by the average molecular velocity

$$E_{kin} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \qquad \frac{1}{2} N_{df} kT = E_{kin}$$

- All thermostats are based on rescaling molecular velocities:

V-rescaling	- simple uniform rescaling of $\{v_i\}$ - does **not** yield canonical ensemble
Berendsen	- weak first-order coupling of v_i to target T - does **not** yield canonical ensemble
Andersen	- periodic v_i replacement with M-B dist^n - correct coord canonical ensemble, **but** unsuitable for for studying dynamics due to v_i discontinuities
Nosé-Hoover	- weak coupling of v_i to target T via fictitious oscillators - correct coord & velocity canonical dist^n and fluctuations*

* for N-H **chains**, single N-H thermostat non-ergodic in certain systems

Barostats

- Pressure is computed from the virial equation

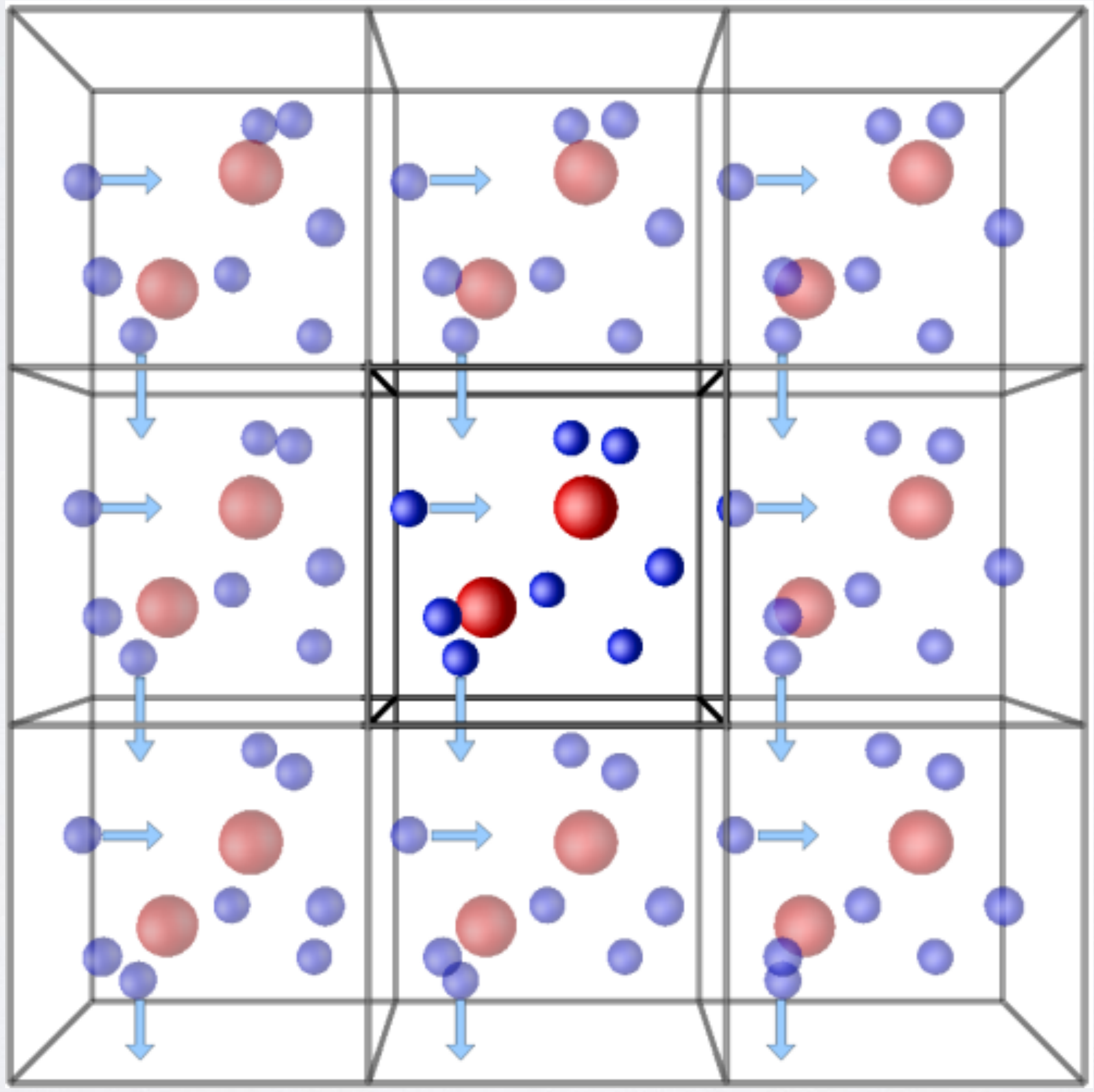
$$\mathbf{P} = \frac{2}{V}(\mathbf{E}_{kin} - \mathbf{\Xi}) \qquad \mathbf{\Xi} = -\frac{1}{2}\sum_{i<j} \boldsymbol{r}_{ij} \otimes \boldsymbol{F}_{ij}$$

- Barostats control pressure by scaling the box volume:

Berendsen	- weak first-order coupling of V to target P - does **not** yield isobaric ensemble
Parrinello-Rahman	- weak coupling of V to target P via fictitious oscillators - similar to Nosé-Hoover T coupling scheme - correct coord & velocity isobaric dist^n and fluctuations

Periodic boundary conditions

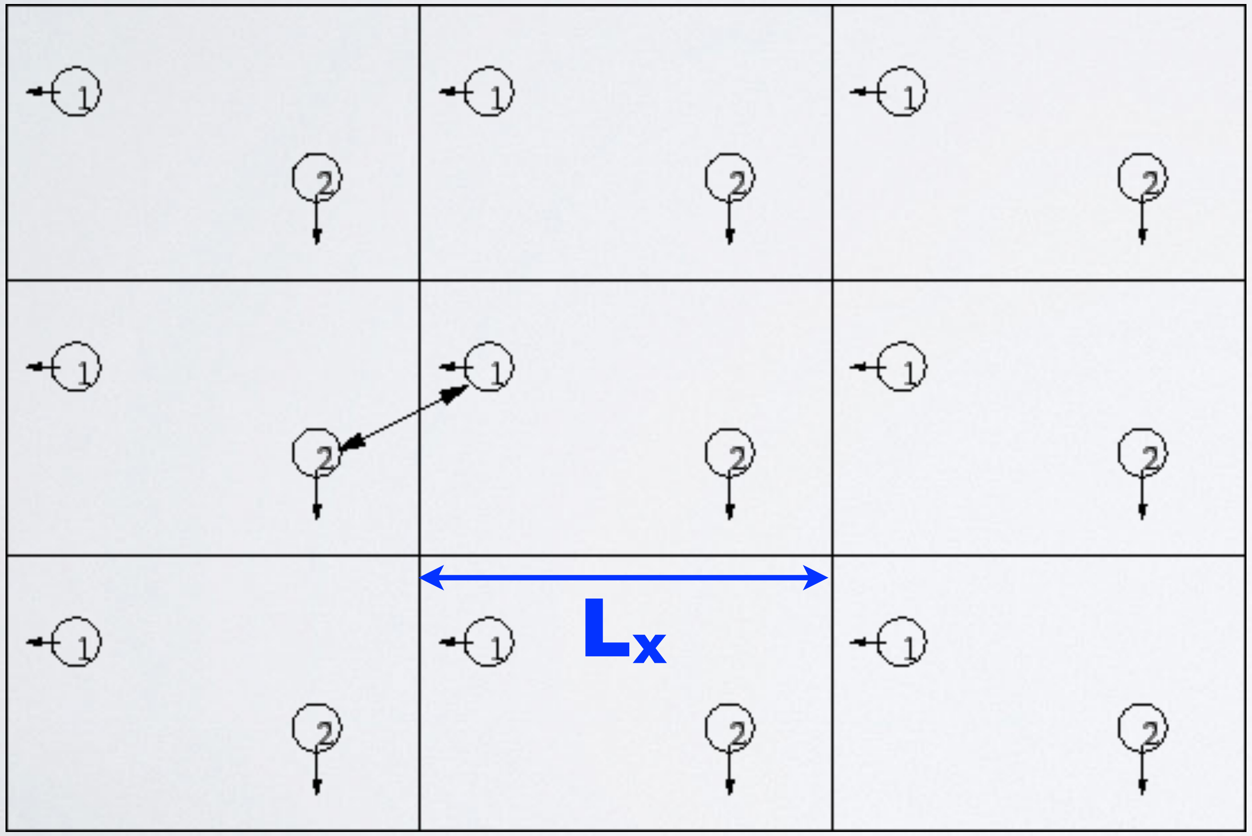
- Can only simulate small (nanoscopic) patch of space
- "Trick" the system into thinking it is infinite by tiling space with periodic replicas of fundamental simulation cell
- Molecules exiting one wall re-enter through the opposite!



http://isaacs.sourceforge.net/phys/images/these-seb/pbc-seb.png

Minimum image convention

- Under PBC, inter-particle distances are measured using the **minimum image convention**
- We must ensure $r_{cutoff} < L/2$ so particles do not interact with multiple images of neighbors

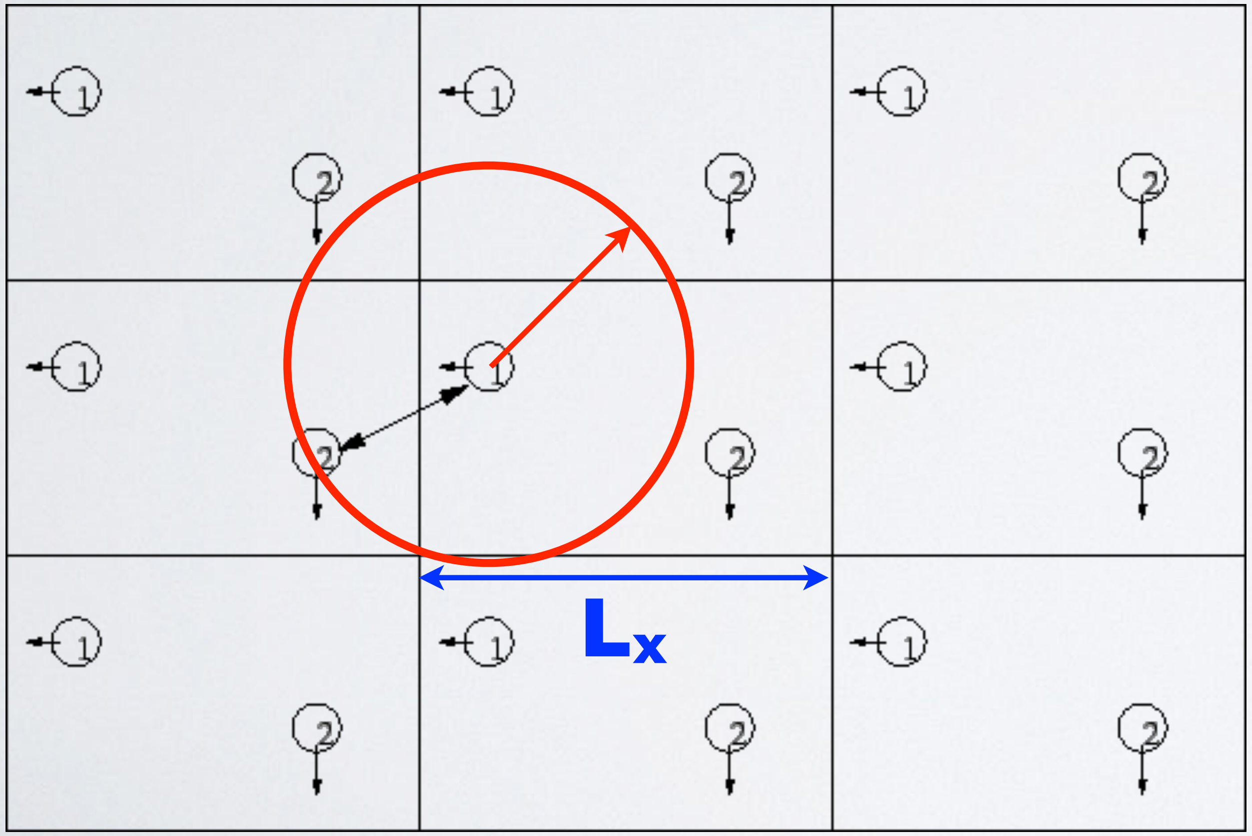


$$\Delta x_{MI} = \Delta x - L_x \operatorname{int}\left(\frac{\Delta x}{L_x}\right)$$

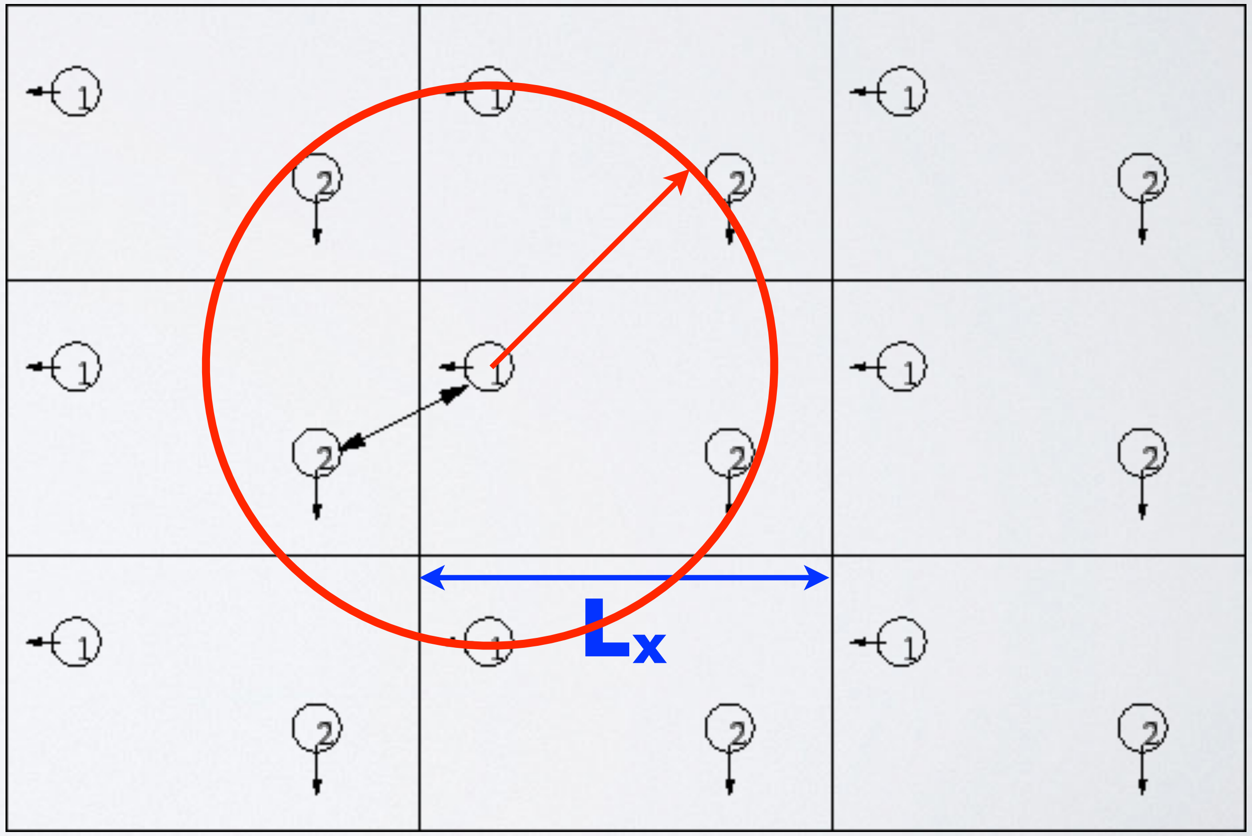
http://www.northeastern.edu/afeiguin/p4840/p131spring04/node41.html

Cut-off vs. box size

- Under PBC, we must enforce $r_{cutoff} < L/2$
- Why? Particles interact with multiple images of the same neighbor - completely aphysical!
- In practice, may also enforce:
  - $r_{cutoff} < L-s$: don't see own tail, s = length of molecule
  - $r_{cutoff} < (L-s)/2$: head & tail of molecule don't interact with same solvent



✔ $r_{cutoff} < ½L$

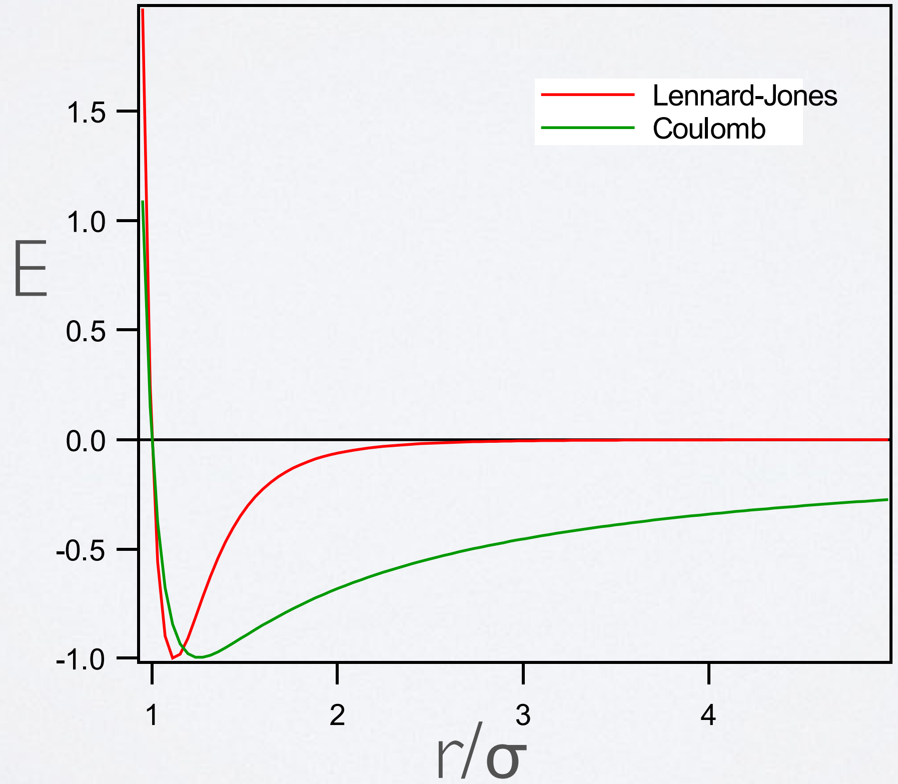


✘ $r_{cutoff} > ½L$

http://www.northeastern.edu/afeiguin/p4840/p131spring04/node41.html

Long-range electrostatics

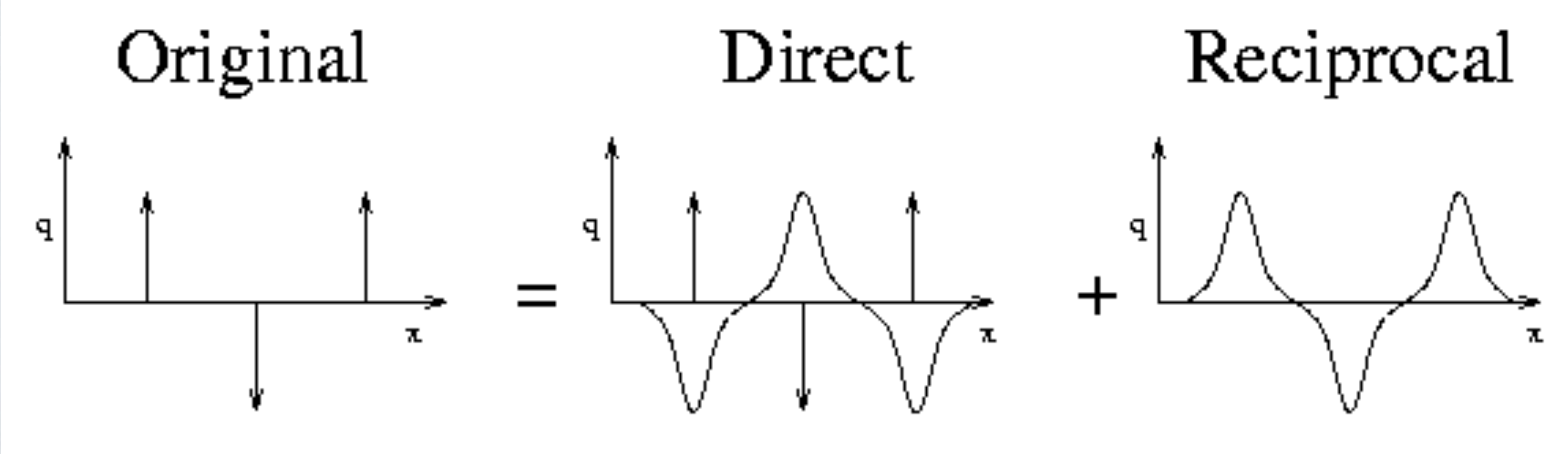
- vdW interaction decays like $1/r^6$, whereas Coulomb is $1/r$
- Coulombic interactions decay very slowly, and long range corrections problematic due to +/- charges
- The absence of a fast algorithm to rigorously treat long range electrostatics was a bugbear in mol sim until 1999



www.eng.buffalo.edu/~kofke/ce530/Text/Long_range_forces_3.doc Darden T, Perera L, Li L and Pedersen L. Structure 7 R55–R60 (1999)

Ewald summation

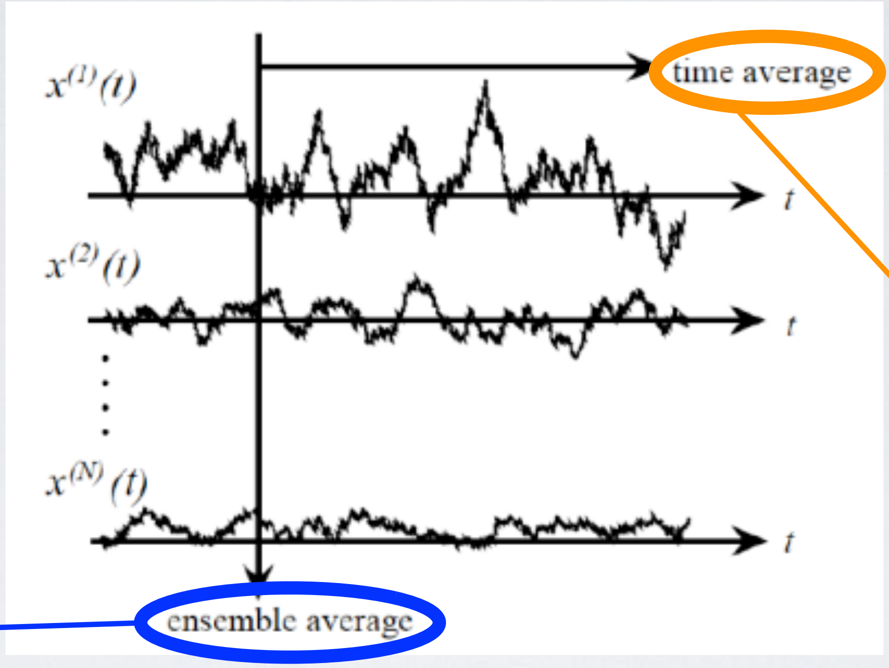
- Ewald summation is an elegant technique to rigorously treat long-range electrostatics and control error tolerance
- Represents electrostatics as unconditionally convergent real-space and reciprocal space components
- Particle Mesh Ewald is a fast implementation of this approach that revolutionized molecular simulation

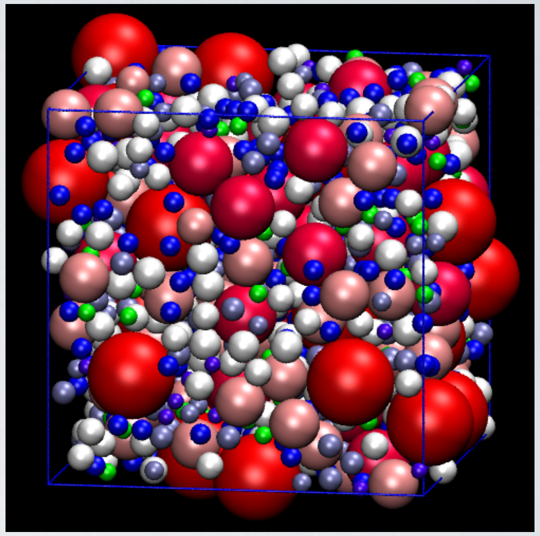


Paul Peter Ewald
b. 1888, d. 1985

Ewald P. Ann. Phys. 369 253 (1921)   Darden T, Perera L, Li L and Pedersen L. Structure 7 R55–R60 (1999)   http://www.nndb.com/people/207/000099907/
https://www.usenix.org/legacy/publications/library/proceedings/als00/2000papers/papers/full_papers/cramer/cramer_html/space.png

Ensemble and time averages





Experiment

Simulation

Ensemble average

- Average over all possible system configurations
- Naturally attained in experiments containing N_{Av} number of particles
- Very hard integral to perform numerically!

$$\langle A \rangle = \int \int dr^N dp^N A\left(r^N, p^N\right) \rho\left(r^N, p^N\right)$$

$$\rho\left(r^N, p^N\right) = \frac{1}{Q} exp\left[-\beta H\left(r^N, p^N\right)\right]$$

$$Q = \int \int dr^N dp^N exp\left[-\beta H\left(r^N, p^N\right)\right]$$

Time average

- Average over a single simulation trajectory
- Approximate time integral by summation

$$\bar{A} = \lim_{\tau \to \infty} \int_{t=0}^{\tau} dt A\left(r^N(t), p^N(t)\right)$$

$$\approx \frac{1}{M} \sum_{m=1}^{M} A\left(r^N(m), p^N(m)\right)$$

http://cssb.biology.gatech.edu/cell_simulation http://www.stanford.edu/~rsasaki/EEAP248/slide1

Ergodic hypothesis

- The **ergodic hypothesis** states that for $\tau \to +\infty$

$$\langle A \rangle = \bar{A}$$

- So we can compute thermodynamic averages from **sufficiently long** MD trajectories
  - Intuition is that long simulations explore all of the important (low energy) terms in the ensemble average
  - How long is long enough is often unknown *a priori* and we rely on internal checks that observables reach steady state

- For **slow processes**, we may need accelerated sampling

Accelerated sampling

- Hardware limits the attainable MD time scales to O(μs), making it hard to study processes with >μs relaxations

- Energetically, the system can be trapped behind large barriers, with the transition an exceedingly rare event

- Accelerated sampling techniques use artificial biases to speed up sampling of conformational space:

umbrella sampling - restrain system to hi E configurations using biasing potentials
replica exchange - use T swaps to accelerate system dynamics at hi T
Hamiltonian exchange - use H swaps to make exploration easier
hyperdynamics - modify H with boost potential to enhance sampling
metadynamics - lay down history dependent potential to flatten H
parallel replica - simulate multiple system copies to accelerate escape
T accelerated - hi T/hi mass coupling of part of system

Specialized MD variants

- **Car-Parrinello MD**
  - ab initio MD (no empirical potential required!)
  - nuclear forces from solution of the electronic problem
  - prohibitively expensive and slow for big systems

- **ReaxFF**
  - reactive MD force field
  - enables classical modeling of chemical reactions

- **GPU enabled MD**
  - massive speedups on commodity graphics cards

- **Implicit field models**
  - trades accuracy for time scale

Car, R. and Parrinello, M. PRL 55 22 2471 (1985)   Nielson, K.D. et al. The Journal of Physical Chemistry A 109 3 493 (2005)

Limitations and Caveats

- No electrons and so no chemical reactions (but ReaxFF)

- No quantum effects (but QM/MM)

- Availability, transferability, and quality of force fields

- Time and length scale limitations

- Statistical significance of single trajectories

- Equilibrated?

Common mistakes

- **Simulation too short (#1 problem!)**
  - answers are not meaningful
  - out of thermodynamic equilibrium
- **Inadequate forcefield**
  - GIGO
- **Δt too large**
  - E not conserved, unstable trajectory
- **System too small**
  - finite size effects
  - hard to model low conc. in small box
- **Missing important physics or chemistry**
  - e.g., salt, surface, impurity
- **Cut-offs too short**
  - improper treatment of long-range interactions

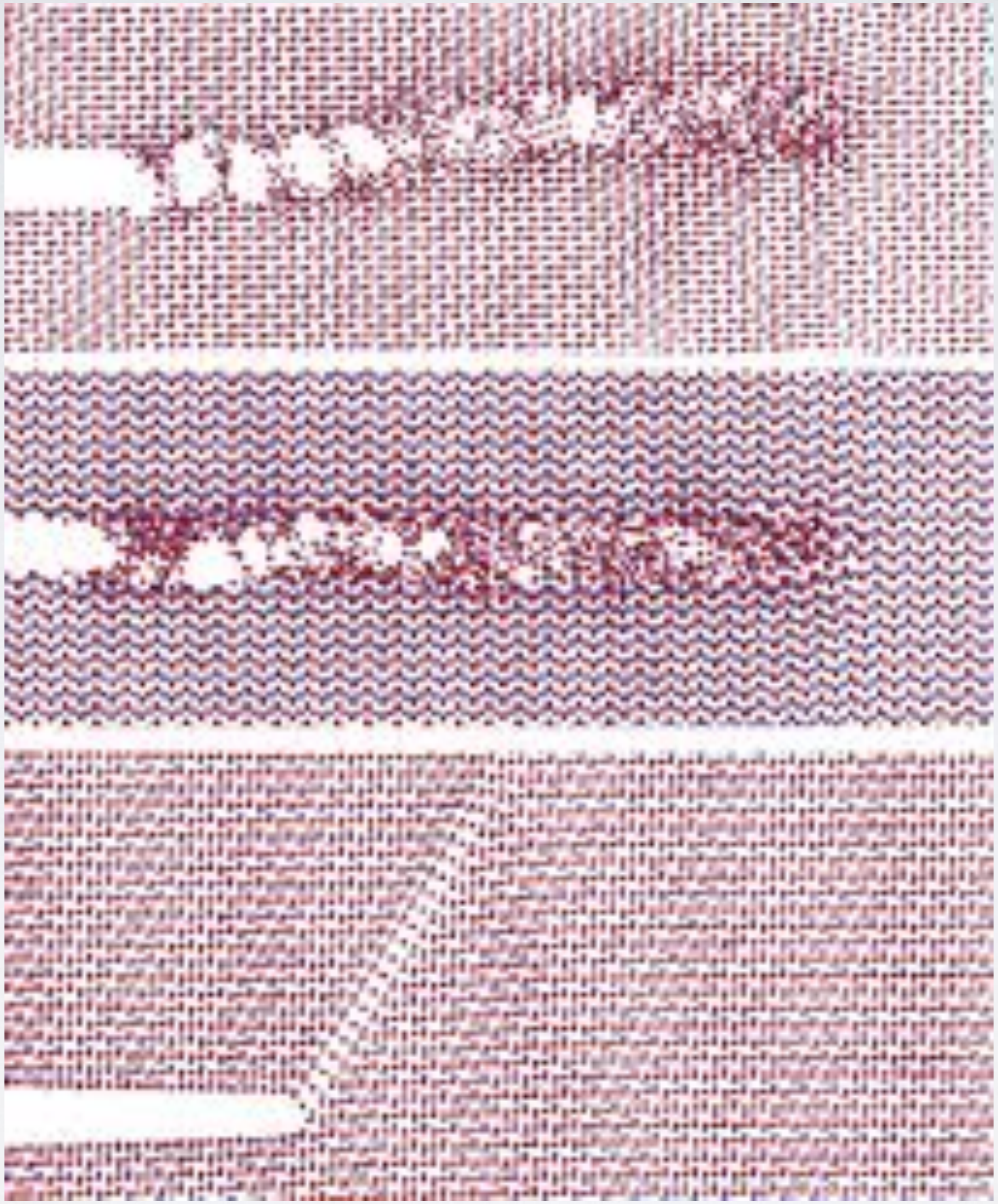
V. Molecular Dynamics Packages

MD software

GROMACS FAST. FLEXIBLE. FREE.	U. Groningen www.gromacs.org	FREE
CHARMM	Harvard www.charmm.org	$600
AMBER	Rutgers *et al.* www.ambermd.org	$400
NAMD Scalable Molecular Dynamics	UIUC www.ks.uiuc.edu	FREE
Desmond D E Shaw Research	D.E. Shaw Research www.deshawresearch.com	FREE
LAMMPS	Sandia National Lab http://lammps.sandia.gov	FREE
HOOMD blue	U. Michigan http://codeblue.umich.edu/hoomd-blue/	FREE
Folding@home distributed computing	Folding@home http://folding.stanford.edu	FREE

VI. Applications

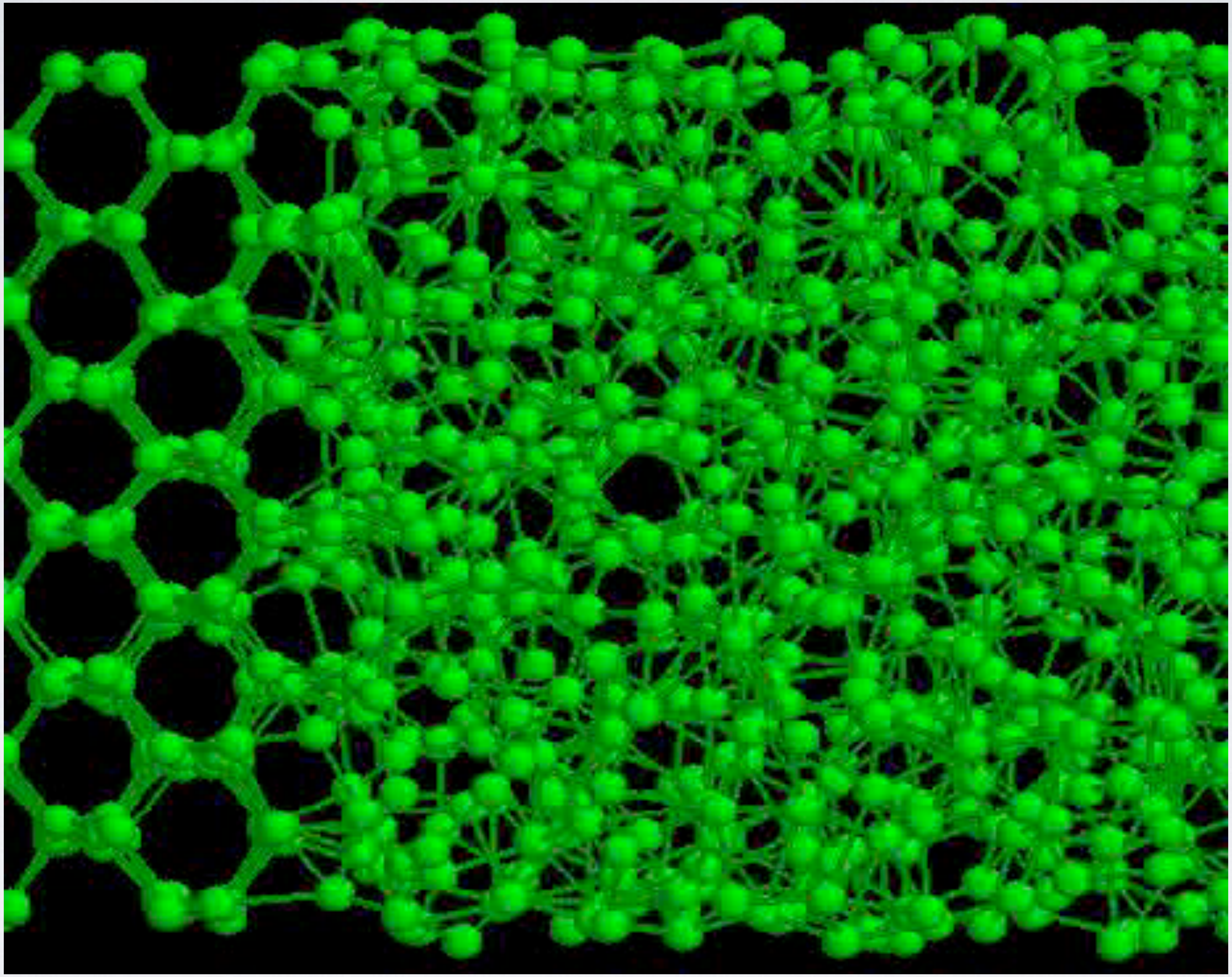
Fracture mechanics



Crack propagation in crystal planes of alumina

http://www.sfb716.uni-stuttgart.de/en/research/subprojects/research-area-b/b2.html

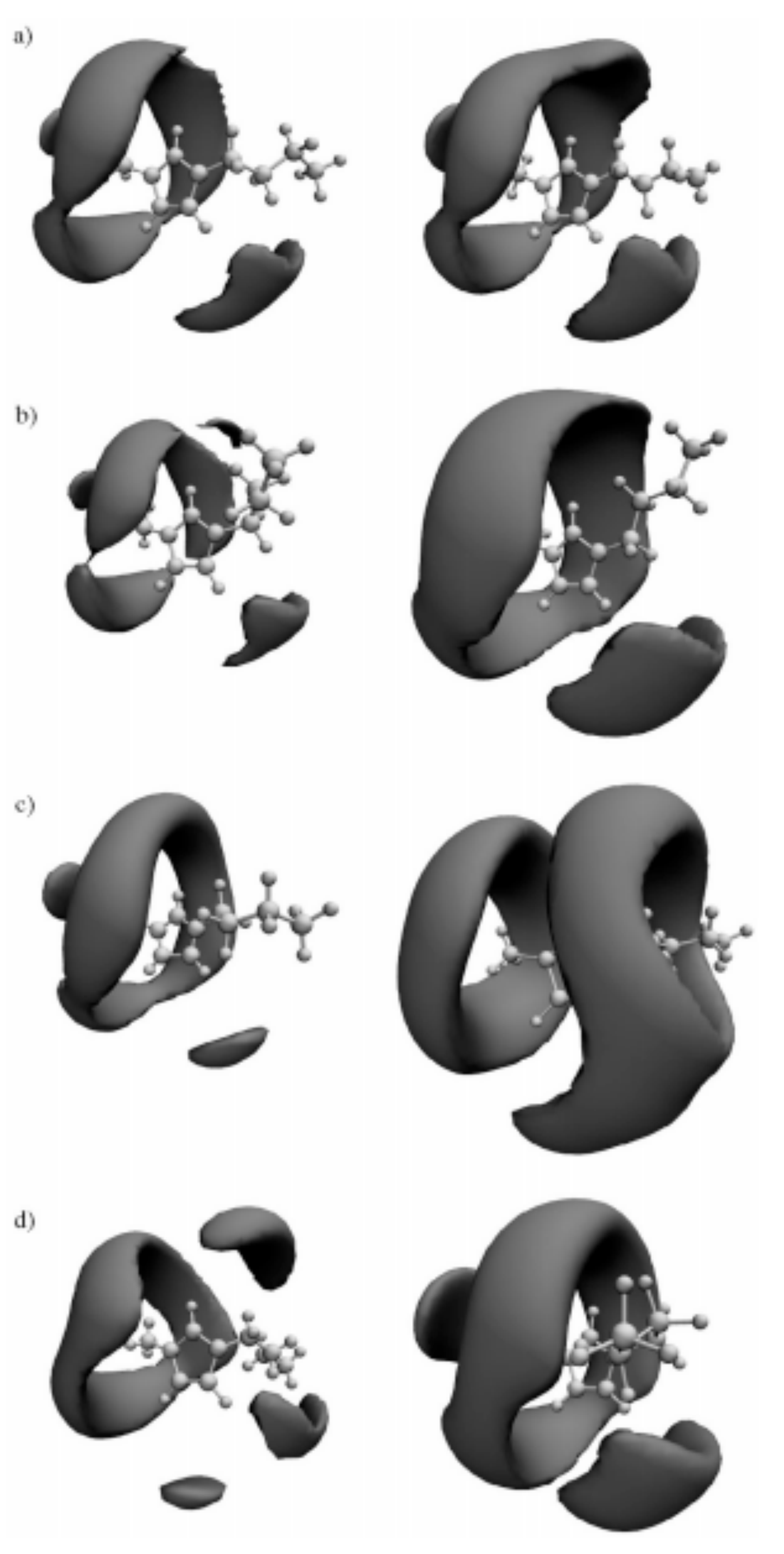
Phase transitions

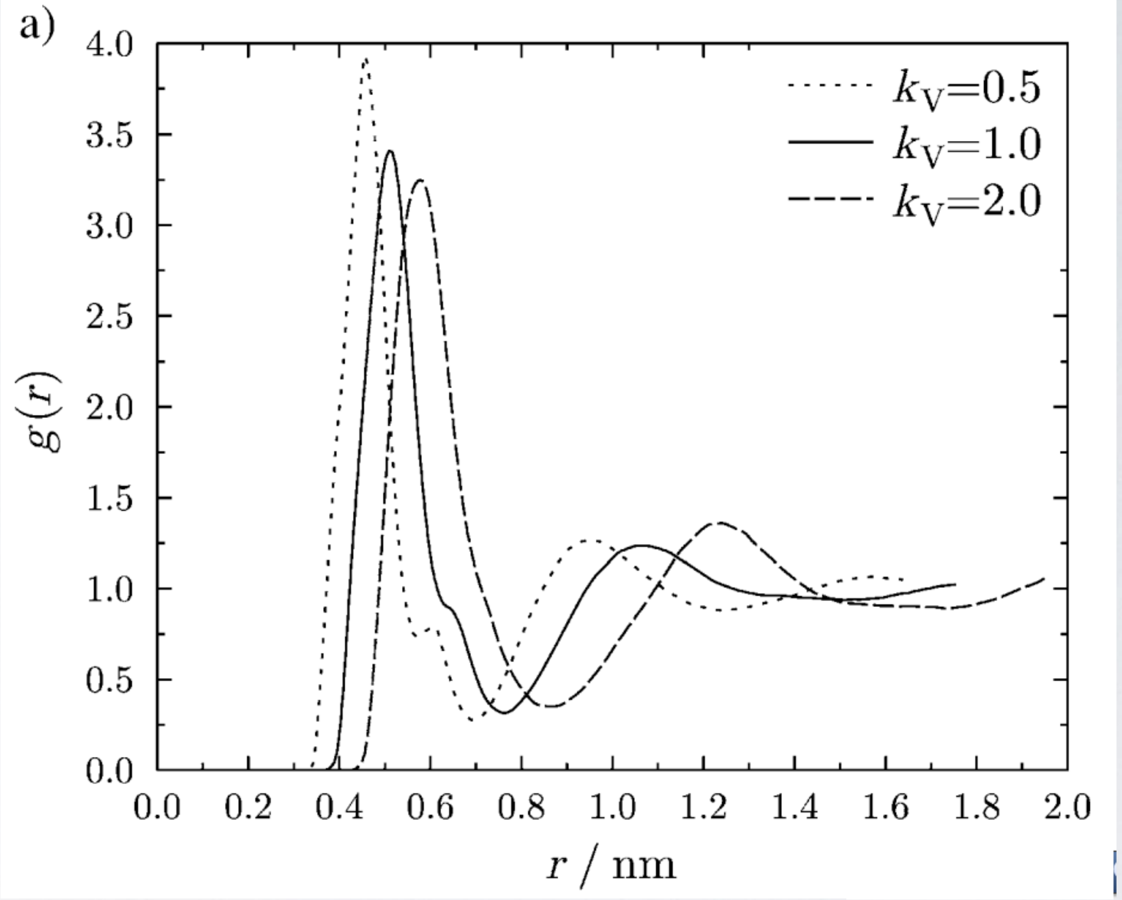


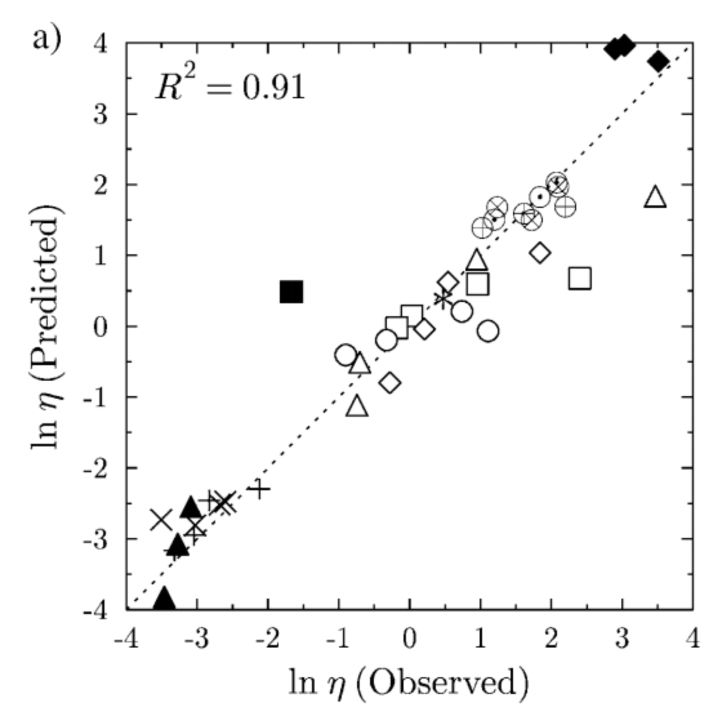
Silicon crystallization

http://www.ele.uva.es/~simulacion/MD.htm

Liquid structure and properties







Structure and properties of [bmim][PF_6] ionic liquid

Butler, S.N. and Muller-Plathe, F. Chem. Phys. Chem. 13 7 1791 (2012)