José R. Valverde
Head Scientific Computing
EMBnet/CNB. CSIC
jrvalverde@cnb.csic.es

Theory of Molecular Dynamics Simulations

Molecular Dynamics
Theory of Molecular Dynamics
simulations

2008-11-10

- Molecular Dynamics
  - History
- Reason for Molecular Dynamics
  - Statistical Mechanics
  - Ensembles
- Basic Molecular Dynamics
  - Classical Mechanics
  - Integration algorithms
- Practical Molecular Dynamics
  - Potential energy functions
  - Special treatments (non-bonded, solvation, PBC)
- Running a simulation
  - The heat-stabilize-measure-analyze cycle

History

- Theoretical study of molecular behavior
  - Fluids
  - Solids
  - Biological macromolecules
- Focus on *Dynamic* behavior
  - Stability
  - Conformational changes
  - Protein folding
  - Molecular interactions
  - Molecular transport
  - Drug design
  - Macromolecular ensembles
  - X-ray and NMR refinement...

- Originally introduced in late 1950's to study atomic interactions as hard spheres.
- Study fluid dynamics
- Development of realistic models (soft spheres)
- Development of realistic potentials (interactions)
- Development of protein potentials
- Today, MD is used for anything from structure refinement to macromolecular simulations.

Book Reviews

Dynamical Theory of Crystal Lattices. By M. BORN and K. HUANG. Pp. viii+420 with 28 figs. Oxford: Clarendon Press (Geoffrey Cumberlege). 1954. Price 50s.

This book has an outstanding value because of its co-ordinated account of the work by Born and his collaborators on the theory of crystals. The last comprehensive review of their work was the article by Born & Maria Göppert Mayer, which appeared in 1933. Since then many

and makes the validity of the adiabatic approximation questionable, the theory of metals is excluded. The scope of the book is further confined to crystals without imperfections. Likewise, surface phenomena are excluded.

The theory given in the book originated in several papers from 1912 and 1913 by Born and Th. von Karman where the mathematical foundation of the theory of the vibrations of simple crystal lattices was given together with its application to the problem of the specific heat

---

Dynamics of Radiation Damage*

J. B. GIBSON, A. N. GOLAND,† M. MILGRAM, AND G. H. VINEYARD
Brookhaven National Laboratory, Upton, New York
(Received July 14, 1960)

Radiation damage events at low and moderate energies (up to 400 ev) are studied by machine calculations in a model representing copper. Orbits of knock-on atoms are found and the resulting damaged configurations are observed to consist of interstitials and vacancies. Thresholds for producing permanently displaced atoms (i.e., interstitials) are about 25 ev in the $\langle 100 \rangle$ direction, 25 to 30 ev in the $\langle 110 \rangle$ direction, and around 85 ev in the $\langle 111 \rangle$ direction. Collision chains in the $\langle 100 \rangle$ and $\langle 110 \rangle$ directions are prominent; at low energies the chains focus, at higher energies they defocus. Above threshold, the chains transport matter, as well as energy, and produce an interstitial at a distance. The range of $\langle 110 \rangle$ chains has been studied in detail.

The basics: Statistical Mechanics

- One experiment does not make science.
  - One *random* experiment much less
  - Molecules exhibit brownian motion
  - We can induce a givenmotion pattern... over the brownian motion
  - *One experiment does not make science*

- In order to translate molecular data to macroscopic observables we need an statistical approach

- Statistical mechanics studies macroscopic processes from a molecular point of view.
  - Provides rigorous mathematical treatment
  - Works from time-independent statistical averages

- **Thermodynamic state** of a system
  - Defines system state through thermodynamic parameters:
    - Temperature (T)
    - Volume (V)
    - Pressure (P)
    - Number of molecules (N)
    - E, U, μ, etc...
- **Mechanical or microscopic state**
  - Defined by atomic positions (q) and momenta (p)
    - Coordinates in phase space
    - A system with N particles has 6N coordinates
    - A point G in phase space defines an state of the system
- **Ensemble**
  - A collection of points in phase space satisfying a given thermodynamic state.

- An ensemble is a time independent collection of configurations with <u>the same thermodynamic properties</u> (e. g. generated by Monte Carlo)

  - NVE (**microcanonical ensemble**)

  - NVT (**canonical ensemble**)

  - NPT (**isobaric-isothermal ensemble**)

  - mVT (**grand canonical ensemble**: fixed chemical potential m (μ), volume and temperature).
    - μ = $\partial U / \partial N$
    - N and U (internal energy) may change as long as μ is constant

10

- Statistical mechanics collects data from an ensemble to compute averages and deduce macroscopic properties
  - <A> (average of A) where
    - $A(p^N, r^N)$ is the observable as a function of particle positions and momenta
    - A depends on the partition function (how configurations are distributed in the ensemble)
    - The partition function depends on all possible states
    - MC goal is to generate enough conformations
- The ergodic hypothesis
  - $<A>_{ensemble} = <A>_{time}$
    - Allows us to evaluate <A> from configurations generated over time using MD
    - MD tries to generate enough configurations to satisfy this equality

The method

- Generate conformations in an ensemble using time dependence:
  - Consider atoms and molecules as a classical system
  - Use classical dynamics to compute movement (µ) and positions (r)
  - Newton's laws:
    - $F_i = m_i \cdot a_i$
    - $F_i = -\frac{1}{2} \nabla_i V$ (where V is the potential energy of the system)
    - $a = dv/dt = -(1/m)(dE/dr)$
    - $v = a \cdot t + v_0 = dr/dt$
    - $r = v \cdot t + r_0 = a \cdot t^2 + v_0 \cdot t + r_0$
  - Therefore we need
    - Initial positions (r_0) and velocities (v_0)
    - A way to compute potential energy changes (∇V)

13

- Start from initial positions and velocities
  - Assign initial positions (e. g. randomly for a fluid)
  - Assign initial velocities (e. g. randomly from a Maxwell-Boltzmann or Gaussian distribution) at a given temperature
  - Select a suitable potential energy function (to compute the gradient of V)
  - Set the daemons loose...
- Evaluate the potential energy and update positions and velocities (momenta)
  - V depends on positions of all particles
  - Too complex to evaluate analytically
  - Needs numerical integration

- Does not use velocities
- At each step ∂t:
  - Compute accelerations at time t (a_t)
  - Use a_t and positions at time t and time $t-\partial t$ to compute new positions (for time $t+\partial t$)
    - $r_{t+\partial t} = 2r_t - r_{t-\partial t} + a_t \partial t^2$
  - Note that $v = (r_t - r_{t-\partial t}) / \partial t$
- Advantages:
  - Easy to implement
  - Moderate requirements
- Disadvantages
  - Moderate precision

- Compute velocities at time $t + \frac{1}{2}\partial t$
- Use these to compute positions at $t + \partial t$
  - $r_{t+\partial t} = r_t + v_{t + \frac{1}{2}\partial t}\, \partial t$
  - Velocities and positions leap (jump) ever each other like a frog
- Advantages
  - Velocities are calculated explicitly
  - But not at the same time as positions

16

- Uses positions, velocities and accelerations at time *t*
- There is no compromise on precision
- Requires more computation

  - $r_{t+\partial t} = r_t + v_t \cdot \partial t + \frac{1}{2} a_t \cdot \partial t^2$

- Closely related to Verlet
  - $r_{t+\partial t} = r_t + v_t \cdot \partial t + (2/3) \cdot a_t \cdot \partial t^2 - (1/6) \cdot a_{t-\partial t} \cdot \partial t^2$
- Advantages
  - More accurate velocities
  - Better energy conservation
- Disadvantages
  - More complex
  - More expensive

The force field
(Potential Energy function)

- An analytical potential function is unfeasible
  - Depends on all particles
  - Depends on interactions among them
- A numerical approximation is used instead
  - Integrated using a suitable method
  - Parameterized for a specific problem
    - Experimental results
    - Quantum mechanical computations
    - Tested against experiment
- No changes in electronic structure are allowed
  - Non bond-breaking/making allowed
- Atoms in molecules (do they exist??) and interactions are modeled as classical particles and forces.

- Potential energy is a function of atomic positions
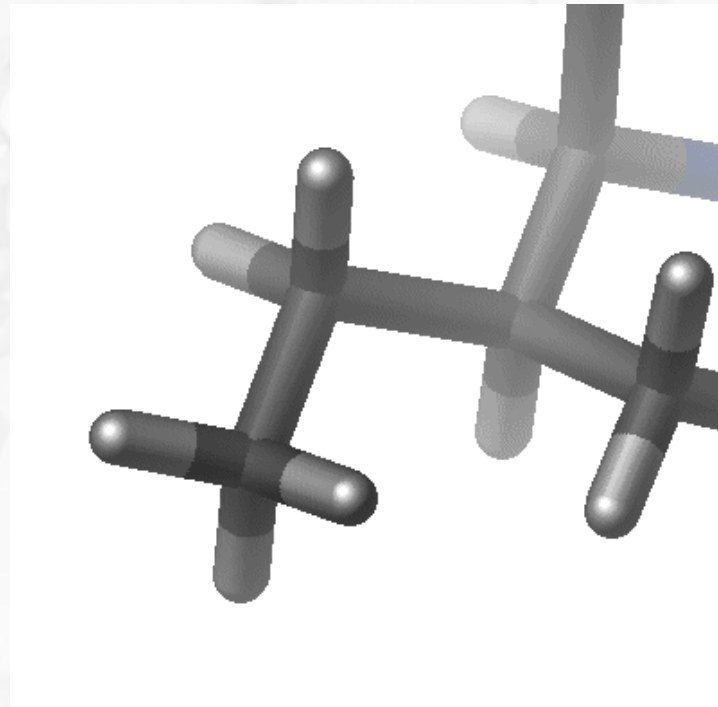 - $V = E_{bonded} + E_{non\text{-}bonded}$

  - $E_{bonded} = E_{bond\text{-}stretch} + E_{bond\text{-}bend} + E_{bond\text{-}rotation}$
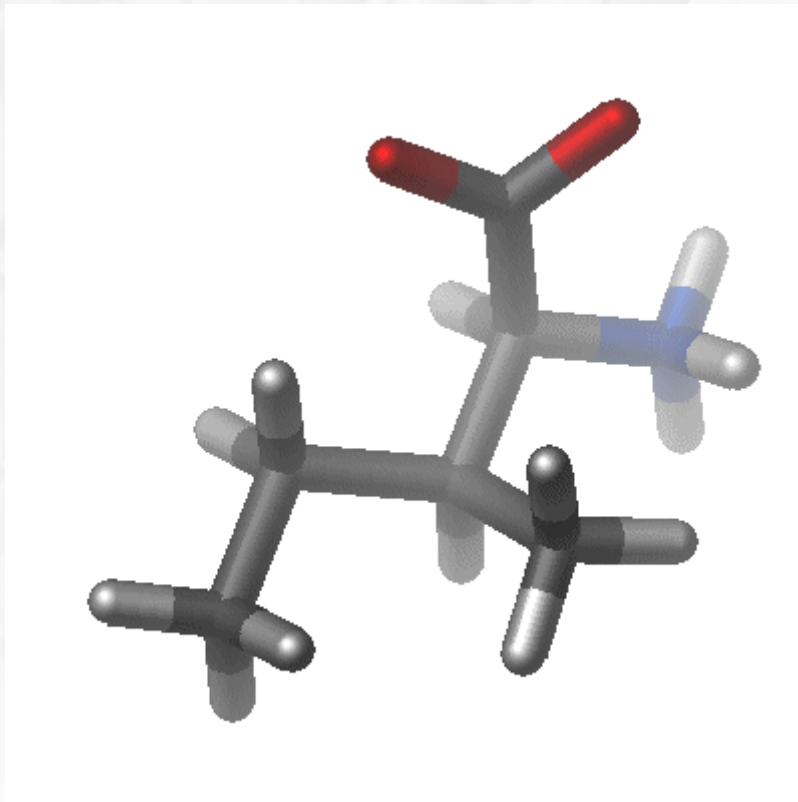
- Modeled as an harmonic potential
- Energy among two atoms joined by a covalent bond
  - Change in Energy as bond lengths deviates from ideal
    - $K_b (b - b_0)^2$
  - Must be summed for all bonds
  - Both K_b and b_0 are bond specific and must be tabulated for each bond type
   - Depends on bond order and
    - atoms bonded

- Depends on changes in the angle between two bonds
  - Involves angle between three atoms
  - Computed as an harmonic potential depending on the ideal bond angle
  - $E = K_\theta (\theta - \theta_0)^2$
  - Must be computed for all bond angles
  - K_θ and θ_0 depend on bond orders and chemical types of the atoms involved
 - Must be tabulated for all combinations
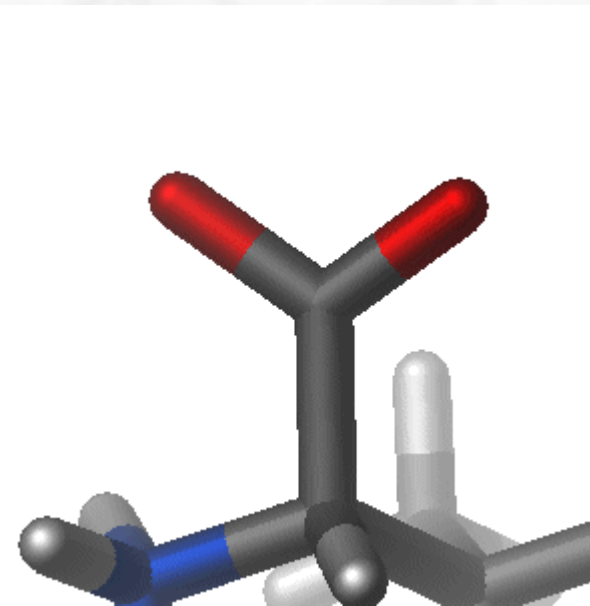
- Involves the effect of torsion around a bond
  - Reflects steric effects between atoms separated by three covalent bonds
  - Described by a dihedral angle and a coefficient of symmetry
  - Modeled as a sine function series
- Some force fields (e. g. CHARMM) add
  - **Urey-Bradley** term
    - atoms separated by two bonds
  - **Improper dihedrals**
    - used to maintain
      - Planarity
      - Chirality
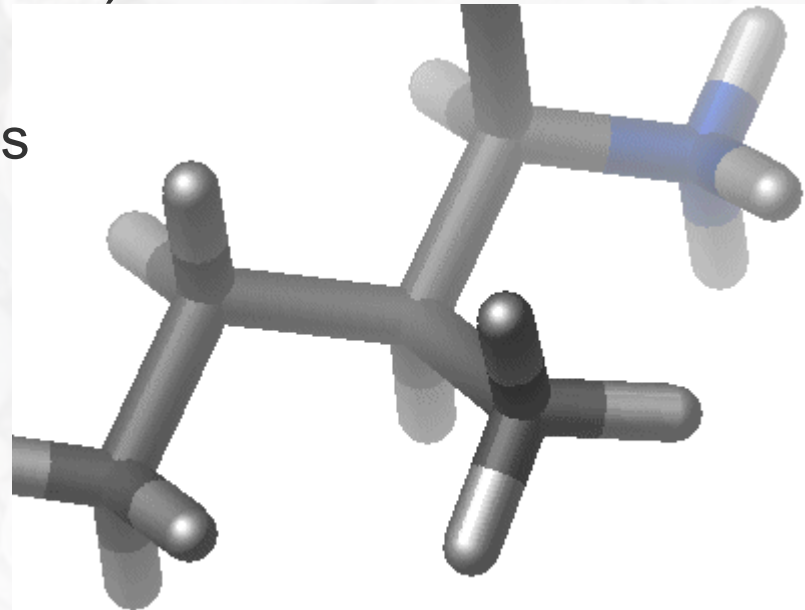
- Sum of all inter-atomic interactions
  - Van der Waals
  - Electrostatic
  - Hydrogen bonding (optional)
  - Must be computed for each atom against all others (N^2)
  - Van der Waals
    - Computed as a Lennard-Jones potential
  - Electrostatic
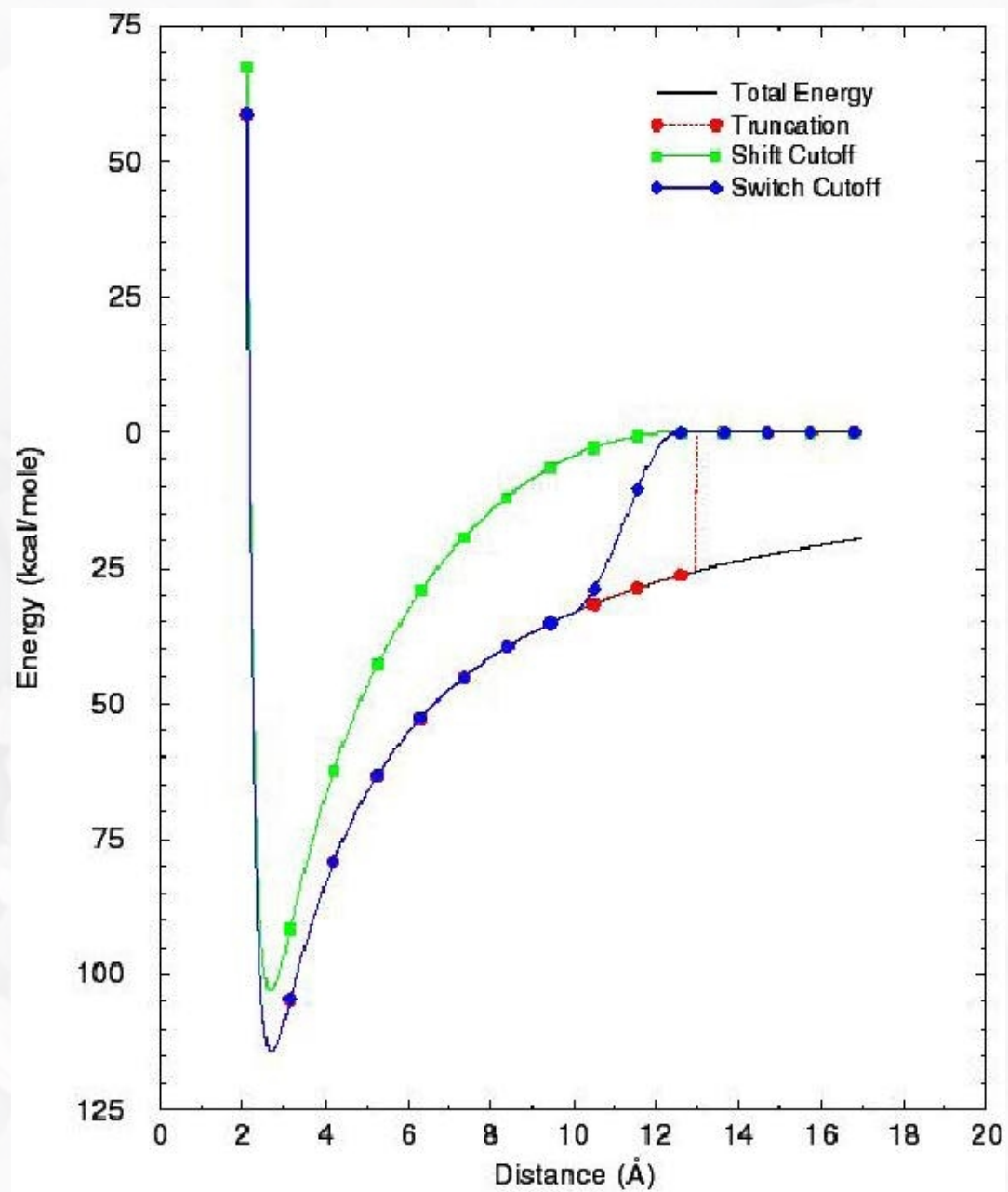   - Computed as classical interaction
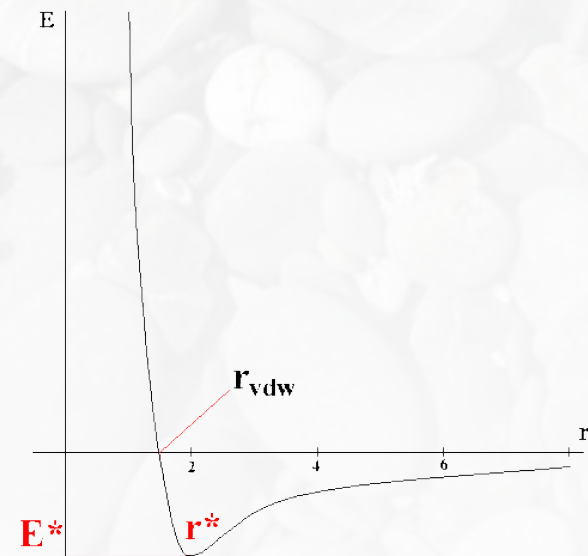
- Potential energy is differentiable against atomic coordinates
  - Gives the value of the force acting on each atom
- An empirical function has limitations in accuracy
  - Fixed set of atom types
    - Same atom acts differently on different environments
    - Definition of **extra atom types**
  - Speedup approximations
    - Additive approximation
      - Interaction of atom with system computed as sum of one atom to one-atom interactions
      - Ignores polarization effects
  - Entropic effects not included in single calculations
    - Minimum E need not correspond to equilibrium E
    - Entropic effects are included in MD simulations

26

- They are the most expensive element in the evaluation of the potential energy function
- Require N^2 evaluations (all atoms against all others)
- Can be reduced by using a **cutoff**
  - Truncation
    - Large energy fluctuations
    - Rarely used
  - SHIFT (modify entire surface)
    - Slight decrease in equlibrium distances)
  - SWITCH (two cutoffs and progressive change)
   - Strong fluctuation in switch regions
    - Avoid for small cutoff regions

- Important in many system
- Increasing cutoff raises computational cost
- Approximations have been developed
  - Proteins or enzymes in crystalline state
    - **Ewald summation**
    - **Particle-mesh Ewald method**
  - Non periodic systems (e. g. enzyme in solution)
    - Multipole expansion
      - Short range component evaluated pairwise
      - Long range component substituted by a multipole
        - **Extended electrostatic model**
   - **Fast Multipole Method**

29

Multipole Approximation
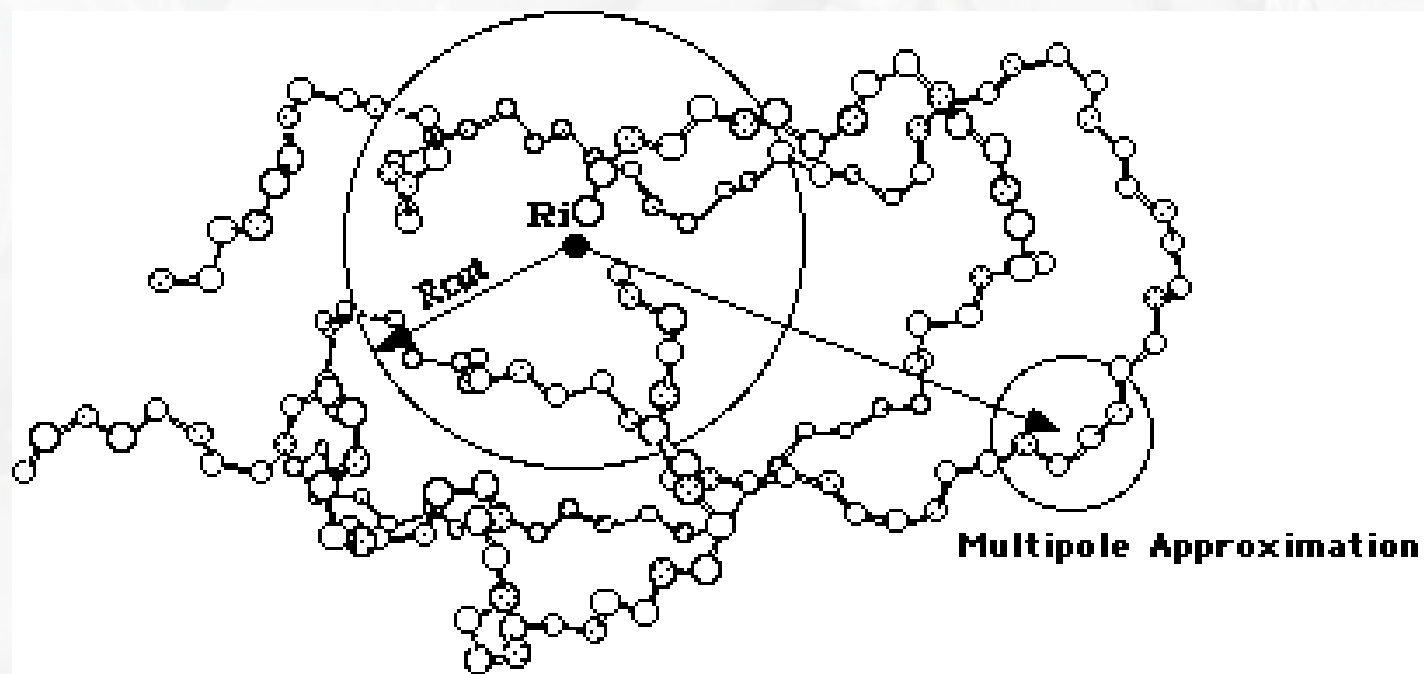
- Has a strong influence
- Involves a potentially large amount of additional molecules (and interactions)
- Approaches
  - Implicit solvent models
    - Substitute solvent by an approximation
      - Effective dielectric constant (e. g. 80 for water)
      - Distance-dependent dielectric coefficient
      - Continuum electrostatic theory
  - Explicit solvent models
    - Include the solvent as an explicit element of the calculation
      - Dielectric constant should be set to 1
   - Requires *boundary conditions* to avoid leakages

- Reduce simulation cost by "replicating" a unit cell many times
  - Replication is virtual
  - Particles exiting the cell on one side re-enter it on the opposite side
  - Interactions are computed considering adjacent cells (modulo cell size)
    - Cutoffs must be used shorter than cell size to avoid a particle seeing its own image in replica cells
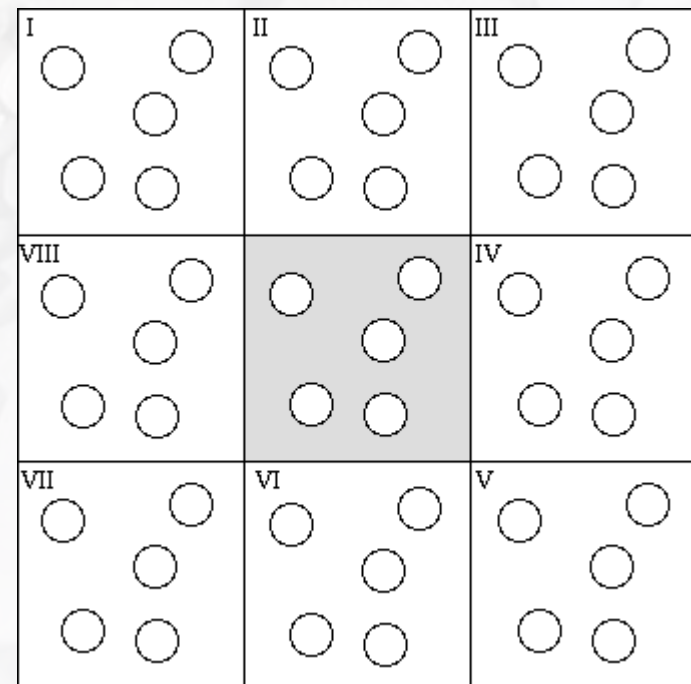
- Avoid using periodic boundary conditions
- Allows inclusion of explicit solvent molecules
  - Build a droplet engulfing the system
  - Use spheric boundary potentials to confine water inside the shell
  - Requires less molecules than periodic boundary conditions
    - Still, a significant amount
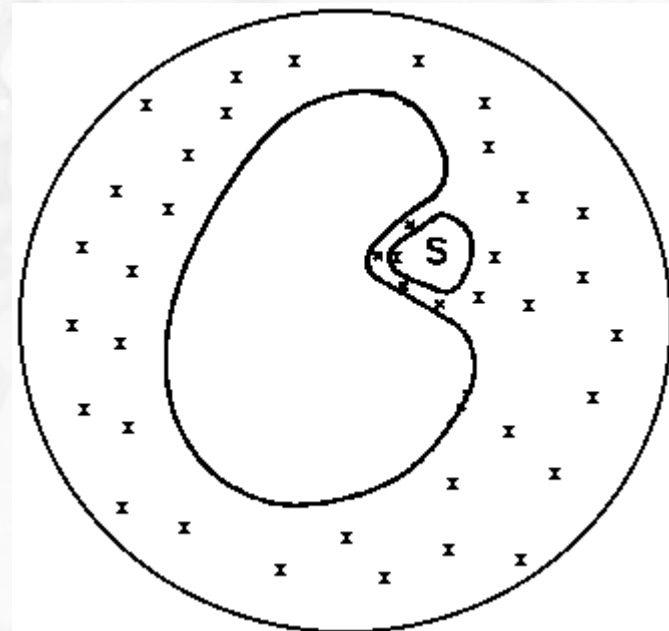
- We can further reduce solvation to the active site
  - When we have a large structure
  - But are interested in only one section
  - Partition the enzyme in two regions
    - Active site/Reaction zone
      - Solvated
      - Allowed to move during the simulation
    - Reservoir zone
      - Unsolvated
      - Movement restricted
        - Harmonically
        - Fixed
- Dramatically reduces cost
 - If we are only interested in one region
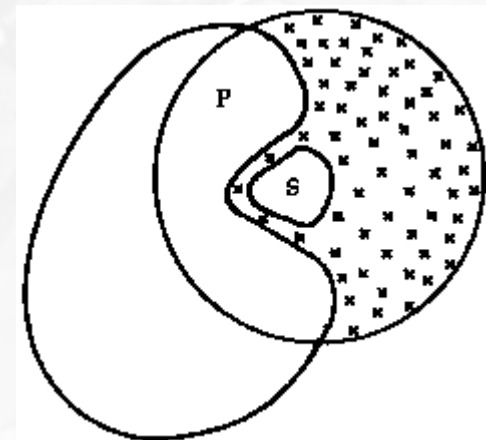
Running a simulation

- Start from initial coordinates and momenta
- Using classical physics and a potential energy function generate a time series of conformations
  - A **trajectory** (the path followed by each atom according to classical mechanics)
  - Usually in the microcanonical ensemble (NVE)
  - Recently isothermal-isobaric (NTP)
- Analyze the conformations (*trajectory*)

- **<u>Requires a suitable starting point</u>**
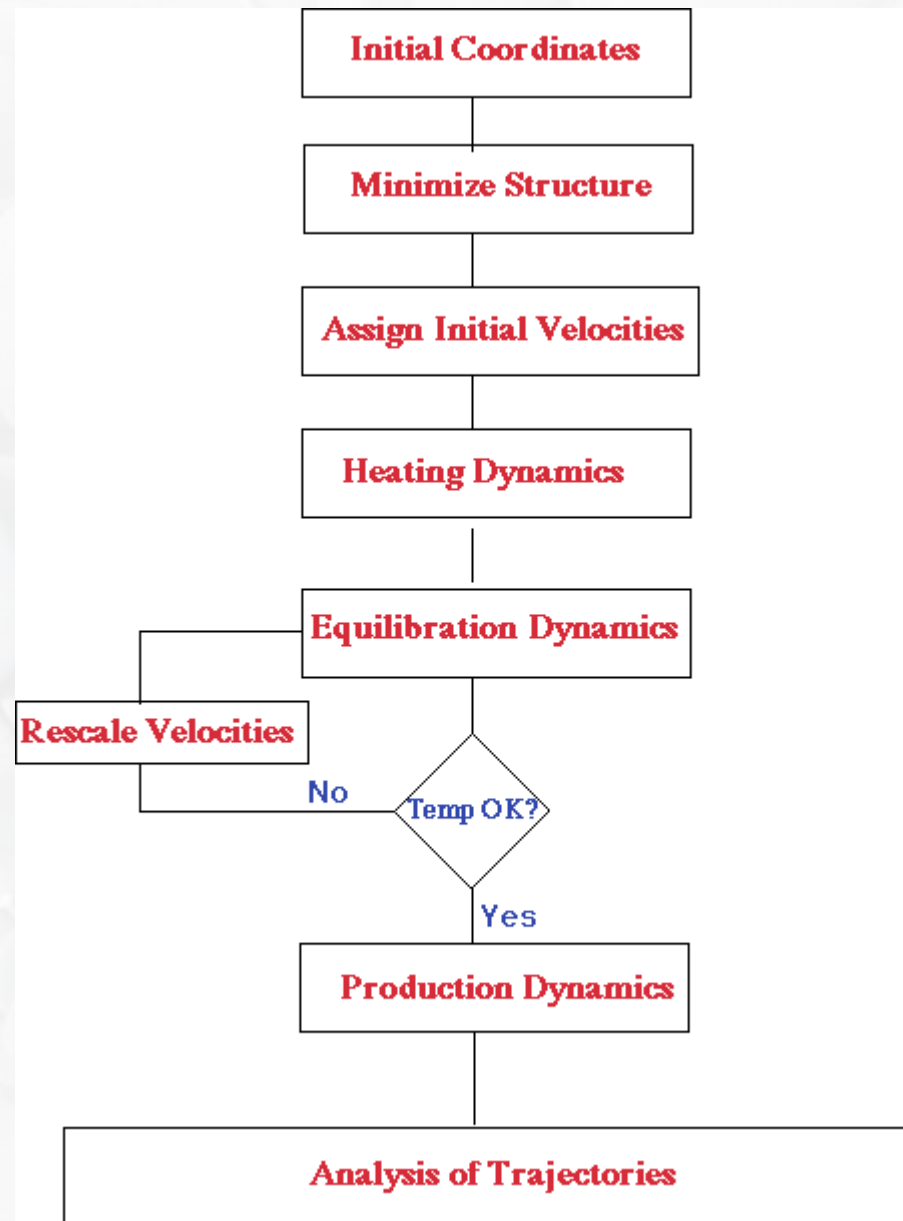
- Select a 3D coordinate dataset
  - E. g. a conformation from a PDB file
  - Simulation will be very sensitive to initial configuration
  - Should be close to the state we want to model
  - Complete the structure if needed
    - Many structures lack H or H bonds
  - It is *often* (not always, THINK!) advisable to optimize the structure
    - Energy minimization/relaxation by molecular mechanics
      - Remove strong Van der Waals (steric) effects
    - Simulated annealing cycles
      - Skip local minima (heat, cook, cool cycles)
  - Add explicit solvent molecules (if needed)
    - Use a minimized box of solvent
    - Minimize solute-solvent complex

- A minimized structure is (supposedly) the most stable structure
  - 0 °K
- Assign initial velocities at a low temperature
- Start the simulation
- At periodic times, velocities are adjusted to a slightly higher temperature
- Until we reach the simulation temperature
  - Take off the freezer and thaw

- Solvated case:
  - If solvent is present, allow solvent to equilibrate first
  - Then equilibrate the complex solvent+solute

- Once the desired temperature is reached
- Keep the system at the desired temperature
- Until the system equilibrates
  - If temperature changes readjust velocities
    - **Nosé-Hoover thermostat**
    - *Berenden velocity scaling*
    - Andersen stochastic collisions
    - Langevin piston
  - Monitor variables
  - Check property stabilization with respect to time

- Goal is to ensure our initial configuration (coordinates, momenta) is as close as possible to the system we want to simulate.

40

- We are now ready to formerly conduct the simulation
- Continue simulation for the time length desired
- Keep system on desired thermodynamic conditions
  - Ensemble
- Collect parameters of interest
  - Save trajectory (coordinates, velocities)
- Monitor progress
  - Simulation proceeds by finite steps
  - Energy *might* not be conserved
    - Cutoff
    - Time step
    - Rounding errors due to finite precison
    - etc...
  - Monitor energy conservation and stability during the time series.

41

- The final step to our goal
- Start from trajectory of the system
  - Graphical visualization
    - Animations
      - Useful to understand conformational changes
    - Time series graphs
      - Energy
      - RMSD and fluctuations
      - Radius of gyration
      - Correlation functions, etc...
  - Compute averages
    - Structural properties
    - Thermodynamical parameters
    - Chemical-physical properties
    - Etc...
  - Transform into macroscopic observables