MODULE 2: MOLECULAR DYNAMICS

Principles and Theory
I. Introduction
What is molecular dynamics?

- A computational microscope
- An experiment on a computer
- A simulation of the classical mechanics of atoms

F = ma

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, steric

- **High-throughput molecular screening**
  - drugs, surfactants, self-assembling materials

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.
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.

![Simulation results showing solid phase, liquid phase, and liquid-vapour phase.]

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

- Currently Professor Emeritus at UC Davis.
**Milestones in MD**

1957
Alder & Wainwright
First MD simulation of hard sphere fluid

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

1974
Rahman & Stillinger
First simulation of liquid water

1977
McCammon et al.
First protein simulation (BPTI) [8.8ps]

1994
York et al.
BPTI hydrated xtal [1ns]

1998
Duan & Kollman
Villin headpiece in water [1μs]

2010
Shaw et al.
BPTI in water [1ms]
III. Basic Principles
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.

\[ \tau_{\text{elec}} \approx 10^{-18} \text{ s} \]
\[ \tau_{\text{nuc}} \approx 10^{-15} \text{ s} \]

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: \( \Lambda_H \sim 1\,\text{Å}, \quad \Lambda_C \sim 0.3\,\text{Å} \)

characteristic atomic separation: \( d \sim 1\,\text{Å} \)

For all but lightest atoms \( d \gg \Lambda \), 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

- 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)
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)
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, $T$.

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

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

The potential energy is typically broken into four parts:

\[
V(\vec{r}) = V_{bonded} + V_{non-bonded} + V_{restraints} + V_{field}
\]
Bonded interactions are based on a fixed list of atoms. They are not exclusively pair interactions, but include \( \text{\textit{body}} \) interactions as well. There are \( \text{\textit{body}} \) interactions, a special type of dihedral interaction. For special purposes, there is an angle potential that uses a fourth-order polynomial:

\[
V_b \left( r_{ij} \right) = \frac{1}{2} k_{ij}^b \left( r_{ij} - b_{ij} \right)^2
\]

\[
V_a \left( \theta_{ijk} \right) = \frac{1}{2} k_{ijk}^\theta \left( \theta_{ijk} - \theta_{ij}^0 \right)^2
\]

\[
V_{rb} \left( \phi_{ijkl} \right) = \sum_{n=0}^{5} C_n \left( \cos(\psi) \right)^n
\]

\[
V_{id} \left( \xi_{ijkl} \right) = \frac{1}{2} k_\xi \left( \xi_{ijkl} - \xi_0 \right)^2
\]
Non-bonded

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

- van der Waals

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

- Coulomb

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

http://guweb2.gonzaga.edu/faculty/cronk/chemistry/images/graph-electrostatic-PE-alt.gif
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 are commonly used to model:
1. external potentials (e.g., electric, magnetic, flow)
2. continuum solvation (no explicit solvent molecules)
Multi-body potential widely used for metallic solids

- **EAM** - Embedded Atom Model
- **MEAM** - Modified Embedded Atom Model

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

\[
E_{\text{total}} = \frac{1}{2} \sum_{i,j}^{N} \Phi_{ij}(r_{ij}) + \sum_{i}^{N} F_{i}(n_{i})
\]

pairwise potential
interatomic separation
local e⁻ density
embedding function

Daw & Baskes PRB 29 12 6443 (1984)
Baskes PRL 59 2666 (1987)
Local $e^-$ density functions

**EAM**

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

**MEAM**

$$n_i = \sum_{j \neq i} \rho_j(r_{ij}) + \frac{1}{2} \sum_{j,k \neq i} f_{ij}(r_{ij}) f_{ik}(r_{ik}) g_i(\cos \theta_{ijk})$$

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

**NIST Interatomic Potentials Repository**

http://www.ctcms.nist.gov/potentials/
Ingredient 3: Integrators

- \([\text{initial atomic coordinates and velocities}] + [\text{force field}] \Rightarrow \text{entire future (and past!)} \text{ modeled by } \mathbf{F}=\mathbf{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 \Rightarrow \text{computers ideally suited to rapid, repetitive calculations.}

- Solving by hand would require thousands of years!
Many possible integration algorithms exist (e.g., explicit/implicit Euler, Gear predictor-corrector, n<sup>th</sup> 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:

\[
\begin{align*}
    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 + O(\delta t^4) \\
    v(t) &= \frac{r(t + \delta t) - r(t - \delta t)}{2\delta t} + O(\delta t^2)
\end{align*}
\]
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

\[
x(t + \delta t) = x(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + O(\delta t^3)
\]
\[
v(t + \delta t) = v(t) + \frac{a(t) + a(t + \delta t)}{2}\delta t + O(\delta t^3)
\]

Leapfrog-Verlet

\[
x(t + \delta t) = x(t) + v \left( t + \frac{1}{2}\delta t \right)\delta t + O(\delta t^3)
\]
\[
v \left( t + \frac{1}{2}\delta t \right) = v \left( t - \frac{1}{2}\delta t \right) + a(t)\delta t + O(\delta t^3)
\]
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.
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 Lyapunov exponents).

- So the simulation trajectories produced by two different machines diverge exponentially in time!

- **How can we possibly trust MD simulation?!?!**
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

START

Initialize Positions and Velocities

Calculate forces for all molecules using Potential

Apply Thermostat and Volume Changes

Update Position and Velocities

Till Termination Condition

Analyze the Data
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.

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) - \( \text{NVT} \)
Isothermal-isobaric - \( \text{NPT} \)
Isenthalpic-isobaric - \( \text{NPH} \)

MD is typically restricted to fixed \( N \)
**Thermostats**

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

\[ E_{\text{kin}} = \frac{1}{2} \sum_{i=1}^{N} m_i v_i^2 \]

\[ \frac{1}{2} N_{df} kT = E_{\text{kin}} \]

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 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
Pressure is computed from the virial equation

\[
P = \frac{2}{V} (E_{\text{kin}} - \Xi) \quad \Xi = -\frac{1}{2} \sum_{i<j} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij}
\]

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!
Under PBC, inter-particle distances are measured using the **minimum image convention**.

We must ensure $r_{\text{cutoff}} < L/2$ so particles do not interact with multiple images of neighbors.

\[
\Delta x_{MI} = \Delta x - L_x \text{ int} \left( \frac{\Delta x}{L_x} \right)
\]
Under PBC, we must enforce $r_{\text{cutoff}} < \frac{L}{2}$

Why? Particles interact with multiple images of the same neighbor - completely aphysical!

In practice, may also enforce:

- $r_{\text{cutoff}} < L - s$ : don’t see own tail, $s = \text{length of molecule}$
- $r_{\text{cutoff}} < \frac{(L - s)}{2}$ : head & tail of molecule don’t interact with same solvent

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

![Graph showing Lennard-Jones and Coulomb interactions](image)

---

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.

Original \[ \text{Direct} + \text{Reciprocal} \]
**Ensemble average**

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

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

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

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

---

**Time average**

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

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

$$\approx \frac{1}{M} \sum_{m=1}^{M} A(r^N(m), p^N(m))$$
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 $O(\mu s)$, making it hard to study processes with $>\mu s$ 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 high $E$ configurations using biasing potentials
- **Replica exchange** - use $T$ swaps to accelerate system dynamics at high $T$
- **Hamiltonian exchange** - use $H$ swaps to make exploration easier
- **Hydredynamics** - 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** - high $T$/high 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

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

- **$\Delta t$ 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

U. Groningen
www.gromacs.org
FREE

Harvard
www.charmm.org
$600

Rutgers et al.
www.ambermd.org
$400

UIUC
www.ks.uiuc.edu
FREE

D.E. Shaw Research
www.deshawresearch.com
FREE

Sandia National Lab
http://lammps.sandia.gov
FREE

U. Michigan
http://codeblue.umich.edu/hoomd-blue/
FREE

Folding@home
http://folding.stanford.edu
FREE
VI. Applications
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
Structure and properties of [bmim][PF$_6$] ionic liquid