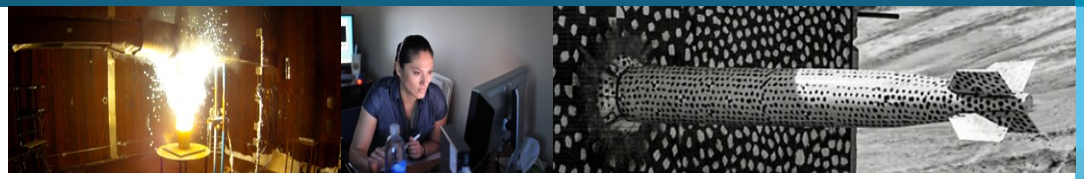
Overview of Molecular Dynamics



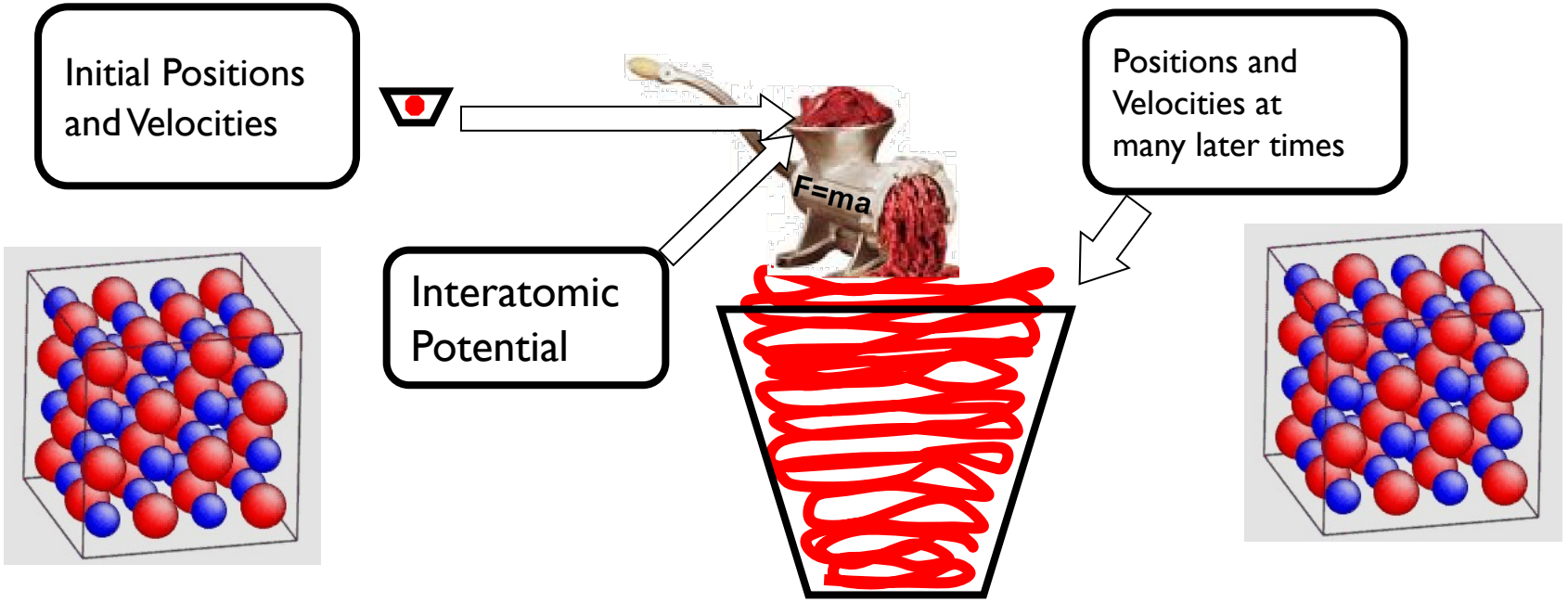
Stan Moore

Virtual LAMMPS Workshop and Symposium 2021

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2021-9693 C

Molecular Dynamics: What is it?



Mathematical Formulation

Classical Mechanics

Atoms are Point Masses: $r_1, r_2, \ldots . r_N$

Positions, Velocities, Forces: r_i, v_i, F_i

Potential Energy Function = $V(r^N)$

$6N$ coupled ODEs

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i$$

$$\frac{d\mathbf{v}_i}{dt} = \frac{\mathbf{F}_i}{m_i}$$

$$\mathbf{F}_i = -\frac{d}{d\mathbf{r}_i} V\left(\mathbf{r}^N\right)$$

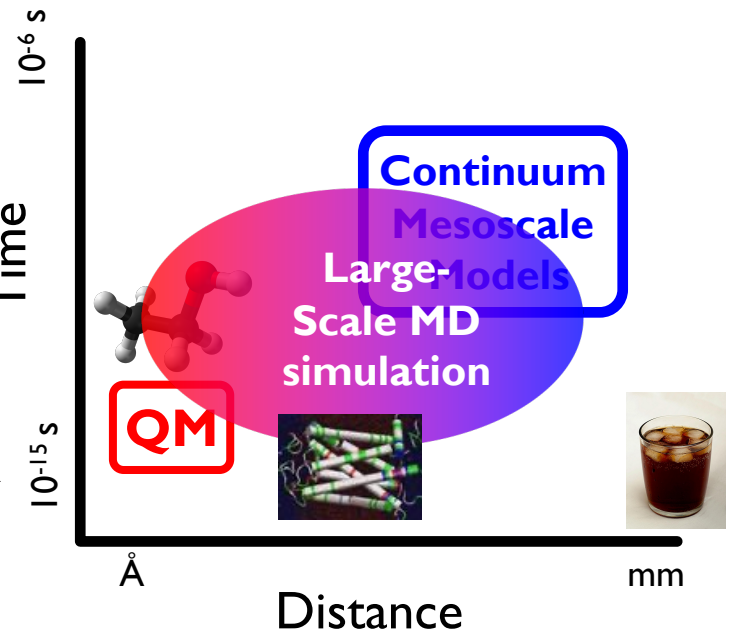
What is MD good for?

Quantum mechanical electronic structure calculations (QM) provide accurate description of mechanical and chemical changes on the atom-scale: 10x10x10~1000 atoms

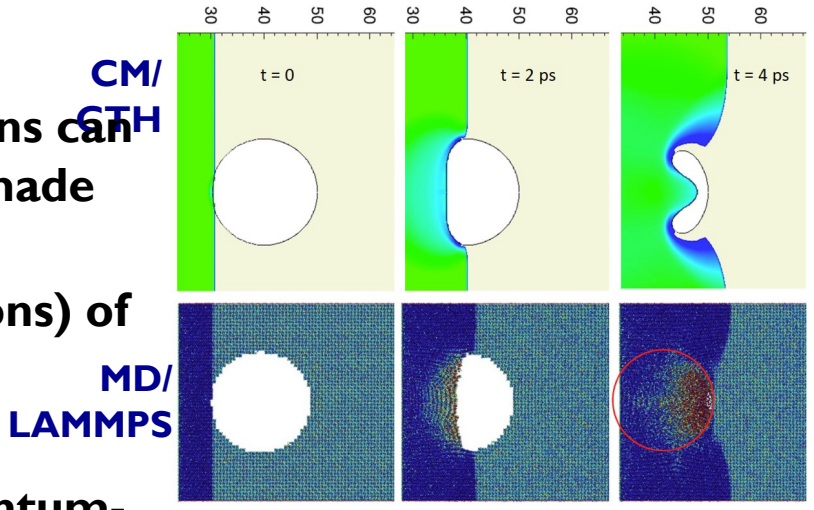
Atom-scale phenomena drive a lot of interesting physics, chemistry, materials science, mechanics, biology…but it usually plays out on a much larger scale

Mesoscale: much bigger than an atom, much smaller than a glass of soda.

QM and continuum/mesoscale models (CM) can not be directly compared.



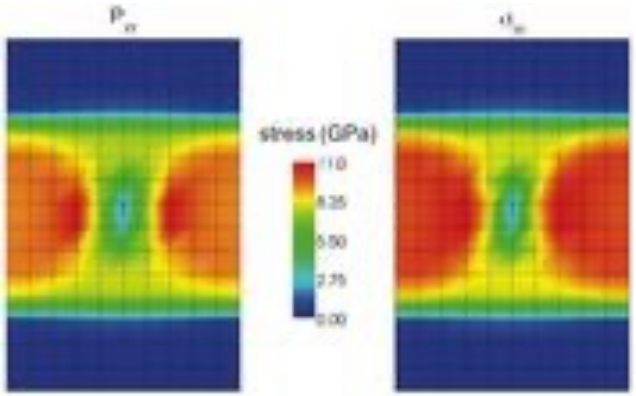
- **Small molecular dynamics (MD) simulations can be directly compared to QM results, and made to reproduce them**
- **MD can also be scaled up to millions (billions) of atoms, overlapping the low-end of CM**
- **Limitations of MD orthogonal to CM**
- **Enables us to inform CM models with quantum-accurate results**

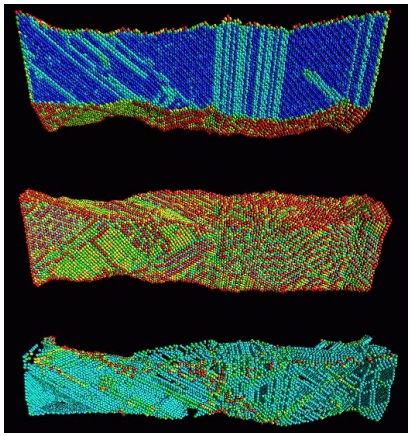


CTH images courtesy of David Damm, Sandia

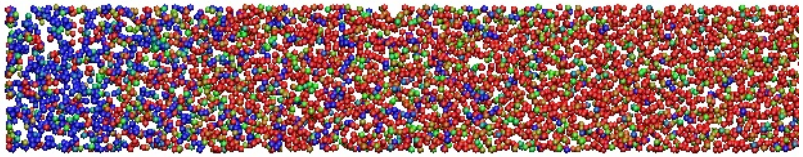
MD Versatility

Coupling to Solid Mechanics



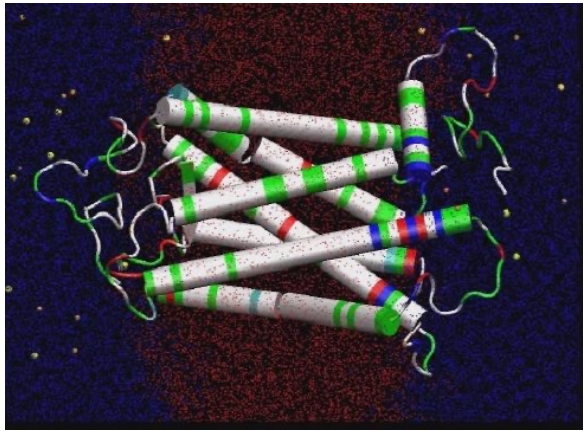


Materials Science

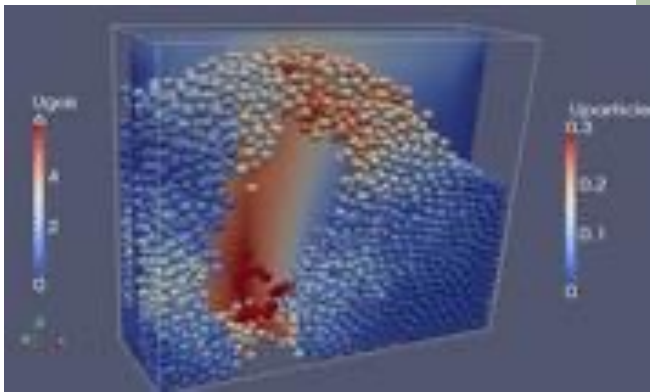


Chemistry

Biophysics



Granular Flow



MD Basics

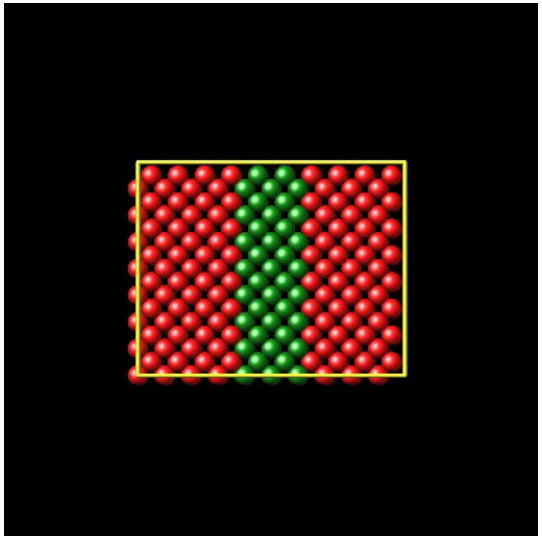
Atoms can be modeled as points (most common), finite-size spheres, or other shapes (e.g. ellipsoids)

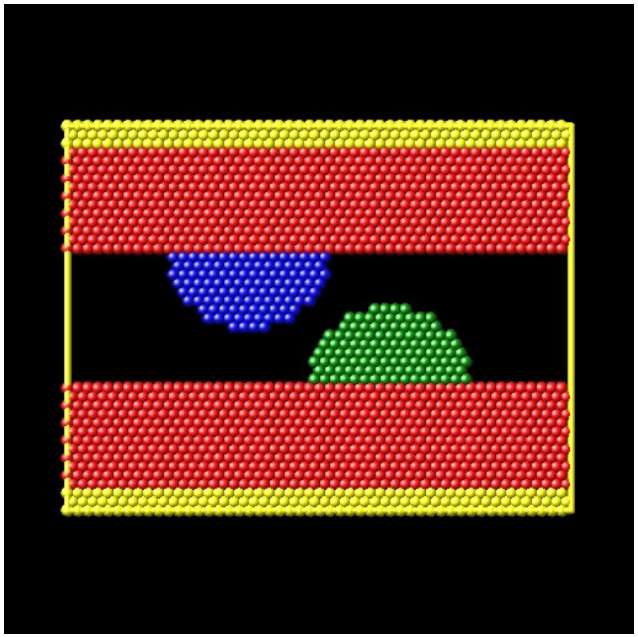
Can model atomic-scale (all-atom model) or meso/continuum scale with MD-like models

Typically use an orthogonal or triclinic (skewed) simulation cell

Commonly use periodic boundary conditions: reduces finite size effects from boundaries and simulates bulk conditions

2D Triclinic





MD Time Integration Algorithm

- Most codes and applications use variations and extensions to the Størmer-Verlet explicit integrator:
- Only second-order : $\delta E = |\langle E \rangle - E_0| \sim \Delta t^2$, but….
- **time-reversible map**: switching sign of Δt takes you back to initial state
- **measure-preserving**: Volume of differential cube (δ**v**,δ**x**) is conserved (but not shape).
- **symplectic:** Conserves sum of areas of differential parallelogram (δ**v**,δ**x**) projected onto each particular (v_i,x_i) plane

Velocity form of Størmer-Verlet

For istep < nsteps :

$$\mathbf{v} \leftarrow \mathbf{v} + \frac{\Delta t}{2}\mathbf{F}$$

$$\mathbf{x} \leftarrow \mathbf{x} + \Delta t\,\mathbf{v}$$

Compute $\mathbf{F}(\mathbf{x})$

$$\mathbf{v} \leftarrow \mathbf{v} + \frac{\Delta t}{2}\mathbf{F}$$

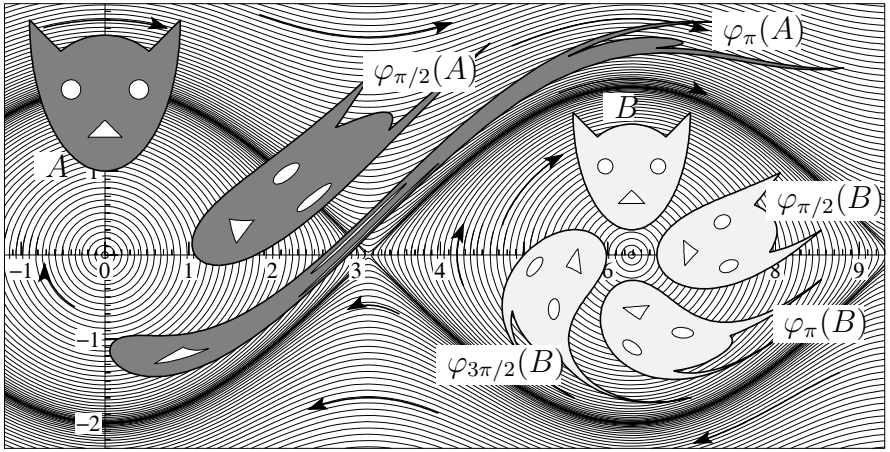
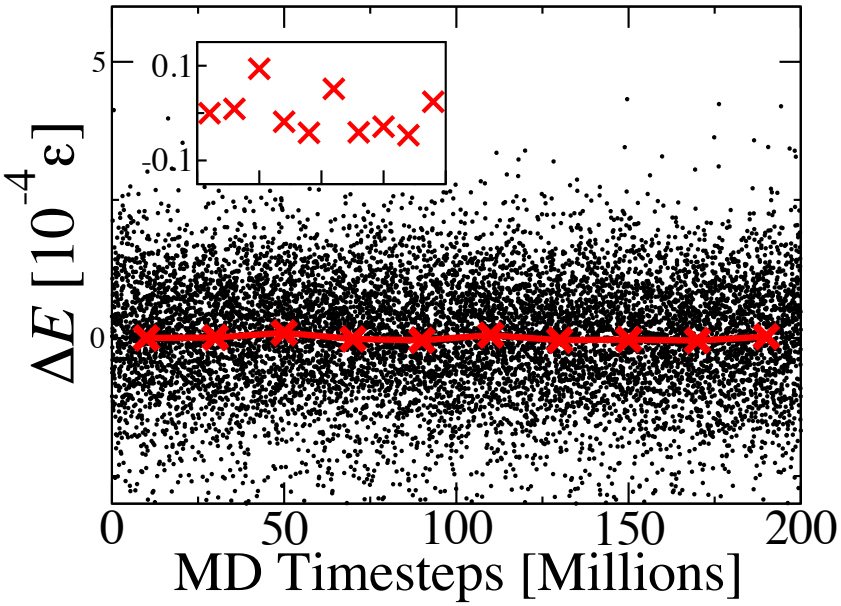


Figure 3: Area preservation of the flow of Hamiltonian systems

Ernst Hairer, Lubich, Wanner, *Geometric Numerical Integration* (2006)

MD Time Integration Algorithm

- **time-reversibility and symplecticity**: global stability of Verlet trumps local accuracy of high-order schemes
- More specifically, it can be shown that for Hamiltonian equations of motion, Størmer-Verlet exactly conserves a "shadow" Hamiltonian and $E\text{-}E_S \sim O(\Delta t^2)$
- For users: no energy drift over millions of timesteps
- For developers: easy to decouple integration scheme from efficient algorithms for force evaluation, parallelization.
- Symplectic high-order Runge-Kutta methods exist, but not widely adopted for MD



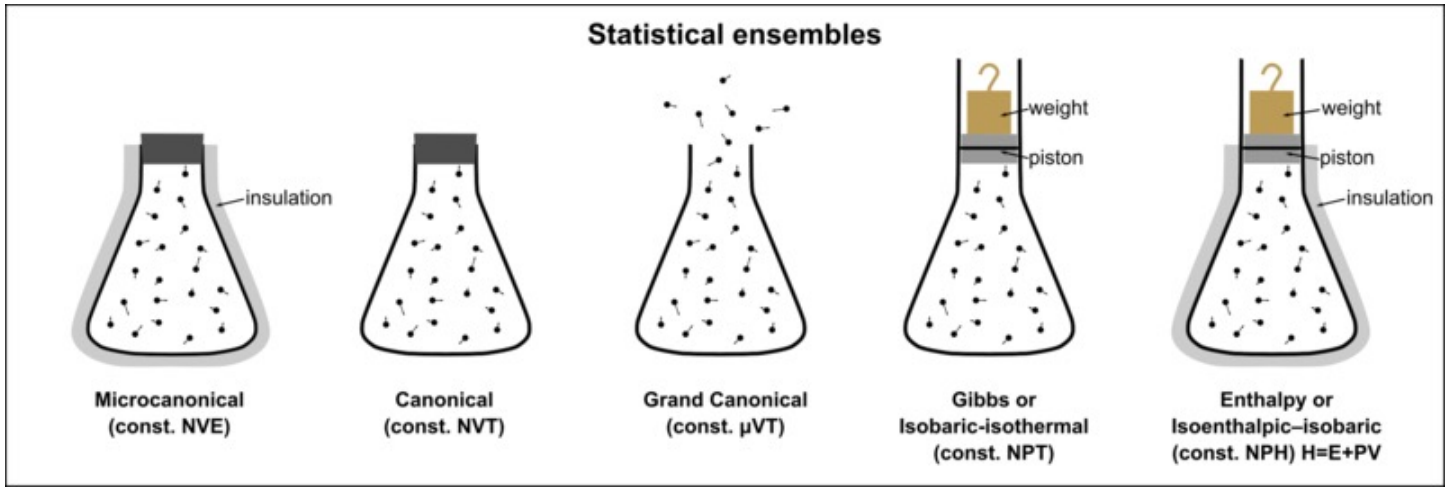
32 atom LJ cluster, 200 million MD steps, Δt=0.005, T=0.4

Statistical Mechanics Basics

Statistical Mechanics: relates macroscopic observations (such as temperature and pressure) to microscopic states (i.e. atoms)

Phase space: a space in which all possible states of a system are represented. For N particles: $6N$-dimensional phase space (3 position variables and 3 momentum variables for each particle)

Ensemble: an idealization consisting of a large number of virtual copies of a system, considered all at once, each of which represents a possible state that the real system might be in, i.e. a probability distribution for the state of the system



Thermostats and Barostats

Using the velocity-verlet time integrator gives the microcanonical ensemble (NVE). How to simulate canonical (NVT) or isothermal-isobaric (NPT) ensembles?

Temperature is related to atom velocities through statistical mechanics, pressure is related to volume of the simulation cell

Could just scale velocities and volume to the exact desired values, but this does not allow for fluctuations with a distribution typical for the ensemble

Instead Nose-Hoover style integrators are commonly used: dynamic variables are coupled to the particle velocities (thermostatting) and simulation box dimensions (barostatting)

Nose-Hoover uses a *damping* parameter specified in time units which determines how rapidly the temperature or pressure is relaxed. If the damping parameter is too small, the temperature/pressure can fluctuate wildly; if it is too large, the temperature/pressure will take too long to equilibrate

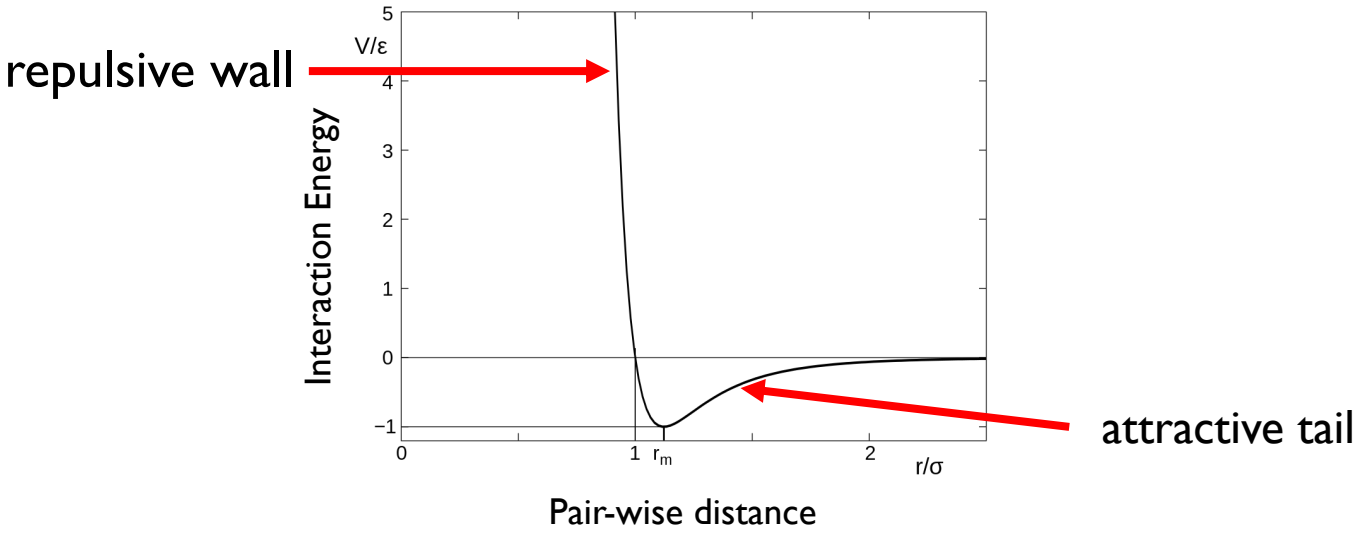
Interatomic Potentials

Quantum chemistry: solves Schrödinger equation to get forces on atoms. Accurate but very computationally expensive and only feasible for small systems

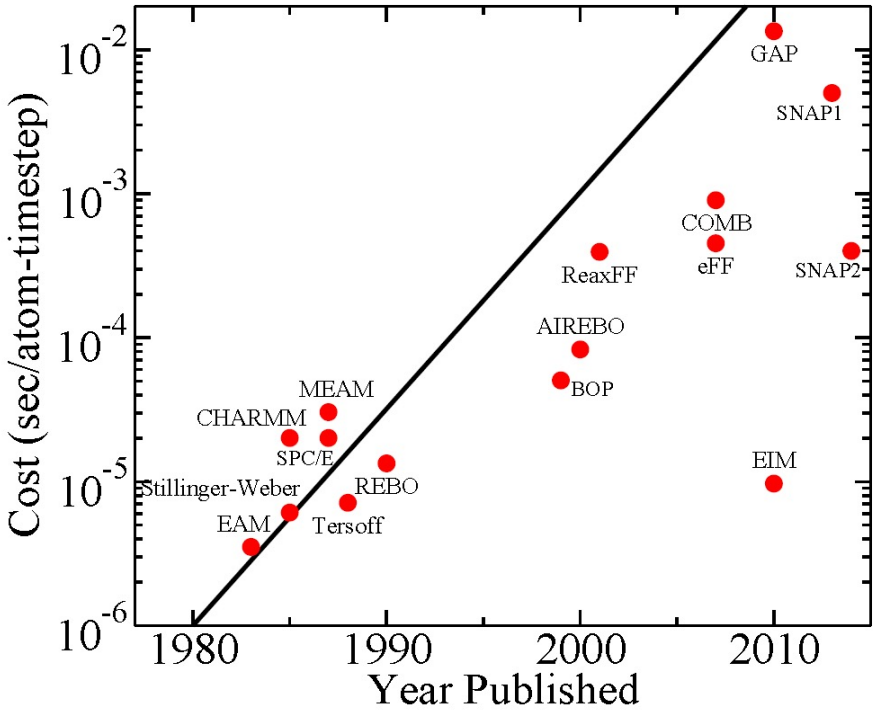
Molecular dynamics: uses empirical force fields, sometimes fit to quantum data. Not as accurate but much faster.

Typically only interact with atoms in a spherical cutoff and only consider pair-wise or three-body interactions

Lennard-Jones Potential



Accuracy = Higher Cost



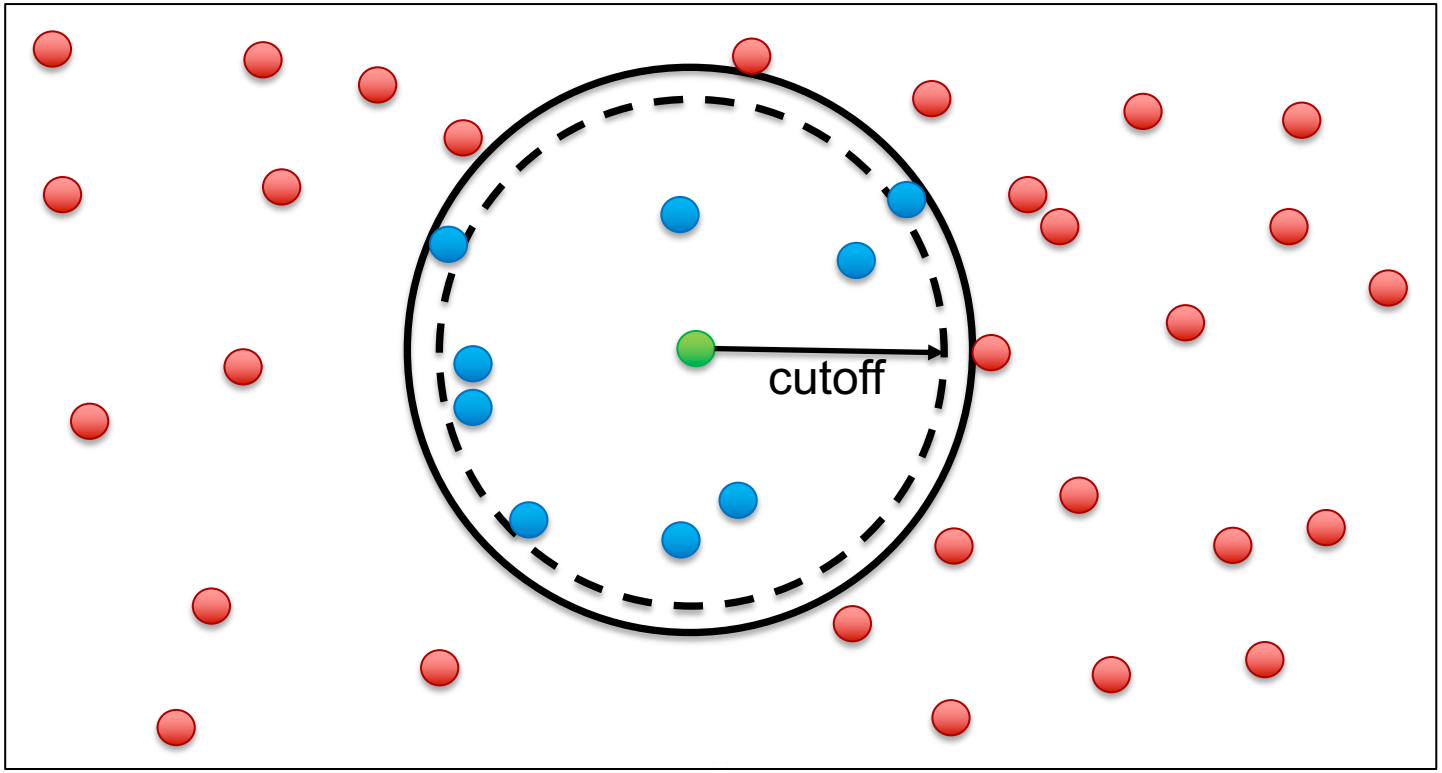
Moore's Law for Interatomic Potentials

Plimpton and Thompson, MRS Bulletin (2012).

Neighbor Lists

Neighbor lists are a list of neighboring atoms within the interaction cutoff + skin for each central atom

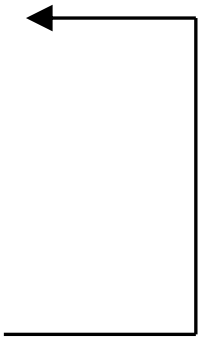
Extra skin allows lists to be built less often



Basic MD Timestep

During each timestep (without neighborlist build):

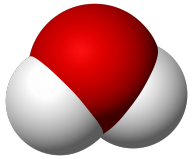
1. initial integrate
2. compute forces (pair, bonds, etc.)
3. final integrate
4. output (if requested on this timestep)



*Computation of diagnostics (i.e. thermodynamic properties) can be scattered throughout the timestep

May also occasionally build neighborlist for diagnostics

Long-Range Electrostatics (Optional)



Truncation doesn't work well for charged systems due to long-ranged nature of Coulombic interactions

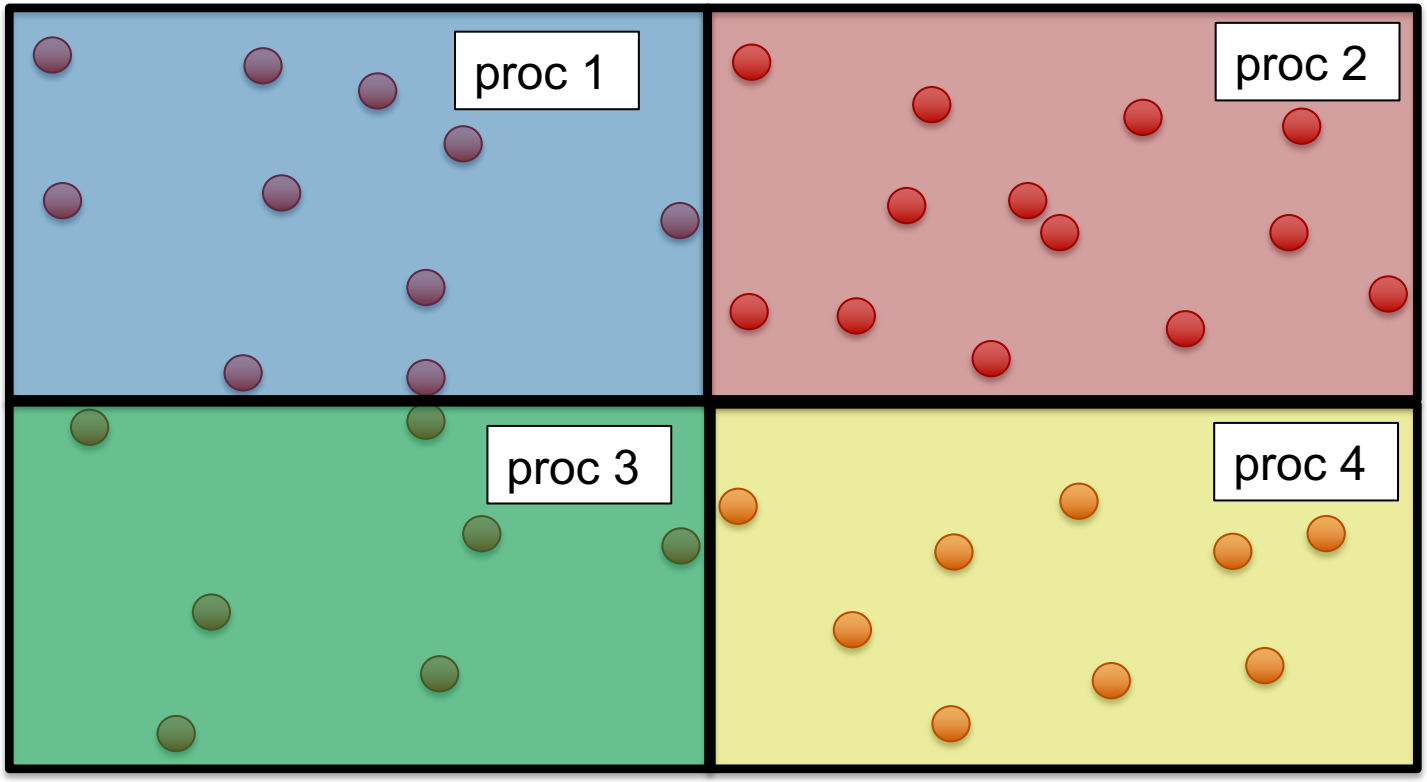
Use reciprocal-space method to add long-range electrostatics:

- Ewald Sum—uses discrete Fourier transform, potentially most accurate, but slow for large systems
- Particle-particle particle-mesh (PPPM) and Smooth particle-mesh Ewald (SPME)—interpolates atom charges to grid and uses fast Fourier Transforms (FFTs), usually fastest

Other real-space methods sometimes used: fast multipole, multilevel summation

MPI Parallelization Approach

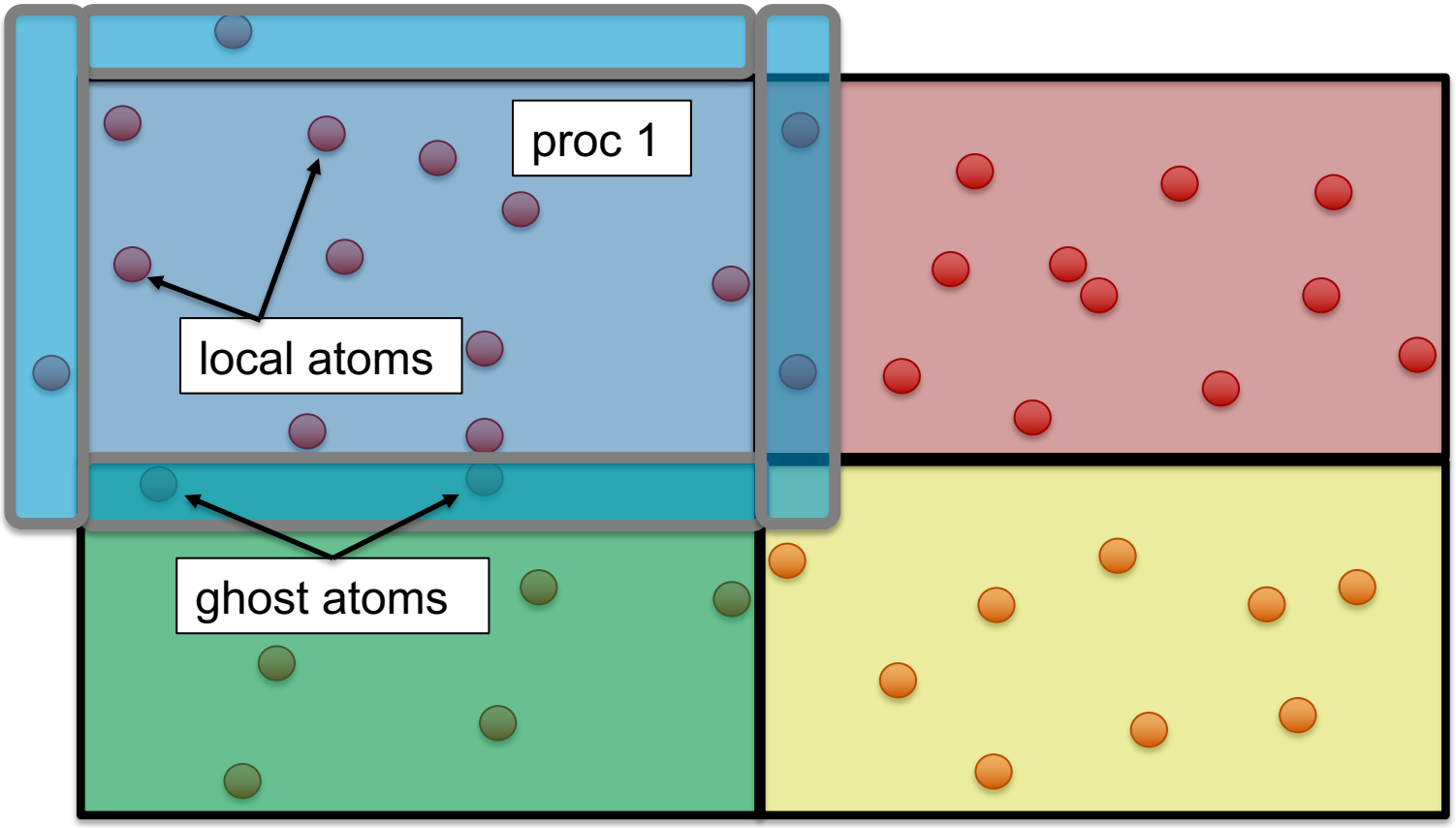
Domain decomposition: each processor owns a portion of the simulation domain and atoms therein



*This method is used by many MD codes (including LAMMPS) use, but there are other strategies as well

Ghost Atoms

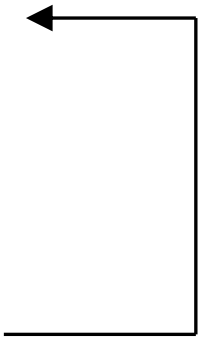
The processor domain is also extended include needed ghost atoms (copies of atoms located on other processors)



Basic MD Timestep

During each timestep (without neighborlist build):

1. initial integrate
2. compute forces (pair, bonds, etc.)
3. final integrate
4. output (if requested on this timestep)



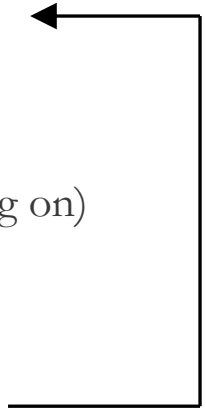
*Computation of diagnostics (i.e. thermodynamic properties) can be scattered throughout the timestep

May also occasionally build neighborlist for diagnostics

Basic MD Timestep with MPI comm

During each timestep (without neighborlist build):

1. initial integrate
2. MPI communication: send atom coordinates to ghost atoms
3. compute forces (pair, bonds, etc.)
4. MPI communication: sum atom forces from ghost atoms (if newton flag on)
5. final integrate
6. output (if requested on this timestep)



*Computation of diagnostics (fixes or computes) can be scattered throughout the timestep

Parallel MD Performance

Strong scaling: hold system size fixed while increasing processor count (# of atoms/processor decreases)

Weak scaling: increase system size in proportion to increasing processor count (# of atoms/processor remains constant)

For perfect strong scaling, doubling the processor count cuts the simulation time in half

For perfect weak scaling, the simulation time stays exactly the same when doubling the processor count

Harder to maintain parallel efficiency with strong scaling because the compute time decreases relative to the communication time

High communication overhead when strong scaling to a few 100 atoms/proc (depends on cost of the force-field)

MD parallelizes well: major parts of timestep (forces, neighbor list build, time integration) can be done in parallel through domain decomposition

MD Codes

There are several freely-available parallel molecular dynamics codes:

CHARMM, AMBER, GROMACS, NAMD, and Tinker are designed primarily for modeling biological systems. AMBER and CHARMM are the original classic codes in this genre. Gromacs, NAMD, and Tinker are more recently developed codes.

DL_POLY includes potentials for a variety of biological and non-biological materials. LAMMPS is focused on materials but versatile. HOOMD is a very fast materials MD code designed to run on GPUs.

NWChem is both a molecular dynamics and quantum code which can model a variety of materials.

We will now learn about LAMMPS in this tutorial.