SCM
Software for
Chemistry &
Materials

Molecular Dynamics: Tips, Tricks, and the Properties of Matter

Tomáš Trnka
trnka@scm.com

May 3, 2022

Introduction

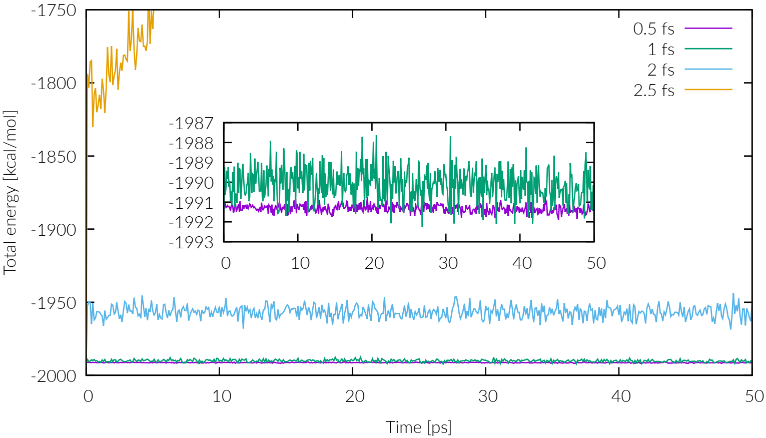
- Doing MD is easy!
  - Open AMSinput, create system, select “Molecular Dynamics”, click Run, done!
- Doing MD correctly is less easy!
  - How do I set all these knobs?
  - Are my results reliable?
  - How do I calculate property X efficiently?

The Heart of MD: The Integrator

- Molecular dynamics: Numerical integration of the equations of motion
  - Propagating positions x and velocities $\dot{x}$ in time based on accelerations $\ddot{x}$ (forces from chosen potential)
- Unlike general ODE integration methods (Runge-Kutta, ...), MD prefers integrators with special properties (time reversibility = energy conservation)
- AMS: velocity Verlet integrator
  - Symplectic method: exactly conserves a perturbed "shadow Hamiltonian"
  - Difference between real and shadow Hamiltonian proportional to $O(h^2)$
  - Global error in positions (over a given time interval) also $O(h^2)$

Integration Stability and Accuracy

300 flexible TIP3P water molecules, 298.15 K



- 0.5 and 1 fs close to each other
- 2 fs already far off and much noisier, but no obvious drift
- 2.5 fs unstable

Integration Stability and Accuracy

- Verlet integrator doesn't produce obvious energy drift even with an excessive timestep
  - Some drift due to roundoff errors only observable on >100 ps timescales
  - Any drift observed in simulations is typically due to inaccurate (noisy) forces
- Symplectic methods become unstable when timestep $h > 0.225T_p$, where T_p is the period of the fastest motion[1]:
  - O – H bond: 10 fs
  - C – C bond: 16 fs (triple) – 20 fs (double)
- Reasonable accuracy requires $h \leq 0.1T_p$

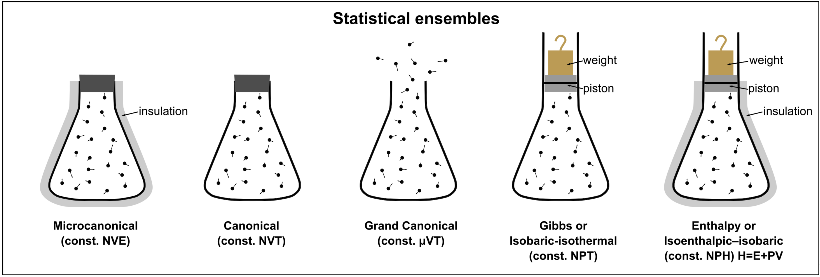
---

[1] Schlick, T., Molecular Modeling and Simulation: An Interdisciplinary Guide, 2nd Ed., Springer 2010, DOI: 10.1007/978-1-4419-6351-2, sect. 14.2.5

Ergodicity

- Ergodic hypothesis: Ensemble average of a property is the same as its time average
    - One long simulation equivalent to many independent short simulations
- An ergodic system spends equal time in equal volumes of the phase space (combinations of positions and momenta)
    - Given enough time, all possible states will be visited
    - There should be no isolated “islands” of phase space disconnected from the rest, no short exactly periodic trajectories

Common Statistical Ensembles



[2]

- Microcanonical ensemble reliable and accurate but not very realistic
- Real-world systems typically in a known temperature and volume/pressure
- Grand canonical MD not in AMS (but GCMC is)

---

[2] © User:Nzjacobmartin, Wikimedia Commons, CC BY-SA 4.0

Berendsen Thermostat

- Emulates the effect of a heat bath by rescaling velocities
- Heat transferred per timestep proportional to

$$\frac{h}{\tau}\,(T_0 - T)$$

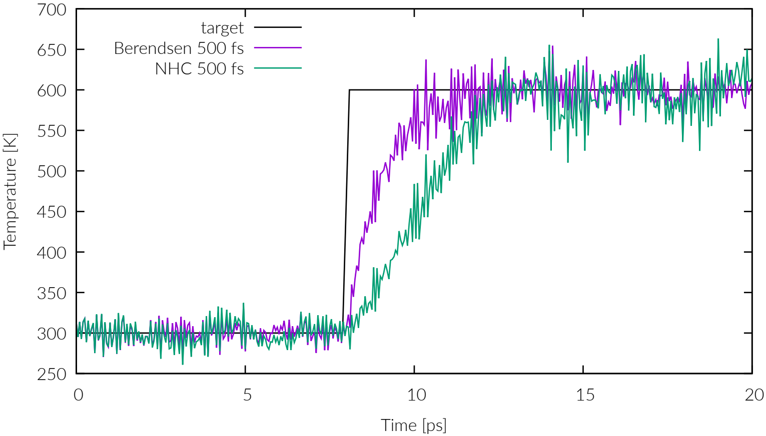
- Advantages:
  - Stable, able to cope with systems arbitrarily far from equilibrium
  - Exponential relaxation (quickly approaches target temperature) with time constant τ
- Disadvantages:
  - Does not generate correct canonical ensemble (fluctuations strongly suppressed)
  - Fundamentally incompatible with temperature replica exchange MD
  - Does not conserve energy
  - Causes severe artifacts when τ is too short (“flying ice cube effect”)

Nosé(-Hoover) Thermostat

- Explicitly simulates heat bath as an additional degree of freedom (3N+1 DoFs: atomic positions plus heat bath energy)
- Original idea by Nosé, reformulated to be more practical by Hoover
- AMS uses “Nosé-Hoover chains” (Martyna, Klein, Tuckerman): set of nested heat baths thermostatting each other (multiple additional DoFs)
- Advantages:
  - Correct canonical ensemble, conserves energy of system + heat bath
  - Relatively safe (artifact-free) across a range of time constants
- Disadvantages:
  - Only works well for systems relatively close to equilibrium
  - Original NH strongly oscillates when disturbed
  - Potentially non-ergodic when τ is too short (real issue for original NH, unlikely to matter with chains for typical systems)
  - Characteristic timescale τ only approximate

Response to Disturbance

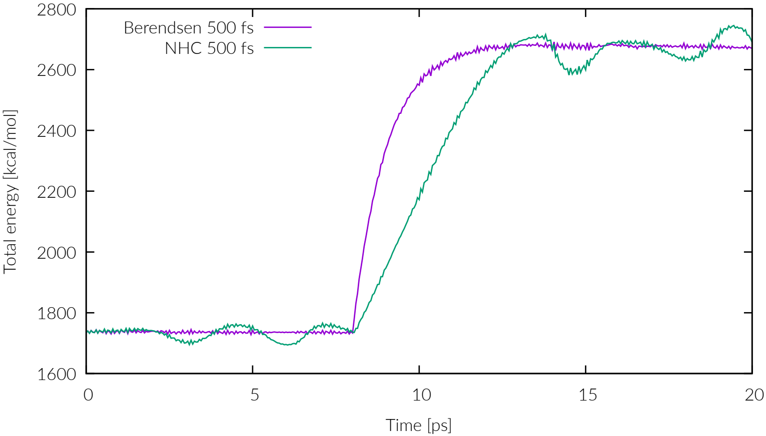
Graphite crystal, 512 atoms, UFF



- Exponential relaxation of Berendsen brings temperature faster to target
- Total energy stays nearly constant in equilibrium with Berendsen, oscillates under NHC

Response to Disturbance

Graphite crystal, 512 atoms, UFF



- Exponential relaxation of Berendsen brings temperature faster to target
- Total energy stays nearly constant in equilibrium with Berendsen, oscillates under NHC

Setting Time Constant

- Aggressive thermostatting (low τ) leads to artifacts:
  - Berendsen: flying ice cube effect (energy pumped into low-frequency modes)
  - Nosé-Hoover: potential non-ergodicity, inaccurate integration of heat bath DoFs
- Time constant of 100 fs a reasonable lower bound if strong thermostatting is required
- Any strong thermostatting can disturb natural dynamics (viscosity, diffusivity, ...)
- For equilibrium production simulations, use high enough τ, comparable with natural relaxation timescales:
  - Liquids: correlation time (2-3 ps for water)
  - Solids: phonon lifetimes (typically > 0.5 ps)

Thermostatting Heterogeneous Systems

- Thermostatting a heterogeneous system as a whole can create temperature gradients
- Typical example: “hot solvent – cold solute problem”
  - Big protein in water: force errors on water molecules higher than on protein chain
  - Unequal heating of both parts
  - Shared thermostat drains heat from both components at the same rate, undercooling the protein
- Solution: use separate thermostats for individual components
  - AMS: define a Region for solute and one for solvent, apply a dedicated thermostat to each region

Berendsen Barostat

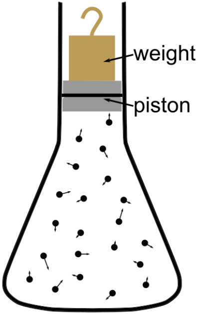
- Similar to Berendsen thermostat
  - Instead of scaling velocities according to temperature, scales volume according to pressure:

$$\frac{h}{K\tau_p}\left(p_0 - p\right)$$

- Exponential relaxation, does not generate correct ensemble (wrong fluctuations), no conserved energy
- Time constant τ_p only valid if the system has the configured bulk modulus K (AMS default: 2.2 GPa for water)
  - For other systems, adjust BulkModulus on Expert AMS panel or rescale τ_p accordingly

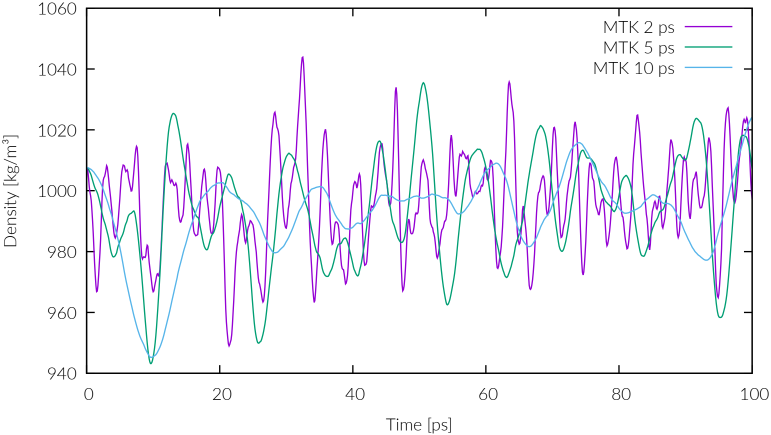
Extended Ensemble Barostats

- Like Nosé-Hoover, simulate a piston instead of a heat bath
- AMS uses the Martyna-Tobias-Klein method (derived from Andersen-Hoover and Parrinello-Rahman-Hoover)
- Correct ensemble with conserved energy
- Can work on any combination of cell dimensions and keep any subset equally scaled
- Special mode to change cell shape while keeping volume constant



Density Fluctuations

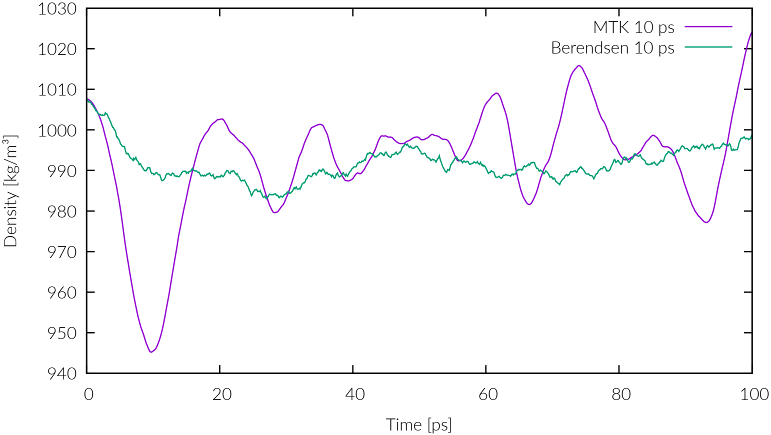
300 flexible TIP3P water molecules, 298.15 K



- Magnitude of fluctuations independent of τ_p
- Period roughly comparable to τ_p
- Berendsen does not fluctuate

Density Fluctuations

300 flexible TIP3P water molecules, 298.15 K



- Magnitude of fluctuations independent of τ_p
- Period roughly comparable to τ_p
- Berendsen does not fluctuate

Barostat Time Constant

- Like for thermostats, use high enough τ_p to minimize disturbance to system
- Barostat should work on a several times slower timescale than thermostat
    - Otherwise barostat will compress/expand system to heat/cool it
- During initial equilibration of a newly prepared system, run NVT first (no barostat) and only add a barostat once temperature is sufficiently close to target

Example Equilibration Protocol

1. Prepare starting geometry (load from file, generate using Edit → Builder for fluids or Edit → Crystal)
2. Optionally run a geometry optimization if starting geometry is very suboptimal
3. Pre-equilibrate near target temperature using NVT with Berendsen and strong coupling (100-500 fs time constant) for 10 ps or more
4. Relax to near equilibrium density in NpT with Berendsen thermostat (τ 1–2 ps) and barostat (τ_p 5 ps for water) for several tens of ps
5. Equilibrate using NHC thermostat and MTK barostat until key properties converge (total and potential energy, density)
6. Alternatively, if preparing for NVT production simulation, equilibrate with Berendsen barostat with a long time constant to reach average density

Restarting or Continuing MD Simulations

- AMS provides two ways to run a follow-up simulation
- Exact restart: “Restart” keyword, “Restart from” option on MD input panel
    - Guaranteed seamless continuation of an interrupted trajectory
    - No input settings should be changed apart from number of steps (and possibly trajectory writing options)
    - Good for recovering from crashes or extending a simulation to more steps
- New simulation starting from final geometry and velocities
    - AMSinput will offer to update geometry and set up InitialVelocities FromFile on MD input panel
    - New simulation independent from original, any input settings may be changed

Trajectory Writing Options

- Most quantities written to trajectory by default
  - Big ams.rkf for long simulations with frequent sampling
- Can be customized on “MD Trajectory Writing” panel:
  - Velocities (if not needed for VACF analysis)
  - Atomic charges (unnecessary for non-polarizable potentials)
  - Bonding tables and molecule analysis (unnecessary for nonreactive FFs)
- Velocity snapshot for restarting/continuation always written every CheckpointFrequency steps

Basic Equilibrium Properties

- Energies, temperature, volume, density: AMSmovie
- Energy + temperature statistics in text output
- Everything on ams.rkf in section “MDHistory” (Python scripting, KFbrowser, **`dmpkf`**)
- Backwards compatibility scripts for old ReaxFF text file formats
  - **`scripting/standalone/reaxff-ams/`**
  - Completely general, not just for ReaxFF

Structure

- Radial distribution function
  - Calculated by **`analysis`** or plotted in AMSmovie (MD Properties → RDF)
  - Doesn’t require long simulations or frequent sampling (few 10s of ps enough)
  - Range limited to half the box thickness

Diffusivity

- Can be calculated in two different ways:
- Green-Kubo approach
  - Integral of the velocity autocorrelation function
  - Requires frequent sampling (tens of fs between samples at most)
  - Many frames needed for converged results (large trajectory with full velocity data)
- Stokes-Einstein-Sutherland approach
  - Linear fit to $MSD(t)$
  - Frequent sampling not needed (hundreds of fs between samples enough), 50–100 ps often enough for basic accuracy
  - MSD data for $t < \tau_c$ (correlation time, 2-3 ps in water) must be discarded (correlated motion instead of random diffusion)
- Significant box size effects
  - Calculated D is underestimated due to interaction of PBC images
  - Roughly 10–15% error for r.t. water in a 20 Å box
  - Correction formula by Dünweg and Kremer[3]:

$$D = D_{\mathrm{sim}} + \frac{2.837 \mathrm{k_B} T}{6\pi\eta L}$$

---

[3] Dünweg, B.; Kremer, K. *J. Chem. Phys.* **1993**, 99, 6983–6997

Viscosity

- Significantly more demanding than diffusivity
- Green-Kubo approach
    - Integral of the ACF of off-diagonal pressure tensor components
    - AMSmovie: MD Properties → ACF, “Pressure tensor“, vector elements 4/5/6 (Voigt order: P_{yz}, P_{xz}, P_{xy})
    - ACF very noisy (>1 ns needed for reliable results), most information at very short times (tens of fs), sampling every <10 fs required (large trajectory)
    - ACF drops to zero at correlation time τ_c, only noise remains for $t > \tau_c$, integral becomes useless (random walk)
    - Picking τ_c subjective, comparing the components can help
- Tutorial available
- Coming soon: NEMD method for viscosity (steady state shearing)

Mechanical Properties

- Young's and shear moduli, Poisson ratio, yield behavior, ...
- Fit to measured stress tensor vs applied strain via MD Deformations feature
  - Tutorial available (includes analysis scripts)
  - Coming soon: Built-in analysis/fitting in AMSmovie
- Many options on MD Deformations panel
  - TargetLength (drive lattice vector lengths to given values, keeping cell shape)
  - LengthVelocity (add an increment to vector lengths every timestep)
  - TargetLattice (drive entire lattice matrix to target matrix)
  - Advanced: StrainRate (apply arbitrary strain on every timestep)
  - Deformation can be once only or oscillatory (sine modulation of lengths/lattice parameters)

Summary

- Use thermostats and barostats with care
  - Berendsen methods good for equilibration
  - Prefer NHC and MTK for production NpT simulations
  - Do not use short time constants unnecessarily, avoid τ < 100 fs
- Use Restart feature for exact continuation, InitialVelocities otherwise
- Monitor basic properties to check for equilibration/convergence
  - AMSmovie can plot time evolution of most parameters, RDF...
  - Alternatives: Python scripting, ReaxFF compatibility scripts
- Check out our tutorials for common properties
  - Diffusion (prefer Stokes-Einstein-Sutherland method, correct for box size)
  - Viscosity (Green-Kubo: long simulations, frequent sampling)
  - Mechanical properties