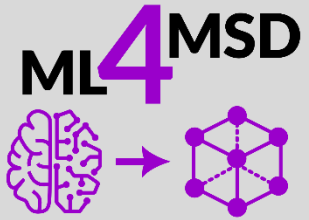# ME 5374-ST

## Machine Learning for Materials Science and Discovery

ML4MSD

Fall 2025

Asst. Prof. Peter Schindler

Lecture 13 – Atomistic Modeling 1, Molecular Dynamics

- Molecular Dynamics (MD) and Its Uses
- Initialization of MD Simulation
- Interatomic Potentials and Calculation of Forces
- Integrators and Evolution of Newtonian Dynamics, Temperature in MD
- Ensembles and Property Prediction
- MD Simulation Choices and Selection of Optimal Time Step
- Overview of MD Procedure and MD Limitations

Computational Materials Science: Accuracy / Scale Tradeoff
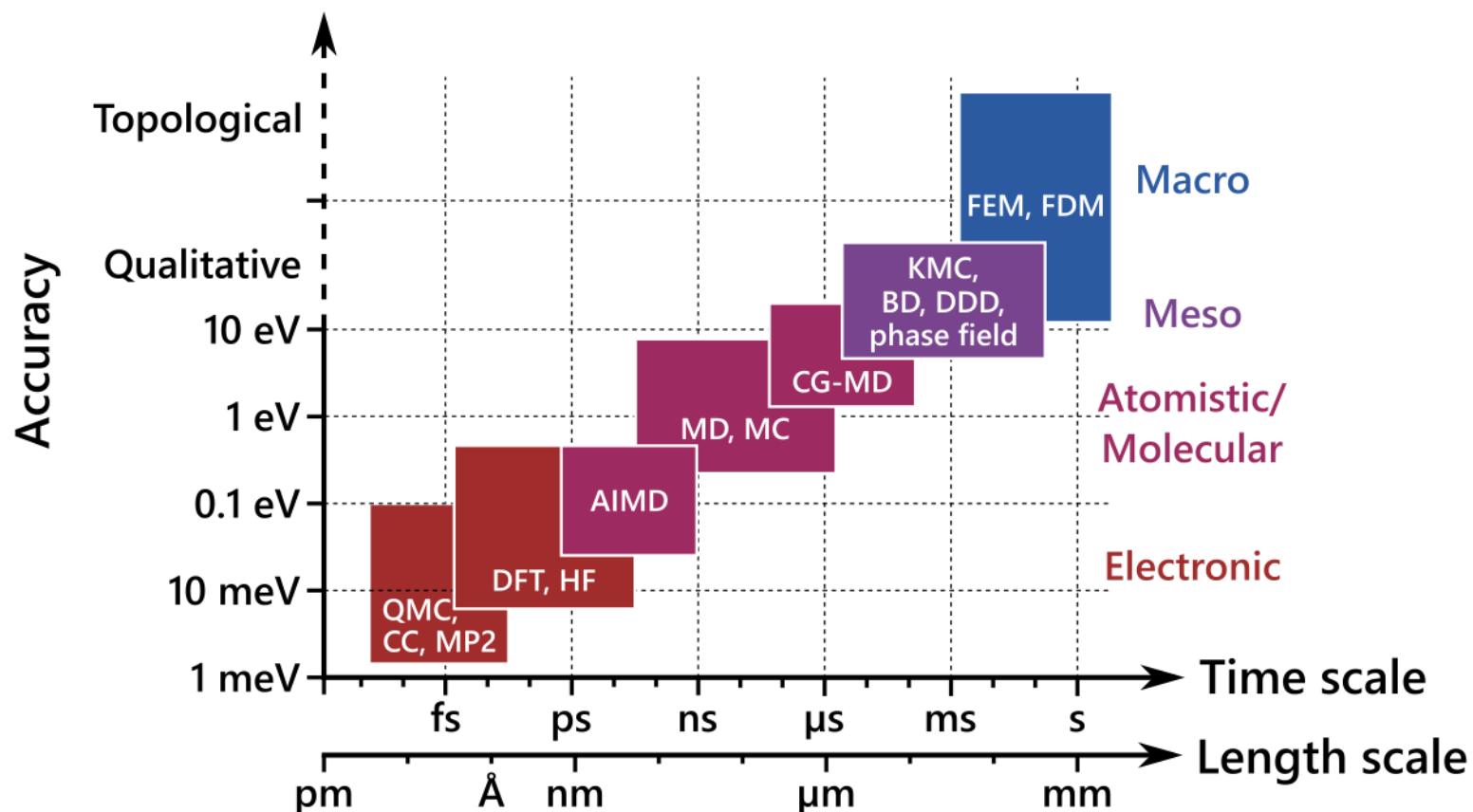
Molecular Dynamics (MD) and Its Uses

Molecular dynamics simulations enable study of atomic dynamics of materials whether they are **solids, liquids,** or **gases** at the atomic level.

How **atoms interact** and move (hence, dynamics) determines many **materials properties.**

- **Calculation of properties of systems at *thermodynamic equilibrium*:**

  Utilizes *ensemble averages* (discussed later)
  Bond lengths, lattice constants, diffusion coefficients, stress, liquid properties, radial distribution functions, charges, vibrational states (including anharmonic effects), density of states for phonons, chemical reactions

- **Calculation of properties of systems *out of equilibrium*:**

  Transport (thermal, chemical, defects, dislocations)
  Vibrational disequilibrium (e.g. collisions, sputtering, radiation damage, etc.)
  Chemical disequilibrium/dynamics (phase separation, combustion, etc.)
  Phase transformations, nucleation, crack propagation

- **Energy minimization methods to seek *global minima***

  Through *simulated annealing*

Main MD Components Needed

1. An initial atomic arrangement and velocities

2. A model for describing the interatomic interactions

3. Calculation of forces

4. Evolution of Newtonian dynamics

5. Prediction of (averaged) properties
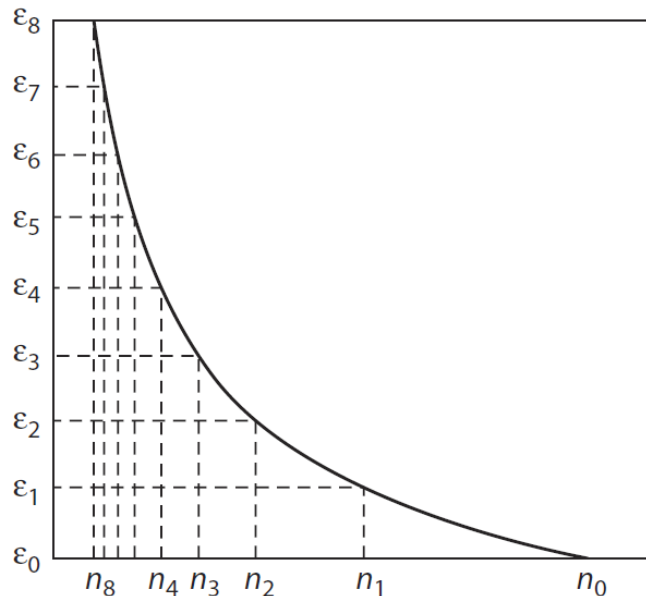
1. Initialization of MD Simulations

Initial positions:
    Depends on system/problem at hand
    For example: Ideal crystal structure (ordered), liquid/gas (randomized pos)

Initial velocities:
    Either randomly initialized or initialized **Maxwell-Boltzmann Distribution**

D. R. Gaskell and D. E. Laughlin "*Introduction to the Thermodynamics of Materials*", Taylor & Francis, 2018.

Asst. Prof. Peter Schindler

$$n_i = \frac{ne^{-\beta\varepsilon_i}}{Z}$$

$$Z = \sum_{i=0}^{r} e^{-\beta\varepsilon_i}$$

$$\beta = \frac{1}{k_B T}$$

Partition Function

Boltzmann constant:

$$k_B = \frac{R}{N_O} = \frac{8.3144(6)}{6.0221 \times 10^{23}} = 1.38065 \times 10^{-23} \, \text{J/K}$$
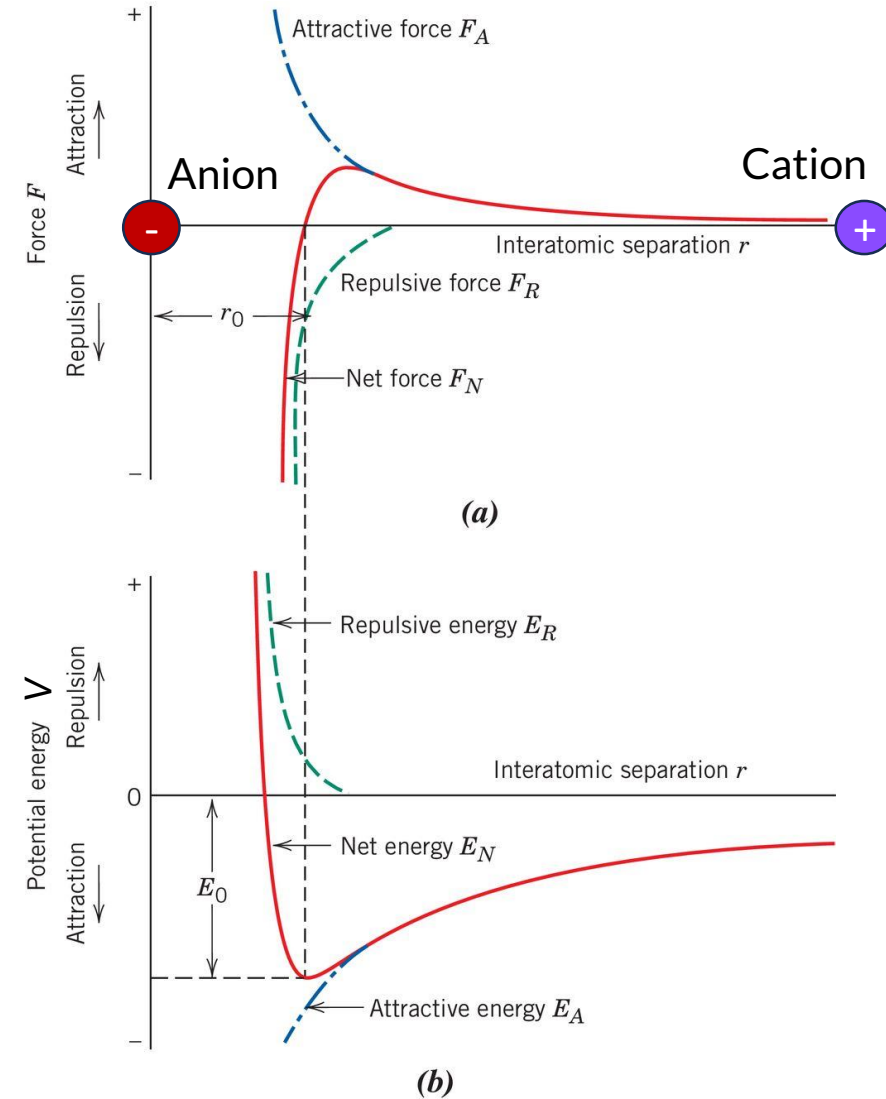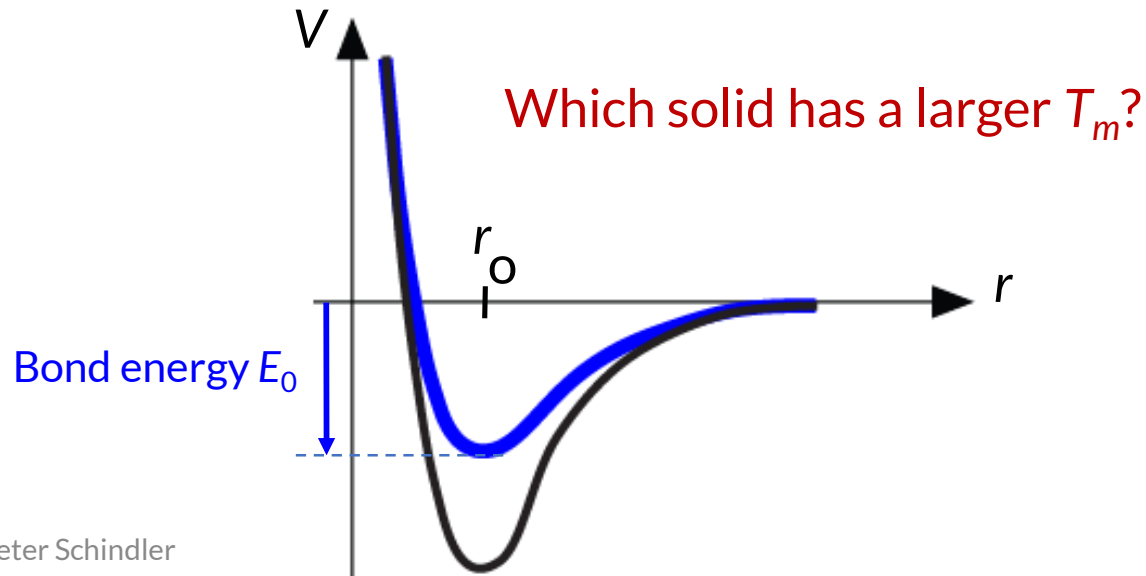
2. Interatomic Potentials

Attractive (e.g., Coulomb, dipole fluctuation)
+
Repulsive (Pauli exclusions) forces

$$V(r') = \int_{\infty}^{r'} F \cdot dr \qquad F = \frac{dV(r)}{dr}$$

Which solid has a larger T_m?

Bond energy E_0

r_o

V

r

(a)

(b)

Anion — Cation +

Asst. Prof. Peter Schindler

Common Two-Body Potentials

- Lennard-Jones - van der Waals interactions (noble gases: Ar, Xe, etc.)

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$

- Morse potentials – covalent interactions

$$V(r) = \epsilon \left(1 - e^{-a(r-r_0)} \right)^2$$

- Coulomb potentials – ionic interactions (NaCl, etc.)

Asst. Prof. Peter Schindler

Failure of Two-Body Interatomic Potentials

Two body potentials **favor high coordination number structures** (fcc, hcp Z=12)
Because cohesive energy is maximized by *maximizing number of bonds.*
Most pure metals are observed in these high coordination structures.

However,...
- Dimensionless material properties show deviation from LJ behavior in metals.
- Suggests **bond energy is nonlinear function of coordination** in metals.

T_m=melting temperature
E_c=cohesive energy
E_v=vacancy energy
C_{12}/C_{44}=ratio of shear elastic coefficients

Property	Cu	Ag	Pt	Au	LJ
$E_c/k_B T_m$	30	28	33	33	13
E_v/E_c	0.33	0.36	0.26	0.25	~ 1
C_{12}/C_{44}	1.5	1.9	3.3	3.7	1

These match well for
noble gases, though!

Dr. Furio Ercolessi

Asst. Prof. Peter Schindler

Beyond Two-Body Interatomic Potentials

Three body potentials to treat covalent bonds:
*   Stillinger-Weber (diamond Si, C, sp3)

Many-body potentials:
• Reactive bond order potentials (REBO) for environment dependent interactions and bond breakage
• Embedded atom method for metals (spherical potential)
• Modified embedded atom method for covalent bonding (nonspherical potential)

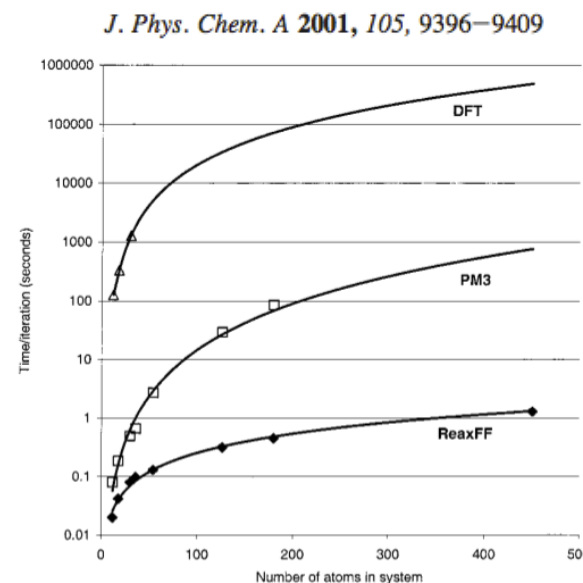Popular state-of-the-art empirically fitted potentials:
ReaxFF

Figure 25. Influence of molecular system size on computer time. Calculations were performed on single- and multimolecular aromatic systems using one processor of a Silicon Graphics Origin-computer (R10000 processor, 195 MHz).

3. Force Calculation

- Need to evolve Newtonian dynamics

$$m_i \frac{d^2}{dt^2}\vec{r_i} = -\frac{\partial}{\partial \vec{r_i}} V(\{\vec{r_i}\}_{i=1,N})$$

- MD equations of motion reduce to $6N$ coupled 1st order ODEs
- **Computationally Expensive**: Interaction of N particles with nearest-neighbor
- *With tricks:* Expense scales as $O(N)$ rather than $O(N^2)$

Need to numerically **integrate over time to evolve atom positions and velocities**

Integration schemes for MD:
Verlet, Velocity Verlet, Leapfrog Verlet
Predictor corrector
Methods based on Trotter expansion (Tuckerman et al, JCP 97, 1990, (1992).)

Asst. Prof. Peter Schindler

4. Dynamics through Force Integration (Verlet)

Taylor Expansions:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \Delta t \cdot \vec{v}_i(t) + \frac{1}{2}\Delta t^2 \cdot \vec{a}_i(t) + \dots$$

$$\vec{r}_i(t - \Delta t) = \vec{r}_i(t) - \Delta t \cdot \vec{v}_i(t) + \frac{1}{2}\Delta t^2 \cdot \vec{a}_i(t) + \dots$$

$$\overline{\vec{r}_i(t + \Delta t) + \vec{r}_i(t - \Delta t) = 2\vec{r}_i(t) + \frac{1}{2}\Delta t^2 \cdot \vec{a}_i(t) + \dots}$$

$$\vec{v}_i(t) = \left.\frac{\mathrm{d}\vec{r}_i(t')}{\mathrm{d}t'}\right|_{t'=t}$$

$$\vec{a}_i(t) = \left.\frac{\mathrm{d}^2\vec{r}_i(t')}{\mathrm{d}t'^2}\right|_{t'=t}$$

$$= -\frac{1}{m_i}\frac{\partial}{\partial\vec{r}_i}V(\{\vec{r}_i\}_{i=1,N})$$

Verlet Integrator $\quad \boxed{\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{1}{2}\Delta t^2 \cdot \vec{a}_i(t) + O(\Delta t^3)}$ (no velocity required!)

$$\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t) = 2\Delta t \cdot \vec{v}_i(t) + O(\Delta t^3)$$

$$\boxed{\vec{v}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2\Delta t} + O(\Delta t^2)}$$

(needed for kinetic energy and T)

Asst. Prof. Peter Schindler

Temperature in MD

- Equipartition of energy over all degrees of freedom.
- The **average kinetic energy per degree of freedom** defines the **temperature** in a classical many-body system:

$$\left\langle \frac{1}{2}mv^2 \right\rangle = \frac{1}{2}k_B T$$

- Instantaneous temperature calculated as $\quad T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_f}$

- Degrees of freedom $N_f=3N-3$ for sys of N particles (3D), fixed total momentum
- Relative fluctuations in T are 1/sqrt(N_f).
  For $N_f=10^2$-10^3, that's about 5-10% in T.

Asst. Prof. Peter Schindler

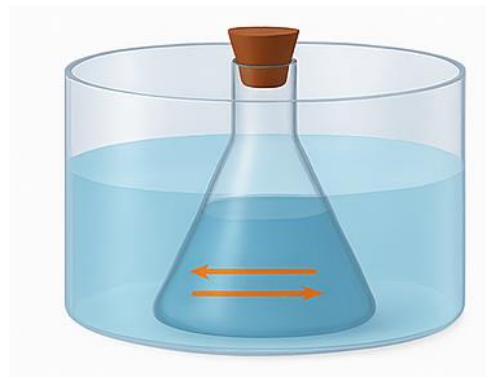# 5. Property Prediction for Thermodynamic Ensembles
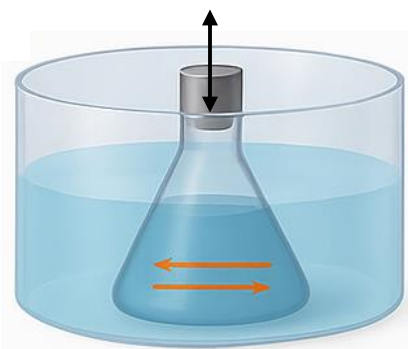

Isolated System
(*N, V, E*)
Microcanonical Ensemble

Closed System
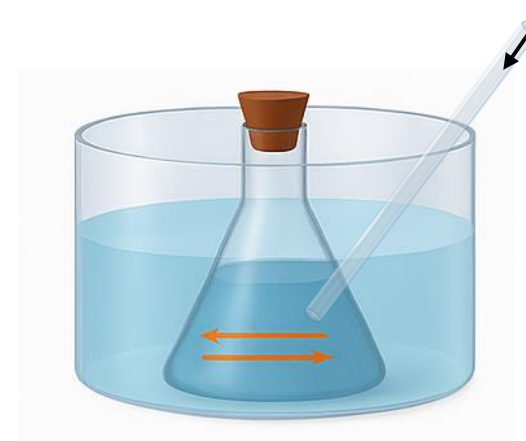in contact with heat bath
(*N, V, T*)
Canonical Ensemble

Closed System
in contact with heat bath
and volume reservoir
(*N, P, T*)
Isothermal-Isobaric Ensemble

Open System
in contact with heat bath
(*μ, V, T*)
Grand-Canonical Ensemble

Asst. Prof. Peter Schindler

Inspired by https://indico.ictp.it/event/8718/material/slides/2.pdf

Calculation of Ensembles Averages

How to calculate the statistical average of some quantity at equilibrium:

1. Pick initial atom positions and velocities
   (establish an initial starting point in phase space).
2. Time evolve system using integration scheme.
   System will likely be out of equilibrium initially.
3. **Run until equilibrium is achieved**.
4. Start averaging the quantity of interest over time.
   Longer times result in smaller uncertainty in the average.

For *ergodic* systems:

$$\langle A \rangle = \frac{\int A \exp(-\beta E)\,\mathrm{d}\vec{r}\,\mathrm{d}\vec{p}}{\int \exp(-\beta E)\,\mathrm{d}\vec{r}\,\mathrm{d}\vec{p}} \quad \Leftrightarrow \quad \bar{A} = \frac{1}{T}\int_0^T A(t)\,\mathrm{d}t$$

Further details for various ensembles: https://ase-lib.org/ase/md.html

MD Simulation Choices

- Initial conditions/structure and size
- Boundary conditions
- Interatomic potentials
- Integrator
- Ensemble
- Time step size
- Temperature ramp/schedule (for annealing studies)
- Property of interest and how to determine it from MD trajectories
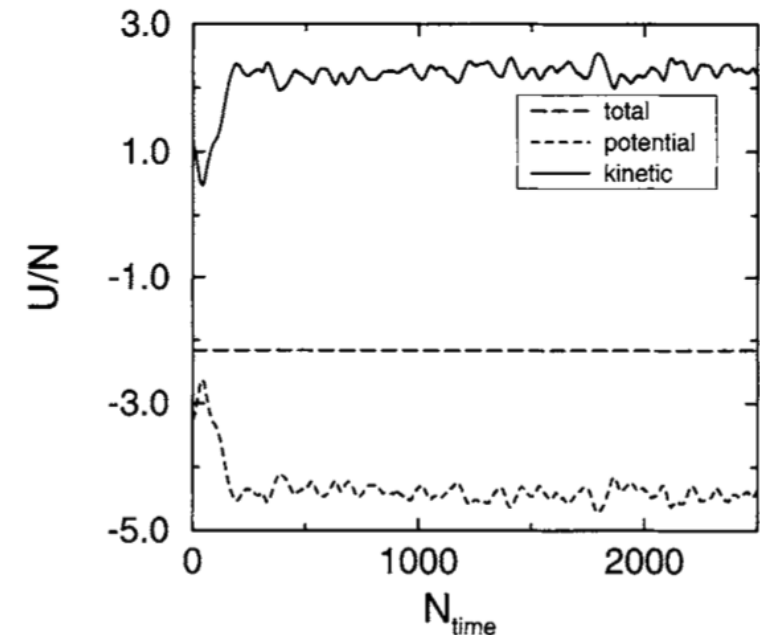
How to Set the Time Step

Balance of accuracy vs. computational cost
- Time step too large: Poor accuracy (poor energy/momentum conservation)
- Timestep too small: Calculation runs slower than necessary
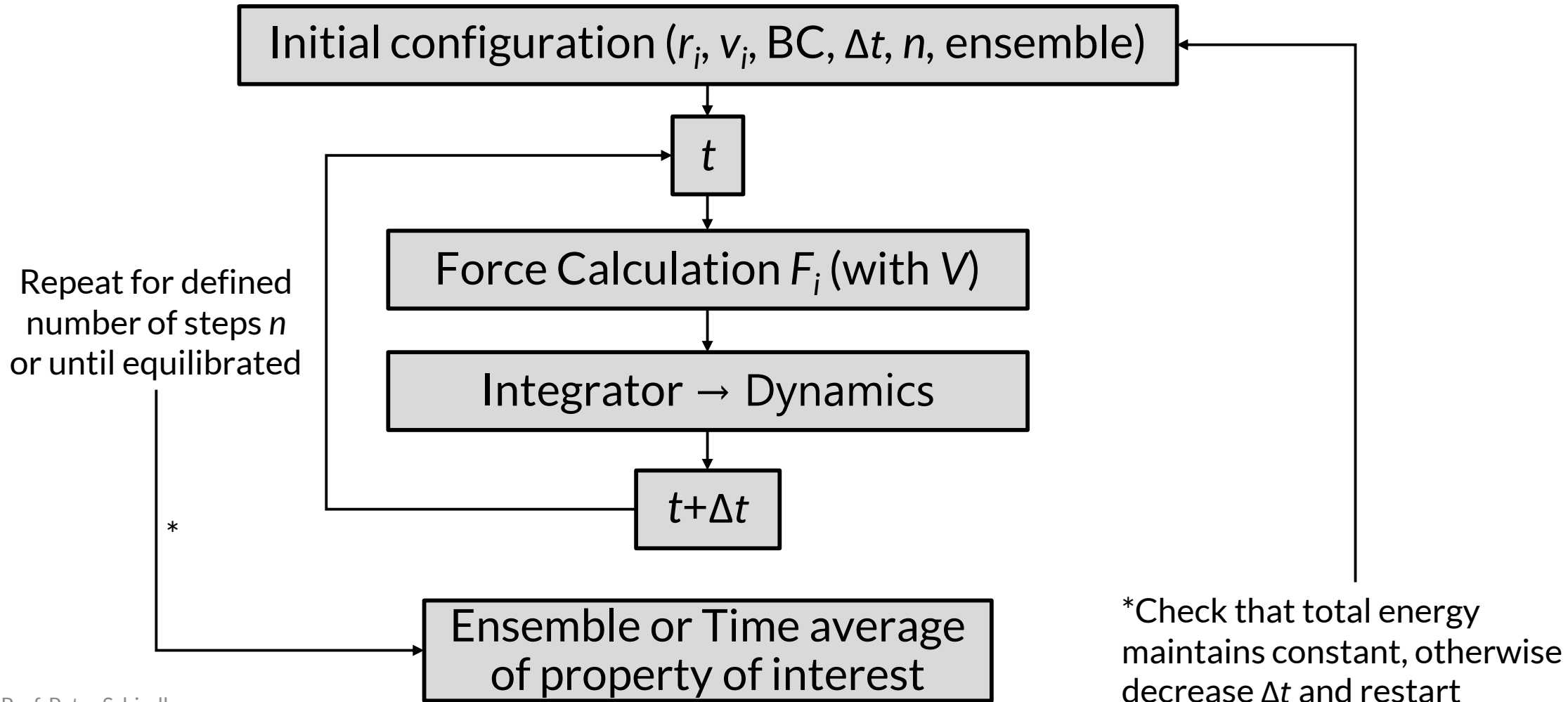- Typical time steps range from 0.1 – 5 fs (femtosecond = 10^{-15} sec)

A good rule of thumb:
The time step should be smaller than 1/10th of the highest frequency atomic vibration period

- Materials like diamond (C) or systems containing H tend to have high vibrational frequencies requiring short time steps
- May need to make it shorter for systems with nonlinear vibrations (high temperatures) or more extreme atomic collisions

Asst. Prof. Peter Schindler

Overview of MD Simulation Procedure

Initial configuration (r_i, v_i, BC, Δt, n, ensemble)

t

Force Calculation F_i (with V)

Integrator → Dynamics

$t+\Delta t$

Repeat for defined number of steps n or until equilibrated

*

Ensemble or Time average of property of interest

*Check that total energy maintains constant, otherwise decrease Δt and restart

Asst. Prof. Peter Schindler

Limitations of MD

- Because it has atomic resolutions, it's computationally expensive to run.

- The atomic dynamics follows Newton's law, meaning it's **classical**.
  In real world, (especially at low T) quantum mechanics dominates.
  Either we do some quantum correction after MD simulations
  or we study materials at high temperatures.
  This threshold of temperature is called *Debye temperature.*

- We can only study **atomic** dynamics – no electronic dynamics.
  For metals, electrons dominate thermal conductivity ($k = k_{phonon} + k_{electron}$).

Computational Limitations: MD vs. DFT

For analytical potentials:

- Typical simulations: Thousands to millions of atoms and 10-100 ps timescales

- Big simulations (need parallel computer):
  Billions of atoms or 1-10 nanosecond timescales

For quantum energy models (DFT):

- Fewer than 100 atoms and 1-10 ps

- Big simulations (need parallel computer):
  Several hundreds of atoms and 100 ps

Lecture Feedback
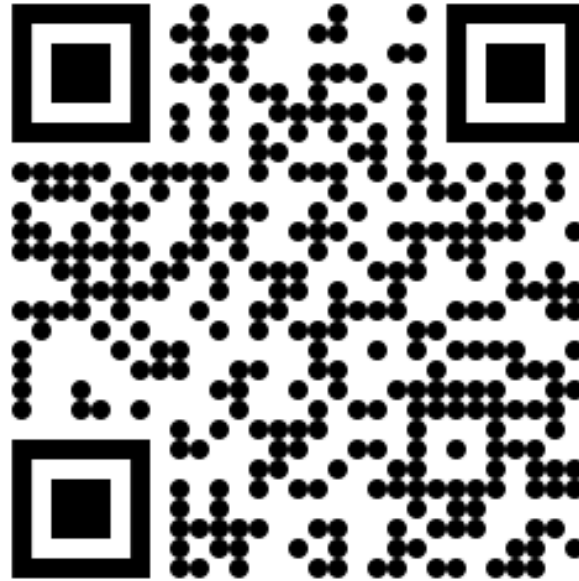
Please, scan the QR code and take a minute to let me know how the lecture was and mention any **feedback/questions**

This form is **anonymous!**