

Practical Course on Molecular Dynamics and Trajectory Analysis

Episode 1: Introduction (environment and data)

Jordi Villà i Freixa

Universitat de Vic - Universitat Central de Catalunya
Facultat de Ciències, Tecnologia i Enginyeries (FCTE)

jordi.villa@uvic.cat

MD Course and Trajectory Analysis
Concepcion, January 2026

- 1 Episode 1: Introduction (environment and data)
 - Overview and scales
 - State variables and states
 - Entropy and free energy
 - Ensembles and partitioning
 - Hamiltonian and dynamics
 - Forces and gradients
 - Partitioning and probabilities
 - Boltzmann distribution
 - Phase space and ergodicity
 - Potential energy surface
 - Force fields
 - Environment and data
 - OpenMM application layer

Episode objectives

- Link thermodynamics, statistical mechanics, and simulation.
- Define variables, state functions, and observables.
- Establish the data flow and working environment.

Time and length scales

$$\Delta t_{\text{vibrations}} \sim 10^{-15} \text{ s},$$

$$\Delta t_{\text{conformations}} \sim 10^{-9} \text{ s to } 10^{-3} \text{ s},$$

$$L_{\text{atoms}} \sim 10^{-10} \text{ m}, \quad L_{\text{proteins}} \sim 10^{-8} \text{ m}.$$

- Atomic simulation resolves ultra-fast scales.
- Statistics link microstates to macroscopic averages.

Microscopic model

$$\text{State} = (\mathbf{r}_1, \dots, \mathbf{r}_N, \mathbf{p}_1, \dots, \mathbf{p}_N),$$
$$\mathbf{p}_i = m_i \dot{\mathbf{r}}_i.$$

- The system is defined by $6N$ phase-space coordinates.
- Potential energy captures the interactions between particles.

From micro to macro

$$\langle A \rangle = \int A(\mathbf{r}, \mathbf{p}) \rho(\mathbf{r}, \mathbf{p}) d\mathbf{r} d\mathbf{p},$$

$$\bar{A} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T A(t) dt.$$

- Ergodicity links time averages with ensemble averages.

Simulation pipeline

- ① System preparation (topology, coordinates, parameters).
- ② Define the Hamiltonian and boundary conditions.
- ③ Time integration and trajectory generation.
- ④ Statistical analysis of observables.

State variables

$$P, V, T, U, H = U + PV, G = H - TS, A = U - TS.$$

- State functions depend only on the macroscopic state.
- Path functions include work W and heat Q .

First law and work

$$dU = \delta Q + \delta W,$$
$$\delta W = -P dV + \sum_i f_i dx_i.$$

- Internal energy changes through heat or work exchange.
- In simulation, δW arises from forces on particles.

Second law and entropy

$$dS \geq \frac{\delta Q}{T},$$
$$S = k_B \ln \Omega.$$

- Ω is the number of microstates compatible with the macrostate.

Thermodynamic potentials

$$G = U + PV - TS,$$
$$A = U - TS.$$

- G minimizes at constant T, P ; A minimizes at constant T, V .

$$dG = -S dT + V dP + \sum_i \mu_i dN_i,$$
$$dA = -S dT - P dV + \sum_i \mu_i dN_i.$$

- Partial derivatives give measurable observables.

Microstates and macrostates

$$S = k_B \ln \Omega, \quad \Omega = \sum_{\text{micro}} 1.$$

- Entropy increases favor states with more configurations.

Free energy and equilibrium

$$\Delta G = \Delta H - T\Delta S,$$

$$\Delta A = \Delta U - T\Delta S.$$

- At equilibrium, $\Delta G = 0$ for processes at fixed T, P .

Chemical potential

$$\mu_i = \left(\frac{\partial G}{\partial N_i} \right)_{T, P, N_{j \neq i}} .$$

- Controls matter exchange and phase equilibrium.

$$P(\text{estado}) \propto e^{-\beta G}, \quad \beta = 1/(k_B T).$$

- Free energy governs the statistical weight of macrostates.

Example: equilibrium $A \rightleftharpoons B$

$$K = e^{-\beta \Delta G},$$
$$\Delta G = -k_B T \ln K.$$

- Relative stability is expressed in terms of ΔG .

Classical ensembles

- Microcanonical (NVE): N, V, E fixed.
- Canonical (NVT): N, V, T fixed.
- Isothermal-isobaric (NPT): N, P, T fixed.

Canonical partition function

$$Z = \frac{1}{h^{3N} N!} \int e^{-\beta H(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}.$$

- Central to deriving free energies and averages.

$$\langle A \rangle = \frac{1}{Z} \int A e^{-\beta H} d\mathbf{r} d\mathbf{p}.$$

- In simulation, they are estimated via time averages.

$$\Delta = \int dV e^{-\beta P V} Z(N, V, T),$$
$$G = -k_B T \ln \Delta.$$

- Includes volume fluctuations controlled by the barostat.

Thermodynamic fluctuations

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2},$$
$$\kappa_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle}.$$

- Fluctuations connect statistics with macroscopic response.

Classical Hamiltonian

$$H(\mathbf{r}, \mathbf{p}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}).$$

- Split into kinetic and potential energy.

Hamilton's equations

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i},$$
$$\dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i}.$$

- Equivalent to Newton's second law.

Newton's equations

$$m_i \ddot{\mathbf{r}}_i = -\nabla_i U(\mathbf{r}).$$

- Force equals the negative gradient of the potential.

Energy conservation

$$\frac{dH}{dt} = 0 \quad (\text{ideal NVE}).$$

- Numerical integration introduces controllable errors.

$$\frac{d\rho}{dt} = \{\rho, H\} = 0 \quad (\text{Liouville equation}).$$

- Probability density is conserved by the Hamiltonian flow.

Forces from the potential

$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}).$$

- In MD, the dominant cost is evaluating U and \mathbf{F}_i .

Pair potentials

$$U = \sum_{i < j} u(r_{ij}), \quad r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|.$$

- Simplify the total energy using pairwise interactions.

Example: Lennard-Jones

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

- Short-range repulsion, medium-range attraction.

Electrostatic energy

$$u_C(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}.$$

- Dominant in biomolecular systems.

Gradients and Hessian

$$\nabla U \equiv \left(\frac{\partial U}{\partial r_1}, \dots, \frac{\partial U}{\partial r_{3N}} \right),$$

$$\text{Hess} = \nabla \nabla U.$$

- The Hessian describes local curvatures (normal modes).

Kinetic/potential separation

$$Z = Z_{\text{kin}} Z_{\text{conf}},$$
$$Z_{\text{kin}} = \prod_i \left(\frac{2\pi m_i}{\beta h^2} \right)^{3/2}.$$

- Z_{conf} depends only on $U(\mathbf{r})$.

Configuration distribution

$$P(\mathbf{r}) = \frac{e^{-\beta U(\mathbf{r})}}{Z_{\text{conf}}}.$$

- Basis for Monte Carlo and thermostatted MD sampling.

Configurational free energy

$$A = -k_B T \ln Z_{\text{conf}} + \text{const.}$$

- Links sampling to ΔA between states.

Velocity distribution

$$P(\mathbf{v}) \propto e^{-\beta \sum_i \frac{1}{2} m_i v_i^2}.$$

- Allows initializing velocities at temperature T .

Equipartition theorem

$$\langle K \rangle = \frac{3N}{2} k_B T.$$

- Each quadratic degree of freedom contributes $\frac{1}{2} k_B T$.

$$P(E) = \frac{1}{Z} e^{-\beta E}.$$

- The exponential weight penalizes high-energy states.

Maxwell-Boltzmann distribution

$$f(v) = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} v^2 e^{-\frac{mv^2}{2k_B T}}.$$

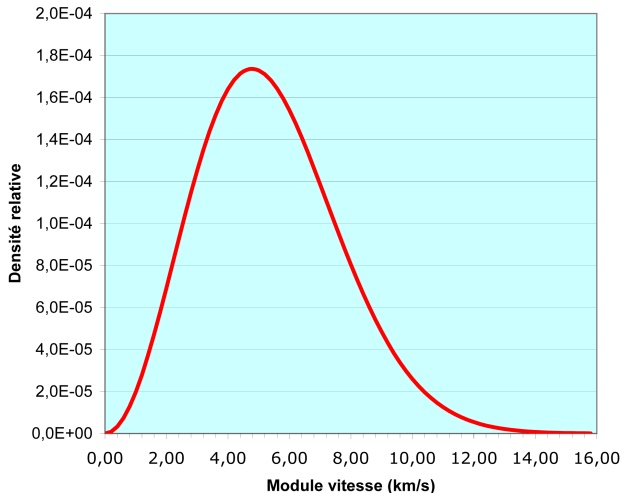
- Characterizes the equilibrium velocity distribution.

Statistical interpretation

- Most particles cluster around v_m .
- The width grows with T .

Graphic example

**Distribution des vitesses
des molécules d'hélium à 5500 K
(Distribution de Maxwell-Boltzmann)**



Simulation application

- Initializing velocities according to $f(v)$ prevents thermal bias.
- Lets us check that the thermostat reproduces equilibrium.

$$\Gamma = (\mathbf{r}, \mathbf{p}) \in \mathbb{R}^{6N}.$$

- Each point represents a full microstate.

Liouville equation

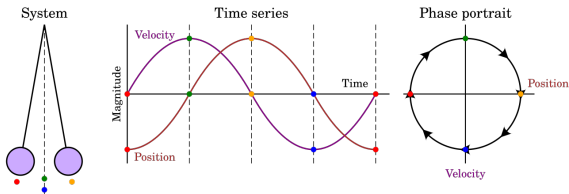
$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0.$$

- Volume in Γ is conserved (Liouville theorem).

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T A(t) dt = \langle A \rangle_{\text{ensemble}}.$$

- Key assumption for replacing ensemble averages with time averages.

Trajectories in Γ



Source: Wikimedia Commons (CC BY-SA 4.0). [2]

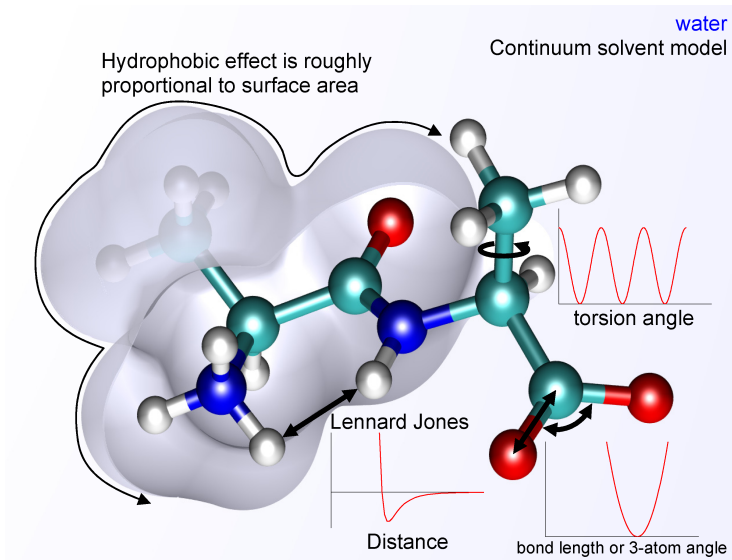
Physical interpretation

- Closed orbits: periodic motion.
- Open orbits: diffusion in high-energy regions.

$$U(\mathbf{r}) : \mathbb{R}^{3N} \rightarrow \mathbb{R}.$$

- Describes the energy landscape where the system evolves.

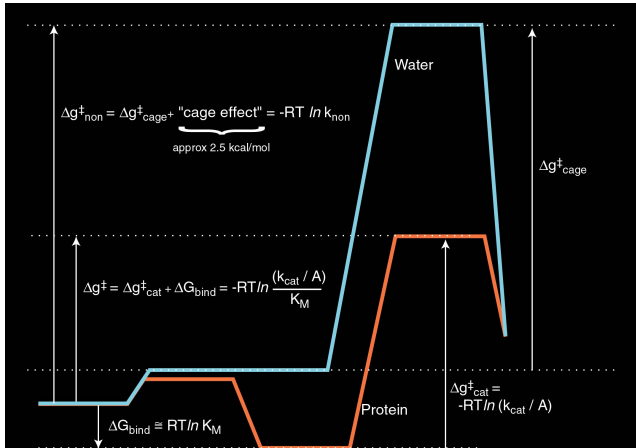
Molecular mechanics potential



Visualizing the typical multi-basined MM surface for a small system.

Enzymes and reactive Hamiltonians

- Enzymatic catalysis reshuffles bonds, so the Hamiltonian must describe bond breaking/forming events beyond fixed-topology potentials.
- Reactive force fields or QM/MM fragments introduce extra terms that depend on electronic reorganization.



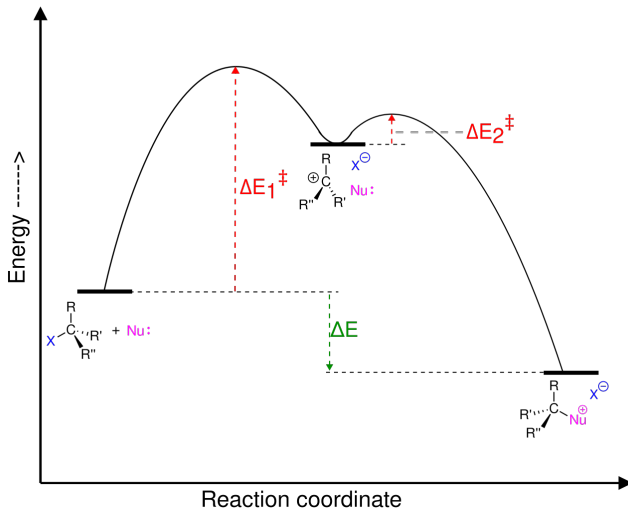
$$\xi = \xi(\mathbf{r}), \quad F(\xi) = -k_B T \ln P(\xi).$$

- Projection that summarizes the dynamics in a key coordinate.

Barriers and metastable states

- Local minima: metastable conformations.
- Barriers control transition scales.

Surface example



Source: Wikimedia Commons (CC BY-SA 4.0). [4]

Dynamics on the surface

- Classical integration produces trajectories on $U(\mathbf{r})$.
- Thermostats ensure canonical sampling.

Potential decomposition

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{nonbond}}.$$

- Separation between bonded and non-bonded terms.

Bonded terms

$$U_{\text{bond}} = \sum_b k_b (r_b - r_b^0)^2,$$

$$U_{\text{angle}} = \sum_a k_a (\theta_a - \theta_a^0)^2.$$

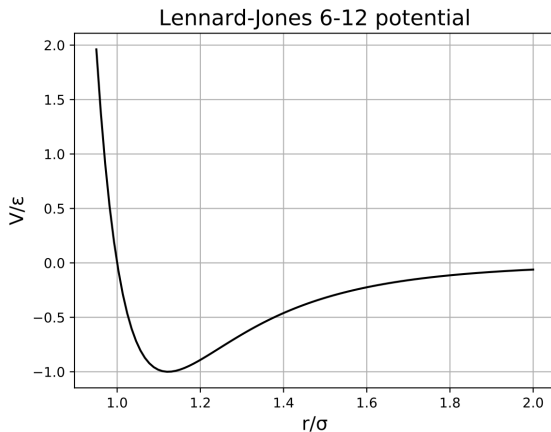
- Harmonic approximation around equilibrium geometries.

Dihedrals and rotation

$$U_{\text{dihedral}} = \sum_d \frac{V_d}{2} [1 + \cos(n_d \phi_d - \gamma_d)].$$

- Control rotational barriers and conformational preferences.

LJ example



Source: Wikimedia Commons (CC BY-SA 4.0). [1]

Parameters and validation

- Force fields are fitted to quantum and experimental data.
- Validation includes densities, energies, and conformations.

Working directory

- Environment variable `COURSE_DIR`.
- Structure: `data/` (inputs) and `results/` (outputs).

- Simple system: alanine.
- Complex system: protein + ligand.

- 1 Download base data.
- 2 Prepare topologies and coordinates.
- 3 Validate the computing environment.

Reproducibility

- Track library versions and random seeds.
- Save input files and scripts.

OpenMM general workflow

- Load PDB/GROMACS/AMBER/CHARMM/TINKER structures with the official scripts (e.g., `simulatePdb.py`, `simulateGromacs.py`, `simulateAmber.py`, and `simulateCharmm.py`).
- Create the Topology, System, and Integrator objects and wire StateReporter/TrajectoryReporter to capture energies, forces, and coordinates.
- Execute integration steps defined by the application layer and save checkpoints with the OpenMM-Setup routines (User Guide §3.5–3.15).
- `simulateAmber.py` accepts `'prmtop' + 'inpcrd'`, while `simulateCharmm.py` and `simulateGromacs.py` respect their native formats.
- `simulatePdb.py` is the traditional starting point for alanine dipeptide or unparameterized protein–ligand systems.
- All scripts expose hooks to tweak reporters and shape output formats that we then process with Python analysis scripts.

Episode summary

- The thermodynamic groundwork connects to statistical sampling.
- The potential defines forces and probabilities.
- The working environment ensures reproducibility.

References I

- [1] AdiPhysics. *Lennard Jones potential graph*. CC BY-SA 4.0. URL: https://commons.wikimedia.org/wiki/File:Lennard_Jones_potential_graph.svg (visited on 01/12/2026).
- [2] Krishnavedala. *Pendulum phase portrait illustration*. CC BY-SA 4.0. URL: https://commons.wikimedia.org/wiki/File:Pendulum_phase_portrait_illustration.svg (visited on 01/12/2026).
- [3] Gilles MAIRET. *Distribution de Maxwell-Boltzmann. Distribution des vitesses de l'hélium a 5500 K*. CC BY-SA 3.0. URL: https://commons.wikimedia.org/wiki/File:Distribution_de_Maxwell-Boltzmann._Distribution_des_vitesses_de_l%27h%C3%A9lium_%C3%A0_5500_K.png (visited on 01/12/2026).

References II

- [4] Unknown author. *Potential Energy Surface for a generic SN1 reaction*. CC BY-SA 4.0. Source: LibreTexts (Organic Chemistry, Morsch et al.)
URL: https://commons.wikimedia.org/wiki/File:Potential_Energy_Surface_for_a_generic_SN1_reaction.svg
(visited on 01/12/2026).