

# 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)

- OpenMM application layer
- Overview and scales
- State variables and states
- Entropy and free energy
- Ensembles and partitioning
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Environment and data

# Notation used throughout the notes

- Phase-space point:  $z = (q, p) \in \mathbb{R}^{3N} \times \mathbb{R}^{3N}$ , with positions  $q$  and momenta  $p$ .
- Potential energy (force field):  $U(q)$ ; forces  $\mathbf{F}(q) = -\nabla_q U(q)$ .
- Hamiltonian:  $H(q, p) = K(p) + U(q)$  with  $K(p) = \sum_i \frac{\|p_i\|^2}{2m_i}$ .
- Temperature  $T$ , Boltzmann constant  $k_B$ , inverse temperature  $\beta = (k_B T)^{-1}$ .
- Time step  $\Delta t$ ; lag time for kinetic models  $\tau$ .

# Mathematical backbone: what MD integrates

We model  $N$  atoms with masses  $m_i$  by Newton/Hamilton equations:

$$\dot{q}_i = \frac{\partial H}{\partial p_i} = \frac{p_i}{m_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} = -\nabla_{q_i} U(q).$$

In practice  $U(q)$  is a *molecular mechanics force field* decomposed into bonded and non-bonded terms,

$$U(q) = \sum_{\text{bonds}} k_b(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} V_n[1+\cos(n\phi-\delta)] + \sum_{i<j} \left( 4 \right.$$

# Suggested figure: molecular mechanics potential energy terms

**Figure:** Example schematic of a molecular mechanics potential energy function (bonded + nonbonded terms). Source: Wikimedia Commons (file used in Wikipedia).

# 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).

# External inputs

- `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 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.

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

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

- 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  $Q$ .

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

- Internal energy changes through heat or work exchange.
- In simulation,  $\delta W$  arises from forces on particles.

## Second law and entropy

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

- $\Omega$  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  $\Delta G$ .

# Classical ensembles

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

# Función de partición canónica

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

- Central para derivar energías libres y promedios.



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

- En simulación, se estiman con promedios temporales.

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

- Incluye fluctuaciones de volumen controladas por el barostato.

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

- Las fluctuaciones conectan estadística y respuesta macroscópica.

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

- Se separa en energía cinética y potencial.

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

- Equivalentes a la segunda ley de Newton.

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

- La fuerza es el gradiente negativo del potencial.

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

- La integración numérica introduce errores controlables.

# Flujo en el espacio de fases

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

- La densidad de probabilidad se conserva en el flujo hamiltoniano.



# Fuerzas a partir del potencial

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

- En MD, el coste dominante es evaluar  $U$  y  $\mathbf{F}_i$ .

# Potenciales de pares

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

- Simplifican la energía total con interacciones binarias.

## Ejemplo: Lennard-Jones

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

- Repulsión a corta distancia, atracción a media distancia.

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

- Dominante en sistemas biomoleculares.

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

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

- El Hessiano describe curvaturas locales (modos normales).

# Separación cinética/potencial

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

- $Z_{\text{conf}}$  depende solo de  $U(\mathbf{r})$ .

# Distribución de configuraciones

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

- Base del muestreo Monte Carlo y MD con termostato.

# Energía libre configuracional

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

- Relaciona el muestreo con  $\Delta A$  entre estados.



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

- Permite inicializar velocidades a temperatura  $T$ .

# Teorema de equipartición

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

- Cada grado de libertad cuadrático aporta  $\frac{1}{2} k_B T$ .

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

- El peso exponencial penaliza estados de alta energía.

# Distribución de Maxwell-Boltzmann

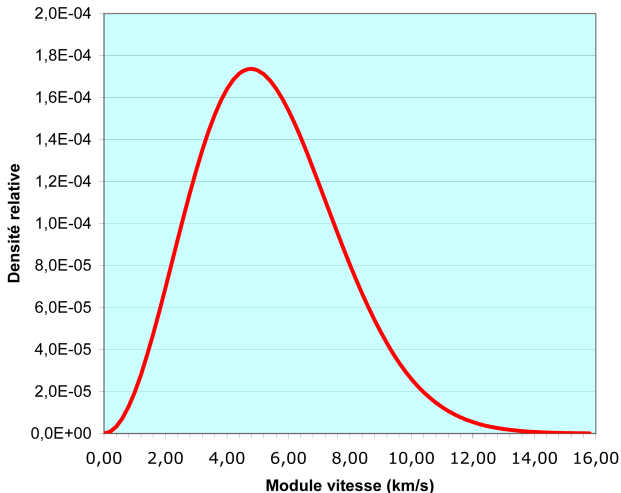
$$f(v) = 4\pi \left( \frac{m}{2\pi k_B T} \right)^{3/2} v^2 e^{-\frac{mv^2}{2k_B T}}.$$

- Caracteriza la distribución de velocidades en equilibrio.

- La mayoría de partículas se agrupan alrededor de  $v_m$ .
- El ensanchamiento crece con  $T$ .

## Ejemplo gráfico

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



- Inicializar velocidades según  $f(v)$  evita sesgos térmicos.
- Permite comprobar que el termostato reproduce el equilibrio.

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

- Cada punto representa un microestado completo.



# Ecuación de Liouville

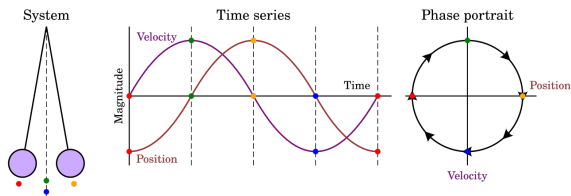
$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0.$$

- El volumen en  $\Gamma$  se conserva (teorema de Liouville).

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

- Supuesto clave para sustituir promedios de conjunto por tiempo.

# Trayectorias en $\Gamma$



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

- Órbitas cerradas: movimiento periódico.
- Órbitas abiertas: difusión en regiones de alta energía.

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

- Describe el paisaje energético donde se mueve el sistema.

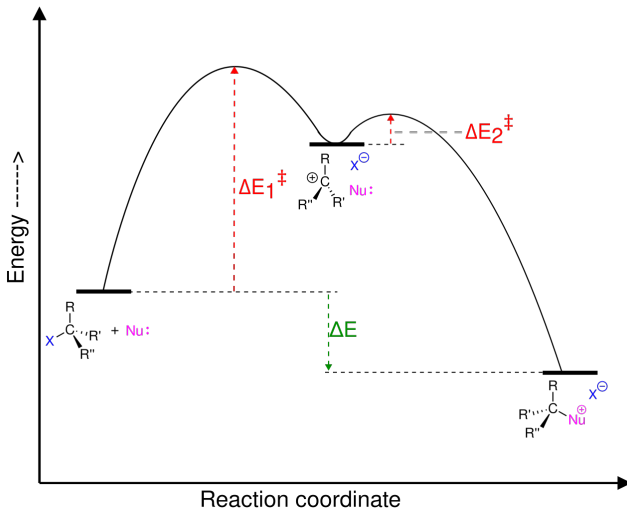
$$\xi = \xi(\mathbf{r}), \quad F(\xi) = -k_B T \ln P(\xi).$$

- Proyección que resume la dinámica en una coordenada clave.

# Barreras y estados metaestables

- Mínimos locales: conformaciones metaestables.
- Barreras: controlan las escalas de transición.

# Ejemplo de superficie



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



# Dinámica sobre la superficie

- Integración clásica produce trayectorias en  $U(\mathbf{r})$ .
- Los termostatos aseguran muestreo canónico.

# Descomposición del potencial

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

- Separación entre términos enlazados y no enlazados.

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

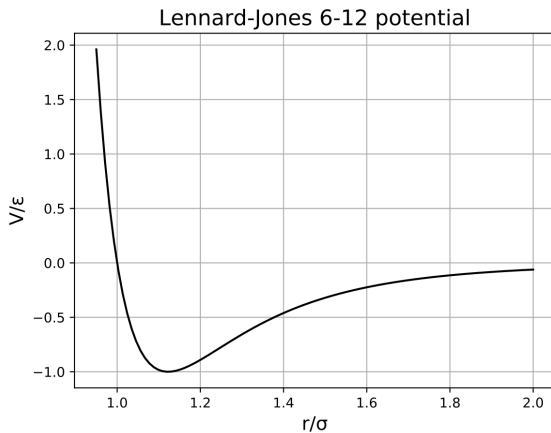
$$U_{\text{angle}} = \sum_a k_a (\theta_a - \theta_a^0)^2.$$

- Aproximación armónica alrededor de geometrías de equilibrio.

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

- Controlan barreras rotacionales y preferencias conformacionales.

# Ejemplo LJ



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

# Parámetros y validación

- Los campos de fuerza se ajustan a datos cuánticos y experimentales.
- La validación incluye densidades, energías y conformaciones.

# Working directory

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

# Course systems

- 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.

# 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](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](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](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](https://commons.wikimedia.org/wiki/File:Potential_Energy_Surface_for_a_generic_SN1_reaction.svg)  
(visited on 01/12/2026).