Questions and Exercises on Statistical Mechanics

Molecular Modeling of Nanosystems – CNE Master

Mario Barbatti

mario.barbatti@univ-amu.fr | *www.barbatti.org*

1. Explain the ergotic hypothesis.

2. The Boltzmann distribution is derived by solving the equation

$$\frac{\partial}{\partial N_j}\left[\ln\left(\prod_i \frac{g_i^{N_i}}{N_i!}\right)+\alpha\left(N-\sum_i N_i\right)+\beta\left(E-\sum_i N_i\varepsilon_i\right)\right]=0$$

Explain each term of this equation.

3. If a gas is equilibrated at 270 K, which fraction of molecules (relative to the ground state) is at the first vibrational excited state 0.02 eV above the ground state? The states are equally degenerated.

4. Consider this extract from a paper. Which type of Gibbs ensemble (microcanonical or canonical) did they adopt in their molecular dynamics (MD) simulations? Why?

MD Simulation

Explicit solvent MD simulations were performed using NAMD 2.7[29] and the CHARMM22 (cmap included) force field.[30] Langevin dynamics with periodic boundary conditions at 300 K was used for all simulations. In NPT simulations, the pressure was maintained with

5. Concerning thermostats in molecular dynamics simulations:
   a) When should we use them?
   b) In the Andersen thermostat, the Newtonian velocity of the atom α may be stochastically replaced by

$$\mathbf{v}_\alpha^{(new)}=\left(v_{\alpha,x}^{(new)},v_{\alpha,y}^{(new)},v_{\alpha,z}^{(new)}\right)=\sqrt{\frac{k_BT}{M_\alpha}}\times\left(r_x,r_y,r_z\right)$$

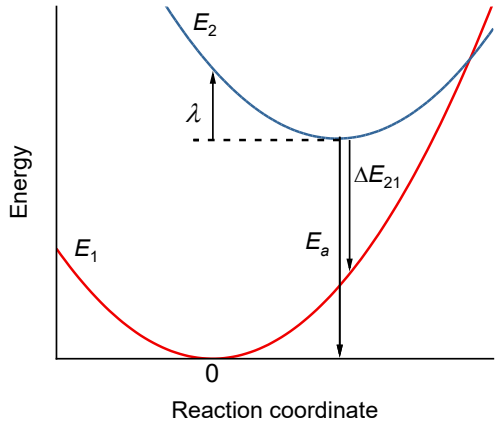
   Explain each term in this equation and the rationale for defining the velocity in this way.

6. The Python code below stochastically samples the value of a molecular bond.
   a) Write explicitly the distribution function employed.
   b) Which type of potential energy function is associated with such a distribution?
   c) In a single execution of this code, how many values will be sampled?
   d) After the sampling, what is the most likely bond value we should have?
   e) What are the minimum and maximum bond values we should have after sampling?

```
def geom_distribution(x, R0, sigma):
        return math.exp(-(x-R0)**2 / (2*sigma**2)) / (sigma*math.sqrt(2*math.pi))

R0 =1.0
sigma = 0.1
Nmax = 10000
(xmin, xmax) = (R0 - 3 * sigma, R0 + 3 * sigma)
(ymin, ymax) = (0, gaussian(R0, R0, sigma))

for i in range(1, Nmax + 1):
        random_x = random.uniform(xmin, xmax)
        random_y = random.uniform(ymin, ymax)
        Probability = geom_distribution(random_x, R0, sigma)

        if random_y < Probability:
                print("Geometry is ", random_x)
```

7. What is the Metropolis algorithm?

8. A molecule is at the minimum of its excited state (E_2). From there, it can either return to the ground state E_1 via internal conversion (radiationless) or fluorescence. The vertical energy gap ΔE_{21} is -1.0 eV, the adiabatic energy E_a is -2.0 eV, and the reorganization energy λ is 0.8 eV. The diabatic coupling V_{12} between the states is 0.2 eV, and the oscillator strength f_{12} between the states is 0.2.



a) Assuming the validity of the Marcus model,

$$W_{\text{int.conv.}} = \frac{2\pi}{\hbar}|V_{12}|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} \exp\left(-\frac{(\lambda + \Delta E_a)^2}{4\lambda k_B T}\right)$$

what is the internal conversion time?

b) Knowing that the expontaneous emission rate is

$$W_{emission} = \frac{2\alpha}{\hbar m_e c^2} \Delta E_{21}^2 f_{12}$$

how long should it take for the molecule to fluoresce?

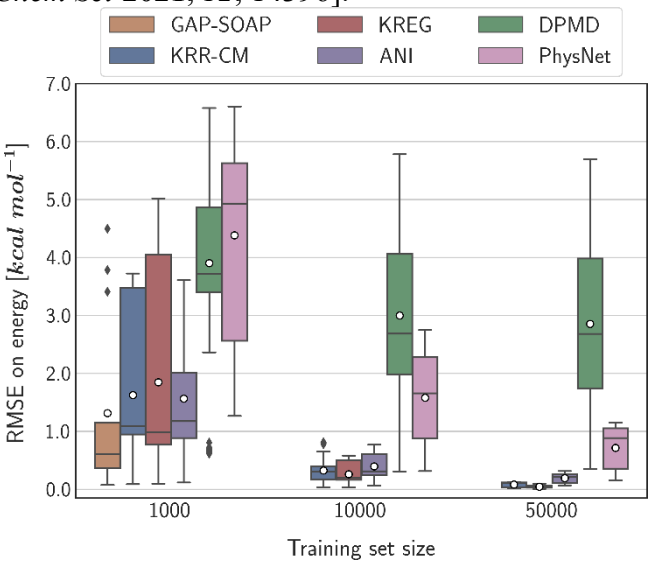
c) Which process should dominate the competition between internal conversion and fluorescence?

9. A molecular dynamics simulation of 1000 trajectories, one observes the following results:

Channel	Proportion of trajectories (%)
Formation of product A	31
Formation of product B	35
No reaction	34

a) What are the margins of error for each of these results, considering a 95% confidence interval?
b) Can we definitely tell which channel dominates the reaction? Why?

10. The graph below reports the root mean square error (RMSE) of several machine learning models for computing the potential energy of molecules as a function of the number of training points [Pinheiro Jr *et al*. *Chem Sci* **2021**, *12*, 14396].



Considering that an acceptable accuracy should have errors below 1 kcal/mol, answer the following:
a) Are there any models that could be used with a small training set of 1000 points?
b) Which models could be used with 10000 points?

11. Ref. [Martinka *et al. J Chem Phys* **2024,** *161*, 174104] shows the following learning curves.

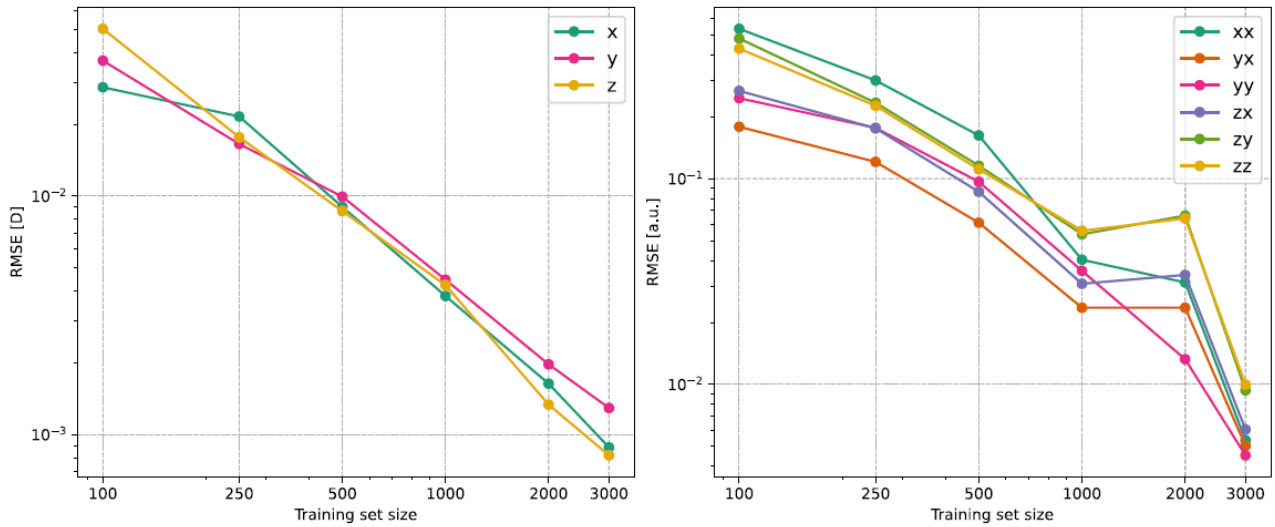


FIG. 2. Learning curves of dipole moment components in Debye (left) and polarizability components in atomic units (right). The accuracy of each model was determined on the test set consisting of 1000 points.

How many points should we have in the training set to have dipole moment predictions better than 10^{-2} D and, at the same time, polarizability predictions better than 10^{-1} au?

12. Consider the loss function used to train a neural network to predict energies (E) and energy gradients (G):

$$\mathcal{L}_k = a\left(E^{(NN)} - E^{(QM)}\right)^2 + b\left(G^{(NN)} - G^{(QM)}\right)^2$$

a) If both energies and gradients are equally important for your simulations, which values of the hyperparameters *a* and *b* would you choose? Why?
b) Suppose that in addition to *E* and *G*, you also want to train the machine using dipole moment data (*D*). How would you change the loss function?