

**Examination D-CHAB Spring 2021:
Statistical Physics for CSE**

17.02.2021 10:00-11:00

HCI J4

**Prof. Sereina Riniker
Prof. Gunnar Jeschke
Prof. Philippe H. Hünenberger**

Problem statements in English

- Write your **name** and **identification (Legi) number** on **each** sheet of paper you hand in.
- The use of laptops, cell phones, calculators, books, course material, *etc.* is **not allowed** (exception: language dictionaries).
- You may use the **question sheets** to report your answers (or part of them).
- Please, clearly **highlight** (*e.g.* underline or frame) your **final answers** to the problems.
- Keep your answers **short**, but **clear**.
- The five problems of the examination have **equal weights** in the final mark.
- You **do not** need to answer **all** questions correctly to obtain a top mark. First try to efficiently solve the easier questions, and then move to the more difficult ones.

1 Concepts and Understanding (S2021.1)

For each of the following items, **answer** the question(s) in a **clear** and **concise** way.

- a. State the two basic tests of “randomness” used to assess the quality of a pseudo-random number generator (producing real values in the interval $[0, 1[$). For each of the two tests, you have to specify the kind of mathematical object to be calculated, and the property this object is expected to satisfy.
- b. The correspondence principle associates quantum-mechanical operators to the variables of classical mechanics.
 - Write the expressions of the operators associated with the position vector, the momentum vector, and the total energy of a particle (in three dimensions).
 - State the quantum-mechanical equation that follows from the classical energy conservation (*i.e.* that the system Hamiltonian function has a constant value) upon applying the correspondence principle, and give the name of this equation.
 - Write the expression for the kinetic-energy operator involved in this equation.
- c. To calculate the relative binding free energies of two ligands A and B to a given protein using molecular dynamics simulations, it is common practice to rely on a thermodynamic cycle involving two alchemical free-energy changes.
 - Draw schematically such a thermodynamic cycle, and write down the resulting equation for calculating the relative binding free energy $\Delta G_{\text{bind}}^B - \Delta G_{\text{bind}}^A$.
 - Explain why it is generally advantageous to calculate the two alchemical free-energy differences instead of the two conformational ones.
 - Give two examples of free-energy calculation methods that can be employed to calculate alchemical free-energy changes.
 - Suggest a specific choice for the end states A and B (possibly unphysical) permitting the calculation of the *absolute* binding free energy of a given ligand.

2 Fundamental Equations (S2021.2)

For each of the following items, **write down** the relevant equation(s), **explain the meaning** of all the involved symbols, **state the SI units** of these quantities, and **answer** the additional questions.

- a. Write the equation for the normalized probability density $p(x)$ corresponding to a normal distribution (Gaussian distribution) in a variable x , of mean μ and standard deviation σ . Additional questions:
 - What condition (equation) does the word “normalized” refer to?
 - What is the relationship between the units of x and those of $p(x)$?
- b. Write down the Newtonian equation of motion (Newton’s second law) for a particle in three dimensions, as well as the equation connecting the force acting on the particle to its potential energy. Additional questions:
 - In what coordinate system is the Newtonian equation of motion valid?
 - Name the three physical quantities that are often conserved in the Newtonian dynamics (*i.e.* are constants of the motion), and specify under which condition each of them is conserved.
- c. The microscopic (instantaneous) configuration of a classical system of N particles is determined by the $3N$ -dimensional Cartesian position and momentum vectors \mathbf{r} and \mathbf{p} of the system.
 - Write an expression for the microscopic observable \mathcal{T} associated with the instantaneous temperature of a configuration.
 - Give the equation connecting \mathcal{T} to its canonical ensemble average T .

3 Derivations (S2021.3)

For each of the following items, **derive** the required expression analytically (*i.e.* it is not sufficient to only give the final result!), and **answer** the additional questions.

- a. Alchemical free-energy calculations have been performed according to a linear coupling scheme

$$\mathcal{H}(\lambda) = (1 - \lambda)\mathcal{H}_A + \lambda\mathcal{H}_B ,$$

where \mathcal{H}_A and \mathcal{H}_B are the Hamiltonians of states A and B , and λ is the coupling parameter. The ensemble averages of \mathcal{H}_A and \mathcal{H}_B were calculated using a set of simulations performed at different λ -values. They were found to be well approximated by straight lines, *i.e.*

$$\langle \mathcal{H}_A \rangle_\lambda = a\lambda + b \quad \text{and} \quad \langle \mathcal{H}_B \rangle_\lambda = c\lambda + d .$$

Using the thermodynamic integration (TI) formula, write down the equation connecting the free-energy change $\Delta G_{A \rightarrow B}$ to the fitting coefficients a , b , c and d .

- b. Within a multiplicative constant (irrelevant here), the classical partition function Z of the canonical (NVT) ensemble can be written

$$Z = \int d\mathbf{x} e^{-\beta\mathcal{H}} \quad \text{with} \quad \beta = \frac{1}{k_B T} ,$$

where \mathbf{x} is the $6N$ -dimensional phase-space vector (coordinates and momenta of the N particles), $\mathcal{H} = \mathcal{H}(\mathbf{x})$ is the Hamiltonian function (total energy), k_B is the Boltzmann constant, and T the absolute temperature. The canonical ensemble average of a microscopic (instantaneous) observable $\mathcal{A} = \mathcal{A}(\mathbf{x})$ can be written

$$\langle A \rangle = Z^{-1} \int d\mathbf{x} \mathcal{A} e^{-\beta\mathcal{H}} .$$

The definitions of the mean-square fluctuation σ_E^2 of the total energy in the canonical ensemble and of the isochoric heat capacity of the system are

$$\sigma_E^2 = \langle \mathcal{H}^2 \rangle - \langle \mathcal{H} \rangle^2 \quad \text{and} \quad C_V = \frac{\partial \langle \mathcal{H} \rangle}{\partial T} .$$

Derive the statistical-mechanical equation connecting σ_E^2 to C_V in the canonical ensemble.

Hint: Before starting, expand the following expression using the product rule

$$\frac{\partial}{\partial \beta} \left(\frac{1}{Z} \frac{\partial Z}{\partial \beta} \right) = \dots .$$

Then, a good way to proceed is: (1) derive expressions relating $\partial Z / \partial \beta$ and $\partial^2 Z / \partial \beta^2$ to $\langle \mathcal{H} \rangle$ and $\langle \mathcal{H}^2 \rangle$; (2) insert these into the definition of σ_E^2 ; (3) simplify the resulting expression using the above expansion result.

4 Explicit Calculations (S2021.4)

For each of the following items, **calculate** the numerical result paying particular attention to the **units**, and **answer** the additional questions.

- a. The single bond between two carbon atoms in a molecule is described in a force field by a harmonic potential-energy term of the form

$$V(b; b_o) = \frac{1}{2}k_b(b - b_o)^2 .$$

Here, the reference bond length is $b_o = 0.15$ nm and the force constant is $k_b = 10^5$ kJ·mol⁻¹·nm⁻². Calculate the value of the potential energy for a bond length $b = b_o + 0.01$ nm and $b = b_o - 0.01$ nm. Additional questions:

- Compare the resulting energy to the thermal energy RT at room temperature, using $R = 8.3$ J·K⁻¹·mol⁻¹ and $T = 300$ K.
- Using this comparison, comment on the typical magnitude of the bond-length fluctuations one should expect in molecular dynamics simulations.

5 Algorithms and Implementation (S2021.5)

For each of the following items, **write the code** of a C++ function performing the required task (or, at least, the pseudo-code; the exactness of your C++ syntax will not be graded), and **answer** the additional questions.

- a. Write a C++ function `MeanSquareDispl` that will calculate the mean-square displacement $s(\tau)$ of a set of N atoms as a function of the lag time τ based on a system trajectory of total duration T according to the equation

$$s(\tau) = \left\langle \frac{1}{N} \sum_{n=1}^N [\mathbf{r}_n(t+\tau) - \mathbf{r}_n(t)]^2 \right\rangle_t ,$$

where $\mathbf{r}_n(t)$ is the Cartesian position vector of particle n in the trajectory frame at time t , and $\langle \cdot \cdot \cdot \rangle_t$ indicates averaging over all possible time origins t from 0 to $T - \tau$. In practice, the trajectory is discretized using a timestep Δt , so that time is specified by an integer index $k = 0, 1, \dots, K - 1$ as $t = k\Delta t$, where K is the total number of trajectory frames (*i.e.* $K\Delta t = T$). The function declaration reads

```
void MeanSquareDispl ( int N, int K, double r[3*N][], double s[] );
```

Provided to the function are the number N of particles, the number K of trajectory frames, and the trajectory array \mathbf{r} , in which $\mathbf{r}[3*\mathbf{n}+\mathbf{i}][\mathbf{k}]$ contains the i^{th} component (0,1,2 for x,y,z) of $\mathbf{r}_n(k\Delta t)$. The function should fill the array \mathbf{s} with the values of $s(k\Delta t)$ for $k = 0, 1, \dots, K - 1$. Additional questions:

- Write the Einstein equation connecting the slope of $s(\tau)$ to an important transport property of the system.
- Explain why, in practice, the curve calculated for $s(\tau)$ will look close to a straight line only for low τ , but will become progressively more “noisy” as τ approaches T
- What should one pay attention to when applying this procedure to simulations carried out under periodic boundary conditions?