# 17 NVE/NVT

**(1 point per correct answer)**

If we run MD in the NVE ensemble, the number of particles in the system, the volume, and the total energy is kept constant.

**Select an alternative**

- ⦿ True ✅
- ○ False

If we run MD in the NVT ensemble, the number of particles in the system, the volume, and the kinetic energy is kept constant.

**Select an alternative:**

- ○ True
- ⦿ False ✅

## 18   MD integrator

**(2 points in total)**

Write two important qualities that a good Molecular Dynamics time-integrator must meet to obtain stable trajectories.

**Write your answer here...**

1) Time-reversible
2) Area-preserving
With these two conditions, there is a Shadow Hamiltonian that is exactly conserved. The total energy remains close to the shadow Hamiltonian, which ensures stable trajectories.

## 19   Good MC moves

**(3 points in total)**

Write three important qualities that a good Monte Carlo algorithm must meet to obtain the correct statistics.
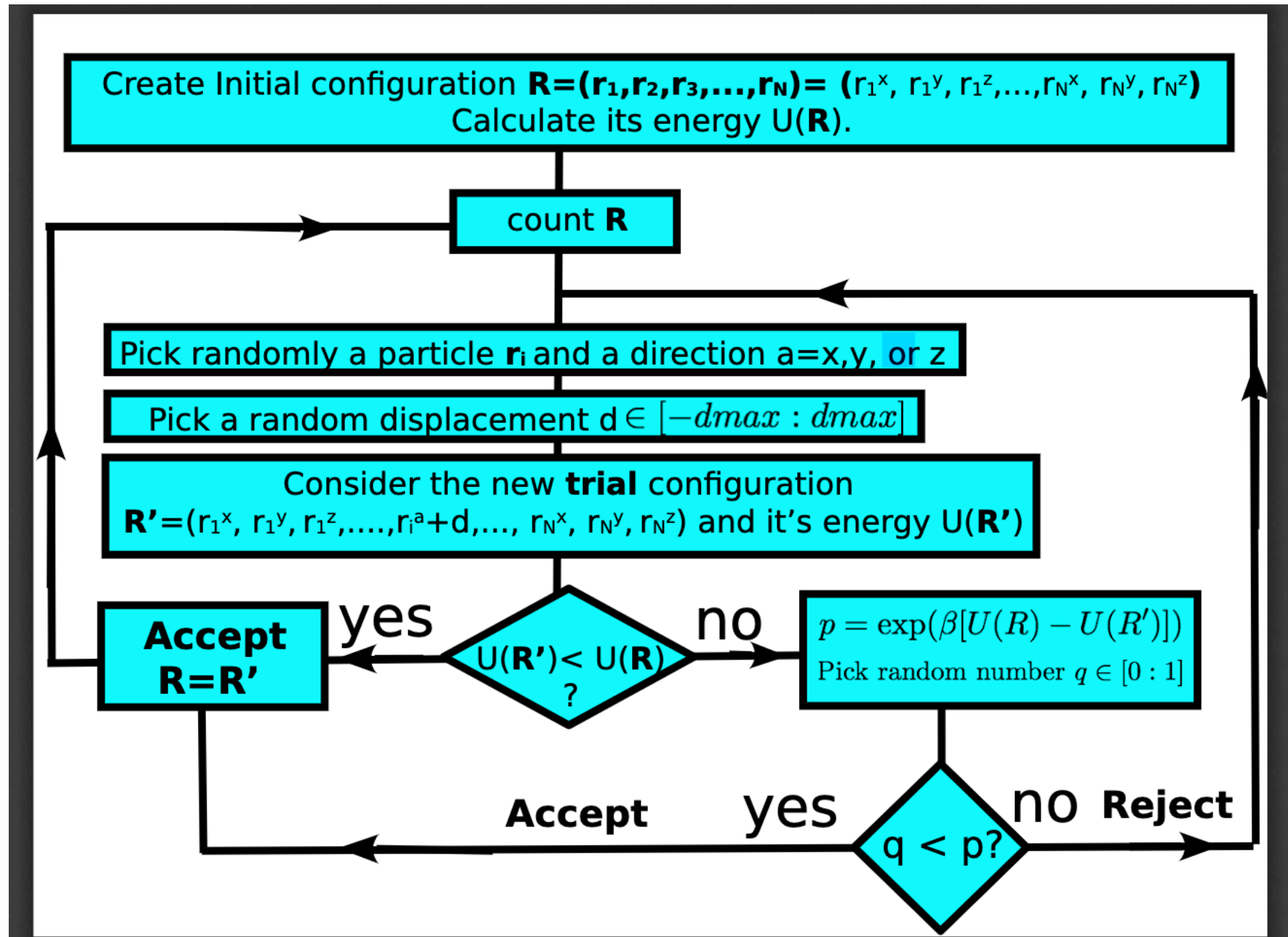
**Write your answer here...**

1) Detailed Balance
2) Markovian
3) Ergodic

Create Initial configuration $\mathbf{R}=(\mathbf{r_1},\mathbf{r_2},\mathbf{r_3},\ldots,\mathbf{r_N})= (r_1{}^x, r_1{}^y, r_1{}^z,\ldots,r_N{}^x, r_N{}^y, r_N{}^z)$
Calculate its energy $U(\mathbf{R})$.

count $\mathbf{R}$

Pick randomly a particle $\mathbf{r_i}$ and a direction a=x,y, or z

Pick a random displacement $d \in \left[-dmax : dmax\right]$

Consider the new **trial** configuration
$\mathbf{R'}=(r_1{}^x, r_1{}^y, r_1{}^z,\ldots,r_i{}^a+d,\ldots, r_N{}^x, r_N{}^y, r_N{}^z)$ and it's energy $U(\mathbf{R'})$

**Accept**
$\mathbf{R=R'}$

yes

$U(\mathbf{R'})< U(\mathbf{R})$ ?

no

$p = \exp(\beta[U(R) - U(R')])$
Pick random number $q \in [0 : 1]$

**Accept**   yes   $q < p?$   no **Reject**

**20** **Flow Chart**

**(3 points in total)**

The flow chart below should represents the Metropolis Monte Carlo scheme.

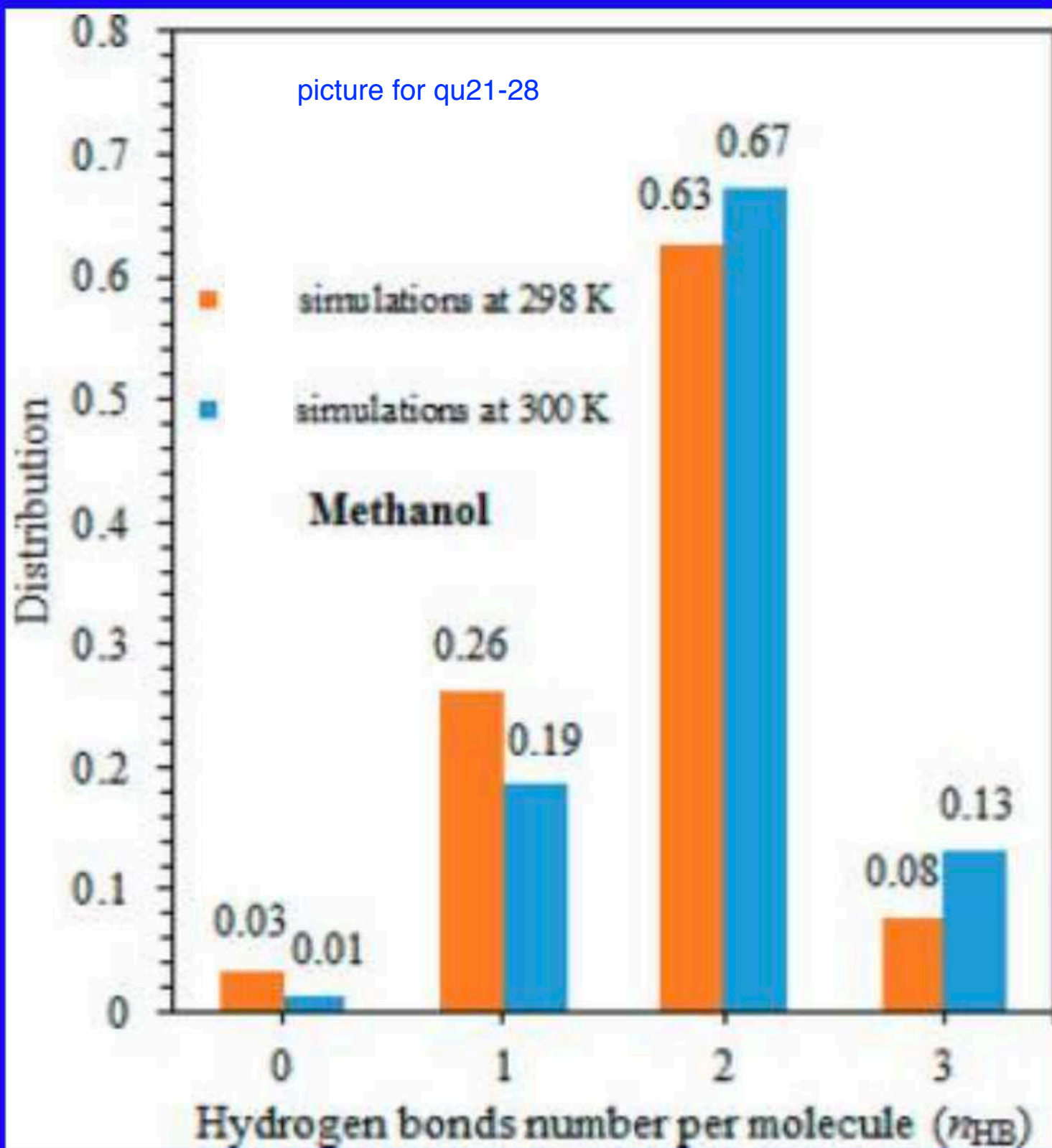There is a mistake. Describe the mistake in the text box below

**Fill in your answer here**

Picture shown at previous page.

The line "no reject" enters back below the box "count R". It is however essential that after a rejection the old configuration is counted again. The line should therefore point inside that box

picture for qu21-28

Legend:
- simulations at 298 K
- simulations at 300 K

Methanol

Values shown: 0.03, 0.01 (at $n_{HB}=0$); 0.26, 0.19 (at $n_{HB}=1$); 0.63, 0.67 (at $n_{HB}=2$); 0.08, 0.13 (at $n_{HB}=3$)

Y-axis: Distribution (0 to 0.8)

X-axis: Hydrogen bonds number per molecule ($n_{HB}$)

## 21  hbonds; MD vs MC

**(1 point per correct answer)**
At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402).
In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid
methanol at 298 K and 300 K.

With what method could they have used to compute the distribution graph at the left?
**Select one alternative**

It are thermodynamics properties (probabilities). It does not relate to a dynamical
property. So both MD and MC are possible

○ with MD

○ with MC

◉ with MD or with MC, both are possible ✓

What would be the most efficient method to compute it? (Just repeat your answer if you selected above that
only one method was possible)
**Select one alternative:**

It is a condensed system (liquid). So MD is more effective

◉ with MD ✓

○ with MC

MD because it is a dynamical property

What method could be used to compute the life time of the hydrogen bond in liquid methanol?
**Select one alternative**

◉ with MD ✓

○ with MC

○ with MD or with MD, both are possible

What would be the most efficient method to compute it? (Just repeat your answer if you selected above that
only one method was possible)
**Select one alternative**

◉ with MD ✓

○ with MC

## 22 hbonds; free energy qualitative

**(1 point per correct answer)**
At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

Say n is the number of hydrogen bonds that a methanol molecule makes in liquid water ($n_{hb}$ in the figure). We can consider the free energy F(n) as function of this parameter.

What would be the lowest free energy in this temperature range (298-300 K)
**Select one alternative**

- ○ n=0          High probability -> low free energy
                 Low probbability -> high free energy
- ○ n=1
- ◉ n=2                                               ✓
- ○ n=3

What would be the highest free energy
**Select one alternative:**

- ◉ n=0                                               ✓
- ○ n=1
- ○ n=2
- ○ n=3

Riktig. 2 av 2 poeng.

## 23 h-bonds; free energy quantitative

**(2 point per correct answer)**
At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

What is the free energy difference between the case n=3 and n=0, $\Delta F=F(3)-F(0)$, at T=298 K and T=300 K?

$\Delta F$ at 298 K

| -2.43 | ✓ | kJ/mol. |

298 K
-kB T* ln(0.08/0.03)= -kB 298 K* ln(0.08/0.03)
-Boltzmann constant*298 Kelvin* ln(0.08/0.03)*Avogadro's constant in kilojoule mol^-1
-2.43020951 kilojoule (mol^(-1))

$\Delta F$ at 300 K

| -6.398 | ✓ | kJ/mol. |

300 K
-kB T* ln(0.13/0.01)= -kB*300 K* ln(0.13/0.01)
-Boltzmann constant*300 Kelvin* ln(0.13/0.01)*Avogadro's constant in kilojoule mol^-1
-6.39785054 kilojoule (mol^(-1))

(Boltzmann constant: 8.31445986 J mol$^{-1}$K$^{-1}$)

Riktig. 4 av 4 poeng.

## 24    hbonds; Correlation function

**(2 points in total)**

At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

The authors also tried to compute the correlation function $C(t) = \langle n(0)n(t) \rangle$ using MD. Here n(t) is the number of hydrogen bonds that a specific methanol molecule has at time t. After an equilibration run they started gathering data. For this, they just followed one methanol molecule and looked how many hydrogen bonds it has at discrete time intervals of 1 ps. They found following sequence t=0 (time at which data is gathered, after the equilibration run): n=2,

t=1 ps: n=1, t=2 ps: n=1, t=3 ps: n=0,  t=4 ps: n=2, t=5 ps: n=3, t=6 ps: n=3, ...

Naturally, the data provided here are way too limited (the actual run would be thousands of ps), but what would be the estimate of C(2 ps) after this preliminary data set of only 7 measurements (at 0,1,..., 6 ps)

C(2 ps)= 2 ✅

t=0 ps: n=2,
t=1 ps: n=1,
t=2 ps: n=1,
t=3 ps: n=0,
t=4 ps: n=2,
t=5 ps: n=3,
t=6 ps: n=3

Riktig. 2 av 2 poeng.

Here one should understand that to compute C(t)=<n(o)n(t)> the t=0 can be shifted

So it is an average of n(0)*n(2), n(1)*n(3), n(2)*n(4), n(3)*n(5), n(4)*n(6)

## 25    Correlation function, 0, inf

These are 2*1, 1*0, 1*2, 0*3, 2*3    or 2,0,2,0, 6 which has an average of 2

**(2 points per correct answer)**

At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

The authors also tried to compute the correlation function $C(t) = \langle n(0)n(t) \rangle$ using MD. Based on the data provided at the left, can you compute $C(0)$ and $\lim_{t\to\infty} C(t)$ at 298 K?

C(0)= 3.5 ✅ (Write -999 if you think you can't compute it from the data provided)

C(∞)= 3.0976 ✅ (Write -999 if you think you can't compute it from the data provided)

Riktig. 4 av 4 poeng.

298 K

| n | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| p. | 0.03 | 0.26 | 0.63 | 0.08 |

<n(0) n(0)>=<n^2>=0.03*0^2+0.26*1^2+0.63*2^2+0.08*3^2=3.5

<n(0) n(inf)> = <n(0)> <n(inf)> = <n> <n> = <n>^2 as they are uncorrelated
<n> = 0.03*0+0.26*1+0.63*2+0.08*3=1.76
1.76^2=3.0976

# 26  hbond; MD-data1

**(4 points in total)**

At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

To get from the MD data a distribution like the one shown at the left, the xyz-atom positions of the methanol molecules (also called $\mathbf{R}^N$ or configuration point) are saved every 1000 time steps. As 1 MD step is 1 fs, this means that the MD data are saved every ps. At the end of the simulation a total simulation time of 10 000 ps or 10 ns is produced.

After this, the hydrogen bond analysis is done on the data that are saved (which are 10 000 configuration points) which finally leads to output like the average number of hydrogen bonds, the distribution function shown at the left etc.

Suppose that disk space is not an issue and saving and analysing a configuration point cost only half the CPU time of doing an MD step. Would it not be better to save every MD step and get 1000 times more data points for only 50% increase in CPU time?

Motivate your answer.

**Fill in your answer here**

No as it won't give 1000 times more uncorrelated data points. Successive MD points, separated by just 1 MD step, are certainly highly correlated. One might consider to save data more frequently, say every 100 MD steps, possibly even every 10 MD steps. This is still a small investment (about 0.5% and 5% more CPU respectively). If the statistical inefficiency of the sampling is larger than 1000 MD steps, however, it won't help much to reduce the statistical errors

Besvart.

## 27   hbond; MD-data2

**(4 points in total)**

At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

To get from the MD data a distribution like the one shown at the left, the xyz-atom positions of the methanol molecules (also called $\mathbf{R}^N$ or configuration point) are saved every 1000 time steps. As 1 MD step is 1 fs, this means that the MD data are saved every ps. At the end of the simulation a total simulation time of 10 000 ps or 10 ns is produced.

After this, the hydrogen bond analysis is done on the data that are saved (which are 10 000 configuration points) which finally leads to output like the average number of hydrogen bonds, the distribution function shown at the left etc.

Besides averaging over the 10 000 configuration points that were saved, we can average over all molecules in the system. There are N methanol molecules. If we assume that the 10 000 configuration points are uncorrelated, this means that we effectively have N*10 000 uncorrelated measurements. Is this correct?

Motivate your answer.

**Fill in your answer here**

No, the number of h-bonds that a certain methanol has in the system correlates with its neighbours. This is obvious as the fact that a methanol molecule accepts or donates a h-bond, one of the neighbours will donate or accept that h-bbond. This alone already creates corresslations. In addition, there are physical phenomena that play a role like polarization: e.g. a methanol molecule that donates a h-bond will be more polarized, leading to higher affinity to accept a h-bond at the other end.

0.03*0+0.26*1+0.63*2+0.08*3=1.76
<n>=1.76
sigma^2=<(n-<n>)^2>=0.03*(0-1.76)^2+0.26*(1-1.76)^2+0.63*(2-1.76)^2+0.08*(3-1.76)^2
=0.4024
sigma=sqrt(0.4024)=0.63435006108

Note we dont need to apply Bessel correction here as we assume that the number of uncorr. data points (that were used to create the distribution graph of the figure) >> 1

epsilon=sigma/sqrt(40000)=0.63435006108/200=0.0031717503

## 28   h-bonds; error

**(5 points in total)**

At the left you see results from an actual molecular simulation study (J. Phys. Chem. B 2016, 120, 3388−3402). In this study the authors analysed the number of hydrogen bonds that methanol molecules make in liquid methanol at 298 K and 300 K.

From the data at the left, we compute the average number of hydrogen bonds at 298 K as follows:
$\langle n \rangle = \sum_{i=0}^{3} p_k k = 0.03 * 0 + 0.26 * 1 + 0.63 * 2 + 0.08 * 3 = 1.76$. What is the error in this number if we know that the number of effectively uncorrelated measurements, on which the data at the left was based, is 40000?

$\epsilon =$   0.00317   ✅   (provide at least 2 significant digits, e.g 0.000000xx )

Riktig. 5 av 5 poeng.

picture for question 29-31
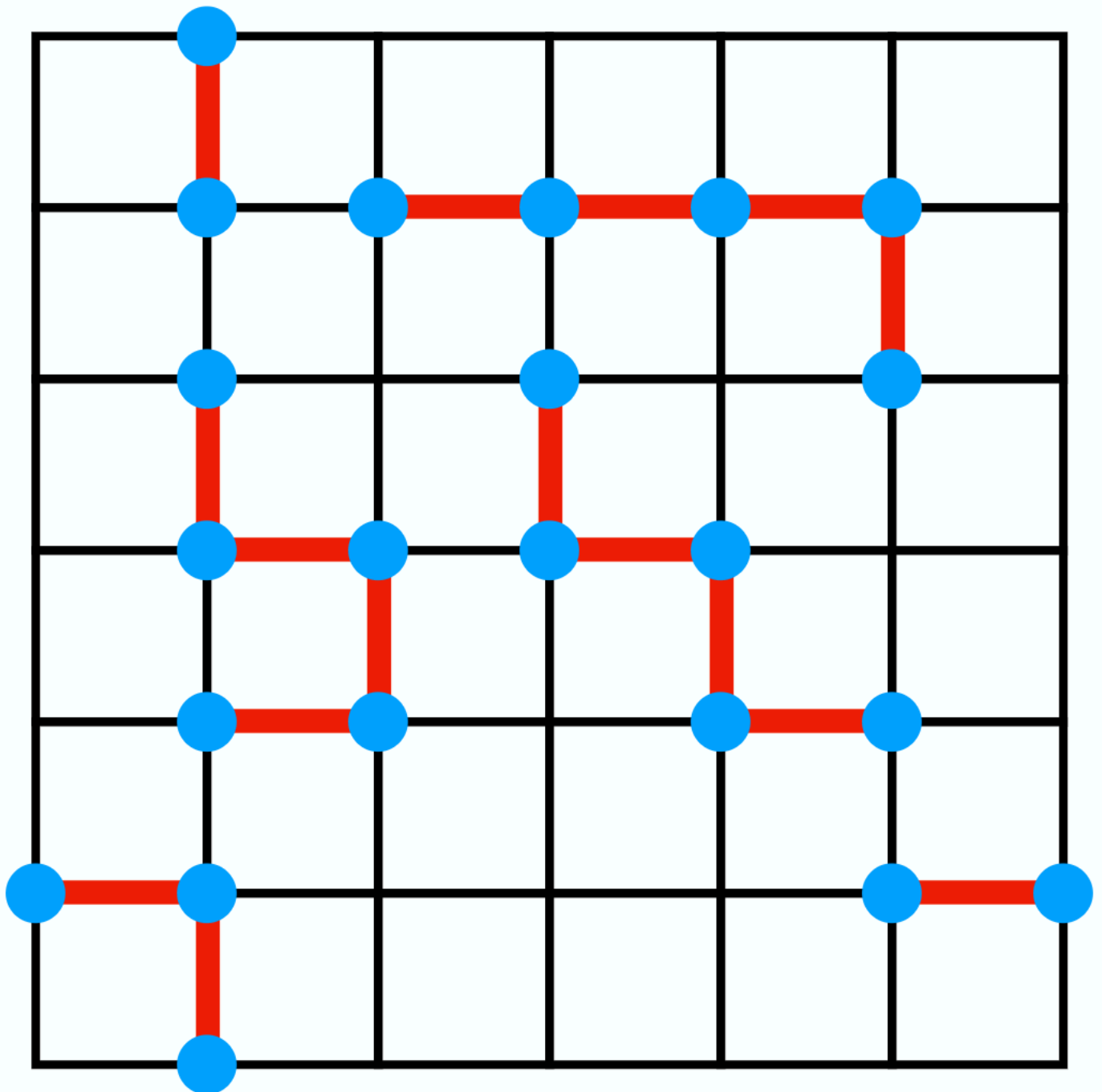
**29** ## 2D lattice model

**(1 point per correct answer)**
We perform a MC simulation to study linear polymers adsorbed on a surface using a 2D lattice model. Periodic boundary conditions are applied. The polymers are based on a coarse-grained model of 5 connected beads. The bonds are fixed and equal to the lattice spacing. A snapshot of the simulation is shown at the left.

How many polymer molecules do you see?

This question tests if one understands Periodic Boundary Conditions. The system is a 6X6 lattice. One molecule exceeds the boundaries (twice) and re-enters from the opposite site.

4 ✓ .

The atom at the bottom and top are periodic images of the same atom. Same for the atom most left and right. So there are 4 molecules, all consisting of 5 atoms

How many lattice points does the model system consist of?

36 ✓

Riktig. 2 av 2 poeng.

**30** ## 2D lattice model2

**(3 points per correct answer)**
We perform a MC simulation to study linear polymers adsorbed on a surface using a 2D lattice model. Periodic boundary conditions are applied. The polymers are based on a coarse-grained model of 5 connected beads. The bonds are fixed and equal to the lattice spacing. A snapshot of the simulation is shown at the left.

The MC algorithm has to follow detailed-balance:
$$\rho(o)\pi(o \rightarrow n) = \rho(n)\pi(n \rightarrow o)$$
where ρ(o) and ρ(n) are the equilibrium probabilities of the old and new state, and $\pi(x \rightarrow y)$ is the probability that the algorithm generates state y starting from state x. All states that have no overlapping beads are assumed to have the same probability. States with overlapping beads (intermolecular or intramolecular) are assumed to have infinite energy and therefore zero probability.

$\pi(x \rightarrow y)$ can be written as the generation probability times an acceptance probability

$$\pi(x \rightarrow y) = P_{gen}(x \rightarrow y)P_{acc}(x \rightarrow y)$$

(Typo in exam question that said "4 beads". Thought, this was an obvious error to identify as the rest of the text always says 5 and also the picture shows polymers with 5 beads)

To make things simple, assume we have one single linear polymer consisting of 5 beads on a 10x10 lattice. We generate a new state of the system by deleting the molecule and growing it at a random position in a random conformation as follows. The first bead is randomly placed on the lattice. The second bead is placed randomly at any of the 4 directions: one lattice point left, right, down, or up with respect to the first bead. The third bead is placed likewise with respect to number 2 with additional condition that the angle between beads 1-2-3 is either 90 degrees or 180 degrees. So bead 3 is placed randomly at one of the 3 possible directions, one lattice point away from number 2. Beads 4 and 5 follow the same rule as bead 3. (This means 4 is one lattice point away from 3, 5 is one lattice point away from 4; angles 2-3-4 and 3-4-5 are either 90 or 180 degrees).

For this system with only one polymer, what is $P_{gen}(o \rightarrow n)$?

Placement of first bead 1/100
placement of nr 2: 1/4
placement of 3,4,5 all 1/3
Total: (1/100)*(1/4)*(1/3)^3=0.00009259259
If we dont label the atoms (implicit assuming that each bead is chemically identical), we can argue that the same configuration can be generated by starting at the end
Then the answer is 2*0.00009259259=0.00018518518

0.0000926 ✓ .

And what is $P_{acc}(o \rightarrow n)$? on average for the one polymer system?

0.926 ✓

Riktig. 6 av 6 poeng.

All trials will be accepted unless you are so unlucky that bead 5 overlaps with bead 1.
This happens if the polymer is forming a closed square. For this to happen the angle 1-2-3 should not be 180 and bead 4 and 5 should pick the right direction to circulate. So that chance is (2/3)*(1/3)*(1/3)=0.07407407407
In that case the move would be rejected. In all other cases accepted. So on average Pacc=1-0.07407407407=0.92592592593

## 2D lattice, text

**(3 points in total)**

What method can you use if there are many polymers in the system and the previously described method almost always creates a configuration with two overlapping beads (as a result $P_{acc}$ becomes very small) ?

**Fill in your answer here**

Configurational Bias MC

## 32   **Optional/Feedback H2019**

**Optional question.** If you have special concerns regarding one or more questions, for instance if it was not clear to you how to interpret the question, then you can write them here.
Please specify the number and title of the question.

Further, if you have time left, we are always happy to receive some words of feedback which you can write here. How did you perceive the exam? Did you find it difficult or not? Any comments are welcome.

**Fill in your answer here**

## 32   **Optional/Feedback H2019**

**Optional question.** If you have special concerns regarding one or more questions, for instance if it was not clear to you how to interpret the question, then you can write them here.
Please specify the number and title of the question.

## 18  Molecular Dynamics

**What is the aim of Molecular Dynamics simulations?**
**Select one alternative:**

○ Efficiently describe the behavior of chaotic systems at long time.

○ Describe the exact evolution of the system in time solving Newton's equations.

● Give an accurate evolution of the system at short time and statistical information at long time.  ✅

○ Give an exact thermodynamic description of the system.

**What is a Molecular Dynamics integrator:**
**Select one alternative**

○ It is a function that calculates integrals.

○ It is a mathematical equation that defines the stable evolution of the system.

● It is a computational algorithm that solves the newton equations of motion for the system.  ✅

**What are the characteristics of a good integrator:**

**Select one alternative**

● It must be time-reversible and area-preserving.  ✅

○ It must respect the ensemble distribution but not the detailed balance.

○ It must be time-irreversible and area-preserving.

○ it must be time-irreversible but not area-preserving.

## 19  MD integrators

Why is it important that an MD integrator follows the characteristics that you have chosen in the previous question? (1 point for the right answer).
**Fill in your answer here**

> If an integrator is both, area preserving and time-reversible the shadow hamiltonian is exactly conserved. The shadow hamiltonian is a perturbation of the true hamiltonian of the system. In particular, the shadow hamiltonian is equal to the true hamiltonian plus a correction term of order $\Delta t^2$ with $\Delta t$ the time step of the integrator.

Name 3 integrators and specify which of them are stable respect to the two mentioned properties. (1 points for each integrator and for the right property).
**Enter text here**

> I mention below some of the integrators and their properties that have been discussed in the lectures. Three of them would have been enough to get the maximum mark.
>
> 1. Euler integrator = It is not are-preserving neither time-reversible. It is then not stable because it does not conserve the shadow hamiltonian.
> 2. Verlet integrator = It is are-preserving and time-reversible. It is stable because it conserves the shadow hamiltonian.
> 3. Velocity verlet integrator = It is are-preserving and time-reversible. It is stable.
> 4. Runga-Kutta = It is not are-preserving neither time-reversible. It is not stable.
> 5. Leap-frog = It is are-preserving and time-reversible. It is stable.

## Molecular Dynamics and Monte Carlo simulation

Select the best simulation method to study the topics in the table. (1 point for each right option).

**You can choose only one option.**

| | Molecular Dynamics | Both methods | Monte Carlo |
|---|---|---|---|
| Computation of Radial Distribution function of a system in a gas phase. | ○ | ⊙ ✓ | ○ |
| Rate of a kinetic reaction | ⊙ ✓ | ○ | ○ |
| Adsorption in nanopourous material (MOFs) | ○ | ○ | ⊙ ✓ |
| Time correlation functions | ⊙ ✓ | ○ | ○ |

## Thermostats

Which one of the following thermostats respect the canonical ensemble (NVT)?
**You can select one or more alternatives:**

- ☑ Nose' Hoover thermostat
- ☐ Berendsen thermostat
- ☑ Langevin thermostat

In our Molecular Simulation we are interested on studying the diffusion properties of a nanoparticle in a viscous solvent at constant temperature. Which one of the above is the most adapt thermostat?
**Select only one alternative:**

- ☐ Andersen thermostat.
- ☑ Langevin thermostat.
- ☐ Nose' Hover thermostat.
- ☐ Berendsen Thermostat.

## 22  Thermostats 2

Briefly explain the reason of your choice. (3 points for the right anwer).
**Fill in your answer here**

> The langevin thermostat alters the newton equations to mimic the friction of a solvent that gives random kicks to the solute around. It respects the Boltzmann distribution. Ideally, one could avoid to explicitly use the solvent if the langevin thermostat is used.

# Metropolis Monte Carlo algorithm

The flow chart below should represents the Metropolis Monte Carlo scheme.

There is a mistake. (3 points for the right answer).

**Can you spot the mistake?**

☐ The relation U(R')<U(R) is wrong.

☐ The number P must be bigger than the random number q.

☐ The configuration is counted after it has been accepted.

☑ The configuration is not counted after it has been rejected.

# Metropolis Monte Carlo 2

Correct the mistake and argument your correction. (5 points for the right answer).
**Fill in your answer here**

When the move is rejected the original configuration needs to be counted again to correctly reproduce the boltzmann distribution, i.e. to give the right statistical weight to the configuration.

## 25    Monte Carlo moves

Mention 3 characteristics of a good Monte Carlo move and briefly explain them. (1 point for each characteristic and 1 point for each right explanation).
**Fill in your answer here**

1. Detailed balance
The number of accepted moves that takes the system from the configuration i to the configuration j must be exactly equal to the number of accepted moves from the configuration j to the configuration i.
Mathematically : rho(0)P(o→ n) = rho(n)P(n→ o) or
density probability of being in state 0 * Transition probability to go from 0 to n = density probability of being in state n * Transition probability to go from n to 0.
2. Ergodic – The moves should be chosen in a way the all the phase space is sampled. Every accessible point in configuration space can be reached with a finite number of MC steps from any other point. In this way the time-averages of the properties are equal to its ensemble averages. The system does not dependent on the initial configuration.
3. Markovian – The future states depend only on the present states and not by previous states. Each configuration should be the same if accessed from different starting point. Loss of memory. The moves follow the Markov chain model. In this respect, the probability of each move depends only on the state attained in the previous event.

## Ensembles

Fill the white gaps with the right espression or acronym. (1 point for each gap correctly filled).

An ensemble is a replica of  microstates  with constant macroscopical variables. MD
simulations naturally generate configurations in the  NVE, microcanonical  ensemble that is
characterized by constant number of  particles ,  volume  and
energy . The MC Metropolis method, instead, in his original version, generates
configurations in the  canonical  ensemble that has constant
particles  volume  and a  temperature  that is
constant on average. A modification of the Newtonian MD scheme with the purpose of generating a
thermodynamical ensemble at constant  temperature  on average is called a thermostat
algorithm. There are many other ensembles that can be reproduced by simulations modifying the original MD or
MC methods. Among them the grand canonical ensemble has constant  chemical potential
volume  and  temperature .

In the text above the first sentence is unclear and it has not be counted in the total mark.
The correct sentence should be:
An ensemble is a collection of replicas of a system called microstates that have the same constant
macroscopical variables but that are microscopically different.

## Properties measured in simulation.

The Mean Square Displacement as a function of time allows to obtain the diffusion constant.
**Select one alternative:**

- ● True ✔
- ○ False

The diffusion coefficient cannot be estimated by the velocity auto-correlation function.
**Select an alternative**

- ○ True
- ● False ✔

The presence of peaks in the radial distribution function depends on the temperature:
**Select an alternative**

- ● True ✔
- ○ False

**22** **MD and MC H2017**

**(1 point per correct answer)**
Tell whether this can be done by MD, by MC, by both, or by neither of the two.

Compute the life-time of a base-pair in a DNA molecule

**Select an alternative**

- ⊙ only by MD ✓
- ○ only by MC
- ○ by both MD and MC
- ○ by neither of the two

Compute the average number of hydrogen-bonds that a water molecule donates or accepts in liquid water at room temperature
**Select an alternative**

- ○ only by MD
- ○ only by MC
- ⊙ by both MD and MC ✓
- ○ by neither of the two

Explanation: 'life time' is a dynamical quantity: MD. The 'average number of hydrogen bonds' is a thermodynamic property: both MC and MD.

Compute the collision frequency between two types of molecules in a gas mixture
**Select an alternative**

- ⦿ only by MD ✓
- ○ only by MC
- ○ by both MD and MC
- ○ by neither of the two

Compute the thermodynamic properties of a system obeying the Gibbs ensembe (µVT) in which the total number of particles in the simulation box is not constant
**Select an alternative**

- ○ only by MD
- ⦿ only by MC ✓
- ○ by both MD and MC
- ○ by neither of the two

Explanation: 'collision frequency' is a dynamical quantity. It indicates how often something happens in time. -> MD.
Fluctuating number of particles: this implies one needs to add or delete particles. This can't be done with MD (which follows Newton's equations of motion).

**23** **MD vs MC H2017**

**(1 point per correct answer)**
What can be studied most efficiently with either MD or MC

**Please match the values:**

|  | MC | MD |
|---|---|---|
| The amount of $CO_2$ adsorbed in a zeolite as function of gas pressure | ⊙ ✓ | ○ |
| Vibrational spectra of liquid water | ○ | ⊙ ✓ |
| Rate of ice crystallization | ○ | ⊙ ✓ |

Explanation:
1) Both can be used but MC is generally most efficient. Non-local moves, Gibbs-ensemble.
2) Vibration is a dynamical quantity: MD
3) Rate is a dynamical quantity: MD

# Area-preserving transformations - H2017

**(1 point per correct answer, -1 for wrong answer)**

Consider transformations $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} x' \\ y' \end{pmatrix}$ and $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$ $\to$ in the 2D and 3D plane and let $b$ be an arbitrary constant.

Which of the following transformations are area-preserving:

A: $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$, B: $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x+by \\ x/b+y \end{pmatrix}$, C: $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x^2 \\ \sqrt{y} \end{pmatrix}$, D: $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{x}\left( b+\frac{x^2}{\sqrt{2}} \right) \end{pmatrix}$, E: $\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \frac{1}{2^{1/3}} \begin{pmatrix} x+y \\ y+z \\ z+x \end{pmatrix}$,

F: $\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x \\ yz \\ \ln y \end{pmatrix}$

**Select one or more alternatives:**

- ☑ A ✅
- ☐ B
- ☐ C
- ☑ D ✅
- ☑ E ✅
- ☑ F ✅

We have to consider the Jacobian matrix.

for 2D: $\mathbf{J} = \begin{pmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{pmatrix}$ and for 3D: $\mathbf{J} = \begin{pmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} & \frac{\partial x'}{\partial z} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} & \frac{\partial y'}{\partial z} \\ \frac{\partial z'}{\partial x} & \frac{\partial z'}{\partial y} & \frac{\partial z'}{\partial z} \end{pmatrix}$

The Jacobian $J$ is its determinant: $J = det(\mathbf{J})$. It is area preserving if $|J| = 1$

$$\text{A: } \mathbf{J} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad J = 0 - 1 = -1 \text{ (yes)}$$

$$\text{B: } \mathbf{J} = \begin{pmatrix} 1 & b \\ 1/b & 1 \end{pmatrix}, \quad J = 1 - 1 = 0 \text{ (no)}$$
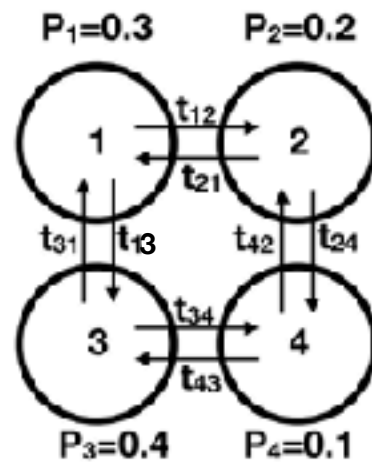
$$\text{C: } \mathbf{J} = \begin{pmatrix} 2x & 0 \\ 0 & \frac{1}{2\sqrt{y}} \end{pmatrix}, \quad J = \frac{x}{\sqrt{y}} \text{ (no)}$$

$$\text{D: } \mathbf{J} = \begin{pmatrix} 2x & 0 \\ \frac{-1}{x^2}(b + \frac{y}{2}) & \frac{1}{2x} \end{pmatrix}, \quad J = 1 \text{ (yes)}$$

$$\text{E: } \mathbf{J} = \begin{pmatrix} \frac{1}{2^{1/3}} & \frac{1}{2^{1/3}} & 0 \\ 0 & \frac{1}{2^{1/3}} & \frac{1}{2^{1/3}} \\ \frac{1}{2^{1/3}} & 0 & \frac{1}{2^{1/3}} \end{pmatrix}, \quad J = 1 \text{ (yes)}$$

$$\text{F: } \mathbf{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & z & y \\ 0 & \frac{1}{y} & 0 \end{pmatrix}, \quad J = -1 \text{ (yes)}$$

$P_1=0.3$   $P_2=0.2$

$P_3=0.4$   $P_4=0.1$

**(1 point per correct answer)**

We consider a system with only 4 states. The probability distribution is given in the figure ($P_i = \frac{e^{-\beta E_i}}{\sum_{j=1}^{4} e^{-\beta E_j}}$). $t_{ij}$ is the probability that starting from state $i$ you will jump to state $j$ in the next step within a MC sampling approach. We will examine which of the following set of jumping probabilities will reproduce this probability distribution ($P_1$, $P_2$, $P_3$, $P_4$).

**Set 1:** $t_{12} = 0.1, t_{31} = 0.15, t_{24} = 0.2, t_{42} = 0.4, t_{43} = 0.2, t_{34} = 0.05, t_{21} = 0.03, t_{13} = 0.04$
**Set 2:** $t_{12} = 0.1, t_{21} = 0.15, t_{24} = 0.2, t_{42} = 0.4, t_{43} = 0.2, t_{34} = 0.05, t_{31} = 0, t_{13} = 0$
**Set 3:** $t_{12} = 0.1, t_{21} = 0.15, t_{24} = 0, t_{42} = 0, t_{43} = 0.2, t_{34} = 0.05, t_{31} = 0, t_{13} = 0$
**Set 4:** $t_{12} = 0.2, t_{21} = 0.3, t_{24} = 0.1, t_{42} = 0.2, t_{43} = 0.2, t_{24} = 0.05, t_{31} = 0.12, t_{13} = 0.16$
**Set 5:** $t_{12} = 0.2, t_{21} = 0, t_{24} = 0.3, t_{42} = 0, t_{43} = 0.6, t_{34} = 0, t_{31} = 0.15, t_{13} = 0$
**Set 6:** $t_{12} = 0.1, t_{21} = 0.1, t_{24} = 0.2, t_{42} = 0.2, t_{43} = 0.2, t_{34} = 0.2, t_{31} = 0.1, t_{13} = 0.1$

**Possible answers:**
A: The MC scheme will sample correctly the probability distribution based on detailed-balance.
B: The MC scheme will sample correctly the probability distribution based on balance, but not detailed balance.
C: Despite being balanced, the MC scheme will not reproduce the probability distribution because the set of jumping rules is not ergodic.
D: The MC scheme will not reproduce the probability distribution because the set of jumping rules is not balanced.

**Please match the values:**

| | A | B | C | D |
|---|---|---|---|---|
| Set 1 | ✓ | | | |
| Set 2 | ✓ | | | |
| Set 3 | | | ✓ | |
| Set 4 | ✓ | | | |
| Set 5 | | ✓ | | |
| Set 6 | | | | ✓ |

# Detailed-balance implies:

$$P_i t_{ij} = P_j t_{ji} \text{ for all } i, j$$

A, B, C, D all obey Detailed-Balance Ergodicity implies that starting from any particular microstate, one should be able to visit all other states eventually.

Set 2 does not allow jumps between state 3 and state 1, but still one can visit state 3 starting from state 1 by jumping via 2 til 4 til eventually 3. So it is ergodic. Also Set 1 and Set 4 are ergodic.

Set 3 does not allow jumps between 1 and 3 and not between 2 and 4. This means that if you start at state 1, for instance, you can never visit state 2 or 4, even not after an infinite number of steps.

Set 5 is not detailed-balance, but it still conserves the probability distribution. To make the 'flea' analogue, suppose there are 300 fleas in state 1, 200 in state 2, 400 in state 4, and 100 in state 4. Then in the next step: 300*0.2=60 fleas will jump out of state 1 (to state 2). At the same time, 400*0.15=60 fleas will jump into state 1 (from state 3). So the number of fleas in state 1 remains the same. The same is true for state 2,3, and 4. So it is balanced, but not detailed-balance

Set 6 is not balance. In the same analogue as the above: 300*0.1=30 fleas will jump out of state 1 while 400*0.1=40 fleas will jump in.

## MC adsorption-1 H2017

**(4 points)**

We want to compute the adsorption energy of argon in a rigid cubic crystal with unit length L at a pressure of 1 bar and a temperature of 800 K. We use a simulation box containing a single unit cell of the crystal and since we assume that at the given pressure and temperature the loading of the crystal is almost always one argon atom per unit cell, we compute our adsorption energy for a NVT ensemble where the number of particles in the system is fixed. This implies that besides the atoms of the crystal unit cell, there is always one argon in the system. Further, since the crystal is rigid we use the common approximation of a fixed lattice: the crystal atoms cannot move. Hence, the only potential energy that we need to compute is the interaction energy of the argon with the fixed atoms in the crystal. We calculate this energy five times using following method:

We take 3 random numbers, $a_1$, $a_2$, and $a_3$, from the [0:1] interval. Then we place the argon molecule at position $(x, y, z) = (a_1 L, a_2 L, a_3 L)$ and compute the interaction energy. This we repeat 5 times. The interaction that we obtain are +0.30 eV, -0.33 eV, -0.33 eV, -0.38 eV, -0.33 eV. What is our best estimate of the adsorption energy? Give your answer here: -0.35 ✅ eV. **Write the answer with two digits after the dot.**

Adsorbtion energy is the average energy that the system gains or loses upon adsorption of the argon atom. Since the crystal atoms are fixed, the only interaction that we need to consider is that of the argon with the crystal atoms. When the argon is outside far away from the crystal its interaction energy will be zero while upon adsorption the energy is likely to go down. Occasionally the interaction energy might be higher than zero due to a collision with the walls of the crystal, but very high positive energies should be considered as rare events which one would probably not observe in a Metroplois MC simulation or in an MD simulation. The algorithm described here is not Metropolis MC! It is just random insertion not trying to avoid overlap of the argon with the crystal. Clearly the first value (+.3 eV) is due to such an overlap. The blind insertion implies that we sample each point in space with equal probability while in practice the argon atom has a higher probability to be found at the low-energy positions. Therefore we need to weight each energy value with the Boltzmann factor.

```
Temperature: 800 K
1/(kbT)=1/(8.6*10^(-5)*800)=14.535/eV

Values:
+.3 eV, -0.33 eV, -0.38 eV,  -0.33 -0.33 eV
(exp(-14.535*0.3)+ exp(-14.535*-0.38)+ 3*exp(-14.535*-0.33))
=613.74875223

(0.3*exp(-14.535*0.3)+(-.38)*exp(-14.535*-0.38)+ 3*(-.33)*exp(-14.535*-0.33))
=-215.052051555

adsorbtion energy: -215.052051555/613.74875223=-0.35

(-215.054688969)/613.738965695
-0.35040090492
```

Both -0.35 as 0.35 would have gotten full points since the sign is merely a convention.

## MC adsorption-2 H2017

**(4 points)**

We want to compute the adsorption energy of argon in a rigid cubic crystal with unit length L at a pressure of 1 bar and a temperature of 800 K. We use a simulation box containing a single unit cell of the crystal and since we assume that at the given pressure and temperature the loading of the crystal is almost always one argon atom per unit cell, we compute our adsorption energy for a NVT ensemble where the number of particles in the system is fixed. This implies that besides the atoms of the crystal unit cell, there is always one argon in the system. Further, since the crystal is rigid we use the common approximation of a fixed lattice: the crystal atoms cannot move. Hence, the only potential energy that we need to compute is the interaction energy of the argon with the fixed atoms in the crystal. If we perform a statistical average of this energy, we should get the experimental verifiable adsorption energy. We calculate this energy using following method:

We take 3 random numbers, $a_1$, $a_2$, and $a_3$, from the [0:1] interval. Then we place the argon molecule at position $(x, y, z) = (a_1 L, a_2 L, a_3 L)$ and compute the interaction energy. This is our first MC move. Then we repeat this, but now we also compute the energy difference between the new configuration minus the energy of the old configuration: $\Delta E$. If it is negative, we accept the new move. If it is positive we take another random number $p$ from the [0:1] interval and we accept the move if $p < e^{-\Delta E/(k_b 800)}$.

Doing so we get the following energies for the different MC moves and $p$ values (for the moves where it was needed):

MC1: $E = +0.30$ eV
MC2: $E = -0.33$ eV,
MC3: $E = -0.40$ eV
MC4: $E = -0.48$ eV, p=0.51
MC5: $E = -0.30$ eV, p=0.21
MC6: $E = -0.35$ eV

To get rid of initialization effect on the average we ignore the first two MC moves. What is our best estimate of the adsorption energy? Give your answer here: **-0.36** ✅ eV. Write the answer with two digits after the dot.

---

$k_B = 8.6 \cdot 10^{-5}$ eV/K. $T = 800$ K. This means $\beta = 1/(k_B T) = 1/(8.6 \cdot 10^{-5} * 800)$ eV$^{-1}$ = 14.53488 eV$^{-1}$.

- MC1: +0.30 eV.
  First move so it is accepted by default.
  Hence, $E_1 = 0.30$ eV

- MC2: -0.33 eV
  Energy goes down, so it is accepted:
  $E_2 = -0.33$ eV

- MC3: -0.40 eV
  Energy goes down, so it is accepted:
  $E_3 = -0.40$ eV

- MC4: +0.48 eV
  Energy goes up
  $\Delta E = 0.88$ eV.
  Chance of accepting this move equals $\exp(-\beta \Delta E) = \exp(-14.53488 * 0.88) = 0.0000028$
  $p = .51$ so move is rejected, we keep the old configuration:
  $E_4 = -0.40$ eV

- MC5: -0.30 eV
  Energy goes up.
  $\Delta E = -0.30 - (-0.40) = 0.10$ eV.
  Chance of accepting this move equals $\exp(-14.53488 * 0.10) = 0.23375$.
  $p = 0.21 < 0.23375$ so move is accepted.
  $E_5 = -0.30$ eV

- MC6: -0.35 eV
  Energy goes down, so it is accepted.
  $E_6 = -0.35$ eV

Ignore MC moves 1 and 2. Take the average of the others: (-0.40-0.40-0.30-0.35)/4=-0.3625 eV. Rounded off to two digits: -0.36. The absolute value 0.36 would have been approved as well. It is basically a convention whether you define adsorption energy with a minus sign or not. It is always assumed that the energy goes down upon adsorption (otherwise it wouldn't adsorb at all).

## Velocity autocorrelation function H2017

**(2 points per correct answer, provide your answer with 2 digits after the dot)**

Consider the velocity correlation function

$$C(t) = \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle$$

of a particle of 10 amu at a temperature of 400 K. What is the value of $C(0)$?

**Write your answer here in km²/sec²**

| 1.00 | ✅ | km²/sec² |

What is the value of $\lim_{t \to \infty} C(t)$?

| 0.00 | ✅ | km²/sec² |

$$C(0) = \langle v_x^2 + v_y^2 + v_z^2 \rangle = 3\langle v_x^2 \rangle = 3\frac{\int_{-\infty}^{\infty} v_x^2 e^{-\beta\frac{1}{2}mv_x^2} dv_x}{\int_{-\infty}^{\infty} e^{-\beta\frac{1}{2}mv_x^2} dv_x} = 3\frac{1}{2}\frac{\frac{1}{\beta\frac{1}{2}m}\sqrt{\frac{\pi}{\beta\frac{1}{2}m}}}{\sqrt{\frac{\pi}{\beta\frac{1}{2}m}}} = \frac{3k_BT}{m} \quad (1)$$

Alternatively, you might remember from thermodynamics that each degree-of-freedom gives $\frac{1}{2}k_BT$. So, also the kinetic energy along $x$: $\langle \frac{1}{2}mv_x^2 \rangle = \frac{1}{2}k_BT$. Using this you end up with the same expression. Putting in the numbers (copy paste the next line in Google (yes, I know you did not have that option during the exam ;-) ))

`3*400 Kelvin*Boltzmann constant/(10 amu) in km^2/second^2 =0.997735183 km^2/s^2`

and you get

`0.997735183 km^2/s^2`

abbreviated as 1.00

For $t \to \infty$, $\mathbf{v}(0)$ and $\mathbf{v}(t)$ become uncorrelated and therfore

$$C(\infty) = \langle \mathbf{v} \rangle^2 = 0$$

# Energies/Free Energies - H2017

**(2 points per correct answer)**
**(provide your answer with 2 digits after the dot. If there is too little information information to give the answer, just write -999).**

Consider two microstates 1 and 2. The energy difference is $E_2-E_1=0.01$ eV. At which temperature is micostate 1 twice as likely as microstate 2?

**Your answer:** 167.42 ✓ Kelvin

At 300 K a certain molecular system with many degrees of freedom has two stable conformations that we will call state A and state B (macrostates). We can assume that the system is at all times in either state A or in state B. The free energy difference between the two states $\Delta F = F_B - F_A$ = -0.02 eV. What is the fraction of time that the system is in state B?

**Your answer:** 0.68 ✓

What is the probability that the system is in state B at 400 K?

**Your answer:** -999 ✓

The probability of a microstate $i$ is proportional to $\propto e^{-\beta E_i}$. Hence, the relative probabilities are equal to $p_1/p_2 = e^{\beta(E_2-E_1)}$ which must be equal to 2.

```
0.01 eV/ln 2/Boltzmann constant
```

gives $167.417864$ kelvin

The difference in free energy at a certain temperature between two macrostates $A$ and $B$ is linked to their relative probability at that temperature

$$p_A(T)/p_B(T) = e^{\beta(F_B(T) - F_A(T))} \text{ or } p_B(T)/p_A(T) = e^{\beta(F_A(T) - F_B(T))}$$

So $P_B/P_A$ equals

```
e^(.02 eV/(Boltzmann constant*300 Kelvin))
```

which gives 2.16763086536. Further we have

$$P_B = 2.16763086536 P_A \text{ and } P_A + P_B = 1 \Rightarrow$$
$$3.16763086536 P_A = 1 \Rightarrow P_A = 1/3.16763086536 \Rightarrow$$
$$P_B = 1 - (1/3.16763086536) = 0.68430664982$$

Last question: The free energy is based on a combination of energy and entropy. The free energy difference at a given temperature tells you what is the relative probability at **that temperature.** However, when the temperature changes normally also the free energy changes and the free energy differences as well. The right answer is therefore that **there is too little information.** In many case, however, assuming that the free energy differences remains the same within a certain temperature range, is a decent approximation (but not always!). If we take that approximation, the answer is 0.64. Students who have given this answer stating that they used this approximation got full points.

# Ising model-Boltzmann statistics H2017
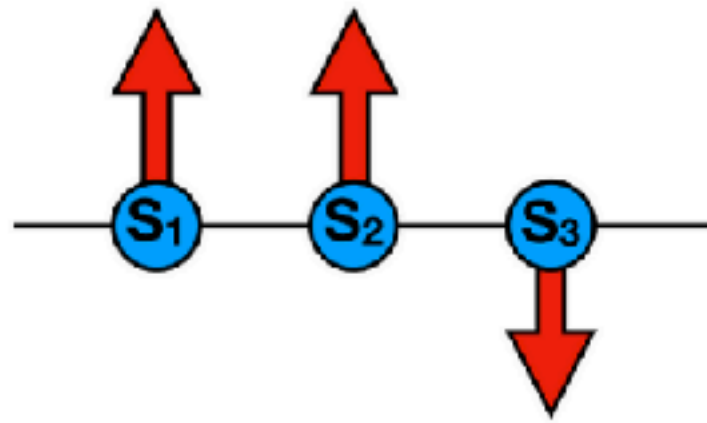


(2 points per correct answer, provide your answer with 2 digits after the dot.)

The Ising model is a famous model to describe magnetic properties of materials. The Ising model consists of nuclei fixed on a lattice which possess spin. The orientation of the nuclei/spins is limited along one direction: it is either up or down. We can also say that the spin of the nuclei is either +1 or -1. The energy of the spin system is given by the interaction of neighboring spins. Depending on the sign of the coupling constant $J$ the neighboring spins tend to be parallel or anti-parallel. Lars Onsager, Nobel prize winner in Chemistry 1968 and former student of the NTNU (then called NTH), became world-famous for his analytical solution of the infinite squared 2D Ising model. To illustrate the concept of discrete Boltzmann statistics and free energy, we will here consider a 1D Ising model with just 3 spins (See figure).

The energy of this 1D Ising model is:

$$U(S_1, S_2, S_3) - \sum_{i=1}^{2} JS_i S_{i+1}$$

with J=-0.1 eV and $S_i$ take the values +1 or -1. The different possible microstates are (1,1,1), (1,1,-1) etc

At 0 K, what is the probability that all spins are aligned (all up or all down)?

Write your answer here:  1  ✓

What is this probability in the limit $T \rightarrow \infty$?

Write your answer here:  0.25  ✓

Note that the coupling J is negative. This implies that the energy is the lowest is all spins are aligned. When T=0 the Boltzmann weights become infinite but the Boltmann weights of the lowest energy state is also infinitely higher than all states with higher energy.  As a result, at lim T=0 the system takes the lowest energy state with 100% certainty.

At lim T=infinite all Boltzmann weights are equal and each state becomes equally likely. There are 2^3=8 states of which 2 are fully aligned (+++ and  - - -). Hence, the chance is 25%

**(4 points, provide your answer with 2 digits after the dot.)**

We consider once more the Ising model with 3 spins:

$U(S_1, S_2, S_3) = \sum_{i=1}^{2} JS_i S_{i+1}$ with J=-0.1 eV

Consider the order parameter $\lambda = \left\| \sum_{i=1}^{3} S_i \right\|$, the absolute value of the sum of the spins.
There are only two possible values that $\lambda$ can take: 3 or 1.

Now we define the free energy $F(\lambda)$ as function of the order parameter. Note that this free energy profile is properly defined except for an arbitrary constant that we can always add to the free energy function. If we set $F(1) = 0$, what is $F(3)$ at 300 K?

**Write your answer here:** -0.18 ✓ eV

---

Consider the following table where for each state we list the energy (divided by coupling constant $J$) and the order parameter $\lambda$:

| State | $E/J$ | $\lambda$ |
|-------|-------|-----------|
| $+++$ | -2 | 3 |
| $++-$ | 0 | 1 |
| $+-+$ | 2 | 1 |
| $+--$ | 0 | 1 |
| $-++$ | 0 | 1 |
| $-+-$ | 2 | 1 |
| $--+$ | 0 | 1 |
| $---$ | -2 | 3 |

The probability for $\lambda = 3$ and $\lambda = 1$ equal

$$P(3) = \frac{2e^{\beta 0.02eV}}{Z}, \quad P(1) = \frac{2e^{-\beta 0.02eV} + 4}{Z}$$

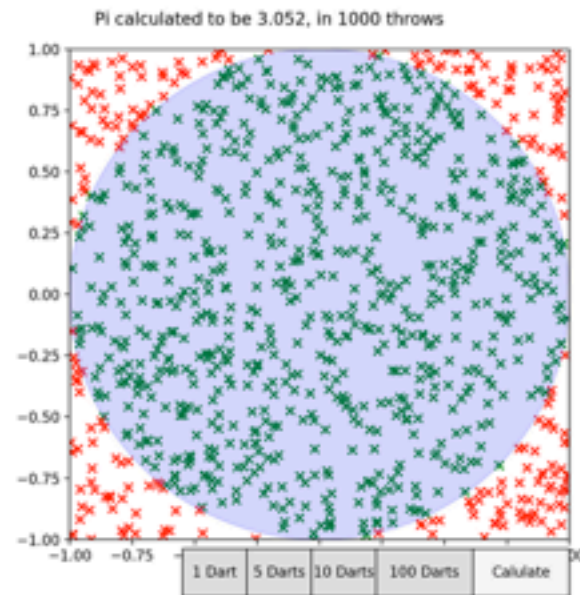with $Z = 2e^{\beta 0.02eV} + 2e^{-\beta 0.02eV} + 4$. Free energy differences are related to the relative probabilities and since $F(1)$ is set to be 0:

$$\frac{e^{-\beta F(3)}}{e^{-\beta F(1)}} = e^{-\beta F(3)} = \frac{P(3)}{P(1)} = \frac{2e^{\beta 0.02eV}}{2e^{-\beta 0.02eV} + 4}$$

$$F(3) = -k_B T \ln \left[ \frac{2e^{\beta 0.02eV}}{2e^{-\beta 0.02eV} + 4} \right]$$

which gives -0.182 eV

Pi calculated to be 3.052, in 1000 throws

**(4 points)**

In the above we see how to compute $\pi$ using Monte Carlo. We basically generate $M$ random pairs $(x, y)$, "throws", where both $x$ and $y$ are random numbers from the $[-1:1]$ interval. Then we compute $r^2 = x^2 + y^2$ and count how often $r^2 < 1$. The number $\pi$ is consequently calculated by:

$$4 \left( \frac{\text{number of } (x,y) \text{ pairs with } r^2 < 1}{M} \right)$$

In the figure you see the result after 1000 throws. Calculate the relative statistical error in percentages

**Write your answer here:** 1.65 ✅ %

Note that the deviation from the True result (3.1415-3.052) is larger than that (2.85%). That is not so surprising as the chance to be outside 3.1415+-1.65% interval is 32%. In practice it can even be higher as there is also an error on the error. In any case, just taking the derivation from the true result would not be appropriate. The error could be very small by plain luck.

An alternative solution, however, would be to calculate the result based on the sample average 3.052/4 instead of the exact average pi/4. When calculating the standard deviation one should then use in principle the Bessel correction (divide by M-1 instead of M) since the sample mean is used instead of the true mean. This has, however, not a significant effect. If the sample mean is used, the relative error would be 1.76. Students who gave this result also received full score.

Initially, we calculate $\pi/4$ and then we get $\pi$ by multiplying with 4. The relative error in the first result $(\pi/4)$ is the same as the relative error in the last result of $\pi$. Let's call $p$ the correct outcome of the initial calculation: $p = \pi/4$. The possible measurements are either 0 or 1. The first has a chance $1 - p$, the second has a chance $p$. Therefore the standard deviation equals

$$\sigma^2 = \left\langle (p_i - p)^2 \right\rangle = p(1-p)^2 + (1-p)p^2 = p(1-p) = \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)$$

The relative error after $M$ uncorrelated measurements is

$$\epsilon_r = \frac{\sigma}{p\sqrt{M}} = \sqrt{\frac{1-p}{pM}} = \sqrt{\frac{1 - \frac{\pi}{4}}{\frac{\pi}{4}1000}} = 1.6529\%$$

**39** **Thermostats H2017**

Monte Carlo usually samples the canonical ensemble, while Molecular Dynamics naturally samples the microcanonical ensemble, unless it is performed with a thermostat. Mention 3 thermostats which allow you to get the correct thermodynamical properties using MD within the canonical ensemble.

**1 point per correct answer**

Format   ▾   **B** *I* U x₂ x² Iₓ   ✂ ▢ ▢   ↶ ↷   ⟳   ⋮≡ ⋮≡   Ω ⊞   ✎ Σ   ✖

Words: 0/200

Possible answers: Andersen thermostat, Nose-Hoover (chains), Langevin.
Not correct is Berendsen thermostat is it is not exactly sampling the canonical distribution. Improvements of the Bussi-thermosthat like the Bussi-thermosthat are correct though

## TKJ4205/KJ8902 Molecular Modelling
## Exam 2016. Questions Part2

**Norges Teknisk-Naturvitenskapelige Universitet (NTNU)**

## Question 1: MD and MC-1

Tell whether this can be done by MD, by MC, by both, or by neither of the two.

**a) Given an initial set of positions and velocities, knowing the positions and velocities 2 nanoseconds later:** by neither of the of the two (Lyapunoc instability)

**b) Knowing how many argon molecules are on average adsorbed per nm2 on a graphene surface after equilibration when a graphene is in contact with an argon gas of 1 bar at room temperature:** by both MD and MC

## Question 2: MD and MC-2

**a) Compute diffusion constants:** by MD

**b) Compute the average number of hydrogen-bonds that a water molecule donates or accepts in liquid water at room temperature:** by both MD and MC

## Question 3: MD and MC-3

**a) Compute the life time of a hydrogen bond:** By MD

**b) Compute the fraction of time that a water molecule has more than 4 hydrogen bonds in liquid water:** By both MD and MC. Since it is the fraction of time that the system is in a certain state, this corresponds to the probability of that state. This probability is a thermodynamic property which can be computed by both MD and MC.

## Question 4: MD vs MC

**What can be studied most efficiently with either MD or MC.**

- **CO2 gas diffusion in the atmosphere:** MD, because it is dynamics

- **The amount of CO2 adsorbed in a zeolite as function of gas pressure:** MC is most efficient, because allows fluctuating number of particles, non-local moves, CBMC, etc.

- **Radial distribution function of liquid water:** Both possible but MD more efficient. Condensed phase, so MC will suffer from the fact that maximum-displacement must be small. MD moves more molecules at the same time. There is also the inertia effect: taking the velocities into account gives a kind of flow which helps to decorrelate more rapidly.

- **Nucleation rate of ice formation:** MD, dynamical property

- **The rate of methane absorption in a zeolite:** If even MC is often more efficient for adsorption, here it is the rate that is asked. Dynamical property, so must be MD

## Question 5: MD integrators

- **The Runga-Kutta is in principle more accurate than velocity-Verlet after only a few MD steps.** True. The Runga-Kutta integrator is higher order scheme (based on a Taylor expansion of higher order in $\Delta t$). This means the systematic error in the new positions and velocities after an MD step deviate less from the true positions and velocities than with velocity-Verlet if the same time step is taken.

- **The integrators which are area-preserving and time-reversible have a shadow-Hamiltonian which is approximately conserved.** False. It is exactly conserved.

## Question 6: Area-preserving transformations

Area preserving operators: A,B,D

## Question 7: Ensemble averages-1

a: True, b: True, c: False

## Question 8: Ensemble averages-2

$$\frac{<A \exp(\beta' E) \exp(-\beta E)>}{<\exp(\beta' E) \exp(-\beta E)>}$$

## Question 9: Metropolis-Hastings

The generation probability is not symmetric in the Metropolis-Hastings algorithm: $P_{\text{gen}}(o \to n) \neq P_{\text{gen}}(n \to o)$

## Question 10: MC adsorption

$k_B = 8.6 \cdot 10^{-5}$ eV/K. $T = 800$ K. This means $\beta = 1/(k_B T) = 1/(8.6 \cdot 10^{-5} * 800)$ eV$^{-1}$ =14.53488 eV$^{-1}$.

- MC1: +0.30 eV.
  First move so it is accepted by default.
  Hence, $E_1$=0.30 eV

- MC2: -0.33 eV
  Energy goes down, so it is accepted:
  $E_2$ =-0.33 eV

- MC3: -0.40 eV
  Energy goes down, so it is accepted:
  $E_3$ =-0.40 eV

- MC4: +0.48 eV
  Energy goes up
  $\Delta E = 0.88$ eV.
  Chance of accepting this move equals $\exp(-\beta \Delta E) = \exp(-14.53488 * 0.88) = 0.0000028$
  $p = .51$ so move is rejected, we keep the old configuration:
  $E_4$=-0.40 eV

- MC5: -0.30 eV
  Energy goes up.
  $\Delta E = -0.30 - (-0.40) = 0.10$ eV.
  Chance of accepting this move equals $\exp(-14.53488 * 0.10) = 0.23375$.
  p=0.21<0.23375 so move is accepted.
  $E_5 = -0.30$ eV

- MC6: -0.35 eV
  Energy goes down, so it is accepted.
  $E_6 = -0.35$ eV

Ignore MC moves 1 and 2. Take the average of the others: (-0.40-0.40-0.30-0.35)/4=-0.3625 eV. Rounded off to two digits: -0.36. The absolute value 0.36 would have been approved as well. It is basically a convention whether you define adsorption energy with a minus sign or not. It is always assumed that the energy goes down upon adsorption (otherwise it wouldn't adsorb at all).

## Question 11: MC efficiency

Large dmax implies a lot of rejections. low dmax means that moves are too similar. Both cause correlations between successive moves.

## Question 12: MC move criteria

**Write four important qualities that a good Monte Carlo move must meet to produce correct Boltzmann statistics. Number these 1,2,3,4 so that you can refer to these numbers in the next question** MC adsorption

i) Detailed-Balance

ii) Markovian

iii) Ergodic

iv) Area preserving (otherwise correct for it in the acceptance move)

## Question 13: Validity of MC move-1

No. Because of detailed balance. If you cross the $x = 0$ line with a big step larger than $d_1$, then it is impossible to reobtain the old configuration again.

## Question 14: Validity of MC move-2

No. This move is not Markovian.

## Question 15: Validity of MC move-3

No. This move is not Ergodic.

## Question 16: 1D potential

$\beta = 1/(0.008314 * 1000) = 0.1203$ mol/kJ. $\exp(-\beta * 1kJ/mol) = \exp(-0.1203) = 0.8867$

- $<x>_{T=1000K} = \frac{\int \mathrm{d}x\, x e^{-\beta U}}{\int \mathrm{d}x\, e^{-\beta U}} = \frac{\int_1^2 \mathrm{d}x\, x + \int_2^3 \mathrm{d}x\, x \times 0.8867}{1 + 0.8867} = \frac{1.5 + 2.5 \times 0.8867}{1.8867} = 1.96$ nm.

- $<U>_{T=1000K} = (0 + 0.8867)/(1 + 0.8867) = 0.47$ kJ/mol

- $C(0) = \frac{\int_1^2 x^2 + \int_2^3 x^2 \times 0.8867}{1.8867} = \frac{(8-1)/3 + 0.8867*(27-8)/3}{1.8867} = 4.21$ nm$^2$

- $C(1ps)$ can't be calculated

- $C(\infty) = 1.96^2 = 3.88$

## Question 17: Free energies

Thermodynamic integration and umbrella sampling

## Question 18: Error analysis-Path Sampling

The error equals the standard deviation divided by the square-root of the number of uncorrelated measurements.

$$\epsilon = \frac{\sigma}{\sqrt{N_u}} \tag{1}$$

The relative error equals

$$\epsilon_r = \frac{\sigma}{p} \frac{1}{\sqrt{N_u}} \tag{2}$$

with $p = P_A(\lambda_{i+1}|\lambda_i) = 0.35$

Say $h_i$ is the $i$-th measurement of the binary function. This implies that $h_i = 1$ with a probability $p$ and 0 with a probability $1 - p$. The average of $h_i$ over all measurements should be $p$. The standard deviation is therefore

$$\sigma = \sqrt{\langle (h_i - p)^2 \rangle} = \sqrt{\langle p(1-p)^2 + (1-p)(0-p)^2 \rangle} = \sqrt{\langle p(1-p) \rangle} \tag{3}$$

and thus

$$\epsilon_r = \sqrt{\frac{1-p}{p}} \frac{1}{\sqrt{N_u}} \tag{4}$$

Hence,

$$N_u = \frac{1-p}{p} \frac{1}{\epsilon_r^2} = \frac{1-0.35}{0.35} \frac{1}{0.06^2} = 515.87 \tag{5}$$

The statistical inefficiency tells you how many MC moves are needed to get a uncorrelated path. Since we generate one million paths, but only 515.87 uncorrelated paths, it means it takes $1000000/515.87 = 1938$ before a new uncorrelated path is created. The statistical inefficiency is, therefore, 1938

# TKJ4205/KJ8902 Molecular Modelling
# Exam 2015. Questions 18-36

**Norges Teknisk-Naturvitenskapelige Universitet (NTNU)**

## Question 18: MD and MC-1

**a) Compute time-correlation functions:** only by MD Because time evolution

**b) Compute the average number of hydrogen-bonds that a water molecule donates or accepts in liquid water at room temperature:** by both MD and MC

## Question 19: MD and MC-2

**a) Given an initial set of positions and velocities, knowing the positions and velocities 1 nanosecond later:** by neither of the two. MC won't work because it requires evolving the equations of motion. MD won't work because of chaos (Lyapunov instability). 1 nanosecond typically involves 1 million time steps which is more than enough for chaos to kick in.

**b) Knowing how many CO2 particles get adsorbed on average inside a MOF per gram MOF from the atmosphere at ambient conditions using a simulation box containing 2X2X2 unit cells of the MOF crystal with periodic boundary conditions in all directions. (MOF=Metallic Organic Framework, which is a nanoporous crystal):** Only by MC. To solve this question one needs to have a simulation in which the number of CO2 molecules are allowed to fluctuate. With MD this would only be possible if one simulates the MOF crystal in contact with a CO2 gas. MC is more flexible since it allows for inserting and deleting CO2 molecules. The question clearly mentions that the simulation box has only one piece of crystal with periodic boundary conditions. Hence, there is no possibility in an MD simulation to see CO2 molecules move outside the crystal. Therefore, only MC can be used.

## Question 20: MD integrators

**a) The velocity-Verlet algorithm ensures that the temperature remains constant.** False. The plain velocity-verlet algorithm keeps the energy (almost) constant. To keep temperature constant the velocity-verlet needs to be combined with a thermostat.

**b) The Euler algorithm is area-preserving.** False.

## Question 21:Ensemble averages-1

**a) <A+B>=<A>+<B>.** True

**b)** $\langle AB \rangle = \langle A \rangle \langle B \rangle$ False. Only if $A$ and $B$ are uncorrelated the above is true. Some students who choose True but explained in the last free text box that it is under the assumption of uncorrelated $A, B$. They got full points anyway.

**c)** $\langle A(0)B(t) \rangle = \langle A(-t)B(0) \rangle$. True

## Question 22: Ensemble averages-2

Andersen and Langevin. NVE has no temperature dependence. Berendsen is based on a velocity rescaling procedure that get the average kinetic energy correct and also some of its fluctuations, but is not exactly equivalent to a correct Boltzmann statistics based constant temperature ensemble.

## Question 23: NVE/NVT

**a) If we run MD in the NVE ensemble, the number of particles in the system, the volume, and the total energy is kept constant:** True

**b) If we run MD in the NVT ensemble, the number of particles in the system, the volume, and the kinetic energy is kept constant.:** False. On average the kinetic energy should be 3/2 N kB T with N the number of particles, but a good thermostat also ensures that the fluctuations in the kinetic energy are statistically correct.

## Question 24: MC move-1

**Write four important qualities that a good Monte Carlo move must meet to produce correct Boltzmann statistics.** A complete answer should contain

- Obeying detailed balance

- Markovian

- Ergodic

- Area preserving

At the last point one could add that Area-preserving is not strictly necessary if one corrects for it in the acceptance-rejection rule

## Question 25: MC move-2

**a) Any MC move obeying super detailed balance is also obeying detailed balance.** True

**b) Any MC move obeying detailed balance is also obeying super detailed balance.** False

## Question 26 (27): MC adsorption

MC adsorption
Boltzmann constant in eV/Kelvin: $8.61733034 \cdot 10^{-5}$ eV/K. Temperature: 800 K.

Adsorbtion energy is the average energy that the system gains or loses upon adsorption of the argon atom. Since the crystal atoms are fixed, the only interaction that we need to consider

is that of the argon with the crystal atoms. When the argon is outside far away from the crystal its interaction energy will be zero while upon adsorption the energy is likely to go down. Occasionally the interaction energy might be higher than zero due to a collision with the walls of the crystal, but very high positive energies should be considered as rare events which one would probably not observe in a Metroplois MC simulation or in an MD simulation. The algorithm described here is not Metropolis MC! It is just random insertion not trying to avoid overlap of the argon with the crystal. Clearly the first value ($+.3$ eV) is due to such an overlap. The blind insertion implies that we sample each point in space with equal probability while in practice the argon atom has a higher probability to be found at the low-energy positions. Therefore we need to weight each energy value with the Boltzmann factor.

```
Temperature: 800 K
1/(kbT)=1/(8.6*10^(-5)*800)=14.535/eV

Values:
+.3 eV, -0.33 eV, -0.38 eV,  -0.33 -0.33 eV
(exp(-14.535*0.3)+ exp(-14.535*-0.38)+ 3*exp(-14.535*-0.33))
=613.74875223

(0.3*exp(-14.535*0.3)+(-.38)*exp(-14.535*-0.38)+ 3*(-.33)*exp(-14.535*-0.33))
=-215.052051555

adsorbtion energy: -215.052051555/613.74875223=-0.35

(-215.054688969)/613.738965695
-0.35040090492
```

Both -0.35 as 0.35 would have gotten full points since the sign is merely a convention. If result was wrong but a correct explanation was given at 27, then 4 points are given.

## Question 28: RMSD of HCl

The plain formula of the RMSD equals

$$\sqrt{\sum_i = dr_i^2/N} \tag{1}$$

Here $N$ is the number of atoms. $dr_i$ is the distance between the position of atom $i$ in the reference state and the same atom position in the present state. However, the RMSD has to be minimized upon rotations and translations.

Say, if H-Cl is the reference state and H*****Cl is the present state, the minimum is achieved whenever the present state is moved on top of the reference state: H*H-Cl*Cl. The distance between the two H's is than $(1.68-1.33)/2=0.175$ Å. The distance between the Cl's is the same. Hence, also the RMSD equals 0.175 Å. Some students, who got an answer that was close but not exact, probably moved the to states together so that one atom exactly overlapped. They got one point.

## Question 29 (30): Error Analysis

```
The average equals:
(0.32+0.37+0.33+0.37+0.36)/5=0.35 eV
The estimated standard deviation:
sqrt(((0.32-.35)^2 + (0.37-.35)^2 + (0.33-.35)^2 + (0.37-.35)^2+ (0.36-.35)^2)/(5-1))
gives
0.02345207879 eV

The absolute error equals
0.02345207879/sqrt(5) = 0.01048808847 eV

The relative error equals
0.01048808847/.35 = 0.02996596705
or 3.0 %

b.
The error should also be
(standard-deviation of the 'raw' data)/sqrt(N)=0.11 eV/sqrt(N)
with N the effective number of uncorrelated data-point.
Note that N<1 000 000 because successive data point are correlated.
Inverting the above gives
N=(0.11 eV/absolute error)^2=(0.11/0.01048808847)^2=110.000000245
So the statistical inefficiency equals 1000000/110.000000245=9090.90907066
approx. 9091
```

Any number between 9041 and 9141 was automatically approved.

## Question 31:

The potntial energy surface is 3N-dimensional function. Calculating the free energy as function of a single or a few order parameters is a way to reduce the dimensionality. In addition, the gas phase approach to estimate reaction barriers becomes meaningless in high dimensional systems and elevated temperatures. A reaction barrier in the gas phase is, for instance, often defined as the difference in potential energy of the transition state (saddle point) and the minimum potential energy point in the reactant state. In the liquid state these points are not relevant since the geometry of the minimum potential energy state would correspond to ice not to liquid water. Therefore it is more relevant to use a language related to probabilities. The free energy as function of a reaction coordinate, for instance, tells you how likely it is to be in the reactant state, in the product state or the transition state (in this language the transition state is the top of the free energy curve, not a single saddle point). If a student mentions two of these points (reduction of dimensionality, loss a meaning of minimum potential energy point for the liquid state, description based on probabilities), full points were given.

## Question 32: Free energy

**a)** Anything can happen. We have too little information

The free energy as function of an order parameter at a given temperature tells you how often this order parameter has a given value (or what is the probability that the system obtains an order parameter equal to a given value). The lower the free energy at a certain value of the order parameter, the more likely it is. This probability involves both potential energy but also the accessible space in the orthogonal directions (one could call these entropic contributions). Depending on these two contributions, the free energy profile might change or not as function of temperature. Since we don't know these, we can't make a prediction.

**b)** $P(X)$ becomes larger than $P(Y)$. The probabilty to be inside a certain interval $[a : b]$ of the order parameter $q$ equals the ratio of two integrals

$$P(q \in [a : b]) = \frac{\int_a^b \exp(-F(q)/k_B T)\mathrm{d}q}{\int_{-\infty}^{+\infty} \exp(-F(q)/k_B T)\mathrm{d}q} \tag{2}$$

Increasing T implies that two states that have different free energies become more similar regarding their probabilities. State X and Y are at low T equally likely. The higher free energy of state X is compensated by having a wider range. In other words, each point q of the order parameter within the X interval has a lower probability than the point q within the Y interval. However, this is yet compensated by the fact that the X interval is larger. If T is increased, the probabilties become more similar and therefore X wins due to its larger accessible space.

## Question 33: Two reaction channels-1

False. The free energy barrier is not the only parameter determining the reaction rate. If a reaction has a low barrier but also a very low transmission coefficient, the rate will be low and possibly lower than another reaction having a high barrier but a transmission coefficient close to 1. People not familiar with the concept of transmission coefficient could also think in terms of Arrhenius behavior. Arrhenius tell us that a reaction rate is proportional to $\exp(-E_a/k_B T)$ with $E_a$ the activation energy which in most cases can be compared to the free energy barrier. However, if the pre-exponential constant is very low, the rate constant of a given reaction might be lower than another reaction having a higher activation energy.

## Question 34: Two reaction channels-2

The barrier will be less than 18 kcal/mol.

After projection, the probability to be in the blue interval is equal to the probability to be in the top blue square plus the probability to be in the bottom blue square. The probability to be in the green interval is the same as the probability to be in the green square before projection. Hence, the probability to be in the blue interval is higher than to be in either the top or bottom blue square, while the probability to be in the green interval is the same as the probability to be in the green square. The relative probability of the blue interval is therefore higher. The free energy barrier must therefore be lower than 18 kcal/mol.

## Question 35: Di-atomic molecule

The free energy equals

$$F(d) - F(r_{eq}) = F(d) = -k_B T \ln(P(d)/P(r_{eq})) \tag{3}$$

where $P(d)$ is the probability that the distance equals a value $d$. This probability is proportional to $\exp(-V(d)/kbT)\cdot$ (number of configurations in which distance is $d$). The number of configurations at a large distance is larger than at a small distance, because if we fix one atom, the other atom can be placed at any position of the sphere with radius $d$ and having the first atom as a center. The number of possibilities is therefore proportional to $4\pi d^2$, the surface of the sphere. Therefore $P(d)$ is proportional to $d^2 \exp(-V(d)/k_B T)$. Since $F(r_{eq}) = 0$ by convention we can write $F(d) = -k_B T \ln(P(d)/P(r_{eq})) = -k_B T \ln[(d^2/r_{eq}^2)\exp(-\{V(d)-V(r_{eq})\}/k_B T)] = -k_B T \ln[d^2/r_{eq}^2] + V(d) = -2k_B T \ln[d/r_{eq}] + V(d) = -2k_B T \ln[d/r_{eq}] + \frac{1}{2}k(d - r_{eq})^2$.

## Question 36: D20

Yes, they should be the same.

Any geometric average can be written as

$$\langle A(\mathbf{R}^N)\rangle = \frac{\int d\mathbf{R}^N\, A(\mathbf{R}^N)e^{-\beta V(\mathbf{R}^N)}}{\int d\mathbf{R}^N\, e^{-\beta V(\mathbf{R}^N)}} \tag{4}$$

As you see, mass is not a part of the equation. This would be different if we want to compute average properties which depend on the velocities or if we want to compute time-correlation functions, diffusion coefficients etc. These are mass dependent.

# Digital exam questions 2015-2019

Per-Olof Åstrand
September 14, 2020

# 1 Quantum chemistry

## 1.1 Eigenvalues

1. Which is the correct eigenvalue to the function $f(x) = Ce^{ax}$, where $C$ and $a$ are constants, for the operator $\frac{\partial^2}{\partial x^2}$? (Exam 2017)

   □ $aC$

   □ $a^2C$

   □ $a$

   ■ $a^2$

2. Which is the correct eigenvalue to the function $f(x) = a\sin(kx)$, where $a$ and $k$ are constants, for the operator $\frac{\partial^2}{\partial x^2}$? (Exam 2015)

   □ $k$

   □ $ak^2$

   □ $-ak$

   ■ $-k^2$

## 1.2 Hamiltonian and exact solutions

3. What is true about the molecular Hamiltonian we normally use in quantum chemical calculations? (Exam 2019)

   □ We include the magnetic field interactions between the particles.

   ■ We regard the nuclei and electrons as *point* particles.

   □ The electrostatic interactions are screened by the dielectric constant of the medium.

   □ The acceleration of the particles is included explicitly in the Hamiltonian.

4. In which coordinate system is the Schrödinger equation for the one-electron atom solved analytically? (Exam 2019)

   ■ Spherical polar coordinates

   □ Cartesian coordinates

   □ Cylindrical coordinates

□ Prolate spheroidal coordinates

5. How many components do we have of a d-orbital? (Exam 2019)

   □ 3
   ■ 5
   □ 7
   □ 9

## 1.3 Orbitals

6. We regard a set of orbitals, $\varphi_i = 1, \ldots, m$. What do we mean by that the orbitals are *normalized*? ($\delta_{ij}$ is the Kroenecker delta.) (Exam 2018)

   □ $\langle \varphi_i | \varphi_j \rangle = \delta_{ij}$
   ■ $\langle \varphi_i | \varphi_i \rangle = 1$
   □ $\langle \varphi_i | \varphi_j \rangle = 0$
   □ $\langle \varphi_i | \varphi_i \rangle = 0$

7. What do we mean by that the orbitals form an *orthonormal* set? (Exam 2018)

   ■ $\langle \varphi_i | \varphi_j \rangle = \delta_{ij}$
   □ $\langle \varphi_i | \varphi_i \rangle = 1$
   □ $\langle \varphi_i | \varphi_j \rangle = 0$
   □ $\langle \varphi_i | \varphi_i \rangle = 0$

8. What is *necessarily* true about the *exact* wavefunction for the molecular Hamiltonian? (Exam 2018)

   □ The states (eigenfunctions) are normalized, and the eigenvalues (energies) have no imaginary part.
   □ The states (eigenfunctions) form an orthogonal set, and the eigenvalues (energies) may have an imaginary part.
   ■ The states (eigenfunctions) form an orthogonal set, and the eigenvalues (energies) have no imaginary part.
   □ The states (eigenfunctions) are normalized, and the eigenvalues (energies) may have an imaginary part.

9. Why do we use *orthonormal* orbitals in molecular orbital theory? (Exam 2018)

   ■ It is convenient since many integrals will be exactly zero.
   □ They have to form an orthonormal set to give a correct description of the molecular wavefunction.
   □ It is a consequence of that the basis set forms an orthonormal set of functions.
   □ We do not use an orthonormal set of orbitals in molecular orbital theory.

## 1.4 Atomic units

10. How does Coulomb's law look like in atomic units for the interaction between two electrons at a distance $r$? (Exam 2016, 2017, 2019)

    ■ $\frac{1}{r}$
    □ $\frac{e^2}{4\pi r}$
    □ $\frac{e^2}{4\pi\epsilon_0 r^2}$
    □ $\frac{e}{4\pi\epsilon_0 r}$

11. Which of the following properties/constants are *not* 1 in atomic units? (Exam 2016, 2017)

    □ $\hbar$
    ■ Speed of light
    □ Mass of electron
    □ $4\pi\epsilon_0$

12. How does the kinetic energy operator for an electron look like in atomic units? (Exam 2017)

    ■ $-\frac{1}{2}\nabla^2$
    □ $-\nabla^3$
    □ $-\nabla^2 - \frac{Z}{r}$
    □ $-\nabla^2 - \frac{1}{2}\nabla^4$

13. What is the energy unit in the atomic unit system? (Exam 2019)

    □ kcal/mol
    □ kJ/mol
    ■ hartree
    □ eV

## 1.5 Born-Oppenheimer approximation

14. Which of the following statements regarding the Born-Oppenheimer approximation is *not* true? (Exam 2017, 2019)

    □ The Born-Oppenheimer approximation results in a separation of the wave function into an electronic wave function and a nuclear wave function.
    □ When solving the electronic problem, the nuclear positions are kept fixed and are treated as *parameters* (in contrast to variables).
    □ When running a quantum chemical calculation, we have to provide the molecular geometry in the input file.
    ■ The Coulomb interaction between the electrons is ignored.

### 1.6   Classical vs quantum mechanics

15. We discussed to which extent classical mechanics is a good approximation for various types of motions by regarding what is measured in various spectroscopic techniques (which measures transitions between quantum mechanical states). (Exam 2017, 2018)
    Please match the values:

    - ■ Infrared spectroscopy - vibrational motion
    - ■ Microwave spectroscopy - rotational motion
    - ■ No spectroscopic technique - translational motion

### 1.7   Molecular orbital theory

16. Which of the following combinations of spatial orbitals, with electrons 1 and 2, has the correct symmetry if the spin part is anti-symmetric? (Exam 2019)

    - ☐ 1s(1) 2s(2) - 1s(2) 2s(1)
    - ■ 1s(1) 1s(2)
    - ☐ 1s(1) 2s(2) - 1s(1) 2p(2)
    - ☐ 1s(1) 2s(2)

### 1.8   Basis sets

17. When improving a basis set, we talk about adding either *polarization* functions or *diffuse* functions, which you did in the exercises. Which type of functions are most important to add for improving the description of respective property? (Exam 2015, 2018)
    Please match the values:

    - ■ polarization functions - geometry
    - ■ diffuse functions - polarizability

18. What type of basis sets did we use in the exercises? (Exam 2019)

    - ☐ Slater-type functions
    - ■ Gaussian functions
    - ☐ Plane waves

19. If we would like to increase the accuracy of the calculation of molecular polarizabilities, what is the best option? (Exam 2019)

    - ■ Add diffuse functions to the basis set.
    - ☐ Add polarization functions to the basis set.
    - ☐ Decrease the size of the basis set.

### 1.9  Hartree-Fock calculations

20. What does the variational theorem state for the Schrödinger equation? (Exam 2019)

    ☐ The approximate dipole moment is always higher than the exact dipole moment.

    ☐ The approximate dipole moment is always lower than the exact dipole moment.

    ☐ The approximate energy is always lower than the exact energy.

    ■ The approximate energy is always higher than the exact energy.

21. What do we lack within the Hartree-Fock approximation? (Exam 2019)

    ■ Electron correlation

    ☐ The kinetic energy of the electrons

    ☐ The Coulomb interaction between the electrons

    ☐ The Pauli principle is not fulfilled.

22. The following results are taken from a Hartree-Fock calculation on the HF molecule for the total energy. Which of the following tables are likely to be correct? Which are for sure wrong? Are there tables where we cannot really tell either away? (Exam 2015, 2017)

| Table 1 | | Table 2 | | Table 3 | |
|---|---|---|---|---|---|
| basis set | E/hartree | basis set | E/hartree | basis set | E/hartree |
| cc-pVDZ | -100.019419 | cc-pVDZ | -100.058021 | cc-pVQZ | -100.067695 |
| cc-pVTZ | -100.058021 | cc-pVTZ | -100.019419 | aug-cc-pVQZ | -100.068568 |
| cc-pVQZ | -100.067695 | cc-pVQZ | -100.067695 | daug-cc-pVQZ | -100.068605 |

    ☐ Table 1 is for sure wrong.

    ■ Table 2 is for sure wrong

    ☐ Table 3 is for sure wrong

    ☐ We do not have sufficient information to make an assessment for at least one of the tables.

23. If we for a calculation on for example the water molecule increases the size of the basis set in the next calculation (but everything else is the same), what is a requirement? (Exam 2019)

    ☐ The dipole moment is closer to the experimental value.

    ☐ The first excitation energy becomes smaller.

    ■ The total energy becomes lower.

24. What is a feature of the restricted Hartree-Fock method as compared to the unrestricted Hartree-Fock method? (Exam 2019)

■ We have two electrons in each spatial orbital.

□ We include relativistic effects.

□ Electron correlation is included.

□ The Born-Oppenheimer approximation is avoided.

### 1.10 Electron correlation

25. When we talk about electron correlation, what is it that actually is correlated? (Exam 2016, 2017, 2019)

   □ The positions of the electrons relative to the nuclei

   □ The motion of the nuclei

   □ The positions of the nuclei

   ■ The motion of the electrons

26. How do we define the electron correlation energy? (Exam 2016, 2017, 2019)

   □ The total energy minus the electrostatic energy

   ■ The exact energy minus the Hartree-Fock energy

   □ The kinetic energy of the electrons

   □ The Coulomb and the exchange energy

### 1.11 Electron correlation methods

27. Which of the following methods *does not* include the correlation energy in quantum chemical calculations? (Exam 2015,2017)

   □ Møller-Plesset second-order perturbation theory (MP2)

   □ Configuration interaction singles and doubles (CISD)

   ■ Hartree-Fock

   □ Complete active space SCF (CASSCF)

28. In which of the following methods is the energy not determined variationally? (Exam 2017)

   ■ Møller-Plesset second-order perturbation theory (MP2)

   □ Configuration interaction singles and doubles (CISD)

   □ Hartree-Fock

   □ Complete active space SCF (CASSCF)

29. For what purpose is Rayleigh-Schrödinger perturbation theory *not* an appropriate choice? (Exam 2019)

   ■ A variational treatment of an orbital expansion of the wavefunction.

☐ Regarding a weak interaction as for example the interaction between a molecule and an electromagnetic field.

☐ Regarding a small correction to an approximate Hamiltonian, as for example correcting the Fock operator.

## 1.12  Density functional theory

30. What is the foundation of density functional theory in quantum chemistry? (Exam 2019)

   ☐ Orbitals

   ☐ The wavefunction

   ☐ The Schrödinger equation

   ■ The Hohenberg-Kohn theorem

31. What is the definition of the electronegativity, $\xi$, in DFT? (Exam 2016)

   ■ $\xi = -\frac{\partial E}{\partial N}$ where $N$ is the number of electrons

   ☐ $\xi = -\frac{\partial E}{\partial V}$ where $V$ is the volume of the system

32. What is the interpretation of the electronegativity in comparison to thermodynamics? (Exam 2016, 2018)

   ■ The electronegativity is minus the chemical potential of the electrons.

   ☐ The electronegativity is minus the pressure of the electrons.

   ☐ The electronegativity is minus the temperature of the electrons.

   ☐ The electronegativity is minus the surface tension of the electrons.

33. When we specify a *functional* in the input to a DFT calculation, which energy terms, within the Kohn-Sham approach, are included in the functional? (Exam 2015, 2018, 2019)

   ☐ Coulomb and correlation energies

   ■ Exchange and correlation energies

   ☐ Coulomb and exchange energies

   ☐ Kinetic and correlation energies

34. If we compare Hartree-Fock theory with Kohn-Sham DFT, which is *not* similar or the same in the two approaches? (Exam 2015, 2018, 2019)

   ☐ Both methods rely on an orbital expansion.

   ☐ The resulting functional in Kohn-Sham DFT may be regarded as an ad hoc modification of the Fock operator.

   ☐ Both methods rely on the variational principle.

   ■ Both methods include electron correlation.

35. Which of the following acronyms is *not* an acronym for a DFT functional? (Exam 2019)

■ MP2

□ B3LYP

□ BLYP

□ PBE

## 1.13 Time-dependent DFT

36. In one of the exercises, we use time-dependent density functional theory (TD-DFT) to calculate electronic excitation energies for azobenzene molecules. (Exam 2018)

   a) When we calculate excitation energies by TD-DFT, what are we actually calculating?

   □ By calculating the energies of the ground state and the excited state, respectively.

   □ By calculating only the energy of the excited state.

   ■ By calculating the poles where the frequency-dependent polarizability diverges.

   b) To what type of spectroscopy are the electronic excitation energies related to?

   ■ UV/VIS spectroscopy

   □ IR spectroscopy

   □ Microwave spectroscopy

   □ NMR spectroscopy

   c) In the figure below, you see one of the molecules studied in the exercise, p-aminoazobenzene, which was compared to the azobenzene molecule. Which of the
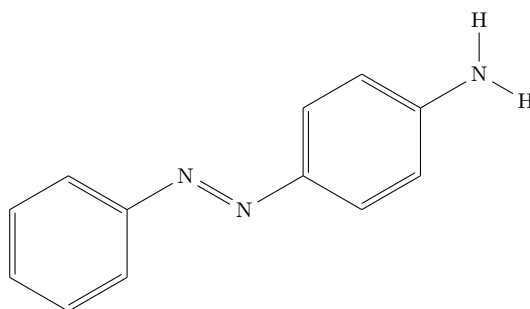


   two molecules, azobenzene or p-aminoazobenzene, are most likely to absorb light in the visual region of the spectrum (considering that most "normal" molecules absorb in the UV region)?

   □ azobenzene

   ■ p-aminoazobenzene

   □ The two molecules absorb light at the same wavelength.

d) In addition to the excitation energy, you also calculated the dipole oscillator strength. What does the dipole oscillator strength correspond to?

- ■ The intensity of a peak in the UV/VIS spectrum.
- □ The wavelength of the light at which the excitation occurs.
- □ The frequency of the light at which the excitation occurs.

### 1.14 Solid state

37. In the lecture on solid state calculations, we briefly discussed relativistic effects. What is true regarding relativistic contributions in quantum chemistry? (Exam 2016, 2017, 2018)

    a)

    - ■ Relativistic effects are important when we have heavy elements in our molecular system.
    - □ Relativistic effects are important when we have light elements in our molecular system.
    - □ Relativistic effects are in practice only important for hydrogen atoms.
    - □ Relativistic effects are important only for metals where we have "conducting" electrons in the system.

    b)

    - □ In a non-relativistic approach, we assume that the speed of light, $c$, is zero.
    - ■ In a non-relativistic approach, we assume that the speed of light, $c$, is infinitely large.
    - □ In a non-relativistic approach, we assume that the speed of light, $c$, is about the same as the speed of the electrons.

38. For which of the following systems would you anticipate that relativistic effects would be crucial to include in a quantum chemical calculation? (Exam 2018, 2019)

    - □ The modeling of hydrogen bonding in the water dimer
    - □ The modeling of the argon dimer
    - □ The formation of a sigma-bond in the methane molecule.
    - ■ The adsorption of a platinum cluster on a carbon surface.

39. What is the definition of the Fermi energy? (Exam 2016)

    - □ It is the difference between the HOMO and LUMO energy levels.
    - □ It is the same as the HOMO energy.
    - □ It is the same as the LUMO energy.
    - ■ It is in the middle between the highest occupied (HOMO) and lowest unoccupied (LUMO) energy levels.

40. What is the band gap? (Exam 2016)

   ■ It is the difference between the HOMO and LUMO energy levels.

   □ It is the same as the HOMO energy.

   □ It is the same as the LUMO energy.

   □ It is in the middle between the highest occupied (HOMO) and lowest unoccupied (LUMO) energy levels.

41. How is the wavefunction represented in the APW model for solid-state calculations? (Exam 2019)

   □ With plane waves

   ■ A combination of plane waves and atomic orbitals.

   □ With atomic orbitals

## 2 Force fields

42. What is true in general about a force field? (Exam 2019)

   ■ It relies on the Born-Oppenheimer approximation.

   □ It relies on molecular orbitals.

   □ It is based on density functional theory.

   □ It is a solution to the Schrödinger equation.

43. Why do we in most cases prefer to use a force field over a DFT functional in molecular dynamics simulations? (Exam 2019)

   □ Force fields are more accurate.

   ■ Force fields are computationally much more efficient.

   □ In contrast to DFT, force fields can be used for all elements in the periodic table.

   □ Force fields has a better foundation in fundamental theory.

44. Force fields are typically divided into different energy terms. Which of the following energy terms describe the intramolecular energy (or bonded interactions/covalent bonds)? (Exam 2019)

   □ The Lennard-Jones potential

   □ Electrostatics and polarization

   ■ Bond stretching and angle bending

   □ The dispersion energy

45. Which of the following models for bond stretching gives the correct dissociation of the energy? (Exam 2015, 2017, 2019)

☐ Harmonic potential with a term for the anharmonicity.

☐ Harmonic potential

■ Morse potential

☐ None of these alternatives

46. What do we call the force-field term used to constrain planar molecules to be planar? (Exam 2019)

■ Improper torsions

☐ Improper angle bending

☐ The Lennard-Jones term

☐ Planar potential

47. Which is the leading electric moment (first non-zero) of the following molecules/atoms? (Exam 2016)

■ Charge - $H^+$ (proton)

■ Dipole moment - water molecule

■ Quadrupole moment - carbon dioxide molecule

■ No electric moment at all - argon atom

48. When we discussed intermolecular interactions, we divided the interaction energy into four major contributions: electrostatics, polarization, dispersion and repulsion. But what do we know about the sign of each terms? (Exam 2016)
a)

■ The electrostatic energy can be either attractive (negative value) or repulsive (positive value).

☐ The electrostatic energy can only be attractive.

☐ The electrostatic energy can only be repulsive.

b)

☐ The polarization energy can be either attractive (negative value) or repulsive (positive value).

■ The polarization energy can only be attractive.

☐ The polarization energy can only be repulsive.

c)

☐ The dispersion energy can be either attractive (negative value) or repulsive (positive value).

■ The dispersion energy can only be attractive.
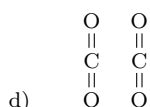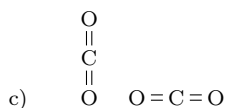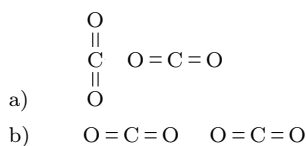
☐ The dispersion energy can only be repulsive.

d)

□ The repulsion energy can be either attractive (negative value) or repulsive (positive value).

□ The repulsion energy can only be attractive.

■ The repulsion energy can only be repulsive.

49. Which energy terms are always attractive (i.e., energy below zero)? (Exam 2019)

   ■ Polarization and dispersion

   □ Electrostatics and polarization

   □ The Lennard-Jones potential

   □ Electrostatics and repulsion

50. Which are the two dominating energy contributions in the interaction between two argon atoms? (Exam 2017)

   □ Electrostatics and dispersion

   □ Electrostatics and repulsion

   ■ Dispersion and repulsion

   □ Polarization and repulsion

51. Which is the dominant attractive interaction energy between an argon atom and an ion?

   □ Electrostatics

   ■ Polarization

   □ Dispersion

   □ Repulsion

52. If we use a force field in a simulation of liquid water, which energy term would dominate? (Exam 2019)

   □ Dispersion

   □ Polarization

   ■ Electrostatics

53. If we use a force field in a simulation of liquid carbon dioxide, which energy term would be the *smallest*? (Exam 2019)

   □ Dispersion

   ■ Polarization

   □ Electrostatics

54. In a force field we use energy terms with a different distance dependence between the atoms. Match the distance dependence for each of the energy terms below. (Exam 2015, 2016, 2017, 2018)

   ■ Dispersion - $1/R^6$

- ■ Repulsion in a Lennard-Jones potential - $1/R^{12}$
- ■ Ion-ion interactions - $1/R$
- ■ Dipole-dipole interactions - $1/R^3$
- ■ Ion-dipole interactions - $1/R^2$
- ■ Dipole-quadrupole interactions - $1/R^4$

55. We refer to molecules without a net charge and a molecular dipole moment but with a quadrupole moment as *quadrupolar* molecules. Which of the following molecules is a quadrupolar molecule? (Exam 2018)

   - □ $H_2O$
   - ■ $CO_2$
   - □ $CH_3F$
   - □ HCOOH

56. Which is the leading (first non-zero) electric moment in the carbon dioxide molecule? (Exam 2019)

   - ■ Quadrupole moment
   - □ Charge
   - □ Dipole moment
   - □ Octupole moment

57. What is the distinction between a second moment and a quadrupole moment? (Exam 2019)

   - □ The second moment is the same as the dipole moment.
   - □ There is no distinction.
   - ■ The quadrupole moment is traceless.

58. What is the distance dependence of a quadrupole-quadrupole interaction? (Exam 2018)

   - □ $1/R^3$
   - □ $1/R^4$
   - ■ $1/R^5$
   - □ $1/R^6$

59. When electrostatic interactions are present, it is the most important contribution to the intermolecular interaction energy. What can we say about the dipole-quadrupole interaction energy? (Exam 2018)

   - ■ It can be either repulsive or attractive.
   - □ It is always attractive.
   - □ It is always repulsive.
   - □ It is always zero.

```
        O
        ‖
        C    O = C = O
        ‖
a)      O
b)      O = C = O    O = C = O


        O
        ‖
        C
        ‖
c)      O    O = C = O


        O    O
        ‖    ‖
        C    C
        ‖    ‖
d)      O    O
```

Which of the four cases has an attractive interaction energy?

60. Regard the four (a-d) different cases for the interaction between two carbon dioxide molecules in the figure below. (Exam 2018)

   ■ case a)

   □ case b)

   □ case c)

   □ case d)

61. Regarding the transferabiltiy of force-field parameters, for which atom-type parameters can it be expected that the transferability is the worst? (Exam 2015)

   ■ Atomic partial charges

   □ Dispersion parameters

   □ Repulsion parameters

   □ Atomic polarizabilities

62. Why is it difficult to model hydrogen bonds as compared to other systems? (Exam 2019)

   □ The interaction energy is highly dependent on the distance between the two molecules.

   ■ The interaction energy is highly dependent on the mutual orientation of the two molecules.

   □ The interaction energy is highly dependent on the short-range repulsion energy.

   □ The interaction energy is highly dependent on the dispersion energy.

63. We used electronegativity equalization as a model for calculating atomic partial charges. Just by looking at the values of the electronegativities we can say some things about the sign and magnitude of the atomic charges. Here we use the Pauling scale for the electronegativities as a guideline: (Exam 2015, 2018)

| Pauling electronegativities ("arbitrary" unit) | |
|---|---|
| H | 2.20 |
| C | 2.55 |
| N | 3.04 |
| O | 3.44 |
| F | 3.98 |

As a reference point: In the HF molecule, the hydrogen charge is approximately $+0.8e$, and the fluorine charge is approximately $-0.8e$.

a) What are the partial charges of the fluorine atoms in the $F_2$ molecule?

☐ 3.98 and 3.98

☐ +1 and -1

☐ -1 and -1

■ 0 and 0

b) Comparing the atomic charges in benzene and carbon dioxide, which of the following statements is true?

■ The carbon charges in the two molecules have different sign.

☐ The hydrogen charges in benzene have a negative sign.

☐ The (absolute) magnitude of the carbon charge in carbon dioxide is much smaller than the carbon charges in benzene.

☐ The oxygen charges in carbon dioxide have a positive sign.

c) Which of the following statements is true (for compounds including the five elements given in the table)?

■ Hydrogen charges are always positive.

☐ Hydrogen charges are always negative.

☐ Nitrogen charges are always positive.

☐ Nitrogen charges are always negative.

d) When the charges for the HF molecule was given above as 0.8 e, atomic units were adopted. What does that imply for the electron charge?

■ The electron charge is -1.

☐ The electron charge is +1.

e) For an uncharged molecule (not an ion), is electroneutrality ensured in the electronegativity equalization method?

☐ No

■ Yes, by adding electroneutrality as a constraint.

☐ Yes, it is fulfilled automatically.

f) What is a reasonable value of the oxygen atomic charge in the water molecule? (Exam 2019)

☐ +0.3
☐ 0.0
☐ -0.02
■ -0.5

g) Comparing the atomic charges in formaldehyde and methane, which of the following statements is true? (Exam 2019)

■ The carbon charges in the two molecules have different sign.
☐ The hydrogen charges in methane have a negative sign.
☐ The (absolute) magnitude of the carbon charge in formaldehyde is much smaller than the carbon charges in methane.
☐ The oxygen charge in formaldehyde has a positive sign.

h) Which of the following statements is true (for compounds including the five elements given in the table)? (Exam 2019)

☐ Fluorine charges are always positive.
■ Fluorine charges are always negative.
☐ Oxygen charges are always positive.
☐ Oxygen charges are always negative.

i) What is a reasonable value of the hydrogen charges in cyclohexane? (Exam 2019)

■ +0.03
☐ +1.0
☐ -1.0
☐ -0.03

64. The Lennard-Jones potential can be written as

$$V = \frac{A}{R^{12}} - \frac{C}{R^6} \ .$$

Which of the following statements are true for the Lennard-Jones potential? (Exam 2016, 2017)

a)

■ It includes two terms: repulsion and dispersion
☐ It includes two terms: repulsion and polarization.
☐ It includes three terms: repulsion, polarization and dispersion.

b)

☐ $A$ has a positive, but $C$ has a negative value
☐ Both $A$ and $C$ have negative values.
■ Both $A$ and $C$ have positive values.
☐ $A$ has a negative value, but $C$ has a positive value.

# 3  Energy minimizations

65. In a geometry optimization, we end up in a minimum on the potential energy surface. How do we know that it is a minimum (in contrast to e.g. a transition state)? (Exam 2016, 2017)

    ☐ The gradient is zero and we have one imaginary frequency.

    ■ The gradient is zero and all frequencies are real

    ☐ The gradient is zero and we have one negative frequency.

    ☐ The gradient is positive and all frequencies are real.

66. Which of the following statements is correct for a transition state? (Exam 2015, 2016, 2017)

    ☐ The gradient is zero and we have one negative frequency.

    ■ The gradient is zero and we have one imaginary frequency.

    ☐ The gradient is positive and all frequencies are real.

    ☐ The gradient is zero and all frequencies are real

67. Which information do we need in each point, $x_k$, in a geometry optimization using Newton-type (Newton-Raphson and Quasi-Newton) of methods? (Exam 2018, 2019)

    ☐ They use only the energy, not the gradient or Hessian.

    ☐ They use the energy and the gradient, not the Hessian.

    ■ They use the energy, gradient and Hessian.

    ☐ They use also the anharmonicity, in addition to the energy, gradient and Hessian.

68. Which of the following statements regarding the Quasi-Newton method are correct? (Exam 2015, 2016)

    ■ The inverse of the Hessian is updated in each optimization step based on the gradient and the energy from the previous step in the geometry optimization.

    ■ It is important to avoid computing the Hessian by a quantum chemical method in each optimization step since Hessian calculations are expensive.

    ☐ The Hessian is computed in each optimization step by taking the second derivative of the potential energy.

    ■ It is important to avoid to take the inverse of the Hessian in each optimization step for large molecules since it can be expensive to calculate the inverse of a matrix.

69. What is the major advantage of the Quasi-Newton method over the Newton-Raphson method? (Exam 2019)

    ■ In the Quasi-Newton method, the inverse of the Hessian is updated in each optimization step based on the gradient and the energy from the previous step in the geometry optimization.

☐ In the Quasi-Newton method, we do not need the Hessian (or the inverse of the Hessian)

☐ In the Quasi-Newton method, we do not need neither the gradient nor the Hessian

☐ The advantage of using the Quasi-Newton method over the Newton-Raphson method is only minor.

70. When we in the exercises do a geometry optimization of a molecule, what do we normally find? (Exam 2019)

■ A nearby local energy minimum

☐ Always the global energy minimum

☐ A transition state

71. After a geometry optimization (i.e., the gradient is zero), we compute the vibrational frequencies and find one imaginary frequency. What does that imply? (Exam 2019)

☐ We found a local minimum

■ We found a transition state.

☐ We found the global minimum.

72. Which model/approximation was discussed in the course for quantum chemical calculations of vibrational frequencies? (Exam 2019)

☐ Rigid rotor

☐ Morse potential

■ Harmonic oscillator

73. Which of the following statements regarding the zero-point vibrational energy (ZPVE) of a molecule are correct? (Exam 2015, 2016, 2018)

a)

☐ The ZPVE is zero at the temperature 0 K.

☐ The ZPVE is zero for diatomic molecules.

■ The ZPVE is always nonzero.

b)

☐ A consequence of the ZPVE is that a molecule is always rotating

■ A consequence of the ZPVE is that a molecule is always vibrating.

☐ A consequence of the ZPVE is that a molecule is vibrating for temperatures over 0 K.

☐ A consequence of the ZPVE is that a molecule is rotating for temperatures over 0 K.

## 4 Solvent models

74. A couple of questions on various solvent models: (Exam 2016, 2018)
    a) What is the Born model?

    - ■ A model for a point charge in the center of a spherical cavity surrounded by a dielectric medium

    - □ A model for a dipole moment in the center of a spherical cavity surrounded by a dielectric medium

    - □ A model for a molecule in a spherical cavity surrounded by a dielectric medium

    - □ A model for a molecule in a cavity with a molecular shape surrounded by a dielectric medium

    b) What is the Onsager model?

    - □ A model for a point charge in the center of a spherical cavity surrounded by a dielectric medium

    - ■ A model for a dipole moment in the center of a spherical cavity surrounded by a dielectric medium

    - □ A model for a molecule in a cavity with a molecular shape surrounded by a dielectric medium

    - □ A model for a molecule in a spherical cavity surrounded by a dielectric medium

    c) What is the polarizable continuum model (PCM)? (Exam 2019)

    - □ A model for a point charge in the center of a spherical cavity surrounded by a dielectric medium

    - □ A model for a dipole moment in the center of a spherical cavity surrounded by a dielectric medium

    - ■ A model for a molecule in a cavity with a molecular shape surrounded by a dielectric medium

    - □ A model for a molecule in a spherical cavity surrounded by a dielectric medium

75. In continuum solvation models, what is the reaction field? (Exam 2019)

    - ■ It is the field on the molecule from the surrounding dielectric medium.

    - □ It is the electric field of the molecule on the surrounding dielectric medium.

76. What is the fundamental approach in QM/MM? (Exam 2017)

    - □ The core is described by a force field and the surroundings by a dielectric medium.

    - ■ The core is described by a quantum chemical method and the surroundings by a force field.

    - □ The core is described by a quantum chemical method and the surroundings by a dielectric medium.

77. The following statements compare continuum solvation models. (Exam 2017)

a) In QM/MM one has to sample many configurations, e.g. from molecular dynamics simulations, to model fluctuations in the solvent whereas it is sufficient to do one calculation with a dielectric model.

- ■ True
- ☐ False

b) QM/MM is better suited than dielectric models to describe hydrogen bonding with the solute.

- ■ True
- ☐ False

c) Both dielectric models and QM/MM have fundamental problems describing short-range exchange-repulsion, i.e. that the Pauli principle is fulfilled for the solute and solvent molecules.

- ■ True
- ☐ False

78. What is the preferred solvation model to describe hydrogen bonding with the surroundings? (Exam 2019)

- ☐ Born model
- ☐ Onsager model
- ☐ PCM
- ■ QM/MM

79. What is the key characteristic of the Langevin dipole model for describing solvation of a molecule?

- ☐ The positions of the surrounding dipoles are optimized, but the orientations are kept fixed.
- ☐ The magnitudes of the dipole moments on the surrounding lattice are optimized.
- ☐ Both the positions and the orientations of the surrounding dipole moments are optimized.
- ■ The dipole moments on the surrounding lattice are free to rotate

80. What do we refer to as an *ab initio* molecular dynamics simulation (where the Car-Parrinello method is one example)?

- ■ The forces between the particles are computed using DFT rather than a force field.
- ☐ It is a hybrid QM/MM method using a combination of DFT and force fields to obtain the interparticle forces

☐ We treat the nuclear motion with quantum mechanics rather than solving Newton's equations for classical motion.

81. For which of the following application areas would you anticipate that the Car-Parrinello method would be crucial to use as compared to "regular" molecular dynamics?

   ■ Heterogeneous catalysis

   ☐ Diffusion in porous media

   ☐ The liquid structure of water

   ☐ The modeling of ice.

## 5   Chemoinformatics

82. SMILES (Exam 2015, 2017, 2019)
    a) Which is the correct SMILES for butane?

   ■ CCCC

   ☐ C1CCC1

   ☐ CC(C)C

   ☐ CC=CC

    b) Which is a correct SMILES for acetamide?

   ☐ CCON

   ■ CC(=O)N

   ☐ CCCN=O

   ☐ CCCON

    c) Which is the correct SMILES for benzene?

   ☐ C1CCCCC2

   ☐ C1CCCCC1

   ☐ c1ccc2

   ■ c1ccccc1

    d) Which is the correct SMILES for propene?

   ■ C=CC

   ☐ C=C=C

   ☐ CCC

   ☐ C(C)C

    e) Which is the correct SMILES for isobutane?

☐ CCCC

☐ CC=CC

■ CC(C)C

☐ C1CCC1

f) Which is the correct SMILES for ethene?

■ C=C

☐ CC=CC

☐ CC(C)C

☐ CC

g) Which is the correct SMILES for cyclohexane?

☐ CCCCCC

☐ C1CCCCC2

■ C1CCCCC1

☐ C(C)CCCC

h) Which is the correct SMILES for toluene?

☐ c1cccccc2

☐ CCCCCCC

■ Cc1ccccc1

☐ CCC(C)CC(C)

i) Which is the correct SMILES for acetone?

☐ CC(=O)N

■ CC(=O)C

☐ CCCO

☐ CCCC

83. What does the acronym QSPR mean? (Exam 2016, 2018)

☐ Qualitative structure-property relations

■ Quantitative structure-property relations

☐ Qualitative structure-principle relations

☐ Quantitative structure-principle relations

84. Which of the following statements are true/false for QSPR? (Exam 2016)
a) In 3D-QSPR (or 3D-QSAR) the descriptors are properties (e.g. the electrostatic potential) placed on a lattice around the molecule.

■ True

☐ False

85. Which of the following properties is a typical 1D descriptor? (Exam 2019)

   - ■ The partition coefficient
   - ☐ Partial atomic charges
   - ☐ Orbital energies
   - ☐ The electrostatic potential on a lattice around the molecule.

86. Which is the advantage of EVA descriptors as compared to COMFA descriptors? (Exam 2019)

   - ■ EVA descriptors are alignment-free.
   - ☐ EVA descriptors map the electrostatic potential on a lattice around the molecule.

# 6  Kinetic Monte Carlo

87. For which type of problem would kinetic Monte Carlo simulations be a useful tool? (Exam 2019)

   - ☐ Geometry optimizations
   - ☐ Quantum chemistry
   - ☐ Molecular dynamics
   - ■ Chemical kinetics

88. Which of the following statements are true for the Kinetic Monte Carlo (KMC) method? (Exam 2016)

   a)

   - ■ KMC is a general method for solving a problem expressed in rates (or rate constants).
   - ☐ KMC is a method to solve the kinetics equation for chemical reactions, but is limited to chemical reactions.

   b)

   - ■ In KMC, we use the number of molecules as the variables, i.e. it is an integer number, 0, 1, 2, etc.
   - ☐ In KMC, we use the concentration of each molecule as the variables.

   c)

   - ☐ We cannot describe adsorption/desorption to/from a surface with KMC, since we cannot express these processes as rates (rate constants).
   - ■ KMC is ideal in modeling heterogeneous catalysis since we can model adsorption/desorption to/from a surface with KMC.

# 7  Protein modeling

89. What is true or false regarding the modeling of protein structure. (Exam 2018)

    a) Hydrogen bonding is important for the protein structure.

    ■ True
    □ False

    b) As for most polymers, a protein has a multitude of equivalent structures that give the function of the protein.

    □ True
    ■ False

    c) The driving force (toward lowering the free energy) in the hydrophobic effect is that water likes water.

    ■ True
    □ False

    d) Biological (genetic) evolution is the reason for that we regard protein families.

    ■ True
    □ False

# TKJ4205/KJ8902 Molecular Modelling
# Exam 12.12.2013, 9.00-13.00

**Norges Teknisk-Naturvitenskapelige Universitet (NTNU)**
**Code A**: All printed and hand-written texts are allowed. All calculators are allowed.
**Contact**: Assoc. Prof. Titus van Erp, Dep. of Chemistry, cell phone: 98048297
**NB**: The tasks are not weighted equally. The weight for each task is given in parenthesis after the task number. The weights sum up to 100.

## Exercise 1 (15)

What is the variational theorem and why is it useful in quantum chemistry? Which of the following quantum chemical methods are variational: Hartree-Fock, the configuration interaction (CI) method, Møller-Plesset second-order perturbation theory (MP2), and Kohn-Sham density functional theory?

> **Solution:** The variational theorem states that for any approximate wavefunction where the energy is determined variationally, the energy of the approximate wavefunction is higher than the exact energy. In molecular orbital theory, we can thus systematically increase the accuracy of the wavefunction (for example by extending the basis set) until convergence of the property we are interested in (as was done in the exercises). The Hartree-Fock method, CI method and Kohn-Sham density functional theory are variational since the energy is minimized for the choice of correlation method and basis set, whereas MP2 is a non-variational method.

## Exercise 2 (10)

Why is hydrogen bonding difficult to model by a standard force field like

$$V = \sum_{I=1}^{N} \sum_{J=I+1}^{N} \frac{q_I q_J}{4\pi\varepsilon_0 R_{IJ}} + 4\epsilon_{IJ} \left( \left( \frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left( \frac{\sigma_{IJ}}{R_{IJ}} \right)^{6} \right)$$

Suggest two approaches to improve the description of hydrogen bonding in force fields.

> **Solution:** The energy of a hydrogen bond is highly orientation-dependent, whereas all the energy terms in the equation above depend only on the distance between atoms. One way to improve the description of hydrogen bonding is to to include angle-dependent terms as in the YETI force field. Another way is to include atomic dipole moments in the force field since dipole-dipole interactions are highly dependent on the relative orientation of the dipole moments.

## Exercise 3 (10)

The two azobenzenes in the figure below
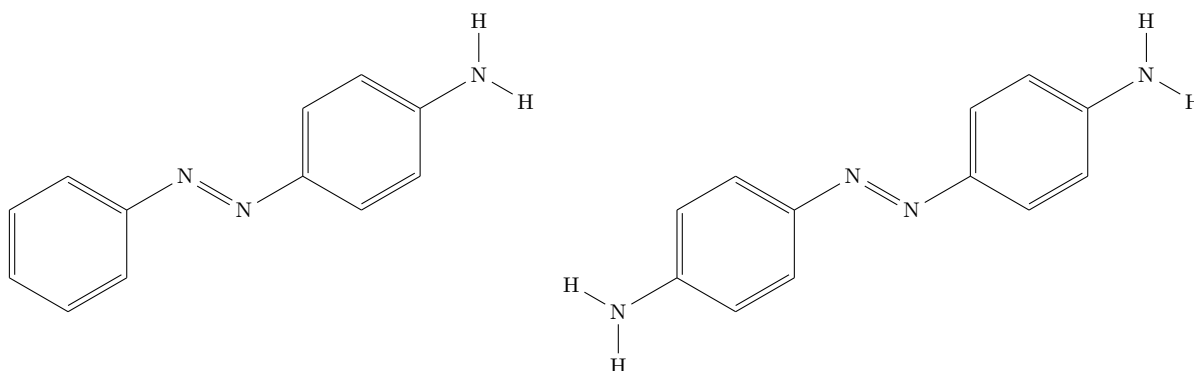
have slightly different colours. In an experiment, your fellow organic chemist does not know which of the two azobenzenes he/she has synthesized, but can you suggest a computational strategy to distinguish between the two molecules based on colour? In some cases, changing solvent may shift the colour of an azobenzene. Suggest different methods to include solvent effects in the calculations.

**Solution:** A colour corresponds to the visual part in a UV/VIS spectrum which in turn corresponds to electronic excitations. We calculate normally excitation energies by time-dependent density-functional theory, as was done in one of the exercises. Solvent effects in quantum chemistry are normally calculated by hybrid classical-quantum models where the core is treated quantum-mechanically and the surroundings is treated by a dielectric continuum (as in the PCM model), force fields (QM/MM) or by Langevin dipoles.

## Exercise 4 (15)

What is the major difference between solid state calculations and quantum chemical calculations on molecular systems? Suggest a method to study catalysis on a surface. Finally, when is relativistic effects important in quantum chemistry, and suggest one approach where relativistic effects may be included approximatively.

**Solution:** In solid state calculations we have *translational symmetry* that needs to be included in the model. To study surfaces, we often use a *slab model*, with a vacuum layer in the "third" direction resulting in a two-dimensional periodic system. Relativistic effects are important for heavy elements (where the speed of the electrons approach the speed of light), and relativistic effects are normally parametrized into *pseudopotentials*.

## Exercise 5 (50: 6, 8, 8, 10, 10, 8)

A complex molecule that is emerged in a solvent has two stable configurations. This molecule consists of two molecular groups that are rigid, but the distance between these two molecular groups can vary. To a very good approximation this complex molecule can be described by a simple model of a diatomic molecule $A - B$ where the "atoms" $A$ and "B" refer to the two rigid molecular groups. Based on DFT calculations, the potential energy as function of separation
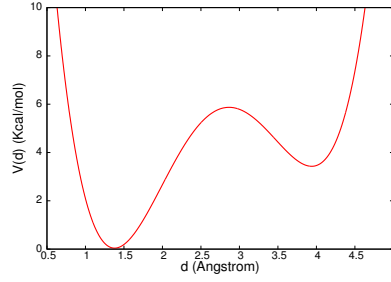
Figure 1: Intra-molecular potential energy as function of separation distance.

distance $d$ of the center of masses was obtained. It is shown in Fig. 1. It clearly shows that there are two stable configurations, a contracted state and an extended state.

i) Since the $A$ and $B$ groups are rigid the configurational space of a single molecule is six-dimensional. Each configuration can be given by the vector $(x_A, y_A, z_A, x_B, y_B, z_B)$. To get a better understanding of the system in a lower dimensional space, you want to study the free energy as function of separation distance $A(d)$ of this molecule. Can you give the relation between the probability distribution function $P(d)$ and the free energy $A(d)$?

> **Solution:** Boltzmann showed that in thermal equilibrium each configuration point has a probability density according to
>
> $$P(\mathbf{R}^N) \propto e^{-\beta V(\mathbf{R}^N)}$$
>
> with $\mathbf{R}^N$ the $3N$ dimensional configuration point, $N$ the number of particles, $V$ the potential energy function, and $\beta = 1/(k_B T)$ with $T$ the temperature and $k_B$ the Boltzmann constant. The free energy $A$ expressed as function of order parameter $d$ has the same characteristics as $V$, but in a lower dimensional space.
>
> $$P(d) \propto e^{-\beta A(d)} \quad \text{or } P(d) = \text{constant}' \, e^{-\beta A(d)}$$
>
> Here, $d$ is a distance which is just a one-dimensional parameter, $P(d)$ is the probability density of this distance (in other words $P(d)\Delta d$ is the probability that the separation between $A$ and $B$ is between $d - \Delta d/2$ and $d + \Delta d/2$). The constant is not known, but if you want to know the relative probabilities that $A$ and $B$ are either separated by $d_1$ or by $d_2$, the the constant cancels.
>
> $$\frac{P(d_1)}{P(d_2)} = e^{-\beta[A(d_1) - A(d_2)]}$$

ii) Suppose that the temperature is sufficiently high so that a NVT molecular dynamics can easily explore all configurations (both extended and contracted state will be sampled). In that case, how can you determine numerically $P(d)$ from a very long MD simulation? What possibilities do you have if the temperature is too low to efficiently sample configuration space (hence, the system will be trapped for very long times in either the extended or contracted state)

**Solution:** In a NVT simulation we should in principle sample the configurational space according the the Boltzmann statistics. $P(d)$ can the be constructed by the *histogram* method. In this method one counts how often $d(t)$ is inside certain intervals (bins). For example, we divide the range $[0 : 6\text{Å}]$ into intervals $[0 : \Delta d], [\Delta d, 2\Delta d], \ldots,$ $[6\text{Å} - \Delta d : 6\text{Å}]$. Each time a bin is visited we increase its counter by 1. The probability density $P(d)$ is afterwards normalized by dividing the counters to the total number of counts and by the bin width $\Delta d$.

At low $T$ we might get trapped in either the contracted state or extended state. $P(d)$ can only be accurately sampled by the straightforward histogram method if the md simulation make sufficient transitions. If this is not the case, we need to apply *rare event* simulation techniques, such as *umbrella sampling, thermodynamic integration,* etc.

iii) Show that for an isolated molecule in the gas phase that is free to rotate and has no interaction with other molecules, the free energy $A(d)$ must be given by

$$A(d) = V(d) - k_B T \ln d^2 + \text{constant}$$

where $k_B$ is the Boltzmann constant and $T$ the temperature in Kelvin. To explain this expression think about the relation of free energy and relative probability and about the orientations that a diatomic molecule can take in 3D. What units must the constant have in Eq. iii?

**Solution:** Suppose that we fix the position of $A$, how "many" positions can $B$ take such that the distance is $d$? Naturally, all these possible positions lie on a sphere of radius $d$ and origin $(x_A, y_A, z_A)$. Each of these positions individually has a probability to occur proportional to $\exp(-\beta V(d))$, but the number of positions is proportional to the surface of this sphere $(4\pi d^2)$. Henceforth:

$$P(d) \propto 4\pi d^2 e^{-\beta V(d)} \text{ or } P(d) = \text{constant}'' d^2 e^{-\beta V(d)} \tag{1}$$

Then, by using the relation between $A(d)$ and $P(d)$ from question 5i:

$$
\begin{aligned}
P(d) &= \text{constant}' \, e^{-\beta A(d)} = \text{constant}'' \, d^2 e^{-\beta V(d)} \Rightarrow \\
-\beta A(d) &= \ln \left\{ \frac{\text{constant}''}{\text{constant}'} d^2 e^{-\beta V(d)} \right\} = \ln \left\{ \text{constant}''' d^2 e^{-\beta V(d)} \right\} \Rightarrow \\
A(d) &= V(d) - \frac{1}{\beta} \ln \left[ d^2 \right] - \frac{1}{\beta} \ln \text{constant}''' \\
&= V(d) - k_B T \ln \left[ d^2 \right] + \text{constant}
\end{aligned}
$$

(Simply use $\ln abc = \ln a + \ln b + \ln c$).

There is something strange with the units in the above expression. $d$ is a length which has the dimension of metres, angstroms, nanometers, or whatever length unit you want to use. The expression $\ln \left[ d^2 \right]$ is therefore not very neat. Still, it is quite common to have something non-dimensionless inside a logarithmic function when free energies are expressed. The reason is that we are mostly interested in free energy differences and then things cancel out. Suppose that we want to use metres as our units of length and define the dimensionless length $\tilde{d} = d/m$. Then $k_B T \ln \left[ d^2 \right] = k_B T \ln \left[ \tilde{d}^2 \right] -$

$k_B T \ln \left[m^2\right]$. $k_B T$ has the right units of energy but $k_B T \ln \left[m^2\right]$ has strange incorrect units and must be cancelled by the constant. It turns out that the constant can be anything of the following: $k_B T \ln \left[am^2\right] + b\text{Joule}$ with $a, b$ dimensionless parameters (In fact, since $\ln \left[am^2\right] = \ln \left[m^2\right] + \ln \left[a\right]$ we can also write constant $= k_B T \ln \left[m^2\right] + b'\text{Joule}$). We can show that in that case everything is correct:

$$A(d) \quad = \quad V(d) - k_B T \ln \left[d^2\right] + k_B T \ln \left[am^2\right] + b\text{Joule} = -k_B T \ln \left[d^2/(am^2)\right] + b\text{Joule}$$

which has units of Joule. In practice, one usually doesn't care about the constant since it cancels out when computing free energy differences. In fact, even setting the constant to zero $A(d) = V(d) - k_B T \ln \left[d^2\right]$ will give right units for free energy differences since $A(d_1) - A(d_2) = V(d_1) - k_B T \ln \left[d_1^2\right] - V(d_2) + k_B T \ln \left[d_2^2\right] = V(d_1) - V(d_2) - k_B T \ln \left[d_1^2/d_2^2\right]$.

Unfortunately, the question was not very well expressed since exercise 5iv shows an equation with $\theta(2.9 - d(0))\,\theta(d(t) - 2.9)$. Due to an latex error the angstrøm symbols dropped out. The 2.9 had to be 2.9 Å. This gives the impression that $d$ is dimensionless or distance divided by 1 Å. In that case the constant just has to be in Joule. Therefore also the answers Joule, eV, Kcal or Joule/mol are approved even though the last option would only be formally correct if $R$ was used instead of $k_B$.

iv) You now want to study the dynamics of this transition in the solvent at a high temperature. At this temperature the transitions occur more frequently and can be studied by long brute force MD trajectories. You want to determine the following correlation function

$$C(t) = \frac{\langle \theta(2.9 - d(0))\theta(d(t) - 2.9)\rangle}{\langle \theta(2.9 - d(0))\rangle} \tag{2}$$

Here, $\theta$ is the Heaviside step-function: $\theta(x) = 1$ if $x > 0$, $\theta(x) = 0$ if $x \leq 0$. The brackets $\langle \ldots \rangle$ denote an NVT ensemble average over the initial conditions (at $t = 0$). The 2.9 Å is the top of the potential curve in Fig. 1 and you define any state with a distance $d$ smaller than this value to be compact, and with a distance $d$ larger than this value to be extended. Show that Eq. 2 is equivalent to

$$C(t) = \langle \theta(d(t) - 2.9)\rangle_{d(0) < 2.9\ \text{Å}} \tag{3}$$

where the subscript means that you only consider configurations with an initial $A - B$ distance smaller than 2.9 Å. Can you argue, based on Eq. 3, that $C(t)$ is in fact nothing else as the conditional probability that the molecule is in the extended state at time $t$ given that it was in the contracted state at time 0.

**Solution:** We know that (*)

$$\langle A \rangle = \frac{\int d\mathbf{R}^N A(\mathbf{R}^N)e^{-\beta V(\mathbf{R}^N)}}{\int d\mathbf{R}^N e^{-\beta V(\mathbf{R}^N)}}$$

Now consider a subspace $\Omega$, a region in configuration space that is smaller than the full space. Then the average of $A$ given the condition that $\mathbf{R}^N \in \Omega$ equals (**)

$$\langle A \rangle_{\mathbf{R}^N \in \Omega} = \frac{\int_\Omega d\mathbf{R}^N A(\mathbf{R}^N)e^{-\beta V(\mathbf{R}^N)}}{\int_\Omega d\mathbf{R}^N e^{-\beta V(\mathbf{R}^N)}}$$

In words, it is the same as before but now you only integrate over the subspace $\Omega$ instead of over the full space. From the first equation (*) it is easy to see that (***)

$$\frac{\langle AB \rangle}{\langle B \rangle} = \frac{\int d\mathbf{R}^N A(\mathbf{R}^N) B(\mathbf{R}^N) e^{-\beta V(\mathbf{R}^N)}}{\int d\mathbf{R}^N B(\mathbf{R}^N) e^{-\beta V(\mathbf{R}^N)}}$$

Now if the function $B(\mathbf{R}^N)$ is such that it is 1 if $\mathbf{R}^N \in \Omega$ and 0 otherwise, then the third equation (***) becomes identical to the second equation (**). And this also proves the equivalence between Eq. 2 and 3 if we take $A = \theta(d(t) - 2.9)$ and $B = \theta(2.9 - d(0))$. Note that $\mathbf{R}^N$ is considered to be coordinates at time $t = 0$ and therefore $d(0)$ is actually a direct function of $\mathbf{R}^N$. Likewise, so is $B$. $d(t)$ is in the first place a function of the coordinates at time $t$. However, since these coordinates also depend on the initial conditions we can write $\mathbf{R}^N(t) = f(\mathbf{R}^N(0))$ where $f$ would correspond to the MD integrator. In other words, also $A$ can be expressed as function of $\mathbf{R}^N(0)$.

Moreover, since $A = 1$ if $d(t) > 2.9\text{Å}$ and 0 otherwise, Eq (**) basically measures how often you are in the extended state at exactly time $t$ if you were inside $\Omega$ (the contracted state; $d < 2.9\text{Å}$) at $t = 0$. More precisely, which fraction of trajectories with initial points inside the contracted state are in the extended state some time $t$ later. Or, it is the probability that the molecule is in the extended state at time $t$ given that it was in the contracted state at time 0.

v) You want to calculate $C(t)$ for $t = 1$ ps. In other words, you are interested in the probability that the molecule is in the extended state if it was in the contracted state 1 ps earlier. To do that you do the following simulation.

   (a) You run a MC simulation of 1000 steps to generate initial configurations with $d < 2.9$ Å. Each of these configurations is saved to the hard disk.

   (b) For each configuration point, we generate random velocities according to the correct temperature distribution

   (c) Starting from each configuration point with corresponding velocities, we run a 1000 step MD trajectory with a time step of 1 fs.

   (d) For each trajectory we sample the value 1 if the end point is in the extended state and 0 otherwise.

So in the end we obtain 1000 data points, each data point is either 0 or 1, and the average gives this conditional probability or $C(1\text{ps})$.

The 1000 measurements are grouped in 5 blocks of 200. The average of each block is calculated. These are: 0.17, 0.07, 0.11, 0.17, and 0.08. Calculate $C(1\text{ps})$ together with its actual statistical statistical error (absolute and relative error) based on the block analysis. We may assume the averages of each block is uncorrelated.

**Solution:** The average of the 5 blocks gives 0.12. The estimator of the standard deviation based on these block equals

$$\sigma^2 \approx \frac{1}{5-1} \sum_{j=1}^{5} (0.12 - av.block_i)^2 \qquad (4)$$

6

The $5 - 1$ instead of 5 is the Bessel correction.

$$\sigma^2 \approx \frac{1}{4}(0.05^2 + 0.05^2 + 0.01^2 + 0.05^2 + 0.04^2) = 0.0023 \tag{5}$$

and

$$\frac{\sigma}{\sqrt{N}} = \sqrt{0.0023/5} = 0.0214 \tag{6}$$

and the relative error: $0.0214/0.12 = 18\%$.

vi) What would have been the expected error in the case that each of the 1000 trajectories was uncorrelated? Could you think of a simple way to reduce the correlations between the trajectories and by that get a more efficient algorithm? Estimate the increase in efficiency.

**Solution:** Let's call $p = C(1\text{ps})$ the probability that a trajectory from the contracted state is in the the extended state at $t = 1$ ps. Each time that we generate a MD trajectory we will sample the value 0 or 1 depending on whether the trajectory end in the contracted state or extended state, respectively. We will sample 1 with a probability of $p$ and zero with a probability of $(1 - p)$. And the average of these 1's and 0's is p as well. This gives for standard deviation of the sampling

$$\sigma = \sqrt{p(1 - p)^2 + (1 - p)p^2} = \sqrt{p(1 - p)} = 0.325 \tag{7}$$

Note that the standard deviation is significantly larger than the standard deviation between the blocks $\sigma_{\text{block}} = \sqrt{0.0023} = 0.048$. This is logical since each block is already an average and, therefore, these average values are closer to each other and closer to $p$. Finally, the error when everything is uncorrelated is: $0.325/\sqrt{(1000)} = 0.0103$. Relative error $0.0103/0.12 = 0.085$ or $8.5\%$. Hence, the actual relative error is twice as high as when all trajectories would be uncorrelated. In the present algorithm each fourth trajectory can be considered uncorrelated. So we can reduce the computational cost significantly by doing 4 MC steps before generating a new MD trajectory. In that case the initial conditions are decorrelated and each MD trajectory can be considered uncorrelated. We should get the same statistical error but with much less CPU time. In the original algorithm we did 1000 MD trajectories and 1000 MC steps. Each MD trajectory was 1000 MD steps. So in total we did 1001000 steps. In the new algorithm we do only 250 MD trajectories and still the same number of MC steps: 250*4=1000. So this gives a total of 251000 steps. The accuracy should be the same. So it is $1001000/251000 \approx 4$ times faster

# TKJ4205/KJ8902 Molecular Modelling
# Exam 19.12.2013, 15.00-19.00

**Norges Teknisk-Naturvitenskapelige Universitet (NTNU)**
**Code A**: All printed and hand-written texts are allowed. All calculators are allowed.
**Contact**: Prof. Per-Olof Åstrand, Dep. of Chemistry, cell phone: 9346 3033
**NB**: The tasks are not weighted equally. The weight for each task is given in parenthesis after the task number. The weights sum up to 100.

## Exercise 1 (30)

In the computational exercises, we do standard "DFT" calculations. Briefly explain and describe the key elements and approximations in the methodology used. (Regard a single point energy calculation, i.e. geometry optimizations or molecular property calculations do not have to be explained).

> **Solution:** In DFT calculations, we use the *Kohn-Sham* (KS) approach to include *electron correlation* approximately. A KS calculation to a large degree resembles a Hartree-Fock (HF) calculation, where the main difference between the KS and HF approximations is that the exchange term in HF is replaced by an *exchange-correlation* functional in KS (for which we have many models, e.g. B3LYP, PBE, etc.). We have thus a modified Fock operator, and the resulting *molecular orbitals* are determined *variationally*. The electron density is thus represented as molecular orbitals, which are expanded in atomic orbitals which in turn are expanded in a *basis set*. The basis set is often represented as atom-centered functions as Gaussian or Slater-type functions.

## Exercise 2 (10)

Give the definition and a physical interpretation of the electronegativity. How can the electronegativity be used in a force field?

> **Solution:** The electronegativity, $\xi$, is defined as
>
> $$\xi = -\left(\frac{\partial E}{\partial N}\right)_{V_{\text{ext}}}$$
>
> where $E$ is the energy, $N$ is the number of electrons, and the differentiation is carried out at fixed external potential, $V_{\text{ext}}$ (e.g a fixed molecular geometry).
>
> The physical interpretation is that the electronegativity is minus the *chemical potential, $-\mu$, for electrons*, i.e. the electrons will be redistributed in the molecule so that the chemical potential is equal everywhere.
>
> This approach has been adopted in a force-field model to calculate atomic charges termed *electronegativity equalization*. In the electronegativity equalization method, the molecular energy, $V$, is given as a sum over all atoms in the molecule,
>
> $$V = \sum_I \left(\xi_I q_I + \frac{1}{2}\eta_I q_I^2 + \frac{1}{2}\sum_{J \neq I}\frac{q_I q_J}{R_{IJ}}\right)$$

which includes two atom-type parameters: $\xi_I$ is the atomic electronegativity and $\eta_I$ is the atomic chemical hardness. The atomic charges are obtained by minimizing the energy,

$$\frac{\partial V}{\partial q_I} = 0$$

with the constraint that the molecular charge is conserved.

## Exercise 3 (10)

What is a molecular descriptor and give examples of different types of molecular descriptors? What is a QSPR model and how are molecular descriptors used in a QSPR model? In which research field has QSPR been commonly employed?

**Solution:** A molecular descriptor is a property of a molecule that is believed to be related to the function/activity of a molecule/material. Molecular descriptors can be either macroscopic or microscopic properties as well as determined either experimentally or from quantum chemical calculations. Normally, we divide molecular descriptors into:

1D: molecular properties (one value per molecule)

2D: atomic properties (one value per atom in the molecule)

3D: COMFA or COMSIA. E.g. in COMFA, the electrostatic potential is calculated on a lattice around the molecule, which is used as descriptors.

QSPR is essentially a linear regression model where the property of interest, $P$, is described as a linear combination of molecular descriptors, $m_i$,

$$P = \sum_{i=1} k_i m_i$$

where $k_i$ are coefficients to be determined.

QSPR was initially denoted QSAR to describe the activity for ligand-binding to enzymes. It has therefore been used a lot in drug design in the pharmaceutical industry.

## Exercise 4 (10,15,15,10)

A complex molecule has two stable configurations. This molecule consists of two molecular groups that are rigid, but the distance between these two molecular groups can vary. To a very good approximation this complex molecule can be described by a simple model of a diatomic molecule $A - B$ where the "atoms" $A$ and $B$ refer to the two rigid molecular groups. Based on DFT calculations, the potential energy as function of separation distance $d$ between $A$ and $B$ was obtained (see figure 1). It clearly shows that there are two stable configurations, a contracted state and an extended state.

a) You plan to study following properties: The equilibrium constant $K$ between contracted and extended state, the rate constant from one state to the other, the average life-time of each state,
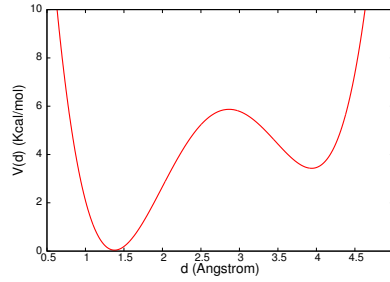
Figure 1: Intra-molecular potential energy as function of separation distance.

the free energy as function of distance $d$. Which of these properties could be calculated with Molecular Dynamics (MD) and with Monte Carlo (MC) simulations, respectively?

b) Now you want to run the simulation in a solvent in an NVT simulation. Make a sketch of how you think the solvent will change the free energy and discuss it (Draw the free energy of the gas phase situation and draw on top of it the free energy in the solvent, and discuss how the minima in the curve will change). Regard two cases i) both the contracted and the extended states are unpolar; ii) the contracted state is unpolar and the extended state is polar. We can assume that in any case no solvent molecule is able to get in between $A$ and $B$.

c) You run a long MD simulation at 300 K with a single $A - B$ molecule and 100 solvent molecules. You start from a situation in which the $A - B$ molecule is in the compact state. After many days of simulating this system you have not observed a single transition to the extended state. What are the conclusions and what would you do?

d) You now study this transition at a much higher temperature. At this temperature the transitions occur more frequently and can be studied by long brute force MD trajectories. You want to determine the following time-correlation function

$$C(t) = \frac{\left\langle \theta(2.9 \text{ Å} - d(0)) \, \theta(d(t) - 2.9 \text{ Å}) \right\rangle}{\left\langle \theta(2.9 \text{ Å} - d(0)) \right\rangle}$$

Here, $\theta$ is the Heaviside step-function: $\theta(x) = 1$ if $x > 0$, $\theta(x) = 0$ if $x \leq 0$. The brackets $\langle \ldots \rangle$ denote an NVT ensemble average over the initial conditions (at $t = 0$). The 2.9 Å is the top of the potential curve in figure 1 and you define any state with a distance $d$ smaller than this value to be compact, and with a distance $d$ larger than this value to be extended.

Except for $t$ being very close to zero, the shape of the correlation function $C(t)$ won't change too much if you choose a slightly different value, e.g 2.7 Å or 3.0 Å. Can you explain why? What can you say about $C(t)$ in the limit $t \to \infty$ in terms of the equilibrium constant?

---

**Solution:** a) All the properties can in principle be calculated with molecular dynamics simulations. The rate constant and average life-times are time-dependent properties and cannot be calculated with regular Monte Carlo simulations.

b) A solvent will organize itself in a way so that it lowers the free energy. So in all cases the free energy surfaces with a solvent will be below the gas-phase surface.

Case i): Excluded volume effects (the work required to create a cavity) will make the difference which will favour the compact state over the extended. The free energy in

---

3

solution will therefore be more lowered for the compact state than for the extended state as compared to the gas phase surface.

Case ii): If present, electrostatics (and polarization) is the most important solvent effect (as discussed for the solvation models in the course). A polar molecule will therefore polarize the surrroundings giving rise to a solvation free energy. Since the extended state is polar, the free energy of the extended state will be more lowered than the compact state.

c) The time-scale of the computer simulations is orders of magnitude smaller than the time-scale of sampling both minima. Solutions to study rare events include free energy perturbation (and thermodynamics integration) techniques for calculating free energy differences, umbrella sampling methods to calculate free energy surfaces (potential-of-mean-force) and path-sampling techniques.

d) Since the barrier energy is high, we will not spend a long time on the barrier and the barrier will be passed very quickly. So when we measure the distances $d(0)$ or $d(t)$ in the simulations, it is highly likely that they are far away from the barrier distance of 2.9 Å and the result will be relatively insensitive to the choice of the barrier distance.

In the limit $t \rightarrow$, we have $\langle a(0)b(t) \rangle = \langle a \rangle \langle b \rangle$. The expectation value in the denominator

$$\left\langle \theta(2.9 \text{ Å} - d(0)) \right\rangle$$

gives the probability to be in the compact state, $p_c$. The time-correlation function, $C(t \rightarrow \infty)$ measures thus the probability to be in the extended state, $p_e$. The equilibrium constant $K$ is the ratio between the probabilities to be in the extended state and the compact state,

$$K = \frac{p_e}{p_c}$$

We also have $p_e + p_c = 1$ leading to

$$C(t \rightarrow \infty) = \frac{K}{1 + K} \ .$$

# TKJ4205/KJ8902 Molecular Modelling
# Exam 10.12.2012, 09.00-13.00

**Norges Teknisk-Naturvitenskapelige Universitet (NTNU)**
**Code A**: All printed and hand-written texts are allowed. All calculators are allowed.
**Contact**: Prof. Per-Olof Åstrand, Dep. of Chemistry, cell phone: 9346 3033
**NB**: The tasks are not weighted equally. The weight for each task is given in parenthesis after the task number. The weights sum up to 100. The expected length of the answer on each task/subtask is around 1/2 page with a normal style of hand-writing, which indicates the expected level of detailness of each answer.

## Exercise 1 (15)

Make a comparison of the Hartree-Fock approximation and the Kohn-Sham approach in density-functional theory. Which are the main similarities and differences?

---

**Solution:** There are many similarities between the Hartree-Fock approximation and the Kohn-Sham (KS) approach. The main components in Hartree-Fock theory are:

- Molecular orbitals are obtained within the LCAO (linear combination of atomic orbitals) approach, where a basis set is used as basis functions for the atomic orbitals.

- The Hamiltonian is replaced with an effective one-electron operator, the Fock operator, in terms of a mean-field approximation.

Although the density is a function of only three coordinates, $\vec{r}$, (whereas the wave function for $N$ electrons is a function of $3N$ coordinates), the density is expanded in terms of atomic orbitals in the KS approach. The molecular orbital approach is thus retained in the KS approachi (and we have $3N$ coordinates).

In the KS approach, the density functional is divided into four contributions,

$$E[\rho(\vec{r})] = E_{\mathrm{KE}}[\rho(\vec{r})] + E_{\mathrm{ext}}[\rho(\vec{r})] + E_{\mathrm{H}}[\rho(\vec{r})] + E_{\mathrm{XC}}[\rho(\vec{r})]$$

where the kinetic energy term and the external energy (i.e., the interaction with the nuclei) correspond to the one-electron term in the Fock operator, the Hartree term corresponds to the Coulomb term in the Fock operator. So the remaining terms are the exchange term in the Fock operator and the exchange-correlation (XC) term in KS theory. Depending on the functional that we choose in a DFT calculation, we include an approximate model for both the correlation and the exchange term. Compared to the Hartree-Fock approximation, we add some electron correlation and we replace the Hartree-Fock exchange with an approximate term (which in some cases include some Hartree-Fock exchange). The energy operator in the KS approach can be expressed as a modified Fock operator, and the same solver can be used to solve the Hartree-Fock and Kohn-Sham equations using a self-consistent field (SCF) approach.

# Exercise 2 (10)

What is a basis set in quantum chemistry? How do we choose a basis set for a quantum chemical computation? What do we mean by *polarization* and *diffuse* functions, respectively? How should a basis set be improved in the calculation of molecular geometries and molecular polarizabilities, respectively?

> **Solution:** A basis set is a set functions (e.g., Gaussian or Slater-type functions (as is used in ADF)) centered on each nucleus in the molecule that is used as basis functions for the atomic orbitals. In quantum chemical computations, we use a hierarchy of basis sets and we are increasing the basis set systematically towards the basis set limit (as you did in the exercises). When we add polarization functions, e.g. by moving from a DZP basis set to a TZP basis set (using the notation in ADF), we for example increase the hydrogen basis from [2s1p] to [3s2p1d] giving an improved description of e.g. covalent bonds improving the accuracy of the molecular geometry. When we add diffuse functions, we add basis functions that improve the tail of the molecular orbitals (i.e., the charge density far away from the molecule), which will improve e.g. the calculation of molecular polarizabilities.

# Exercise 3 (10,10)

a) What are the differences between methods for finding the global and a local minimum on a potential energy surface, respectively? Considering the geometry optimizations carried out in the exercises: are they global or local minimizations?

b) Compare the Quasi-Newton method to the Newton-Raphson and Steepest descent methods, respectively, and point out where the Quasi-Newton method has its advantage compared to the two other methods. If the Quasi-Newton method is not available, which of the two other methods would be preferred for a large-scale (i.e., each calculation of the molecular energy is expensive) quantum chemical calculation and a force-field optimization of a macromolecule, respectively? Motivate the answer.

> **Solution:** a) In a local method, the goal is to find the "nearest" local minimum irrespective of it is the global minimum or not. Given a starting point on the potential energy surface near the local minimum of interest, the minimum is reached by an iterative procedure (e.g. simplex, steepest-descent, conjugate-gradient, Newton-Raphson, etc.) where the energy and sometimes the gradient and the Hessian are calculated in each point.
>
> A global optimization method aims at finding all "interesting" local minima including the global minimum. It requires therefore a procedure where many structures are included, and genetic algorithms is an example of such a method.
>
> In all geometry optimizations in the exercises, a local optimization procedure was used. (In one case, you actually did a global optimization by generating starting points for all minima by hand.)
>
> b) The Quasi-Newton method uses an update scheme for the inverse of the Hessian. In the Newton-Raphson method, the gradient gives the direction of the move in the iteration scheme whereas the inverse of the Hessian gives information on the length of the move. For a quadratic potential surface, the Newton-Raphson method would give the minimum in one

iteration. So for a general potential surface, which is not quadratic, it is only important to have an accurate Hessian close to the minimum. In the Newton-Raphson method, the Hessian and its inverse is calculated in each iteration which is avoided with the Quasi-Newton method. In the steepest-descent method, a fixed parameter is used for the step length (i.e. the inverse of the Hessian when comparing to the Newton-Raphson method) which leads to slow convergence (not quadratic) close to the minimum, so the Quasi-Newton method offers an improvement.

In a large-scale (where each energy calculation is expensive), we would like to avoid a large number of iterations. If the analytical Hessian is available, the Newton-Raphson method is therefore preferred if the Quasi-Newton method is not available. For a geometry optimization of a macromolecule (irrespective of if we use a force field or a quantum chemical method), we can normally not afford to store the Hessian and getting its inverse is expensive. Therefore, the steepest-descent method is preferred if the Quasi-Newton method is not available.

# Exercise 4 (15)

What is an *umbrella potential* in Monte Carlo simulations, and how can it be used? What is the potential of mean force (explain in words), and how can an umbrella potential be used to calculate the potential of mean force?

**Solution:** An umbrella potential is an extra potential energy added to a Monte Carlo simulation for artificially removing a (free) energy barrier during a simulation. The purpose is to improve the sampling rate in high-energy regions and pass a high energy barrier so that both sides of the barrier is sampled according to the statistics of the ensemble. Some examples where it can be used includes conformational analysis (passing high barriers in the rotation of side-chains in molecules), passing reaction barriers, and the study of repulsive interactions (e.g. the interaction of two cations in solution).

The potential of mean force (PMF) is a free energy surface, i.e. how the free energy depends on a coordinate in the system (e.g. a reaction coordinate, a dihedral angle, or an intermolecular distance). It is defined in terms of the radial distribution function. One way to use an umbrella potential to calculate the PMF is to use the window-sampling technique, where the umbrella potential restricts the sampling to a restricted part (a "window") of the reaction coordinate. In addition, it is realized that the optimum umbrella potential is minus the PMF. Therefore, one can use an adaptive method where the umbrella potential is updated during the simulation so that it converges to minus the PMF.

# Exercise 5 (5,15)

a) What do we mean by *transferability* of force-field parameters and why is that a concern?

b) What is an atomic charge? Do atomic charges show a large degree of transferability? Using electronegativity equalization, discuss how atomic charges can be obtained from (to a large extent) transferable atom-type parameters. For example, explain how the carbon charges in benzene ($C_6H_6$) of around $q = -0.08$ and the carbon charge in formaldehyde ($H_2CO$) of around $q = +0.50$ can be obtained with the same atom-type parameters.

**Solution:** a) With transferability of force-field parameters, we mean to what extent the same values of the parameters can be used for different types of molecular systems (proteins, unpolar polymers, surfaces, solid state, etc.). The concern is that we cannot expect that a given force field can be used for our chosen application without validation, either by doing a literature study or by comparing to quantum chemical calculations. For example, it cannot *a priori* be expected that a force field developed for proteins can be used to model other types of systems.

b) If the molecular charge distribution is divided into atomic contributions, we refer to that as atomic charges. An atomic charge consists of the charge of the nucleus and the part of the electron distribution assigned to the nucleus. Atomic charges by themselves show a very small degree of transferability, which is examplified by the molecules given in this exercise.

Electronegativity equalization is based on that charge is allowed to flow between the atoms so that the electronegativity ("chemical potential for electrons") is equal everywhere in the molecule. In this model, it turns out that the atomic charges to a large extent depend on the difference in electronegativity between neighbouring atoms, where the atomic electronegativities, $\xi_i$, can be regarded as transferable atom-type parameters. If $\xi_H < \xi_C < \xi_O$ (or the reverse order, depending on the definition of the electronegativity), the example in the text can be explained since we then get a charge a transfer from carbon to hydrogen as well as from oxygen to carbon. The magnitude of the charge transfer is regulated by the electronegativity difference as well as the atomic chemical hardness parameters.

# Exercise 6 (20)

Your fellow organic chemistry student comes around and asks you about a problem he/she has: "I have a molecule here, but I do not know which is the most stable conformation in solution. Also, the conformation seems to change when I change the polarity of the solvent." You realize that molecular modelling would be useful and you vigorously offer to do some calculations. Which strategies (more than one) would you consider/suggest?

**Solution:** We need to combine what we know about conformational analysis and the calculation of solvent effects.

I would probably start by doing a conformational analysis of the molecule in the gas phase using DFT calculations. Depending on the size of the molecule, I would use a systematic search routine looking at all relevant minima for a small molecule or e.g. genetic algorithms if the molecule is too large for a systematic search. It is reasonable to start with a gas-phase study since it can be expected that the solvent contribution to the preferred conformation is relatively modest, but important, so only low-energy conformations in the gas-phase will be relevant also in solution.

To include solvent effects, I would start with a continuum solvation model like the polarizable-continuum model (PCM) or the COSMO approach (used in the exercise) with a cavity shaped according to the molecule, also here using DFT calculations. If hydrogen bonding is important for the solvation, I would either include the important solvent molecules explicitly in the DFT calculation (if it is not too expensive computationally) or use as hybrid quantum-mechanical molecular-mechanics (QM/MM) model.

4

An alternative for studying solvent effects is to do free-energy simulations, i.e. the free-energy difference between two conformations is the property of interest. Using either free-energy perturbation techniques (or its companion methods) or potential-of-mean-force calculations would be useful methods. However, free-energy simulations would probably be less straightforward to do as compared to hybrid DFT-continuum solvent models.

# TKJ4205/KJ8902 Molecular Modelling
# Exam 15.12.2011, 09.00-13.00

**Norges Teknisk-Naturvitenskapelige Universitet (NTNU)**
**Code A**: All printed and hand-written texts are allowed. All calculators are allowed.
**Contact**: Prof. Per-Olof Åstrand, Dep. of Chemistry, cell phone: 9346 3033
**NB**: The tasks are not weighted equally. The weight for each task is given in parenthesis after the task number. The weights sum up to 100. The expected length of the answer on each subtask is around 1/2 page with a normal style of hand-writing, which indicates the expected level of detailness of each answer.

## Exercise 1 (15,15)

a) Describe the general features of the Hartree-Fock model in quantum chemistry (without mathematical derivations). How is it built up and what are the approximations?

b) How is electron correlation defined, and what is our physical interpretation of electron correlation? How is electron correlation included in molecular orbital and density-functional theory calculations, respectively?

---

**Solution:** a) Starting at the Born-Oppenheimer approximation, we need to consider the model/approximation for both the Hamiltonian and the wave function. The molecular wave function is represented as a *Slater determinant* of a product of *molecular orbitals* where the molecular orbitals are approximated as a *linear combination of atomic orbitals*. The atomic orbitals are in turn expanded in a *basis set*.

The Hamiltonian is approximated in a *mean-field approximation* as a *Fock operator*. Since the Fock operator includes the orbitals (which is the result from solving the Hartree-Fock equations), we get a set of coupled one-electron equations which has to be solved self-consistently.

b) The electron correlation energy is defined as *the exact energy minus the Hartree-Fock energy*. The physical interpretation of electron correlation is that the *motion* of the electrons is correlated, and this is not described in the Hartree-Fock approximation.

In molecular-orbital (wave function) methods, electron correlation is included by noting that the exact wave function (within a given basis set) can be expanded in terms of the solutions of the Hartree-Fock equation,

$$\psi_i^{\text{exact}} = \sum_\mu C_{\mu i} \psi_\mu^{\text{HF}}$$

and this can be done in different ways: CI expansion, MCSCF wave functions, Møller-Plesset perturbation theory, etc.

In DFT, correlation is included in terms of the exchange-correlation functional, which may be regarded as an *ad hoc* modification of the Fock operator.

---

## Exercise 2 (15)

The solvation effect on a molecule can be calculated by a "full quantum chemical calculation" by including the solute and all the solvent molecules on the same level of theory. Discuss three

solvation models where the solvent is treated with a more simple model. What are the main features of each model, their respective advantages and disadvantages as compared to a "full quantum chemical model"?

> **Solution:** Three common solvation models:
>
> To describe the solvent with a *dielectric continuum* in a *self-consistent reaction-field model*. The shape of the cavity may be described by putting a sphere around each atom of the molecule as done in the *polarizable continuum model (PCM)*. The advantage of the method is that it is computationally efficient (basically no extra cost). The disadvantages are that specific interactions (e.g. hydrogen bonding) is difficult to model and that in principle there is no unique definition of the shape and size of the cavity.
>
> To use a *hybrid quantum-mechanical and molecular-mechanics (QM/MM) model*, where the surrounding molecules are described with a force field. Compared to a dielectric continuum model, it gives an improved description of short-ranged electrostatic interactions (e.g. hydrogen bonding). However, a single configuration of the solvent molecules does not exist, so we have average over a set of solvent configurations generated e.g. by molecular dynamics simulations. As for the dielectric continuum models, it (normally) does not include exchange-repulsion and dispersion contributions.
>
> In the *Langevin dipole* model, the solvent is represented by a set of freely rotatable dipole moments placed on a lattice around the molecule. It has basically the same advantages and disadvantages as a dielectric continuum model.

## Exercise 3 (15)

What do we mean by liquid structure (e.g. compared to gases and solids)? How is it calculated in molecular simulations?

> **Solution:** An *ideal* gas do not have any structure, i.e. since interactions between the gas molecules are ignored the relative positions of the molecules are not correlated. In solids, the relative positions of the atoms are correlated on a macroscopic scale, i.e. we have translational symmetry (ignoring quasi-crystals). In a liquid, we talk about a *local* structure, i.e. we have a short-range structure noted by first and second solvation shells whereas further away from the molecule, the liquid may be described as a dielectric continuum (i.e. structureless).
>
> In simulations, we calculate liquid structure by the *radial distribution function*.

## Exercise 4 (10,10,10,10)

Your fellow student comes around and asks you about some problems he/she has. You realize that molecular modelling would be useful and you vigorously offer to do some calculations. Which strategy would you suggest in the following cases:

a) To predict the positions of the peaks in a UV/VIS spectrum of a coloured liquid.

b) To calculate the diffusion constant of small molecules (e.g. $CO_2$) in a porous medium (e.g. a zeolite).

c) To calculate which is the most stable conformation in a small organic molecule (10-20 atoms) and in a macromolecule (e.g. a biopolymer with thousands of atoms), respectively.

d) To suggest which molecule in a series of ligands ($L_1$, ..., $L_{30}$) that is most efficient in inhibiting an enzyme.

---

**Solution:** a) The peaks in a UV/VIS spectrum refers to the positions of the electronic excitation energies, which we may obtain from *time-dependent density functional theory (TD-DFT)* calculations. Since it is a liquid, we may use one of the solvation models in Exercise 3 to include the effects of the surrounding molecules in a liquid.

b) The diffusion constant is a transport property (time-dependent property) so we need to use *molecular dynamics simulations* and get the appropriate average from a *time-correlation function*. It is probably a very large simulation, including thousands of atoms, so we will probably have to use a suitable *force field* for calculating the interatomic forces.

c) For a relatively small molecule, the preferred choice is probably to use a *systematic search* where all local minima are generated and to use a *quantum chemical method*, e.g. DFT, to calculate the energy of each local minimum structure. For a macromolecule, the preferred choice is probably to use a method like *genetic algorithms* combined with inexpensive energy calculations using a *force field* to obtain the most stable conformation.

d) The method referred to as *docking* is the preferred choice. Docking consists of a *conformational search* of the ligand (and in most cases not of the protein) and a *free energy calculation* of the ligand-protein interaction. The binding free energy may be calculated in different ways: *free-energy perturbation techniques* or simpler methods based on *scoring functions* and/or *QSAR*.

---

**Theory Questions Molecular Simulations Set 1 (taken from digital exams 2015/2016)**

## Question 1: MD and MC-1

**a) Compute time-correlation functions:** only by MD because time evolution

**b) Compute the average number of hydrogen-bonds that a water molecule donates or accepts in liquid water at room temperature:** by both MD and MC

## Question 2: MD and MC-2

**a) Given an initial set of positions and velocities, knowing the positions and velocities 1 nanosecond later:** by neither of the two. MC won't work because it requires evolving the equations of motion. MD won't work because of chaos (Lyapunov instability). 1 nanosecond typically involves 1 million time steps which is more than enough for chaos to kick in.

**b) Knowing how many CO2 particles get adsorbed on average inside a MOF per gram MOF from the atmosphere at ambient conditions using a simulation box containing 2X2X2 unit cells of the MOF crystal with periodic boundary conditions in all directions. (MOF=Metallic Organic Framework, which is a nanoporous crystal):** Only by MC. To solve this question one needs to have a simulation in which the number of CO2 molecules are allowed to fluctuate. With MD this would only be possible if one simulates the MOF crystal in contact with a CO2 gas. MC is more flexible since it allows for inserting and deleting CO2 molecules. The question clearly mentions that the simulation box has only one piece of crystal with periodic boundary conditions. Hence, there is no possibility in an MD simulation to see CO2 molecules move outside the crystal. Therefore, only MC can be used.

## Question 3: MD integrators

**a) The velocity-Verlet algorithm ensures that the temperature remains constant.** False. The plain velocity-verlet algorithm keeps the energy (almost) constant. To keep temperature constant the velocity-verlet needs to be combined with a thermostat.

**b) The Euler algorithm is area-preserving.** False.

## Question 4: Ensemble averages-1

**a)** $\langle A + B \rangle = \langle A \rangle + \langle B \rangle$. True

**b)** $\langle AB \rangle = \langle A \rangle \langle B \rangle$ False. Only if $A$ and $B$ are uncorrelated the above is true.

**c)** $\langle A(0)B(t) \rangle = \langle A(-t)B(0) \rangle$. True

## Question 5: Ensemble averages-2

Andersen and Langevin. NVE has no temperature dependence. Berendsen is based on a velocity rescaling procedure that get the average kinetic energy correct and also some of its fluctuations, but is not exactly equivalent to a correct Boltzmann statistics based constant temperature ensemble.

## Question 6: NVE/NVT

a) If we run MD in the NVE ensemble, the number of particles in the system, the volume, and the total energy is kept constant: True

b) If we run MD in the NVT ensemble, the number of particles in the system, the volume, and the kinetic energy is kept constant.: False. On average the kinetic energy should be 3/2 N kB T with N the number of particles, but a good thermostat also ensures that the fluctuations in the kinetic energy are statistically correct.

## Question 7: MD and MC-1

Tell whether this can be done by MD, by MC, by both, or by neither of the two.
a) Given an initial set of positions and velocities, knowing the positions and velocities 2 nanoseconds later: by neither of the of the two (Lyapunoc instability)

b) Knowing how many argon molecules are on average adsorbed per nm2 on a graphene surface after equilibration when a graphene is in contact with an argon gas of 1 bar at room temperature: by both MD and MC

## Question 8: MD and MC-2

a) Compute diffusion constants: by MD

b) Compute the average number of hydrogen-bonds that a water molecule donates or accepts in liquid water at room temperature: by both MD and MC

## Question 9: MD and MC-3

a) Compute the life time of a hydrogen bond: By MD

b) Compute the fraction of time that a water molecule has more than 4 hydrogen bonds in liquid water: By both MD and MC. Since it is the fraction of time that the system is in a certain state, this corresponds to the probability of that state. This probability is a thermodynamic property which can be computed by both MD and MC.

## Question 10: MD vs MC

What can be studied most efficiently with either MD or MC.

- **CO2 gas diffusion in the atmosphere:** MD, because it is dynamics

- **The amount of CO2 adsorbed in a zeolite as function of gas pressure:** MC is most efficient, because allows fluctuating number of particles, non-local moves, CBMC, etc.

- **Radial distribution function of liquid water:** Both possible but MD more efficient. Condensed phase, so MC will suffer from the fact that maximum-displacement must be small. MD moves more molecules at the same time. There is also the inertia effect: taking the velocities into account gives a kind of flow which helps to decorrelate more rapidly.

- **Nucleation rate of ice formation:** MD, dynamical property

- **The rate of methane absorption in a zeolite:** If even MC is often more efficient for adsorption, here it is the rate that is asked. Dynamical property, so must be MD

# Question 11: MD integrators

- **The Runga-Kutta is in principle more accurate than velocity-Verlet after only a few MD steps.** True. The Runga-Kutta integrator is higher order scheme (based on a Taylor expansion of higher order in $\Delta t$). This means the systematic error in the new positions and velocities after an MD step deviate less from the true positions and velocities than with velocity-Verlet if the same time step is taken.

- **The integrators which are area-preserving and time-reversible have a shadow-Hamiltonian which is approximately conserved.** False. It is exactly conserved.

# Question 12: Area-preserving transformations

Area preserving operators: A,B,C, E, F

# Question 13: Ensemble averages-1

a: True, b: True, c: False

# Question 14: Balance, Detailed-Balance, Ergodicity

Answers:
Set1:A, Set2: A, Set3: C, Set4: A, Set5: B, Set6: D

Esplanantion:
Detailed balance implies:

$$P_i t_{ij} = P_j t_{ji} \text{ for any pair } i, j \tag{1}$$

Detailed balance holds for Sets: 1,2,3, and 4.

But we have also to consider Ergodicity:
Ergodicity implies that starting from any particular microstate, one should be able to visit all other states eventually. Set 2 does not allow jumps between state 3 and state 1, but still one can visit state 3 starting from state 1 by jumping via 2 til 4 til eventually 3. So it is ergodic. Also Set 1 and Set 4 are ergodic. Set 3 does not allow jumps between 1 and 3 and not between 2 and 4. This means that if you start at state 1, for instance, you can never visit state 2 or 4,

even not after an infinite number of steps.

So of Sets 1-4: All of them are A except Set 3 which is C because it is not ergodic.

Set 5 has no detailed-balance, but it still conserves the probability distribution. To make the "flea" analogue, suppose there are 300 fleas in state 1, 200 in state 2, 400 in state 4, and 100 in state 4 (these numbers are taken to be proportional to $P_i$). Then in the next step: $300 \times 0.2=60$ fleas will jump out of state 1 (to state 2). At the same time, $400 \times 0.15=60$ fleas will jump into state 1 (from state 3). So the number of fleas in state 1 remains the same. The same is true for state 2,3, and 4. So it is balanced, but not detailed-balance
So Set5 has B as an answer.

Set 6 is not balanced. In the same analogue as the above: 300*0.1=30 fleas will jump out of state 1 while 400*0.1=40 fleas will jump in. So the answer is D for this set.

# Question 15: MC adsorption 1

Adsorbtion energy is the average energy that the system gains or loses upon adsorption of the argon atom. Since the crystal atoms are fixed, the only interaction that we need to consider is that of the argon with the crystal atoms. When the argon is outside far away from the crystal its interaction energy will be zero while upon adsorption the energy is likely to go down. Occasionally the interaction energy might be higher than zero due to a collision with the walls of the crystal, but very high positive energies should be considered as rare events which one would probably not observe in a Metroplois MC simulation or in an MD simulation. The algorithm described here is not Metropolis MC! It is just random insertion not trying to avoid overlap of the argon with the crystal. Clearly the first value (+.3 eV) is due to such an overlap. The blind insertion implies that we sample each point in space with equal probability while in practice the argon atom has a higher probability to be found at the low-energy positions. Therefore we need to weight each energy value with the Boltzmann factor.

```
Temperature: 800 K
1/(kbT)=1/(8.6*10^(-5)*800)=14.535/eV
Values:
+.3 eV, -0.33 eV, -0.38 eV,  -0.33 -0.33 eV
(exp(-14.535*0.3)+ exp(-14.535*-0.38)+ 3*exp(-14.535*-0.33))
=613.74875223
(0.3*exp(-14.535*0.3)+(-.38)*exp(-14.535*-0.38)+ 3*(-.33)*exp(-14.535*-0.33))
=-215.052051555
adsorbtion energy: -215.052051555/613.74875223=-0.35040090492
```

Both -0.35 as 0.35 would have gotten full points since the sign is merely a convention.

# Question 16: MC adsorption 2

$k_B = 8.6 \cdot 10^{-5}$ eV/K. $T = 800$ K. This means $\beta = 1/(k_B T) = 1/(8.6 \cdot 10^{-5} * 800)$ eV$^{-1}$ =14.53488 eV$^{-1}$.

- MC1: +0.30 eV.
  First move so it is accepted by default.
  Hence, $E_1$=0.30 eV

- MC2: -0.33 eV
  Energy goes down, so it is accepted:
  $E_2$ =-0.33 eV

- MC3: -0.40 eV
  Energy goes down, so it is accepted:
  $E_3$ =-0.40 eV

- MC4: +0.48 eV
  Energy goes up
  $\Delta E = 0.88$ eV.
  Chance of accepting this move equals $\exp(-\beta \Delta E) = \exp(-14.53488 * 0.88) = 0.0000028$
  $p = .51$ so move is rejected, we keep the old configuration:
  $E_4$=-0.40 eV

- MC5: -0.30 eV
  Energy goes up.
  $\Delta E = -0.30 - (-0.40) = 0.10$ eV.
  Chance of accepting this move equals $\exp(-14.53488 * 0.10) = 0.23375$.
  p=0.21<0.23375 so move is accepted.
  $E_5 = -0.30$ eV

- MC6: -0.35 eV
  Energy goes down, so it is accepted.
  $E_6 = -0.35$ eV

Ignore MC moves 1 and 2. Take the average of the others: (-0.40-0.40-0.30-0.35)/4=-0.3625 eV. Rounded off to two digits: -0.36. The absolute value 0.36 would have been approved as well. It is basically a convention whether you define adsorption energy with a minus sign or not. It is always assumed that the energy goes down upon adsorption (otherwise it wouldn't adsorb at all).