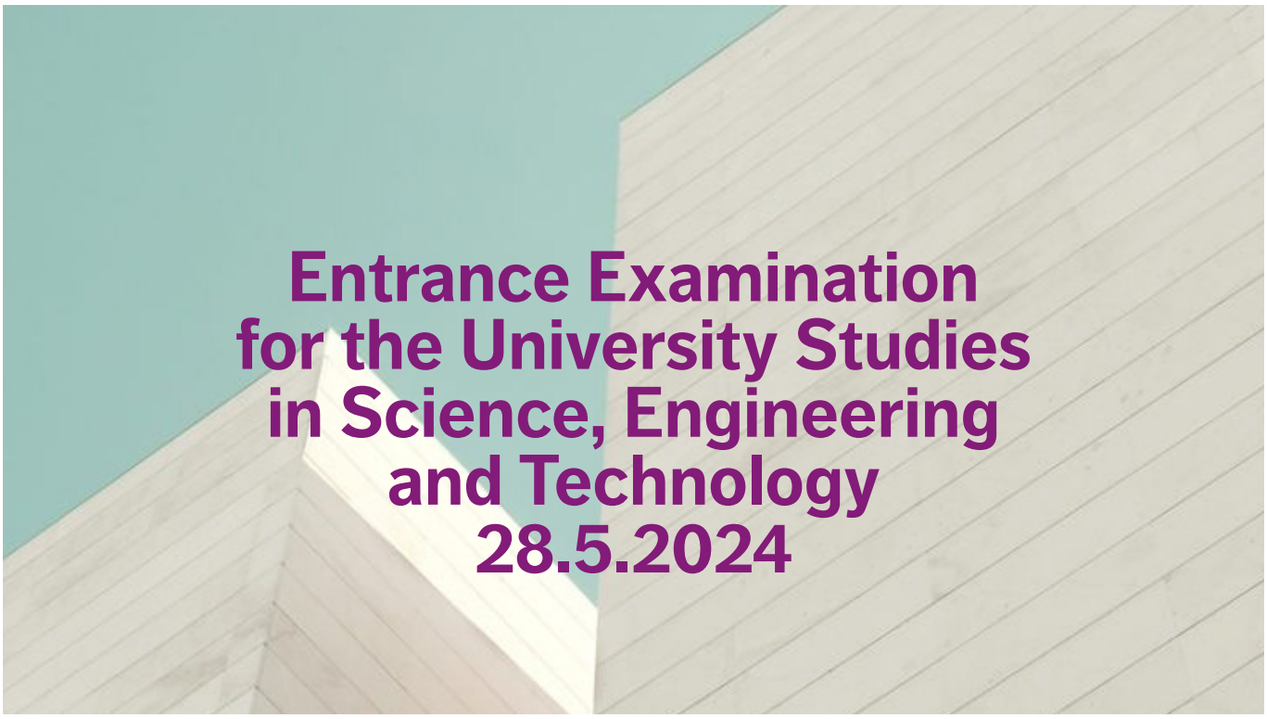
dia DIPLOMI-INSINÖÖRI- JA ARKKITEHTIKOULUTUKSEN YHTEISVALINTA
DIPLOMINGENJÖRS- OCH ARKITEKTUTBILDNINGENS GEMENSAMMA ANTAGNING



Entrance Examination for the University Studies in Science, Engineering and Technology 28.5.2024

QUESTIONS

INSTRUCTIONS

This entrance exam has three (3) obligatory mathematics questions. In addition, there is an optional section with questions in physics, chemistry and creative problem solving within the field of technology. In the optional section you must answer three (3) questions.

You can get six (6) points for each question. The maximum number of points in the entrance exam is 36.

Please write all your answers in the answer booklet. After the examination, you may take the question booklet with you.

QUESTION BOOKLET
Do not answer here.

Justify your answers in all mathematics questions.

Mathematics | Question 1.
Report the answers in exact values unless stated otherwise.

a) Solve the equation $\frac{x+1}{x-1} = \frac{x-1}{x+1}$. (1 p.)

b) Solve the equation $(x-2)(e^x - 3) = 0$. (1 p.)

c) Determine the derivative function of $f(x) = e^{1-x}$. (1 p.)

d) Determine $\int x^{-2}\,dx$. (1 p.)

e) One solution to $\cos x = \frac{1}{2}$ is $x = \frac{\pi}{3}$. Find all solutions of the equation $\cos x + \cos(x + 2\pi) = 1$. (1 p.)

f) In an online lottery, the winning probability of each lot is always 1%, regardless of how many lots have been sold earlier. What is the probability of at least one win, if 10 lots are purchased? Give the answer as a percentage to two decimal places. (1 p.)

Mathematics | Question 2.

Stina's moped car runs on fuel with 1 volume percentage of lubricating oil. However, the gas station sells only fuel without lubricating oil, and fuel mixed with 5 volume percentages of lubricating oil.

a) Stina intends to fill up the empty 10 liter tank. How much Stina needs to refuel each fuel to make the mixture suitable for her moped car? (2 p.)

b) Stina mistakenly refueled with 6 liters of fuel, which does not contain the lubricating oil. She no longer wants to fill up the tank, but just enough so that the mixture is suitable. How much does she need to refuel with the other fuel? (4 p.)

Mathematics | Question 3.

a) For which value of the constant $a > 0$ the equation $2x^2 + a = \sqrt{x}$ has exactly one real solution? (3 p.)

b) The side profile of a skating ramp is bounded by the curve $y = \frac{1}{9}x^2$ from above, the x axis from below, and the lines $x = -6$ and $x = 6$ from the vertical sides. The side of the ramp is painted starting from the left, and the paint runs out when 7 surface units have been painted. The right edge of the painted area is then vertical. How far is it from the left edge of the ramp? (3 p.)

QUESTION BOOKLET | Do not answer here.

Physics | Question 1.

Answer the sub-questions 1–4. In multiple-choice questions, select one answer (A–D). Correct answer: 1 p. Wrong answer, no answer, or multiple answers: 0 p.

On windy and sunny days, renewable energy sources generate more electric energy than is needed. The excess energy can be stored, for example, as heat in a sand battery. In the sand battery, sand inside an insulated reservoir can be heated up to 600 °C using electric energy. Energy is thus left to be used later on windless and cloudy days. The efficiency of the sand battery can be considered good, since 90 % of the energy stored into it can be utilized. The specific heat capacity of sand is 0.84 kJ/(kg·K) and its density is 1.5 kg/dm^3.

1. While operating for one year, the sand in a sand battery (1 p.)

   A. only releases energy.
   B. only receives energy.
   C. neither releases nor receives energy.
   D. alternately releases and receives energy.

2. On frosty days, on average 4.0 kW power is needed to heat a single-family house. How many single-family houses can be heated with a sand-battery in which 8.0 MWh energy is stored? (1 p.)

   A. 83
   B. 75
   C. 2000
   D. 1800

3. Inside the sand, there are pipes made of steel. At 50 °C temperature the outer diameter of the pipes is 100 mm and the inner diameter is 94 mm. Air flows inside the pipes enabling heat transfer to the battery and from the battery. When the temperature of the battery rises from 50 °C to its maximum temperature, (1 p.)

   A. the outer and inner diameters of the pipes do not change.
   B. the outer and inner diameters of the pipes decrease.
   C. the outer and inner diameters of the pipes increase.
   D. the outer diameter of the pipes increases, and the inner diameter of the pipes decreases.

4. Instead of sand, the energy can be stored also in water. Examine how big a water reservoir is needed, so that it can release the same amount of energy as the sand battery. Use the following information in the calculations: The maximum temperature of the sand battery is 595 °C and its volume is 71 m^3. The maximum temperature of the water reservoir is 91 °C. When using either battery, the temperature can drop to 49 °C, and the efficiencies of the batteries are the same. You can assume that the density of water is 1.00 kg/dm^3 and the specific heat capacity of water is 4.19 kJ/(kg·K). (3 p.)

Physics | Question 2.

Based on the reference material, answer the sub-questions 1–4. In multiple-choice questions, select one answer (A–D). Correct answer: 1 p. Wrong answer, no answer, or multiple answers: 0 p.

A laser light propagating in air is described as a sinusoidal wave whose electric field is

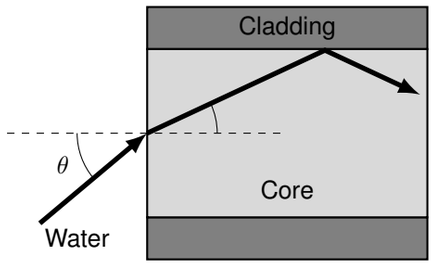
$$E_y(x,t) = 12.5 \cdot 10^3\,\mathrm{V/m} \cdot \sin[(1.2 \cdot 10^7 \mathrm{rad/m})x - (3.6 \cdot 10^{15}\mathrm{rad/s})t].$$

1. What is the wavelength of this laser light? (1 p.)
   A. $7.5 \cdot 10^7\,\mathrm{m}$
   B. $5.2 \cdot 10^{-7}\,\mathrm{m}$
   C. $1.9 \cdot 10^6\,\mathrm{m}$
   D. $8.3 \cdot 10^{-8}\,\mathrm{m}$

2. What happens to the wavelength of a laser light when it propagates from air to quartz glass? (1 p.)
   A. The wavelength does not change.
   B. The wavelength lengthens.
   C. The wavelength shortens.
   D. The wavelength either lengthens or shortens depending on the angle of incidence.

3. What is the power of a laser, if the cross-sectional area of its beam in air is $1.2 \cdot 10^{-6}\,\mathrm{m}^2$? (1 p.)
   A. $0.25\,\mathrm{W}$
   B. $12.5 \cdot 10^3\,\mathrm{W}$
   C. $2.0 \cdot 10^{-6}\,\mathrm{W}$
   D. $8.3 \cdot 10^{-10}\,\mathrm{W}$

4. The beam produced by a laser hits the end of an optical fiber from water at an angle of θ. The core of the fiber is surrounded with a cover called a cladding. At visible light range, the refractional index of the core is 1.50 and the refractional index of the cladding is 1.46. The light must not refract from the core in order not to attenuate significantly. How large can the angle of incidence θ be at most, so that the light propagates inside the core without refracting into the cladding? (3 p.)



QUESTION BOOKLET | Do not answer here.

Physics | Question 2. Reference material: Describing laser light as a wave

Laser light is an electromagnetic wave whose frequency and phase remain constant. It is possible to describe an electromagnetic wave with electric and magnetic fields, and both fields are always present in a wave.

Laser light can be modeled as a sinusoidal wave. For a wave propagating in the positive x direction the electric and magnetic fields are of the form:

$$\begin{aligned} E_y(x,t) &= E_{\max}\sin(kx-\omega t) \\ B_z(x,t) &= B_{\max}\sin(kx-\omega t) \end{aligned} \tag{1}$$

This description is of the same type as used for mechanical sinusoidal waves. In the above expressions (1) the wave number $k = 2\pi/\lambda$ for a wave with wavelength λ. The angular frequency $\omega = 2\pi f$ can be obtained from the frequency f.

The electric and magnetic fields of the expression (1) can be given in a vector form $\bar{\boldsymbol{E}} = E_y\hat{\boldsymbol{j}}$ and $\bar{\boldsymbol{B}} = B_z\hat{\boldsymbol{k}}$. The fields of a wave propagating in x direction are drawn in the adjacent Figure 1 at a certain time.

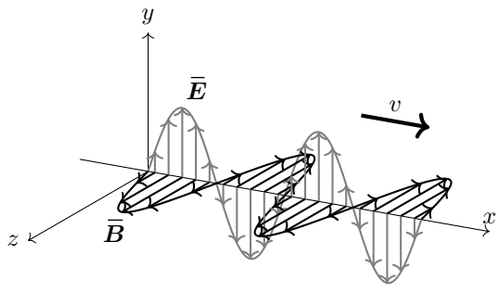


Figure 1. *A sinusoidal electromagnetic wave propagating in $+x$ direction at a certain time.*

The electric and magnetic fields in a wave are always perpendicular to each other but also to the propagation direction of the wave. The wave described in the expressions (1) is polarized, since the electric field vector points always in the y direction and the magnetic field vector in the z direction. $E_{\max}$ and $B_{\max}$ are called field amplitudes and correspond to the maximum values of the fields. The amplitudes are connected by the propagation speed v of the wave: $E_{\max} = vB_{\max}$.

The intensity I measures the energy content carried by the wave per time and area units. Its SI unit is thus $\mathrm{W/m^2}$. For a wave propagating in the vacuum, the intensity of a sinusoidal wave can be calculated directly using its electric or magnetic field amplitude, e.g.

$$I = \frac{1}{2}\epsilon_0 c {E_{\max}}^2 \tag{2}$$

where constant $\epsilon_0 = 8.854\cdot 10^{-12}\,\mathrm{J/(V^2\cdot m)}$ is called the permittivity of the vacuum. The above expression is very accurate also for a wave propagating in air.

The propagation speed of the wave can be calculated in the same manner as for mechanical waves: $v = \lambda f$. In the vacuum the speed of light is known to be $c = 2.998\cdot 10^8\,\mathrm{m/s}$. When light propagates inside a material its speed decreases, even though its frequency remains the same. The propagation speed depends on the index of refraction of the material in a simple manner: $v = c/n$.

When a laser beam propagates from one material into another it refracts at the boundary according to the law of refraction: $n_1\sin(\theta_1) = n_2\sin(\theta_2)$. The angle of refraction can be calculated using the law of refraction if the angle of incidence and the indices of refraction on both sides of the boundary are known. For visible light, the indices of refraction are for air 1.00, for water 1.33, for glass about 1.5 and for diamond 2.4.

If the angle of incidence is too large, the beam cannot refract at the boundary, but instead it experiences total internal reflection. This borderline case corresponds to a situation where the angle of refraction is $90°$.

The light produced by a laser can propagate a large distance in an optical fiber without its intensity getting much smaller. Because the fiber cannot be straight, the light propagates inside the fiber by constantly reflecting from the boundary between the core and the cladding. In order for the intensity not to decrease, the reflections have to be total internal reflections, i.e., the beam must not refract into the cladding.

QUESTION BOOKLET | Do not answer here.

Chemistry | Question 1.

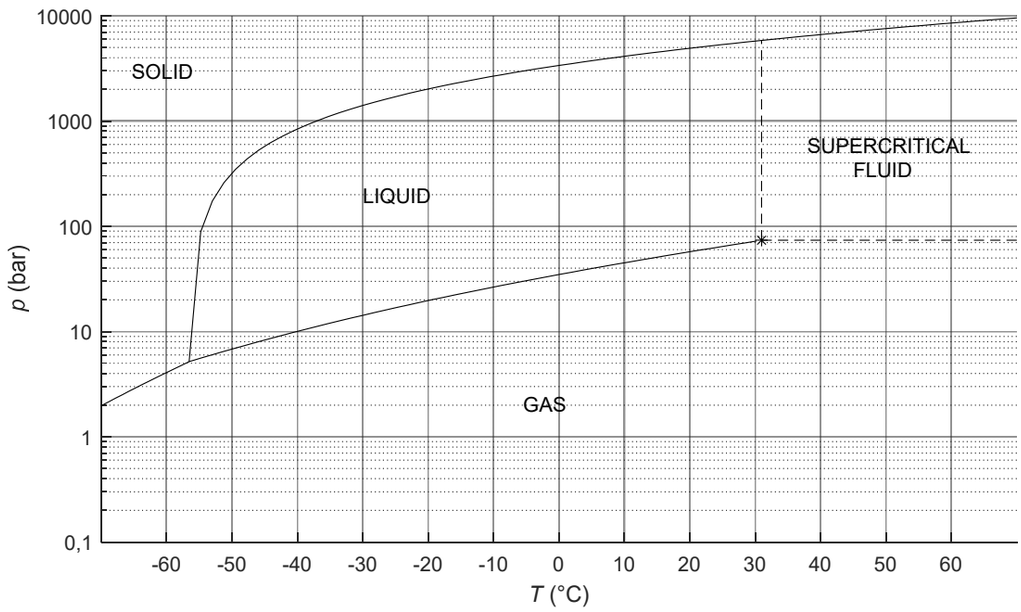
Answer the sub-questions 1–4. In multiple-choice questions, select one answer (A–D). Correct answer: 1 p. Wrong answer, no answer, or multiple answers: 0 p.

You can use the attached periodic table of the elements when answering the questions. Please note that the decimal separator used in the periodic table of elements is comma (,).

Constants: $R = 8.31446 \frac{Pa \cdot m^3}{mol \cdot K} = 0.0831446 \frac{bar \cdot dm^3}{mol \cdot K}$; 0 °C = 273.15 K

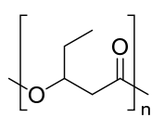
1. A phase diagram is a graph that shows the state of matter of a substance at a given temperature and pressure. The graph shows that a phase transition can take place by changing the temperature and/or the pressure. Study the phase diagram of carbon dioxide below and deduce how carbon dioxide that is at a temperature of 20 °C and a pressure of 1 bar can be liquefied, *i.e.* converted to liquid state.

(1 p.)



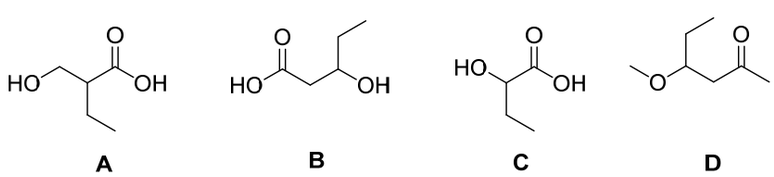
A. By decreasing the temperature by 80 °C and increasing the pressure by 5 bar.
B. By decreasing the temperature by 50 °C and increasing the pressure by 10 bar.
C. By keeping the temperature at 20 °C and increasing the pressure by 70 bar.
D. None of the alternatives above is correct.

2. Polyhydroxyalkanoates, such as poly-3-hydroxyvalerate (PHV), are polyesters in their chemical structure. Carbon dioxide can be utilized as one of the starting materials in their production. Which of the structural formulas (A-D) describes the monomer that can be used in the production of PHV?



PHV

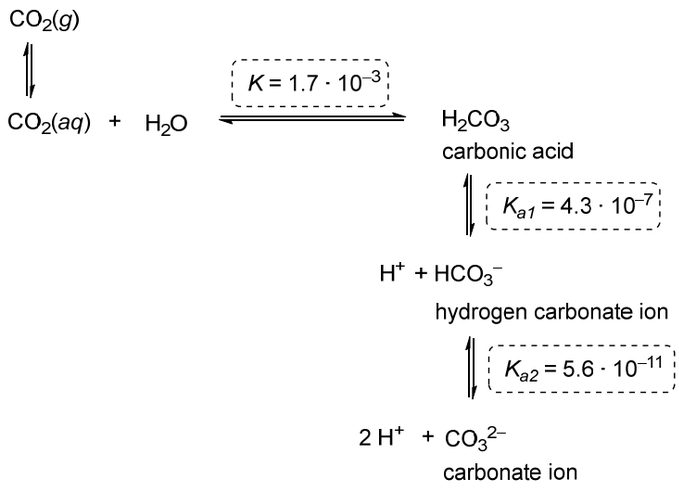
(1 p.)



A **B** **C** **D**

QUESTION BOOKLET | Do not answer here.

3. Carbon dioxide dissolves in water to some extent, and in water it forms different products as described in the scheme below (K is the equilibrium constant for a reaction step in question).



Which of the following statements is correct? (1 p.)

A. Most of the dissolved carbon dioxide exists as hydrogen carbonate ions in the aqueous solution.
B. Most of the dissolved carbon dioxide exists as carbonate ions in the aqueous solution.
C. There are more hydrogen carbonate ions than carbonate ions in the aqueous solution.
D. Carbon dioxide exists completely in the form of carbonic acid in the aqueous solution.

4. Plants produce glucose (M = 180.16 g/mol) and oxygen in the photosynthesis reaction using carbon dioxide in the air according to the chemical equation 6 $CO_2(g)$ + 6 $H_2O(l)$ → $C_6H_{12}O_6(s)$ + 6 $O_2(g)$.

a) How many cubic meters of pure carbon dioxide is needed to produce 1.00 kg of glucose? Assume that carbon dioxide is at a temperature of 25.0 °C and a pressure of 101 325 Pa. Carbon dioxide can be assumed to behave ideally. Justify your answer.
b) The carbon dioxide content in the air is 0.04 percent by volume. How many cubic meters of air is needed to produce 1.00 kg of glucose? Assume that air is at a temperature of 25.0 °C and a pressure of 101 325 Pa. Justify your answer.
c) The photosynthesis reaction consumes 1882 kJ energy per one mole of oxygen (O_2) formed. How much energy is required to produce 1.00 kg glucose? Justify your answer.

(3 p.)

Chemistry | Question 2.

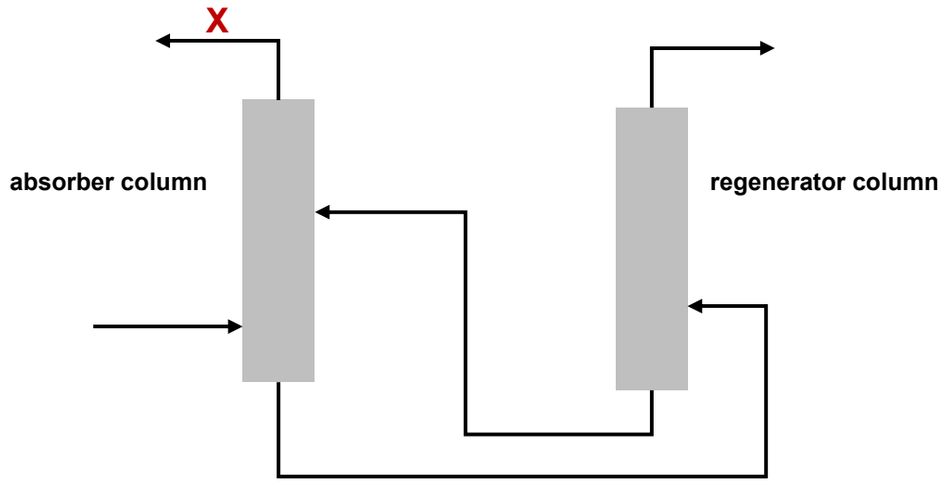
Based on the reference material, answer the sub-questions 1–5. In multiple-choice questions, select one answer (A–D). Correct answer: 1 p. Wrong answer, no answer, or multiple answers: 0 p.

You can use the attached periodic table of the elements when answering the questions. Please note that the decimal separator used in the periodic table of elements is comma (,).

1. The absorption of carbon dioxide taking place in the amine scrubber is (1 p.)
   A. a chemical and reversible absorption.
   B. a chemical and irreversible absorption.
   C. a physical and reversible absorption.
   D. a physical and irreversible absorption.

2. Study the reaction between monoethanolamine and carbon dioxide that takes place in the absorber column. One MEA molecule clearly reacts with carbon dioxide, but what is the role of the other MEA molecule in the reaction? (1 p.)
   A. It acts as an acid and donates a proton.
   B. It acts as a base and donates a proton.
   C. It acts as an acid and accepts a proton.
   D. It acts as a base and accepts a proton.

3. A simplified process flow diagram describing an amine scrubber is shown below.



   Which of the following best describes the position X? (1 p.)
   A. The captured carbon dioxide
   B. The flue gas from an industrial plant
   C. The treated flue gas
   D. The monoethanolamine solution

4. Draw the structural formula of the intermediate formed in the reaction between monoethanolamine and carbon dioxide. Hint: The intermediate is similar to the carbamate product, but it is a zwitterion. (1 p.)

5. Write the balanced reaction equations for the carbonation of the following minerals. The state symbols are not required.
   a) forsterite
   b) chrysotile (2 p.)

QUESTION BOOKLET | Do not answer here.

Chemistry | Question 2. Reference material: Capture and mineralization of carbon dioxide

Gaseous water and carbon dioxide are examples of greenhouse gases in the atmosphere. The natural greenhouse effect caused by these gases enables the average temperature that is favourable for life. However, the content of carbon dioxide in the atmosphere has increased significantly due to the human activity. Since more carbon dioxide is released into the atmosphere than is bound into the biomass by the photosynthesis, chemical and engineering tools are required to curb the increase of the carbon dioxide content in the atmosphere.

Direct capture of carbon dioxide from the air using the available technology is not yet economical, but large amounts of carbon dioxide can be captured from industrial flue gases. However, the capture is not sufficient, but the carbon dioxide still needs to be stored and/or utilized for example in the production of fuels or chemicals. Figure 1 describes in a simplified manner how carbon dioxide from flue gases can be absorbed* by using an amine scrubber, and how the captured carbon dioxide can be stored as carbonates.

**Absorption* is a physical or chemical phenomenon, where atoms, molecules or ions enter liquid, gas or solid material. The absorption is either reversible or irreversible.

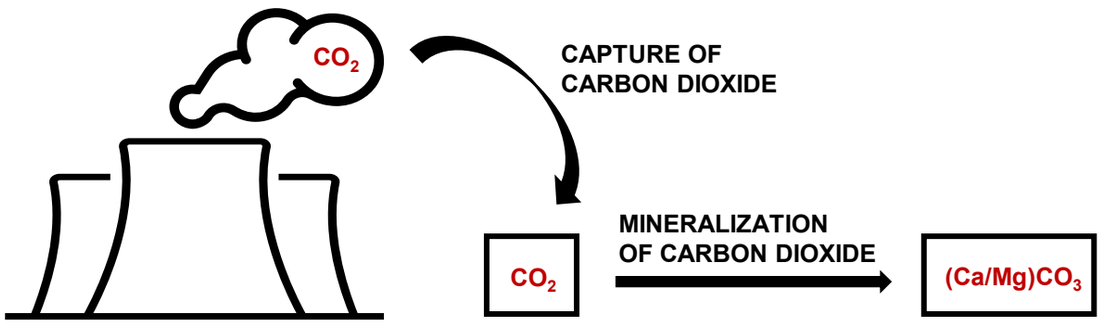
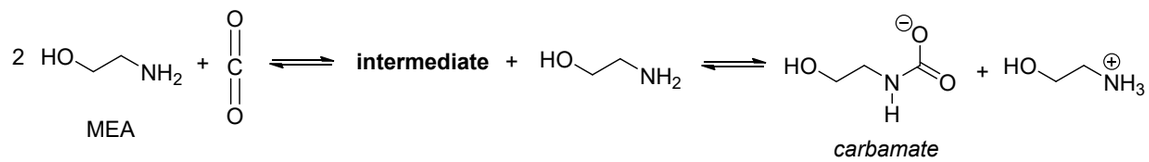


Figure 1. *Capture and storage (mineralization) of carbon dioxide.*

Capture of carbon dioxide by amine scrubber

Amine scrubbing is a process that allows the capture of even 90% of the carbon dioxide in the flue gas. Different amines, such as monoethanolamine (MEA) that is an amino alcohol, are utilized in the process. In the amine scrubber, the carbon dioxide containing flue gas is led into an absorber column, where carbon dioxide reacts with MEA approximately at a temperature of 50 °C. A carbamate product is formed in the reaction *via* an intermediate, as is shown in *Scheme 1.* The carbamate solution is led from the absorber column to a regenerator column, where the temperature is approximately 120 °C. At this temperature, the N-C bond is broken, and the released carbon dioxide can be captured. The amine solution can, in turn, be recycled back to the absorber column.



Scheme 1. *The reaction between monoethanolamine and carbon dioxide.*

Storage of carbon dioxide by mineralization

Carbonation of minerals refers generally to a reaction of a calcium or magnesium containing mineral with carbon dioxide, resulting in the formation of metal carbonates, *i.e.* CO_3^{2-} salts. Minerals utilized in carbonation include for example wollastonite ($CaSiO_3$), forsterite (Mg_2SiO_4) and chrysotile ($Mg_3Si_2O_5(OH)_4$). In addition to metal carbonates, silicon dioxide (SiO_2) and, in some cases water, are formed in the carbonation reactions. The products formed are chemically stable and they can be further utilized in different processes.

Periodic table of elements

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 **H** 1,008					järjestysluku kemiallinen merkki atomimassa	1 **H** 1,008	atomnummer / atomic number kemiskt tecken / chemical symbol atommassa /atomic mass										2 **He** 4,003
2	3 **Li** 6,941	4 **Be** 9,012											5 **B** 10,81	6 **C** 12,01	7 **N** 14,01	8 **O** 16,00	9 **F** 19,00	10 **Ne** 20,18
3	11 **Na** 22,99	12 **Mg** 24,31											13 **Al** 26,98	14 **Si** 28,09	15 **P** 30,97	16 **S** 32,07	17 **Cl** 35,45	18 **Ar** 39,95
4	19 **K** 39,10	20 **Ca** 40,08	21 **Sc** 44,96	22 **Ti** 47,87	23 **V** 50,94	24 **Cr** 52,00	25 **Mn** 54,94	26 **Fe** 55,85	27 **Co** 58,93	28 **Ni** 58,69	29 **Cu** 63,55	30 **Zn** 65,38	31 **Ga** 69,72	32 **Ge** 72,63	33 **As** 74,92	34 **Se** 78,96	35 **Br** 79,90	36 **Kr** 83,80
5	37 **Rb** 85,47	38 **Sr** 87,62	39 **Y** 88,91	40 **Zr** 91,22	41 **Nb** 92,91	42 **Mo** 95,94	43 **Tc** (98)	44 **Ru** 101,07	45 **Rh** 102,91	46 **Pd** 106,42	47 **Ag** 107,87	48 **Cd** 112,41	49 **In** 114,82	50 **Sn** 118,71	51 **Sb** 121,76	52 **Te** 127,60	53 **I** 126,90	54 **Xe** 131,29
6	55 **Cs** 132,91	56 **Ba** 137,33	57-71	72 **Hf** 178,49	73 **Ta** 180,95	74 **W** 183,84	75 **Re** 186,21	76 **Os** 190,23	77 **Ir** 192,22	78 **Pt** 195,08	79 **Au** 196,97	80 **Hg** 200,59	81 **Tl** 204,38	82 **Pb** 207,2	83 **Bi** 208,98	84 **Po**	85 **At**	86 **Rn**
7	87 **Fr**	86 **Ra**	89-103	104 **Rf**	105 **Db**	106 **Sg**	107 **Bh**	108 **Hs**	109 **Mt**	110 **Ds**	111 **Rg**	112 **Cn**	113 **Nh**	114 **Fl**	115 **Mc**	116 **Lv**	117 **Ts**	118 **Og**

Lantanoidit/ Lantanoider/ Lanthanides	57 **La** 138,91	58 **Ce** 140,12	59 **Pr** 140,91	60 **Nd** 144,24	61 **Pm**	62 **Sm** 150,36	63 **Eu** 151,96	64 **Gd** 157,25	65 **Tb** 158,93	66 **Dy** 162,50	67 **Ho** 164,93	68 **Er** 167,26	69 **Tm** 168,93	70 **Yb** 173,05	71 **Lu** 174,97
Aktinoidit/ Aktinoider/ Actinides	89 **Ac**	90 **Th** 232,04	91 **Pa** 231,04	92 **U** 238,03	93 **Np**	94 **Pu**	95 **Am**	96 **Cm**	97 **Bk**	98 **Cf**	99 **Es**	100 **Fm**	101 **Md**	102 **No**	103 **Lr**

QUESTION BOOKLET | Do not answer here.

Problem Solving | Question 1.

Answer the sub-questions 1 to 4. In multiple-choice questions, select one alternative. Correct answer: 1 p. Wrong answer, no answer, or multiple answers: 0 p.

Electricity transmission connections between regions have been marked with arrows on the map in Figure 1. A number at the tip of an arrow indicates the maximum electricity transmission capacity in megawatts (MW) from a region to another. For example, the maximum electricity transmission capacity from the region NO4 to the region SE1 is 700 MW, and the maximum capacity from the region SE1 to the region NO4 is 600 MW. Some arrows have a single number in the middle, indicating that the maximum transmission capacities for both regions are equal.

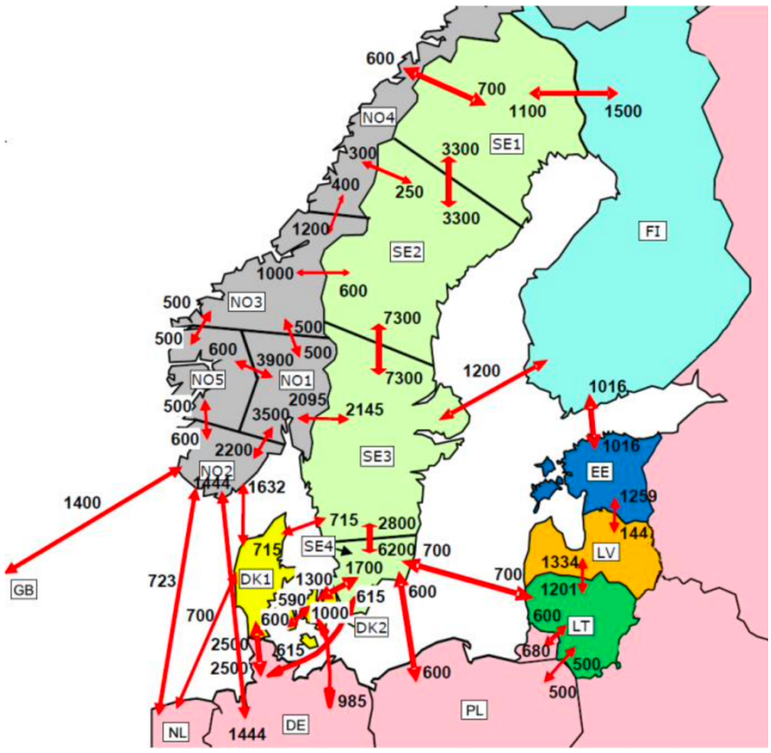


***Figure 1**. Maximum electricity transmission capacities between regions in megawatts.*

1. If one electric stove requires an average power of 10 kilowatts (kW), what is the maximum number of electric stoves that can simultaneously be heated with imported electricity in Finland? (1 p.)

   A. 517 600
   B. 371 600
   C. 331 600
   D. None of the alternatives above is correct

2. It has been estimated that the peak power demand during a cold winter in Finland is approximately 14 400 MW. Let the domestic electricity generation capacity be at most 11 300 MW. Due to cold weather, assume that the electricity transmission capacity from northern Sweden to Finland is limited to 1200 MW and to 1000 MW from southern Sweden to Finland. What percentage of the total capacity of the electricity transmission connections would be in use in Finland during peak consumption if all domestic electricity generation capacity is in use and no electricity is transmitted from Finland to neighbouring countries? (1 p.)

   A. Approximately 88%
   B. Approximately 91%
   C. Approximately 96%
   D. Approximately 101%

QUESTION BOOKLET | Do not answer here.

3. Assume that an additional capacity of 1500 MW is built in Finland with 4 MW wind turbines. Assume that the investment cost of an onshore wind turbine is 1.5 million €/MW, and that the investment cost of an offshore turbine is 25% higher than an onshore turbine. How much the total investment costs are if 40% of the new wind turbines must be built offshore and the rest onshore? (1 p.)

   A. Approximately 620 million euros
   B. Approximately 2.8 billion (2800 million) euros
   C. Approximately 2.6 billion (2600 million) euros
   D. None of the alternatives above is correct

4. Network-like structures can be represented by graphs, which consist of vertices and edges. Figure 2 shows a graph with three vertices NO4, SE1 and FI representing the regions in Figure 1 and edges representing the transmission connections between the regions.

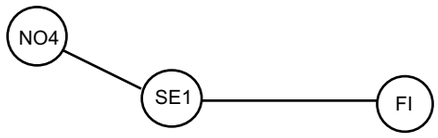


***Figure 2**. An example of a simple graph.*

   a) Draw a graph corresponding to all the regions of Finland, Sweden and Norway in Figure 1 and the electricity transmission connections between them. (1 p.)

   For simplicity, assume in b) and c) that one region is managed by one electricity company.

   b) Can all regions in Finland, Sweden and Norway be managed by three electricity companies so that, for reasons of competition, no company manages two adjacent regions with a transmission connection between them? Justify your answer with a graph. (1 p.)

   c) Assume that the regions in Finland, Sweden and Norway are managed by two electricity companies. What is the minimum number of transmission connections to be removed so that neither company manages two adjacent regions with a transmission connection between them? Justify your answer. (1 p.)

Reference: Reliability of electricity supply in 2023 (in Finnish), Energy authority

Problem Solving | Question 2.

Answer the sub-questions 1 to 5 using the material provided. In multiple-choice questions, select one alternative. Correct answer: 1 p. Wrong answer, no answer, or multiple answers: 0 p.

1. Approximately how many text comparisons are needed to find a key using a suffix array from a text consisting of 1000 characters? (1 p.)

   A. 10
   B. 100
   C. 1000
   D. 10000

2. The memory of a computer consists of bytes. Assume that each character of a text is represented by one byte and each integer by eight bytes. How many bytes are approximately needed to store an n-character text and its suffix array to the memory of a computer? (1 p.)

   A. $8n$
   B. $9n$
   C. $8n + n^2/2$
   D. $9n^2$

3. Assume that a text and its suffix array are stored in a slow memory, from where a single piece of information can be read every 0.01 s. The piece of information is either an integer, or a section of text that is long enough for a text comparison. From how long a text a key can be found in one second? (1 p.)

   A. From about a single page of the text (about a thousand characters).
   B. From about the text of a single thick book (about a million characters).
   C. From about the text content of the Finnish National Library (about a trillion (10^{12}) characters).
   D. From a much longer text than the previous choices.

4. Construct the suffix array of text ”`baabacacc`”. (1,5 p.)

5. The suffix array of text ”`caabbabbacaca`” is `[ 13 2 3 6 11 9 5 8 4 7 12 1 10 ]`. List all the text comparisons that are done to search the key ”`abbc`” from the suffix array. (1,5 p.)

QUESTION BOOKLET | Do not answer here.

Problem Solving | Question 2. Reference material: Suffix array.

Suffix arrays are used in information systems, among other purposes, as directory structures for text, for data compression, and for finding various recurring segments. Consider a long text T, consisting of n characters. The *suffix* starting from the position 1 is the whole text. Respectively, the suffix starting from the position i consists of the characters starting from the i'th position until the end of the text. The suffix array presents all suffixes of the whole text, thus containing n elements. The suffix array contains only the starting positions of each suffix in the original text, not the entire suffix as a text. The suffix array is sorted so that it displays all starting positions of the suffixes in their alphabetical order.

Table 1 presents the suffix array of the text "abcccaabba". The column "suffix" does not actually exist but it is presented in Table 1 for clarity. The suffix array of an n-character text T is thus an array of integers of length n, containing the numbers $1..n$ in the order dictated by the alphabetical order of the suffixes. Creating and using it also requires the text T.

Table 1. *The suffix array of text "abcccaabba".*

	suffix array	suffix
1	**10**	a
2	**6**	aabba
3	**7**	abba
4	**1**	abcccaabba
5	**9**	ba
6	**8**	bba
7	**2**	bcccaabba
8	**5**	caabba
9	**4**	ccaabba
10	**3**	cccaabba

Once the suffix array of the text T has been constructed, *binary search* can be used to find another shorter text from it. The shorter text is called a *key*. In binary search, text comparisons are required to find the key. A text comparison $A < B$ is true if and only if text A precedes text B in alphabetical order. For example, the text "bbc" precedes alphabetically the texts "bcb" and "bbca".

Binary search begins with the entire array (the search range from positions $1..n$). During each step of the binary search, we denote the start of the search range by the letter s, the end of the range by the letter e, and the midpoint of the range by the letter m. In each step of the binary search, the key is compared to the suffix starting from the position of the text indicated by the number at the midpoint ($m \leftarrow (s+e)/2$). If the division does not result in an integer, the result is rounded down to the nearest integer. If the suffix starting from the position of the text indicated by the number at the midpoint precedes the key in alphabetical order, the search continues from the latter part of the suffix array; otherwise, it continues from the first part. The search is continued from the first part by moving the end point to the midpoint ($e \leftarrow m$), and from the latter part by moving the start point to the position following the midpoint ($s \leftarrow m+1$).

Text comparisons and reducing the size of the search range continue until the start and end points meet. Finally, one last text comparison is performed to determine if the key is found at the position of the text indicated by the number at position s in the suffix array. If the key appears at multiple positions in the text, the binary search will find one of them.

Algorithm 1 describes the binary search in an exact form. Notation Text[i..] denotes the suffix of the text starting from position i. Notation SuffixArray[i] refers to the number at position i in the suffix array.

Algorithm 1. BINARY SEARCH(Text, SuffixArray, Key)

```
s ← 1      e ← n          ▷ set s and e to the start and the end of suffix array
While s < e , do          ▷ repeat until there is only one position left
    m ← ⌊(s + e)/2⌋       ▷ set m to the midpoint of s and e, round down if needed
    If Text[SuffixArray[m]..] < Key , then   ▷ compare part of the text and Key
        s ← m + 1         ▷ continue search on the latter part
    else
        e ← m             ▷ continue search on the first part
If Text[SuffixArray[s]..] matches Key , then
    Key is found from position SuffixArray[s].. of Text
else
    Key does not exist in Text
```

QUESTION BOOKLET | Do not answer here.