

ENVIRONMENTAL PHYSICS METHODS LABORATORY PRACTICES



Foundations of Environmental Science Lecture Notes Series

A környezettan alapjai

A környezetvédelem alapjai

Környezetfizika

Bevezetés a környezeti áramlások fizikájába

Környezeti ásványtan

Környezeti mintavételezés

Környezetkémia

Környezetminősítés

Környezettudományi terepgyakorlat

Mérési adatok kezelése és értékelése

Bevezetés a talajtanba környezettanosoknak

Environmental Physics Methods Laboratory Practices



Eötvös Loránd University
Faculty of Science

ENVIRONMENTAL PHYSICS METHODS LABORATORY PRACTICES

Editor:

Ákos Horváth

Associate Professor, Institute of Physics

Authors:

Máté Csanád

Assistant Professor, Institute of Physics

Ákos Horváth

Associate Professor, Institute of Physics

Gábor Horváth

Associate Professor, Institute of Physics

Gábor Veres

Assistant Professor, Institute of Physics

Reader:

Árpád Zoltán Kiss

Professor Emeritus, Hungarian Academy of
Sciences, Institute of Nuclear Research



2012

COPYRIGHT: © 2012-2017, Dr. Máté Csanád, Dr. Ákos Horváth, Dr. Gábor Horváth,
Dr. Gábor Veres, Eötvös Loránd University Faculty of Science

Reader: Dr. Árpád Zoltán Kiss

Creative Commons NonCommercial-NoDerivs 3.0 (CC BY-NC-ND 3.0)

This work can be reproduced, circulated, published and performed for non-commercial purposes without restriction by indicating the author's name, but it cannot be modified.

ISBN 978-963-279-551-5

PREPARED under the editorship of [Typotex Kiadó](#)

RESPONSIBLE MANAGER: Zsuzsa Votisky

GRANT:

Made within the framework of the project Nr. TÁMOP-4.1.2-08/2/A/KMR-2009-0047, entitled „Környezettudományi alapok tankönyvsorozat” (KMR Foundations of Environmental Science Lecture Series).

Nemzeti Fejlesztési Ügynökség
www.ujszachenyiterv.gov.hu
06 40 638 638



A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

KEYWORDS:

Environmental physics, environmental radiation, noise, acoustics, infra sound, natural radioactivity, solar energy, polarized light, dosimetry, ionizing radiation, radon, gamma-spectroscopy, positron emission tomography.

SUMMARY:

In this book we overviewed 17 laboratory practices in the subject of environmental physics. Our measurements mainly covered the area of environmental radiations starting from the acoustic waves, electromagnetic radiation hazard, visible light and going into the area of radioactivity: X-rays, gamma-spectroscopy, annihilation radiation, Cherenkov-radiation, alpha- and beta-spectroscopy. These exercises are good examples for those students who intend to work in laboratories using these spectroscopic or other environmental physics methods.

There are of course lots of areas in environmental physics that were not covered here, but these exercises are adjusted to the technical possibilities of the Environmental Center at Eötvös Loránd University, Budapest.

Preface

This work is based on three laboratory practices that environmental major students have been studying for about 10 years at Eötvös Loránd University (ELTE). We describe 17 lab practices and give 4 more chapters as an introduction to the fields. Several laboratory practices have a Hungarian description in a book edited by Prof. Dr. Ádám Kiss with coauthorship of our colleagues: Panni Bornemisza, Gyula Pávó, Ottó Csorba, Ferenc Deák, Botond Papp, András Illényi and the authors of this book. However, a few new measurement were added to the list and this book covers already the actual material for the laboratory measurement of three lab teaching units: Environmental Physics, Environmental Physics Methods and Radiation Physics. The introductory chapters of this book are also helpful for the students of the master course: „Environmental radiation”.

New measurement of electromagnetic radiation hazards, microwave radiation, infrasound, Cherenkov-radiation, radon exhalation, solar heat collector and polarized light pollution widened our topics in the broad field of environmental physics.

Especially, the English description of the „Infrasound wave detection” laboratory practice is well established on a work of Gábor Szeifert, Gábor Gelencsér and Gyula P. Szokoly at the Department of Atomic Physics of ELTE. The chapter about the „Solar heat collector” is based on the Hungarian description and undergraduate research thesis (TDK) of Edina Juhász and Veronika Pongó.

This English text is for those student who study at ELTE e.g. as Erasmus students in environmental sciences or in physics or by any reason they prefer English language description.

Budapest, March 2012.

The authors

CONTENTS

Chapter I.

<i>Mechanical waves in the environment (Cs.M.)</i>	7
1. Mathematics of waves	8
2. Mechanical waves	23
3. Investigation of environmental noise (ZAJ).....	31
4. Infrasound wave detection (INF)	41

Chapter II.

<i>Nonionizing electromagnetic radiation</i>	50
5. An overview of the electromagnetic waves (Cs.M.)	51
6. Low-frequency electromagnetic radiation of common household devices (ESZ) (Cs.M.)	60
7. Microwave radiation of household devices (MIK) (Cs.M.).....	67
8. Solar air heat collector (NKO) (H.Á.).....	73
9. Polarized light pollution (POL) (H.G.)	82

Chapter III.

<i>Environmental radioactivity I. (Measurements using X-rays and gamma-radiation)</i>	109
10. The basic physical principles of radioactivity (H.Á., V.G.).....	111
11. Heavy metal contents determined by X-ray fluorescence analysis (NFS) (H.Á.)	118
12. Film dosimetry (FDO) (H.Á.)	124
13. Thermoluminescence dosimetry (PTL) (Cs.M.)	133
14. Gamma spectroscopy with scintillation detectors (NAI) (H.Á.).....	138
15. Radioactivity of natural soil and rock samples (TAU) (V.G.)	144
16. Annihilation radiation and positron emission tomography (PET) (V.G., Cs.M.)	155

Chapter IV.

<i>Environmental radioactivity II. (Particle radiations) (H.Á.)</i>	166
17. Tritium content of water samples (TRI).....	167
18. Investigation the Cherenkov-radiation from potassium beta-decay (CSE).....	177
19. Radon measurements of indoor air (LEV)	185
20. Measurement of radon concentration in water (RAD).....	192
21. Measurement of radon exhalation from rocks and soil samples (REX)	199
Summary	206

Chapter I.

MECHANICAL WAVES IN THE ENVIRONMENT

Most of the people live in cities in our modern life, in an area that lacks of the tranquility of nature and contrariwise it is full of traffic, industry and noisy machines even in the living space at home or in workplaces. The noise has been a significant factor only in workplaces but our life changed to the direction where anybody is exposed to this annoying environmental fact. High level noise can have health effect but low level noise can also make our concentration less or just disturb us in our everyday activities that will result in a less efficiency, wrong decisions or lower level of work. The noise is already regulated in several parts of the life. To understand the meaning of these regulations and to understand how we can reduce the risk we need to know the behaviour of the mechanical waves that is the physical substance of the noise.

However, the waves are present in more topics in environmental physics. The same phenomenon appears in the field of electromagnetic radiation that has several very different types from the radio waves to the ionizing gamma-radiation. There are a lot common in that radiation that can be described in mathematical form. Therefore this chapter begins with a short summary of the mathematics of the waves (that knowledge will be used later at the laboratory practices using electromagnetic and ionizing radiation, as well) and an introduction to acoustic waves. The two practices are the basic measurements of noise, acoustic waves that are audible and directly important for people to live. The other laboratory practice will investigate infrasound that has lower frequency than people can hear it. But infrasound also appears as a consequence of industrial works and some everyday machines can radiate this, too. Infrasound also has importance in simple environmental researches and in those procedures that monitor the environment or those processes that are helpful to understand what is going on Earth.

1. Mathematics of waves

1.1 Introduction

Waves are present in our everyday environment, in form of sound, noise, earthquakes, electro-smog (the field generated by network frequency electric devices) radio waves, mobile phones or any other kind of wireless communication, but also in form of light, heat, ultraviolet radiation, X-ray or CT machines, or even radioactivity. Thus it is a very important part of environmental physics to understand the nature of radiation and waves.

The question arises first, how one could mathematically formalize waves, or what are waves from a mathematical point of view. In the next subsections, we will review the general form of waves, the wave equation and its solutions.

1.1.1 Radiation and waves

Radiation is nothing else than wave propagation, we will see that any kind of radiation is equivalent to a wave. But what are waves? Let us take any spatially changing function, denoted by $f(x)$. Here f can be any physical quantity that may depend on space x . Examples might be the displacement of a string (on any bowed or plucked string instrument, e.g. a violin or a guitar), but also the water level in a bowl of water or in a lake or sea. An everyday example for such an f value would be the density of cars on a highway; a more physical example is the local density of air when sound is propagating in it. A high-school example is the density of spirals in a spring. Figure 1.1 shows images of the above examples.

If we have such a function $f(x)$, then it is clear that it is a time-varying function, note for example the water level in case of surface waves in a lake. Thus we may denote our function as $f_t(x)$, where the index t denotes the time. In the simplest case we have a permanent propagating wave. How does such a propagating wave work?

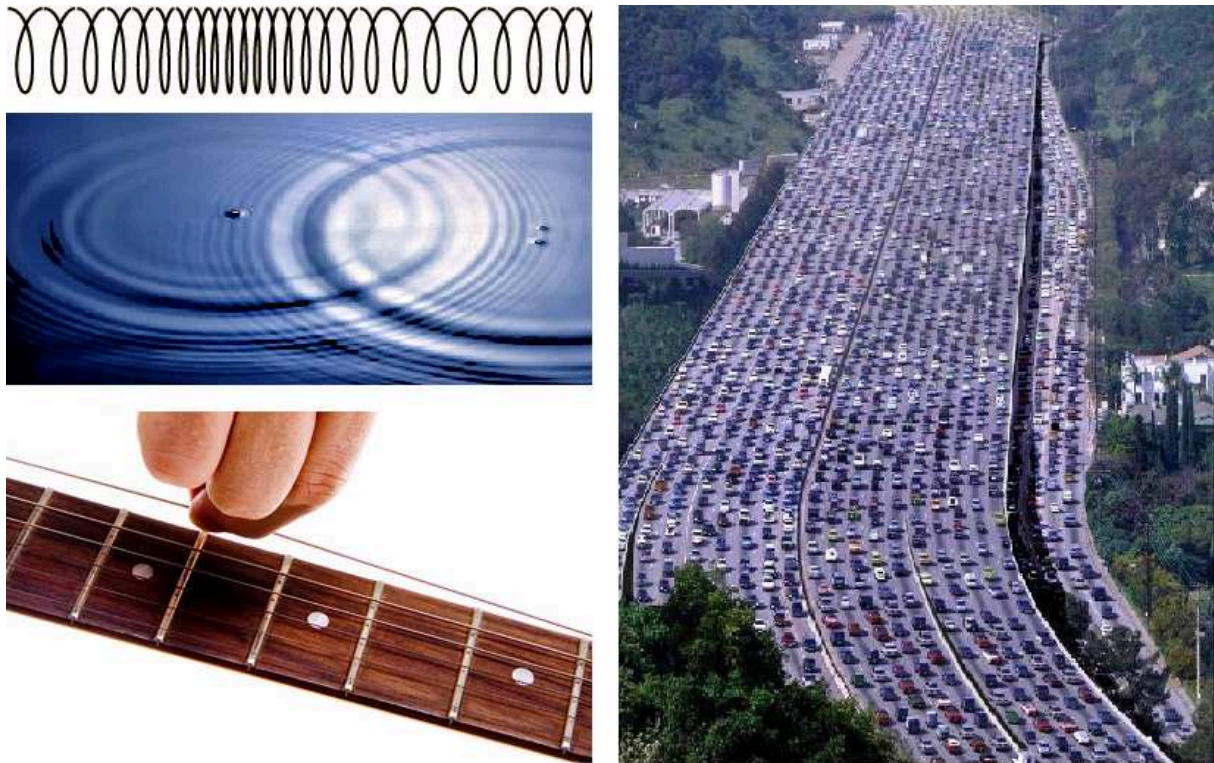


Figure 1.1. Examples of spatial functions that may behave as waves.

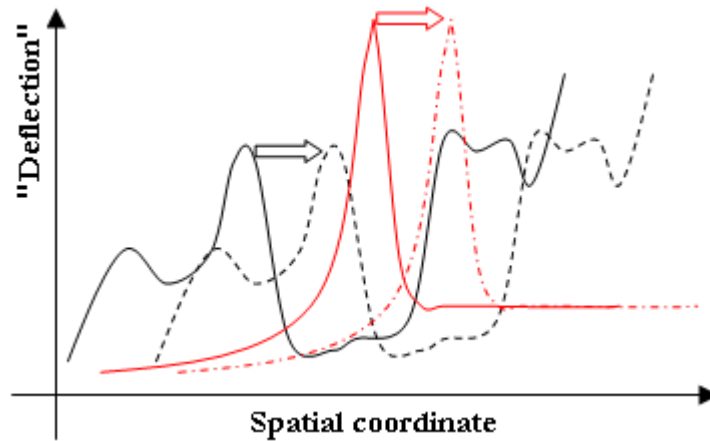


Figure 1.2. Examples of propagating waves. The dashed line represents the same physical quantity after some given time. Note that the two curves are just moved by a fixed distance.

Let us rely ourselves on the highway example now. The spatial coordinate x is then our location on the highway, while the value f is the density of cars at a given point. Let there be a traffic jam, i.e. let the car density be very large at a given section. The value of maximal density is unchanged however, just its location is moving away. This may be formulated mathematically as follows.

As noted above, the location of the maximal density depends on time. Then the following equality holds:

$$f_t(x) = f_{t+\Delta t}(x - \Delta x),$$

i.e. the density at time t and location x is the same as some Δt time later at a point behind by a distance of Δx . I.e. if the traffic jam now is here, then one hour (Δt) later it will be back by five kilometres (Δx). As the time is an equivalently important argument of the function, from now on we will note $f_t(x)$ as $f(t,x)$. This way, the above equation is modified as:

$$f(t + \Delta t, x) = f(t, x - \Delta x).$$

Figure 1.2 shows examples of propagating waves.

1.2 Basics of wave propagation

The equations in the previous sections are thus the basis of wave propagation. If, however, our function f depends on time and space not separately, but through a simple formula of $x-ct$, then the equations are automatically fulfilled. I.e. any $f(x-ct)$ or $g(x+ct)$ function describes a propagating wave. Let us check how this is possible. Substitute one of the above functions in the above equation (it works similarly for the other one):

$$\begin{aligned} f(t + \Delta t, x) = f(x - \Delta x, t) \text{ is transformed to} \\ f(x - c(t + \Delta t)) = f(x - \Delta x - ct), \end{aligned}$$

which is automatically fulfilled if

$$c\Delta t = \Delta x \quad \text{or} \quad c = \frac{\Delta x}{\Delta t}.$$

This means that the constant c here is the velocity of the wave propagation. As a summary, one might say that at a given time (let this be $t=0$) the density snapshot is $f(x)$, while at a time t it is $f(x-ct)$. Clearly, $g(x+ct)$ is a similar wave but with an oppositely directed velocity.

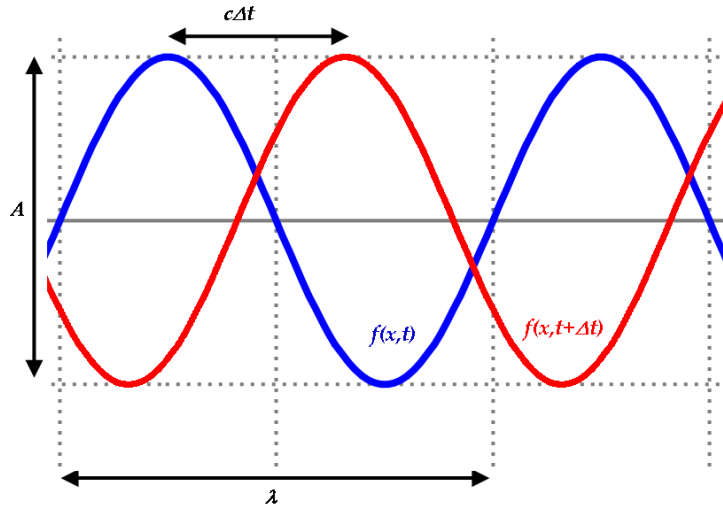


Figure 1.3. Examples of periodic and propagating waves. Clearly, after a time Δt a full period is reached if $\lambda = c\Delta t$. Thus the period is $T = \lambda/c$.

1.2.1 Periodic waves

We were talking about general waves until now. A usual wave, however, is periodic in space. If anyone would imagine a sea wave then it is definitely a periodic wave. Periodicity may be described mathematically as follows. A function f is periodic if a value λ exists such as

$$f(t, x) = f(t, x + n\lambda)$$

for any location x and integer n . If this is true, λ may be defined as the wavelength of the given periodic function or wave. It is interesting that if the above spatial periodicity holds, then there will also be a periodicity in time and vice versa. Periodicity in time would mean that a value T exists such as

$$f(t, x) = f(t + nT, x)$$

for any time t and integer n . According to the previous equation for propagating waves, for $\Delta x = \lambda$ and $\Delta t = T$ and thus $c = \Delta x / \Delta t$ i.e. $c = \lambda / T$ one gets:

$$f(x - \lambda - ct) = f(x - c(t + T)), \text{ thus, due to the spatial periodicity}$$

$$f(x - ct) = f(x - c(t + T)), \text{ i.e. there is a periodicity in time, } T$$

This means that if we have a spatial periodicity then we also have timely periodicity and the two are connected by $c = \lambda / T$, i.e. velocity is wavelength over period. [Figure 1.3](#) shows this behaviour. The periodic time or period T means that after this duration, the function (e.g. the density distribution) is exactly the same. One of such periods may be called an oscillation. We may also introduce the quantity called frequency. This is the number of oscillations happening during one second. Frequency is thus defined as $\nu = 1/T$, i.e. $\nu = c/\lambda$.

As usual mathematical functions we use are dimensionless; instead of $f(x-ct)$, one should use dimensionless quantities in the argument of the function (note that the logarithm or sine or exponential of 1 cm is meaningless). Thus we will use a constant k to remove the spatial dimension, i.e. from now on our wave will be described by the function $f(k(x-ct))$. Let us open the brackets, and define another quantity: $\omega = kc$. This way our wave will look like

$$f(t, x) = f(kx - \omega t).$$

Note furthermore that the origin (zero point) of time and space can be chosen arbitrarily, thus in general a phase φ may also be introduced, thus the general shape will be $f(kx - \omega t + \varphi)$. Then the basic constants and the connections between them will be:

$$k = \frac{\omega}{c}, \quad \lambda = Tc, \quad v = \frac{c}{\lambda},$$

The general wave shape is a sine wave with arbitrary amplitude. Thus the most general sine wave is

$$f(t, x) = A \sin(kx - \omega t + \varphi),$$

the wavelength of this sine function is 2π however, and that of $\sin(kx)$ is $2\pi/k$, thus here

$$\lambda = \frac{2\pi}{k}, \quad T = \frac{2\pi}{\omega}, \quad v = \frac{\omega}{2\pi}.$$

hence k is called the wave number. [Figure 1.3](#) shows these values in case of a periodic function.

1.2.2 Three dimensional plane waves

We were discussing the simplest case until now, where there is only one spatial dimension, and the propagating or oscillating quantity (the “wave”) is also a scalar quantity. Let us make one step further where there are three spatial dimensions. Let the coordinates be x , y and z . In this case our function will be modified to

$$f(t, x, y, z) = f(k_x x + k_y y + k_z z - \omega t + \varphi),$$

where k is now a vector, and its scalar product with the coordinate vector is in the argument of the function. If time is constant, then the function f is constant also if

$$k_x x + k_y y + k_z z = \text{const.}$$

This is true for planes orthogonal to the vector $k = (k_x, k_y, k_z)$ (note the rules of scalar product of vectors). Because of this, the above formula is called a plane-wave. The propagation velocity can be calculated exactly the same way as before, it will however be a vector quantity here:

$$\mathbf{c} = \frac{\omega \mathbf{k}}{|\mathbf{k}|^2}, \quad \text{or} \quad \mathbf{k} = \frac{\omega \mathbf{c}}{|\mathbf{c}|^2},$$

if both \mathbf{k} and \mathbf{c} are vectors. Because of this, \mathbf{k} is called the wave number vector.

We of course may define a new coordinate system, where the direction of the vectors \mathbf{k} and \mathbf{c} are in the direction of the the new x' axis. In this new coordinate system $\mathbf{k} = (k, 0, 0)$. [Figure 1.4](#). shows such a plane wave with a new coordinate system. It is clear from this that plane waves can be regarded as simple one-dimensional waves – everything is constant in the other y and z directions.

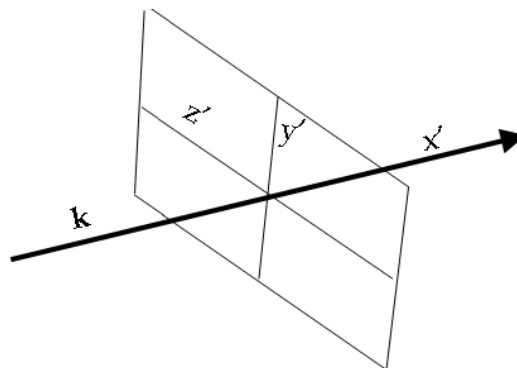


Figure 1.4. A plane wave and its coordinate system. The value of the wave is constant on the noted y' - z' plane, hence the name plane wave.

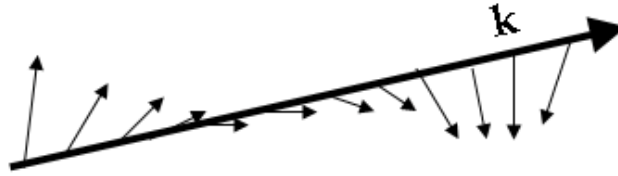


Figure 1.5. A circularly polarized vector wave. The vector is always orthogonal to the direction of propagation, and is rotating around this direction as axis.

1.2.3 Vector waves

Another complication may be if the value of the oscillating quantity is not a scalar (as in case of density or displacement waves) but a vector quantity. It is a little bit more difficult to understand but also a quantity like $\mathbf{E}=(E_x, E_y, E_z)$ may be oscillating or propagating like a wave. Even though this propagation might happen in three dimensions, in case of plane waves, only one dimension has to be regarded. Thus a simple sine wave of the vector quantity \mathbf{E} could be mathematically formulated as follows:

$$E_x = E_{x0} \sin(kx - \omega t + \varphi_x),$$

$$E_y = E_{y0} \sin(kx - \omega t + \varphi_y),$$

$$E_z = E_{z0} \sin(kx - \omega t + \varphi_z),$$

where of course the wave number k and also ω are independent of the direction (otherwise it would not be one wave but a superposition of several waves). The phase of each component might be different however. The value of each phase characterizes the spatial propagation of the vector. One simple example is, if

$$E_x = 0 \text{ (the wave is orthogonal to the direction of propagation)}$$

$$E_y = E_{y0} \sin(kx - \omega t),$$

$$E_z = E_{z0} \sin(kx - \omega t + \pi/2),$$

i.e. $E_{x0}=0$, $\varphi_y=0$ and $\varphi_z=\pi/2$. This is called a circularly polarized wave, as shown in Figure 1.5.

1.3 Oscillation

1.3.1 The harmonic oscillator

Before going into details of the origin of waves, let us discuss the more basic phenomenon of oscillations. Harmonic oscillators are present everywhere in our world: not only springs, clocks, antennas or radio sources contain them, but they are the basic building block of any material object: atoms or molecules in any solid objects are performing harmonic oscillations, and the temperature of the object corresponds to the strength of these oscillations. One could go even further in particle physics, but that is not the subject of present note.

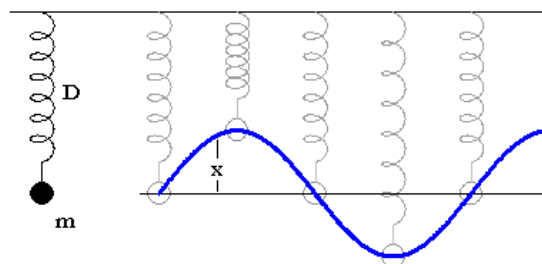


Figure 1.6. A simple harmonic oscillator: a body of mass m hanging on a spring

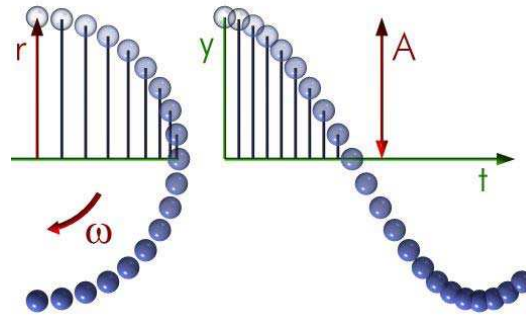


Figure 1.7. Illustration of the connection between harmonic oscillation and circular rotation

The simplest example of harmonic oscillators is a spring. The spring force, according to Hooke's law is proportional to the extension of the spring, i.e. $F = -Dx$ with D being the spring constant, while x is the extension. [Figure 1.6](#) shows such a system. According to Newton's law of motion, acceleration times mass is equal to the total force, thus

$$m\ddot{x} = -Dx, \text{ i.e.}$$

$$\ddot{x} = -\frac{D}{m}x,$$

with \ddot{x} being the acceleration and m the mass of the body moved by the spring. This is a differential equation, where a function $x(t)$ is searched the second timely derivative of which is proportional to the function itself. This differential equation can be solved and the result is:

$$x(t) = A\sin(\omega t + \varphi) \quad \text{with} \quad \omega = \sqrt{\frac{D}{m}},$$

where φ (the phase) and A (the amplitude) are arbitrary. Note that the parameter ω is called angular frequency, as the $x(t)$ function presented above is similar to the x coordinate of a circular motion. This is illustrated on [Figure 1.7](#). Thus the period of the motion is

$$T = 2\pi\sqrt{\frac{m}{D}}.$$

This is the harmonic oscillator, one of the most basic phenomena in physics. The reason for that is the following: differential equations describing nature are mostly second order linear equations. The solution of such equations is usually an exponentially increasing function or a harmonic oscillator – the former leads to infinitely large values is thus not very common in Nature. Mathematically, the negative sign in the right hand side of the equation of motion ensures the result to be an oscillation; this negative sign comes from negative feedback. Negative feedback thus causes oscillation, while positive feedback causes limitless increase, leading to a cut-off or a crash of the system.

1.3.2 Damped and driven harmonic oscillation, resonance

In the above discussed case the oscillations were happening just by themselves, and were going on forever. In more realistic cases, there might be damping or friction however. Both slow down the motion of the system, as they act with a force opposed to the direction of motion. This force is usually proportional to the velocity (at least in case of damping), thus the equation of motion is modified to:

$$m\ddot{x} + c\dot{x} + Dx = 0$$

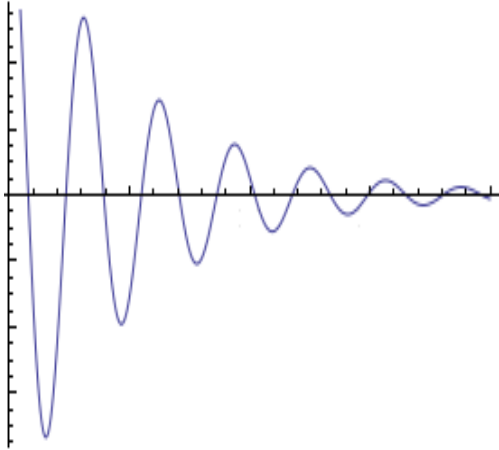


Figure 1.8. Curve of a damped oscillation. The frequency of the oscillation is approximately the same as in case of a simple harmonic oscillator, but the amplitude is decreasing exponentially.

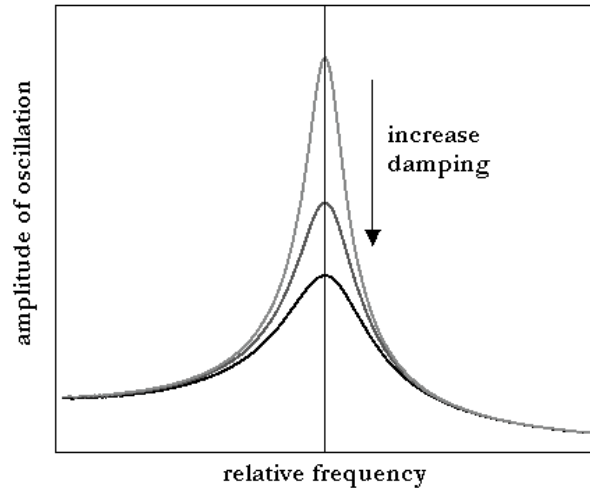


Figure 1.9. Output amplitudes in case of a driven oscillation. If the driving frequency is close to the resonance frequency (noted by a vertical line in the middle), the amplitude grows strongly. Without damping it grows beyond limits.

The solution of this differential equation is a so-called damped oscillation equation. If the damping is not very strong (which would stop the system from performing even one single oscillation), the motion is an oscillation with gradually decreasing amplitude. The frequency of the motion will be almost the same, but, as mentioned before, the amplitude will decrease exponentially with a function of $\exp(-zt)$ with z being a damping factor, being proportional to the constant of the damping force c . Figure 1.8. shows a damped oscillation curve.

In case of driven oscillations, besides the negative feedback force and the damping force, there is a driving force as well. One of the best examples is that of a playground swing, where the driving force is that of the person pushing the swing. Another example is when one tries to shake a tree, or when soldiers are marching over a bridge. In each case, the driving force is also a harmonic function of time. In these cases, resonance may occur. Resonance is the phenomenon where the amplitude of the system starts to grow, even though the driving force is of constant amplitude. It is beyond the limits of present note to discuss the details of resonance, but basic idea is the following. The driving force transmits packets of energy to the oscillator. The oscillator has an own typical frequency ($2\pi\sqrt{D/m}$ in case of a spring), this is called the resonance frequency. If the frequency of the driving force is near the resonance, the oscillator can absorb the energy packets almost completely. Thus the response of the system depends on the relative frequency of the driving, compared to the resonance frequency of the system. If the two are equal, response amplitude will be very large, even beyond any limits if there is no or very small damping – this is called the resonance catastrophe. Figure 1.9 shows response curves of driven oscillation.

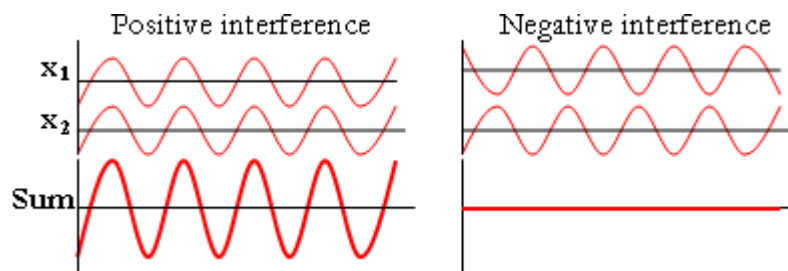


Figure 1.10. Illustration of positive and negative interference

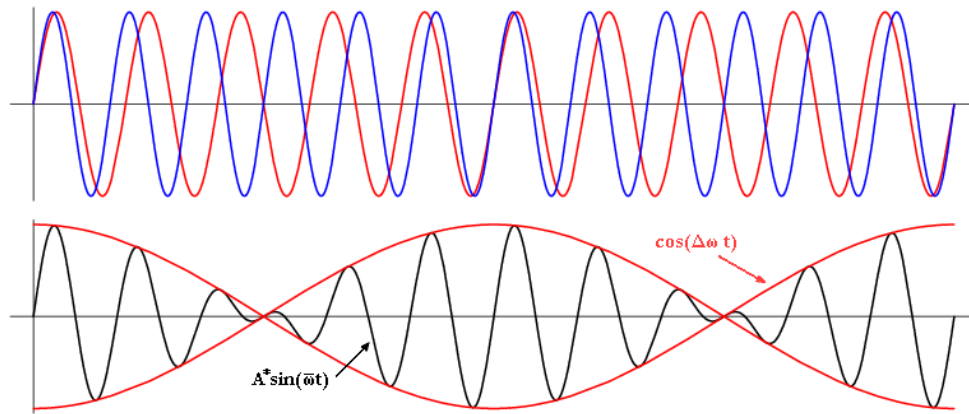


Figure 1.11. Illustration of beat in acoustics

1.3.3 Superposition of oscillations

It is important to discuss the superposition of oscillations. In case of springs this would mean that two independent springs are affecting one single object. More important is the case when two waves are superposed on each other: e.g. two water waves meet in a lake, or electromagnetic waves reach the same point. There are several possibilities for such a superposition. The simplest is if the two waves have the same frequency but might have a different amplitude or phase, i.e.

$$x_1(t) = A_1 \sin(\omega t + \varphi_1) \quad \text{and} \quad x_2(t) = A_2 \sin(\omega t + \varphi_2),$$

with x_1 and x_2 being the two oscillations. The superposition of these two waves is a similar one, $x(t) = A \sin(\omega t + \varphi)$, where the amplitude A and the phase φ can be calculated as follows:

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(\varphi_1 - \varphi_2)},$$

$$\varphi = \arctan\left(\frac{A_1 \sin \varphi_1 + A_2 \sin \varphi_2}{A_1 \cos \varphi_1 + A_2 \cos \varphi_2}\right).$$

We can immediately see from this that there is a maximum enhancement (or constructive interference) if $\varphi_1 = \varphi_2$, while there is destructive interference if $|\varphi_1 - \varphi_2| = \pi$. [Figure 1.10](#) illustrates these possibilities (with $A_1 = A_2$ there). There is a more complicated possibility, when the frequency of the oscillations or waves is different. Let us investigate such a case, with equal amplitudes and zero phases however. Here

$$x_1(t) = A \sin(\omega_1 t) \quad \text{and} \quad x_2(t) = A \sin(\omega_2 t).$$

The sum of these waves is the following:

$$x(t) = A^* \sin\left(\frac{\omega_1 + \omega_2}{2} t\right), \quad \text{where}$$

$$A^* = 2A \cos\left(\frac{\omega_1 - \omega_2}{2} t\right)$$

is the new, time dependent amplitude. This is interesting if the frequency of the two waves is almost the same, in which case the new amplitude is almost constant. Figure 1.11 shows this effect. The observed phenomenon in this case is called beat in acoustics. It is observed when two tones are close in pitch but not identical: the resulting tone has the average frequency (pitch), but the amplitude is oscillating very slowly. For example, if one tone has the frequency of 440 Hz and the other has 441 Hz, then the resulting tone will be of frequency

440.5 Hz, but the volume will oscillate with a frequency of 0.5 Hz, i.e. the period of volume oscillation will be two seconds. This effect is often used to tune instruments, as the vicinity of unison can be recognized by this beat effect.

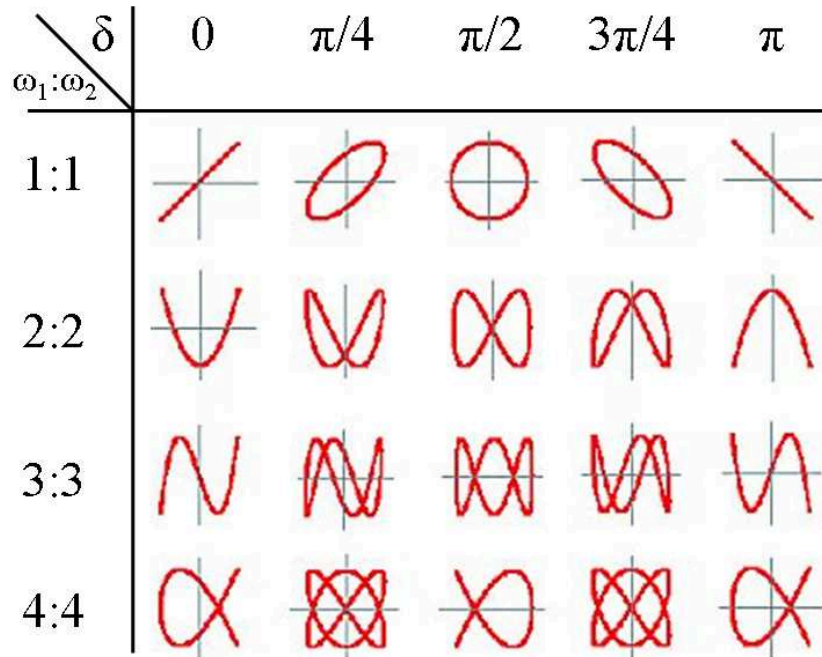


Figure 1.12. Superposition of two orthogonal oscillators with the same amplitude. Depending on the relative phase, the resulting shape is an ellipsis, a circle or a simple line. If the frequencies of the two oscillators are not equal, more complicated figures arise, the Lissajous curves.

Another important possibility of interference or superposition is when the two superimposed waves are orthogonal to each other, i.e. $x(t)=A\sin(\omega t)$ and $y(t)=B\sin(\omega t+\delta)$. The superposition of the two will produce a curve with the geometric equation of

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} - 2\frac{xy}{AB}\cos\delta = \sin^2\delta,$$

which corresponds to a rotated ellipsis in general. It is a non-rotated ellipsis if $\delta=\pi/2$ (or even a circle if $A=B$), and it is a straight line if $\delta=0$. Figure 1.12 shows more examples of superposition, also in the case if the frequency of the two oscillators is not the same.

1.4 Origins of the wave equation

One of the simplest waves can be made by a series of springs which are oscillating with a phase shift increasing with the distance measured from the first spring. Let us denote the expansion of each string by y , and the index of the springs by i , then the expansions as a function of time are:

$$y_i(t) = A\sin(\omega t - \varphi_i),$$

where the phase shift shall be proportional to the distance, i.e. $\varphi_i=kx_i$. Here k is the proportionality constant. If substituted back, and changing i index to x variable, we get a regular wave equation:

$$y(t, x) = A\sin(\omega t - kx).$$

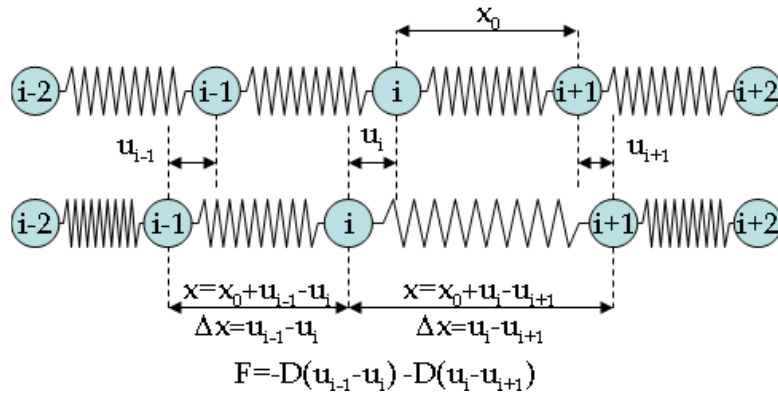


Figure 1.13. A simple system of springs and masses. The equation of motion of this system will be also the wave equation, as detailed in the text.

This is, however, a “fake” wave, since no information is propagating, just the phases of the springs show a wave-like behaviour. A simple but real example is that of series of springs. Let all springs with spring constant D be connected, with objects of mass m between them, indexed by i . The displacement of each mass will be denoted by u_i . The original distance of the masses is x_0 , thus the length of a string between the i^{th} and the $(i-1)^{\text{th}}$ mass is

$$x_{i,i-1}(t) = x_0(t) + u_{i-1}(t) - u_i(t) \quad \text{i.e.} \quad \Delta x_{i,i-1}(t) = u_{i-1}(t) - u_i(t).$$

The force acting upon the i^{th} mass is thus, from Hooke’s law:

$$F = -D \cdot [u_{i-1}(t) - u_i(t)] - D \cdot [u_i(t) - u_{i+1}(t)],$$

as two springs are connected to any single mass. Figure 1.13 shows this setup.

The original location of the masses is proportional to their index, $x_i = i \cdot x_0$. Thus instead of $u_i(t)$, we may write $u(t, x)$ as well. In this case, the spatial derivative of $u(t, x)$ may be written as

$$u'(t, x) = \frac{\partial u(t, x)}{\partial x} \approx \frac{u(t, x) - u(t, x - x_0)}{x_0} = \frac{u_i(t) - u_{i-1}(t)}{x_0},$$

(where the prime denotes spatial derivative) and similarly for the second spatial derivative:

$$u''(t, x) = \frac{\partial^2 u(t, x)}{\partial x^2} \approx \frac{u'(t, x) - u'(t, x - x_0)}{x_0} = \frac{[u_i(t) - u_{i-1}(t)] - [u_{i-1}(t) - u_{i-2}(t)]}{x_0^2}$$

i.e. the force from Hooke’s law is $F = Dx_0^2 \cdot u''(t, x)$

Thus the equation of motion will be, according to Newton’s law

$$m\ddot{u}(t, x) = Dx_0^2 \cdot u''(t, x) \quad \text{or alternatively} \quad \frac{\partial^2 u(t, x)}{\partial t^2} = \frac{Dx_0^2}{m} \frac{\partial^2 u(t, x)}{\partial x^2}.$$

This is the classic form of wave equation, a partial differential equation, generally written as

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2},$$

thus the wave propagation velocity in the above case will be $c^2 = Dx_0^2 / m$ or $c = \sqrt{Dx_0^2 / m}$.

One might imagine the waves propagating in this system: we displace one mass and then this displacement spreads over the whole system as a wave.

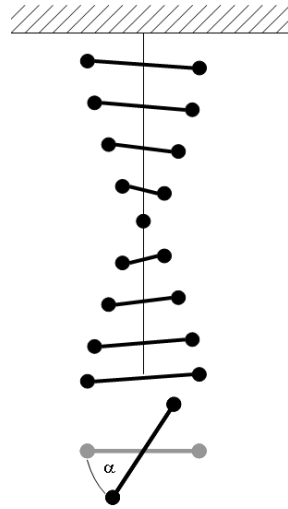


Figure 1.14. Bars fixed on a torsion wire. Torque is proportional to turning angle α .

Another example is when bars with weights are fixed on a torsion wire. The torsion wire behaves like a spring, except that torque (M) has to be used instead of force, turning angle (α) instead of expansion and moment of inertia (Θ) instead of mass. Newton's law of motion in case of circular motions is

$$M = \Theta \ddot{\alpha}$$

where $\ddot{\alpha}$ is the angular acceleration, the second time derivative of the angle. For a torsion wire, the torque is proportional to the angle, i.e.

$$M = G\alpha .$$

From here on, exactly the same derivation has to be performed as in the case of springs. The torque acting on one bar results from the relative angle of the one below it and the one above it, thus

$$M = -G \cdot [\alpha_{i-1}(t) - \alpha_i(t)] - G \cdot [\alpha_i(t) - \alpha_{i+1}(t)] .$$

If we assume that the bars are equally placed, and the height difference between two is h , then the location of the i^{th} bar is $x_i = i \cdot h$, thus again $\alpha_i(t)$ may be replaced by $\alpha(t, x)$. Then the above equation, similarly to the previous case, may be transformed to contain spatial derivatives:

$$M = Hh^2 \cdot \alpha''(t, x) .$$

The equation of motion will be finally

$$\Theta \ddot{\alpha}(t, x) = Gh^2 \cdot \alpha''(t, x) \quad \text{or alternatively} \quad \frac{\partial^2 \alpha}{\partial t^2} = \frac{Gh^2}{\Theta} \frac{\partial^2 \alpha}{\partial x^2} ,$$

and this is again a wave equation with wave propagation velocity $c^2 = Gh^2/\Theta$. The waves in this case are due to the torsion of the bars, i.e. this might be called a torsional wave.

Another example where the general wave equation turns out to be the governing principle of the motion of the system is the case of the continuity equation. Let us take a continuous medium with locally varying density n and velocity field v . An example for this might be when shockwaves are propagating in a medium after an explosion. We cannot detail the calculations, but the above statement can be formulated mathematically as follows. The time derivative of density plus the spatial derivative of density times flow velocity is zero, i.e.

$$\frac{\partial n}{\partial t} + \frac{\partial vn}{\partial x} = 0 .$$

In case of a constant flow, this means $\dot{n} = -vn'$. The spatial derivative of this gives $\dot{n}' = -vn''$, while from the rearranged version one gets $\dot{n}' = -\ddot{n}/v$. Putting it all together, the result is

$$\frac{\partial^2 n}{\partial t^2} = v^2 \frac{\partial^2 n}{\partial x^2}$$

This is again the general wave equation, and waves are propagating with the constant v velocity. Another example for such a system would be the number of cars on a highway, if all cars have the same velocity. This is the simplest form of a wave; a spatial function is just shifted in space as time elapses.

We have reviewed some systems the principal differential equation of which is a general wave equation. Now let us see how this partial differential equation can be solved, what kind of solutions we can find.

1.5 Solutions of the wave equation

Let our varying quantity be $f(t,x)$ (this might be a displacement, a turning angle, a density or any other physical quantity that depends on time and space). Then the general form of the wave equation (in one spatial dimension) is

$$\frac{\partial^2 f}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial x^2}.$$

In order to find a solution of the above equation we have to introduce new variables. This will be $a=x+ct$ and $b=x-ct$. Clearly, $\partial a/\partial x=1$, $\partial b/\partial x=1$, $\partial a/\partial t=c$, $\partial b/\partial t=-c$. Thus with these new variables, the first derivatives may be expressed, using the chain rule of composite functions

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{\partial f}{\partial a} \frac{\partial a}{\partial t} + \frac{\partial f}{\partial b} \frac{\partial b}{\partial t} = \frac{\partial f}{\partial a} + \frac{\partial f}{\partial b}, \\ \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial a} \frac{\partial a}{\partial x} + \frac{\partial f}{\partial b} \frac{\partial b}{\partial x} = c \frac{\partial f}{\partial x} - c \frac{\partial f}{\partial x}. \end{aligned}$$

The second derivatives can be calculated in a similar way, which we do not detail here, but the result is (try to double check it):

$$\begin{aligned} \frac{\partial^2 f}{\partial t^2} &= \frac{\partial^2 f}{\partial a^2} + 2 \frac{\partial^2 f}{\partial a \partial b} + \frac{\partial^2 f}{\partial b^2}, \\ \frac{\partial^2 f}{\partial x^2} &= c^2 \frac{\partial^2 f}{\partial a^2} - 2c^2 \frac{\partial^2 f}{\partial a \partial b} + c^2 \frac{\partial^2 f}{\partial b^2}. \end{aligned}$$

Now this might be substituted into the wave equation, and the result:

$$\frac{\partial^2 f}{\partial a \partial b} = 0$$

There is a very general solution for this equation:

$$f(a,b) = F(a) + G(b)$$

with $F(a)$ and $G(b)$ being arbitrary functions. In terms of time and space, the general solution is then the following:

$$f(t,x) = F(x - ct) + G(x + ct).$$

Thus we have found a general solution to the wave equation. It may be not surprising, but this form agrees with the functions mentioned as general forms of waves at the very beginning of [Subsection 1.2](#). Function F is an advanced wave (going forward in space) while G is a retarded wave (going backward in space). As also mentioned in 1.2.1, our functions F and G have to be dimensionless. In order to have that, we shall introduce the quantity k with dimension m^{-1} . Let us also introduce $\omega=kc$, and then our function will be

$$f(t, x) = F(kx - \omega t) + G(kx + \omega t)$$

Everything we discussed before was in case of one spatial dimensions. However, in reality we have to deal with more than one dimension, usually three. Let us see how we can use the above result in three dimensions.

1.5.1 Plane waves

Let the spatial coordinates be x, y and z . Then the differential equation is modified to

$$\frac{\partial^2 f}{\partial t^2} = c^2 \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \right) = c^2 \Delta f$$

where the common Laplace differential operator or Laplacian

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

was used. A similar derivation can be performed as above, and the result is the same, but our new functions F and G will also be defined in three dimensions (readers are encouraged to double check if these functions fulfill the above three dimensional wave equation:

$$\begin{aligned} F(kx - \omega t) &= F(k_x x + k_y y + k_z z - \omega t), \\ G(kx + \omega t) &= G(k_x x + k_y y + k_z z + \omega t), \end{aligned}$$

where $k=(k_x, k_y, k_z)$ is now a vector with $c^2 k^2 = \omega^2$. The scalar product of k with the coordinate vector (x, y, z) is in the argument of the function. At a given time F and G are constant also if

$$k_x x + k_y y + k_z z = \text{const.}$$

This is true for planes orthogonal to k , as discussed in 1.2.2 also. Thus the symmetry property. As mentioned there as well, a suitable coordinate system has to be chosen where $k=(k,0,0)$, and then we are back at the one dimensional case.

1.5.2 Spherical waves

There are also waves originating from point-like sources, or one should rather say that almost all waves are originating from point like sources (e.g. an earthquake, a light bulb, an antenna or an explosion). For such waves, the above plane wave solution is not suitable. The symmetry property of these waves is not that they would be constant on parallel planes, but they are constant on concentric spheres, with the source being in the center. In order to find such solutions, one has to derive the spherical form of the wave equation. This means that one has to use spherical coordinates of radius r , azimuth angle φ and inclination θ . The coordinate transformation from (x, y, z) to (r, φ, θ) can be written as

$$\begin{aligned} x &= r \sin \theta \cos \varphi \\ y &= r \sin \theta \sin \varphi \\ z &= r \cos \theta \end{aligned}$$

The symmetry property is that the function depends only on r , but not on φ or θ .

It is beyond the scope of present note to discuss the general wave equation in spherical coordinates. However, if the above symmetry property holds, then the derivatives with respect to φ or θ vanish. This makes the wave equation quite simple:

$$\frac{\partial^2 (rf)}{\partial t^2} = c^2 \frac{\partial^2 (rf)}{\partial r^2}$$

i.e. it is a simple one dimensional wave equation for r times f . The solution for this is well known: $rf = F(kr - \omega t) + G(kr + \omega t)$. Thus the solution for $f(t, r)$ is:

$$f(t, r) = \frac{F(kr - \omega t)}{r} + \frac{G(kr + \omega t)}{r}.$$

Thus spherical waves look exactly like plane waves, but their amplitude is decreasing inversely proportional with the distance from the center. This means that in case of a compression wave, the amplitude of the density oscillations is ten times smaller if the distance from the center is ten times higher. We will see later that the intensity of a wave is usually proportional to the square of the oscillation amplitude, thus wave intensity of spherical waves from point like sources usually decreases with the squared distance from the center.

1.6 Fourier series

We have seen that solutions of the different forms of wave equations contain arbitrary functions. As noted in 1.2.1, waves are usually periodic functions, thus in reality, the arbitrary functions F and G should be also periodic functions. However, in mathematics, there is a very important theorem about periodic functions about the existence of Fourier series of such functions. A Fourier series is a decomposition of any periodic function into the sum of an infinite set of simple oscillating functions, in particular sine and cosine functions. For the sake of simplicity, let us assume F has the period length (wavelength) of 2π . In this case, it can be decomposed as

$$F(x) = \sum_{n=0}^{\infty} [a_n \cos(nx) + b_n \sin(nx)],$$

where the coefficients a_n and b_n can be calculated as

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} F(x) \cos(nx),$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} F(x) \sin(nx).$$

[Figure 1.15](#) shows an example. The function used there is a periodic function which is 1 if $-\pi < x < 0$ and -1 if $0 < x < \pi$. This can easily be decomposed into Fourier harmonics, we do not give the functions of the series but show the progression of the approximation in the mentioned figure.

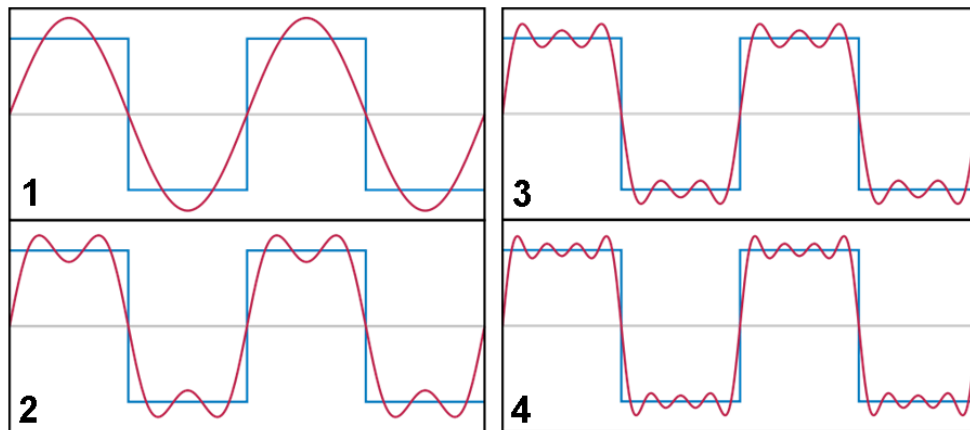


Figure 1.15. Fourier series of a periodic function. After four approximations the sum converges significantly towards the function, i.e. the difference between the sum and the original function is not very large.

This means that even though the general solution of the wave equation contains arbitrary functions, we have to deal with the harmonics only. Or otherwise, any other periodic function may be decomposed into the sum of harmonics, thus we will deal only with these harmonic components. In case of plane waves, the general solution is then

$$f(t, x, y, z) = A \sin(kx - \omega t)$$

with k being the length of the wave number vector pointing in direction x (we have chosen the coordinate system according to the direction of the wave propagation). For spherical waves the general solution is

$$f(t, r) = \frac{A}{r} \sin(kr - \omega t),$$

where k is the wave number, a scalar. In both cases, the wave number and the frequency are arbitrary, but they are connected with the so-called dispersion relation $kc = \omega$ with c being the wave propagation velocity.

Real solutions add up from many harmonic components of the above form, each with a different frequency. The spectrum of a radiation or wave is then the amplitude of all harmonic components as a function of their frequency.

It is important to know that a similar theorem exists for non-periodic functions as well. For them the theorem states that for any function $F(x)$ there exists an unambiguous function $\hat{F}(k)$ the role of which is similar to the Fourier-coefficients a_n and b_n mentioned above in case of periodic functions. Instead of a sum then an integral has to be used, and $\hat{F}(k)$ is called the Fourier spectrum of the function $F(x)$. If the function is periodic, then the previous coefficients are retained, i.e. the Fourier spectrum is discrete and has non-zero values only in case of integer arguments. It is again beyond the scope of present note to go into details, but the readers are encouraged to study Fourier-transforms in textbooks or also on the web.

The important message of this section is that regular waves or radiations can be decomposed to a sum of harmonic functions, and the amplitude of the components as a function of their frequency is the spectrum of the radiation. An important consequence is also that one has to deal with harmonic functions only as any other function can be decomposed to them in form of a Fourier series.

2. Mechanical waves

2.1 Introduction

In the previous section the mathematics of waves was discussed. Now let us see one of the main applications of this, the physics of mechanical waves. In case of mechanical waves the particles of a medium displace, thus these are called displacement waves. In most cases this entails the change in density as well, in which case we are talking about compression waves. In both cases, the wave is basically a disturbance moving through the medium.

There are three main types of mechanical waves: transverse, longitudinal and surface waves. In the next parts of this subsection we will detail important examples of waves.

2.1.1 Surface waves

Let us discuss surface waves first. Surface waves propagate along the surface between two media, usually a fluid and a gas. It turns out that in case of surface waves the motion of particles of the surface describes a circle: they are moving up and down but also forward and backward. A typical example is that of ocean waves. Passive drifting of objects on the waves is also due to this circular motion – as the drifting object floats on top of the water, it always feels only one direction of motion from the circular movement. The most important property of surface waves is their velocity. In deep water the wave velocity c is

$$c = \sqrt{\frac{g\lambda}{2\pi} + \frac{2\pi\alpha}{\rho\lambda}}$$

with λ being the wavelength, α the surface tension of water, ρ the density and g the gravitational acceleration constant. In shallow water however, the physics of surface waves is different, and there the wave propagation velocity is

$$c = \sqrt{\frac{g\lambda}{2\pi} \tanh\left(\frac{2\pi d}{\lambda}\right)} \approx \sqrt{gd}$$

in case of $d \ll \lambda$, where d is the depth of water. Clearly, in both cases, wave velocity depends strongly on wavelength. Usually fast wave corresponds to long wavelengths, thus in case of a tsunami or a storm, the waves that first arrive on the coast are those with largest wavelengths.

2.1.2 Longitudinal and transverse waves

In case of transverse waves, the motion of the particles is perpendicular to the direction of wave propagation. An example for this is an oscillating string of a musical instrument, or the so-called audience wave (sometimes called Mexican wave) formed in a stadium when successive groups of audience briefly raise their arms or stand up. In case of longitudinal waves, the vibration of medium particles is parallel to the wave propagation direction. This is the type of wave when the density of the medium varies in time and locally, thus longitudinal mechanical waves may be also referred to as compression waves. Everyday examples of compression waves include sound waves and some type of seismic waves too.

2.1.3 Speed of mechanical waves

In gases and fluids, mechanical waves are generally longitudinal waves due to the weakness of shear forces compared to forces originating from pressure differences. (There are exceptions like the great atmospheric waves.) In solid state objects however, transverse waves may also appear. In gases or fluids however, we are always referring to compression waves,

which can also be referred to as pressure waves as with oscillation of density comes the oscillation of pressure as well. We will discuss this for the case of acoustic waves. Let us note here that the velocity of pressure waves (which may also be called sound velocity, as these waves are sound waves basically), in case of small density fluctuations, can be calculated according to the simple law

$$c = \sqrt{\frac{\partial p}{\partial \rho}}$$

i.e. the partial derivative of pressure with respect to density. This equation can be transformed to contain the bulk modulus or incompressibility. Its definition is (using volume V)

$$K = -V \frac{\partial p}{\partial V},$$

which is positive as pressure increases if the volume decreases. Note that the volume is inversely proportional to the density. In case of a gas, the bulk modulus is approximately given as $K=\gamma p$, with γ being the so-called adiabatic index of the gas (same as the heat capacity ratio or the isentropic expansion factor). Using this, the Newton-Laplace equation gives the velocity as a function of the bulk modulus, and thus as a function of the adiabatic index:

$$c = \sqrt{\frac{K}{\rho}} = \sqrt{\gamma \frac{p}{\rho}}.$$

Note that for ideal gases, $pM=\rho RT$, where M is the molecular mass, R is the molar gas constant (8.314 J/mol·K) and T is the temperature in Kelvin. Thus the above formula may be written as

$$c = \sqrt{\frac{K}{\rho}} = \sqrt{\gamma \frac{RT}{M}}$$

In air, velocity of longitudinal or pressure waves (sound velocity) is experimentally well known. At 0 °C temperature its value is 331 m/s, while for other temperatures, the experimental formula of

$$c = 331 \text{ m/s} \cdot \sqrt{\frac{T}{273.15 \text{ K}}}$$

holds, where T is the temperature in Kelvin. At 20 °C temperature (293.15 in Kelvin) the speed of sound is 343 m/s, at 30 °C it is 349 m/s. This is clearly compatible with the above formula, with $\gamma=1.4$ and $M=29$ g/mol for air (double check this result, i.e. substitute the above values into the formula with the adiabatic index and the temperature). Note that his formula will fail, however for very large temperatures, as the adiabatic index is also weakly depending on temperature (at 1000 °C the adiabatic index goes down to $\gamma=1.365$, which is quite near 1.4).

The speed of sound in different gases depends on their adiabatic index (which does not vary very strongly, typical values are between 1.2 and 1.6 for almost any gas at any temperature), but more importantly it depends on the molar mass. The molar mass of air was roughly 29 g/mol, while that of hydrogen gas is only 2 g/mol, thus speed of sound in hydrogen is much larger, as shown above. It may reach values of approximately 1300 m/s. Note also that for monatomic gases, the adiabatic index is about 20% larger, thus the speed of sound in a monoatomic gas is 10% larger compared to a non-monoatomic gas having the same molar mass (c.f. CO₂ and Ar).

In liquids, the speed of sound may also be computed according to the Newton-Laplace equation (the version with the bulk modulus). In fresh water at 25 °C, the speed of sound is roughly 1500 m/s, i.e. much larger than in air. It depends however, on pressure (which depends on depth), salinity and temperature (depending also on depth). This is a very important topic for sonar's and acoustic undersea communication.

In solids longitudinal and transverse waves may travel too. Their velocity is different also. The velocity of longitudinal waves (sound waves, generating volumetric deformations) can be expressed with bulk modulus K and shear modulus G as

$$c = \sqrt{\frac{K + \frac{4}{3}G}{\rho}},$$

whereas the velocity of transverse waves (waves generating shear deformations) is

$$c = \sqrt{\frac{G}{\rho}}.$$

Clearly, longitudinal waves are thus faster, while both are proportional to the stiffness of the material. Sound waves in steels have a velocity around 5-6000 m/s, in diamond it is 12000 m/s, while only a little bit more than 1000 m/s in lead.

2.1.4 Seismic waves

The most important example of waves in the earth are seismic waves, generated by earthquakes. In case of an earthquake, there are both transverse and longitudinal waves. The latter are quite faster, and they are called P-waves (primary waves), as they arrive first at a given point. Transverse waves are slower, they are called S-waves (secondary waves), and they arrive at a given point later than P-waves.. Typical velocities for a P-wave are between 5-8 km/s, while S-waves are about 60% of that roughly. Time difference is usually on the order of a minute. Figure 2.1 shows the velocities of S- and P-waves in different layers of the Earth. Figure 2.2 shows the waves originating from an example earthquake, and how they are propagating through the different layers of the Earth. [Figure 2.3](#) shows a seismogram of an earthquake as measured by a seismograph (velocities of ground motion are plotted on the vertical axis).

This velocity difference is a very important tool for geologists, as from the time difference of arrival of the different waves, the exact focus of the earthquake can be located. To determine the location with a high accuracy, several measurement points have to be used.

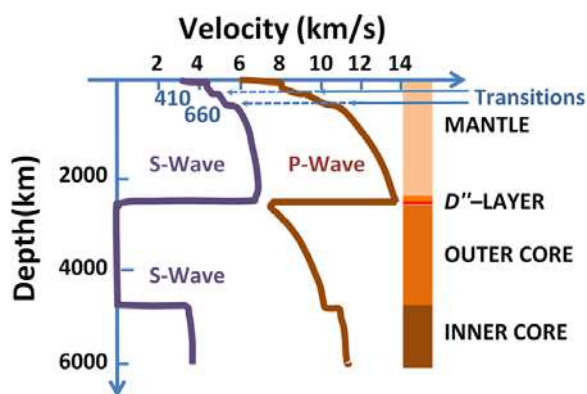


Figure 2.1. Velocity of seismic waves in several different layers of the Earth, as a function of depth (image from Wikipedia).

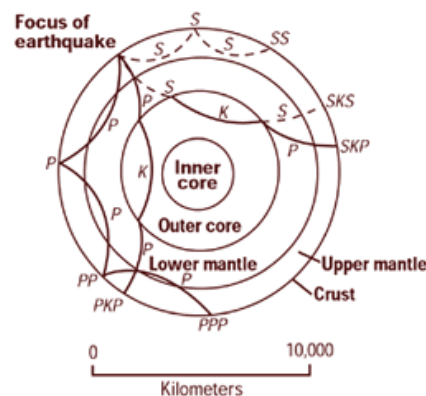


Figure 2.2. Different waves originating from an earthquake are propagating through the Earth (image from Wikipedia)

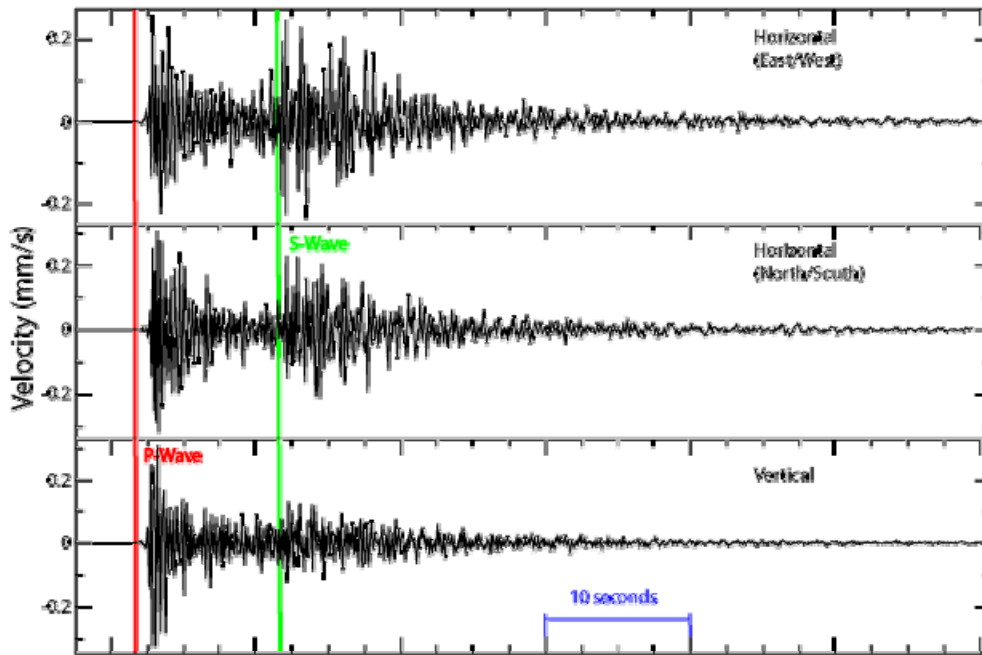


Figure 2.3. Seismogram showing S-waves and P-waves

2.2 Acoustic waves

Sound is a series of acoustic waves, waves of pressure propagating through a compressible medium such as water or a gas. During its propagation the particles of the medium are oscillating, and due to this oscillation, pressure and density are also fluctuating as a function of spatial coordinates and time. Let us see how these acoustic waves can be described with simple physical principles, using what we know about waves.

2.2.1 Physics of sound waves

Let the displacement of a given particle described by the following wave:

$$u = u_0 \sin(\omega t - kx)$$

where u_0 is the maximal displacement of particles. Clearly, velocity of these particles is the time derivative, thus

$$v = \omega u_0 \sin(\omega t - kx).$$

The local pressure of the medium can then be calculated:

$$p = \rho c \omega u_0 \sin(\omega t - kx)$$

where ρ is the density of the medium. Intensity of the sound is the energy flowing through a given area, or alternatively pressure times velocity. The average intensity is half of the maximum, and can be calculated as:

$$I = \frac{1}{2} p_{\max} v_{\max} = \frac{1}{2} \rho c \omega^2 u_0^2 = \frac{1}{2} \frac{p_{\max}^2}{\rho c}.$$

i.e. intensity is proportional to the square of pressure.

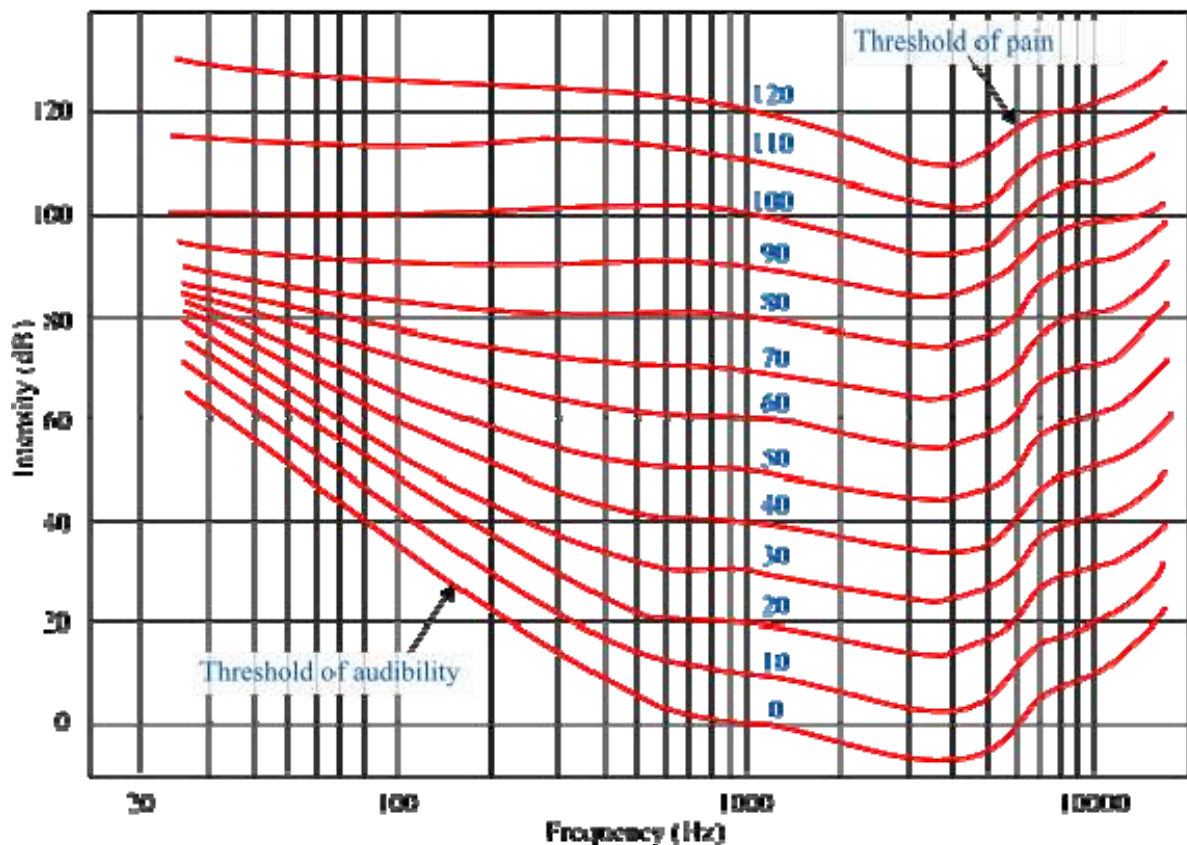


Figure 2.4. Curves of constant phon values as a function of frequency. Phon and decibel are equal at 1 kHz, whereas 0 phon corresponds to more than 60 dB at 20 kHz.

The important physical quantities of sound are thus pressure or intensity. Note, however, that our perception of sound is logarithmic, i.e. if a ten times stronger sound feels louder by one unit, then a hundred times stronger sound feels louder by two units. Thus a quantity reflecting this logarithmic perception has to be introduced. The name of this new quantity is sound pressure level, and the unit is decibel or dB (after A. G. Bell), and its definition is:

$$d = 20 \cdot \log_{10} \frac{p}{p_0} = 10 \cdot \log_{10} \frac{I}{I_0}$$

because pressure and intensity are squared proportional. Here I_0 and p_0 are introduced to make the argument of the logarithm dimensionless. They are chosen to reflect the limit of audibility ($I_0=10^{-12}$ W/m² or $p_0=2 \times 10^{-5}$ Pa at 1 kHz), and the corresponding decibel value is 0 dB. A ten times more intensive sound is of 10 dB then, while a sound with ten times higher pressure amplitude is of 20 dB. Note that the constants in the definition reflect the audibility threshold only at 1 kHz, thus a 0 dB sound might not even be heard if it is at a different frequency. Thus a unit called phon is also used sometimes, where the dividing constant I_0 is the audibility threshold for the given frequency. In that case, 0 phon is always the quietest audible sound. Figure 2.4 shows constant phon curves as a function of frequency. Clearly the audibility threshold is at 0 dB at 1 kHz, but at more than 60 dB at 20 Hz. The pain threshold is also plotted on the figure, this is the threshold where the sound causes already pain in the auditory system. At 1 kHz, this corresponds to 120 dB.

More details about this subject are given in the description of the lab course “Environmental Noises”.

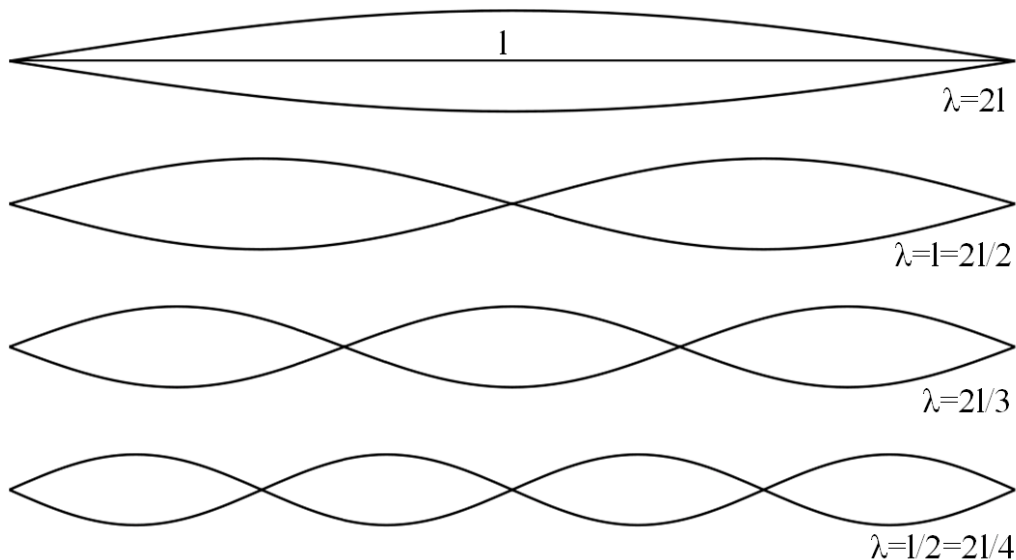


Figure 2.5. Vibrations of a string. Possible wavelengths can be described as $2l/n$.

2.2.2 Sources of sound, resonators

Sources of sound are usually vibrating objects, such as a string, a drum surface, a vocal chord or the surface of loudspeaker. These objects were hit by a perturbation or are moved by a driving force and the frequency of the vibration is the resonance frequency of the given object. Usually, however, they have many resonance frequencies thus the vibration can be described by an extended spectrum (see [Section 1.6](#)). The simplest sound source is a string or a chord. Frequencies produced by strings are, each representing a different vibrating mode,

$$f_n = \frac{nc}{2l} \quad \text{i.e.} \quad \lambda_n = \frac{2l}{n}$$

where l is the length of the string, c is the wave velocity on the string and n is an arbitrary number. The main sounded frequency is the first one, $f=c/2l$, the others are its overtones. Figure 2.5 shows details of this. It is interesting furthermore that the wave velocity can be calculated as

$$c = \sqrt{\frac{F}{A\rho}}$$

where F is the stress force acting on the string, A is its cross-section area and ρ its density. Clearly the frequency of a string can thus be enhanced by increasing its tension – this is how musical instruments are toned.

Source of sound may be also a membrane, as in case of drums or loudspeakers. The vibrating modes of a membrane are significantly different from those of a string. Ernst Chladni, a German physicist and musician discovered the law describing the frequency of vibrating modes of membranes. We do not detail this here, but readers are encouraged to look up “Chladni figures” on the Internet. [Figure 2.6](#) shows a simple example.

Besides the vibrating object, a resonator is also needed, as this will amplify the sound (enhance the sound level). The vibrating string or surface produces a spectrum of vibrations. The resonator is then driven by this vibration, i.e. it behaves as a driven oscillator (see [Section 1.3.2](#)). The resonator may be the body of a guitar or a violin, the mouth cavity (in case of speaking) or the enclosure of a loudspeaker. These objects have a range of resonance frequencies, and if the range of driving frequencies (the vibration of the sound source)

coincides with it (or the two ranges have common frequencies in them), resonance occurs and the sound level is significantly enhanced.

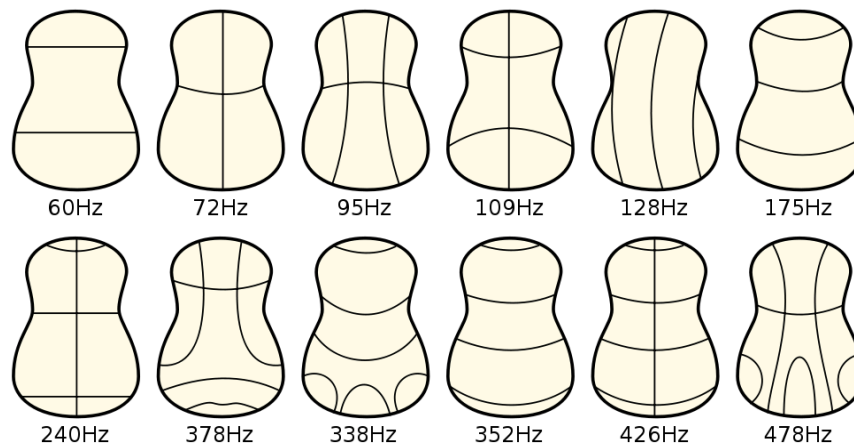


Figure 2.6. Chladni vibration modes of a guitar plate. Image from Wikipedia.

The specific tone of a given musical instrument or of a person is determined by the resonance frequencies of the resonator, i.e. the body of the instrument or the mouth cavity of the person. On a string for example, all f_n frequencies are sounded, but the resonator enhances them differently, shaping thus the tone of the instrument.

In case of wind instruments, air enters into a flow-control valve attached to a resonant chamber, usually a cylindrical pipe opened at the far end. The resonance frequencies can be easily calculated here, we just quote the result here without details:

$$\lambda_n = \frac{4l}{(2n-1)} \quad \text{i.e.} \quad f_n = \frac{c \cdot (2n-1)}{4l}$$

i.e. the wavelength of the principle tone is $4l$, the next one is $4l/3$, and so on. Note that the frequency depends here on the speed of sound as well. This depends on the temperature of air in case of regular musical instruments, or on the medium itself (note the “helium speech”).

2.2.3 Sound propagation

Sound is propagating in medium like a regular wave. It is reflected or absorbed by different surfaces; this is of crucial importance in noise protection, see description of the lab course “Environmental Noises”. Acoustic absorption happens if a given material transforms the acoustic energy of a sound wave into another form, usually heat. It is a method of sound attenuation, where the amplitude of the wave is reduced as a function of distance through the medium. Attenuation of intensity is generally described by the Beer-Lambert law (it is originally used in case of light, but holds for almost any type of attenuation):

$$I(x) = I_0 \exp(-\alpha x)$$

where I_0 is the original intensity (when entering the medium), x is the distance covered in the medium, and α is the attenuation coefficient. This coefficient usually depends on the frequency even in case of a given medium. In case of ultrasound absorption in the human body at 1 MHz frequency, α values sweep the range from 0.002 (water) through 1 (muscle) to 10 (bones). Absorption is sometimes given also as a ratio of absorbed intensity to incident intensity, i.e. $a = I_{\text{absorbed}}/I_0$. Typical values for the absorption coefficient a in this case are: 0.01 (3 mm plywood panel), 0.1 (6 mm cork sheet), 0.3 (12 mm hardwood), 0.6 (100 mm mineral wool).

Sound wave can also be bent around slits, openings or edges, this is called diffraction. It is governed by the Huygens–Fresnel principle: propagation of a wave can be described considering every point on a wave-front as a point-like source for a secondary spherical wave. In this note we do not detail this principle or wave diffraction but encourage the reader to look it up in physics textbooks or on the Internet.

3. Investigation of environmental noise (ZAJ)

3.1 Sound and noise

Sound and noise are of the same age as the human race, they were there always along our civilization and development. Speech and music are among the most important aspects of human life, but the unpleasant feeling of noise was also always part of human communities. The indispensable need for silence to pursue creative acts often comes up already in the thoughts of Greek and Roman philosophers. Noise in this regard can be defined as any unwanted or disturbing sound.

The world around us is more and more noisy. The industry requires more and more energy, more and more powerful and noisy machines; the development of transportation causes a large increase in the number and velocity of vehicles. To save time and trouble, there are more and more noisy machines in our homes as well. The classic and folk music of our predecessors was replaced by the loud pop, rock and metal music, which are amplified by enthusiasts by even more powerful amplifiers.

The urbanization corresponding to our modern age caused a sudden increase in the noise-load of the urban population; this is mainly caused by large traffic and noises from the neighborhood in large residential areas. In taller buildings or offices the elevators and the ventilating devices represent new noise-sources.

Many regard noise as a necessary “accessory” of civilization, which one cannot, therefore should not fight. Others think that noise in our lives is beyond endurance, thus the available noise reduction is not enough, and the noise sources themselves should be eliminated from our surroundings.

Clearly, noise as phenomenon has to be dealt with, as it is an important condition of our environmental state. Noise levels can be kept at a level at which healthy people can pursue an untroubled lifestyle. We have to know the technical and legal resources by which – with due technical knowledge and appropriate behaviour – the noise sources working around us can be reduced to such levels. First of all, however, we have to acquaint ourselves with the nature of sound.

For environmental physics, sound is such a mechanical vibration (wave), which is propagating in a flexible or elastic medium and it reaches the human ear, causing a sensation of sound. Mechanical vibrations here mean that particles of the given medium (mainly air, but might be water or carbon dioxide as well) are in motion. Alternatively, density of the medium is changing as a function of position and time. Change of density in air means also the change in pressure, and this is what causes the sensation in our brain, reaching our eardrum and mediated by our hearing nerves. As for the sensation, person-by-person variability is quite large however. Thus the same phenomenon might cause a different sensation for different people. There is a large uncertainty connected to the whole issue; for advancement an objective investigation (i.e. measurement) of sound or noise strength is needed.

In our days many areas from traffic industry (cars, airplanes, noises inside and outside these) through architecture (noise load of homes, offices) to energetics (power plants, transformers, and simple electronical devices) are concerned by noise and its measurement. Acoustic aspects have to be respected at the construction of almost any building, and the radiation principle ALARA (as low as reasonable achievable, i.e. reaching the lowest possible radiation level that makes meaningful activity still possible) makes its way into noise management. Thus we definitely have to know some simple theoretical and experimental facts and relations of the physics of sound.

The goals of present lab course are to recall our basic physical knowledge in the field of acoustics, and, by some simple measurements, to gain introspection into the challenge of measuring sound parameters. The tasks to perform are assembled such that we get to know a simple handheld sound measuring device. We will follow the limits of our measuring possibilities, and by asking and answering some interesting questions we will be more acquainted with the not in the least easy field of acoustics.

3.2 Introduction to the basics of acoustics

3.2.1 Amplitude of sound pressure, sound intensity, decibel

Sound strength depends on the pressure fluctuations of the medium, the amplitudes of the pressure waves. Latter is called the pressure amplitude of sound. Sound intensity is proportional to the square of pressure, we will see later, why. The smallest sound pressure that can be heard by the human ear is called audibility threshold. The upper limit of audible sounds is the pressure that already causes pain, this is called the pain threshold. The range between the two limits is twelve orders of magnitude large, i.e. the sound that causes pain is 10^{12} times i.e. one trillion stronger than quietest sound.

It is very useful to look at the orders of magnitude when investigating sound strength, because – as biologists claim – stimulation and perception, i.e. the physical strength of sound and the generated sensation are connected exponentially/logarithmically. This means that if a *ten times* stronger sound feels louder *by one unit*, then a *hundred times* stronger sound feels louder *by two units*, and a *thousand times* stronger sound feels louder *by three units*.

This phenomenon can be formalized by introducing a new physical quantity, the Sound Pressure Level (SPL), instead of sound intensity or pressure amplitude. This new parameter should follow the scale of our perception, therefore it is logarithmic. The unit of the SPL is decibel or dB, named after the American inventor of the telephone, Alexander Graham Bell (1847-1922). The definition of SPL in decibel units is:

$$d = 20 \cdot \log_{10} \frac{p}{p_0}$$

or

$$d = 10 \cdot \log_{10} \frac{I}{I_0}$$

because pressure and intensity are squared proportional. Here I_0 is the limit of audibility (10^{12} W/m² at 1 kHz), and the corresponding sound pressure is $p_0 = 3 \times 10^{-5}$ Pa. I.e. the audibility limit is 0 dB (as it was already mentioned), the thousand times more intensive sound is 30 dB, and the million times more intensive is 60 dB (think about it).

3.2.2 Basics of the physics of sound

In the previous parts we were talking about frequency, pressure and intensity. Let us clear the meaning of these quantities. Sound is a mechanical wave, i.e. during its propagation the particles of the medium are oscillating. Due to this oscillation, some volumes are tighter; some looser packed with particles, depending on the particles moving into or out of the volume. When there are more particles in a given volume, then the density is clearly higher there. Because of this, however (due to thermodynamics) the pressure is also higher there.

Thus sound waves mean that the local density of the medium is changing in time in a periodic way. Besides that the waves depend of course on the coordinates as well, i.e. the density has a maximum at a given time.

Let us try to follow that with simple physical equations. The deflection of the particles at a given time, at a given space-coordinate is given by

$$\xi = \xi_m \sin\left(2\pi f\left(t - \frac{x}{c}\right)\right),$$

where ξ is the deflection of a particle at space x at time t , while ξ_m is the maximal deflection (this corresponds to the strength of sound), and f is the frequency, c is the sound velocity.

Then the velocity of the particles is

$$v = v_m \cos\left(2\pi f\left(t - \frac{x}{c}\right)\right),$$

where v is velocity, v_m its amplitude (clearly $v_m = \xi_m \cdot 2\pi f$ from kinematics). From this, pressure can also be calculated, the result is:

$$p = p_m \sin\left(2\pi f\left(t - \frac{x}{c}\right)\right),$$

where p_m is the pressure amplitude, and $p_m = \rho_0 c v_m$, if ρ_0 is the average density of the medium. Intensity of the sound is the energy flowing through a given area, and this can be calculated as:

$$I = \frac{1}{2} \rho_0 c v_m^2 = \frac{1}{2} \frac{p_m^2}{\rho_0 c}.$$

Unit of intensity is W/m^2 , and the stimulation threshold – corresponding to the earlier mentioned $p_0 = 3 \times 10^{-5}$ Pa quantity – in these units is 10^{12} W/m^2 , as it can be calculated from sound velocity (340 m/s at 20 °C) and air density (1.2 kg/m^3 at 20 °C). This relation is very different at different temperatures or media, i.e. sound velocity in water is much larger, 1440 m/s.

Frequency of the sound corresponds to the pitch of sound; low frequency sounds are “low-pitched”, while high frequency sounds are “high-pitched”. The regular musical note A has the frequency of 440 Hz, i.e. here the air is oscillating 440 times per second. The musical note C (middle C) below that has a frequency of roughly 260 Hz (There are more than one scales, the frequency of the C₄ note of the Equal Tempered Scale is for example 261.63 Hz and the corresponding C₅ note has double frequency 523.25 Hz). Frequency of high C is twice that, i.e. roughly 520 Hz, that of high A is 880 Hz. Clearly one octave corresponds to a doubling in frequency, while two octaves correspond to a four times higher frequency. A 1024 times higher frequency means ten octaves difference.

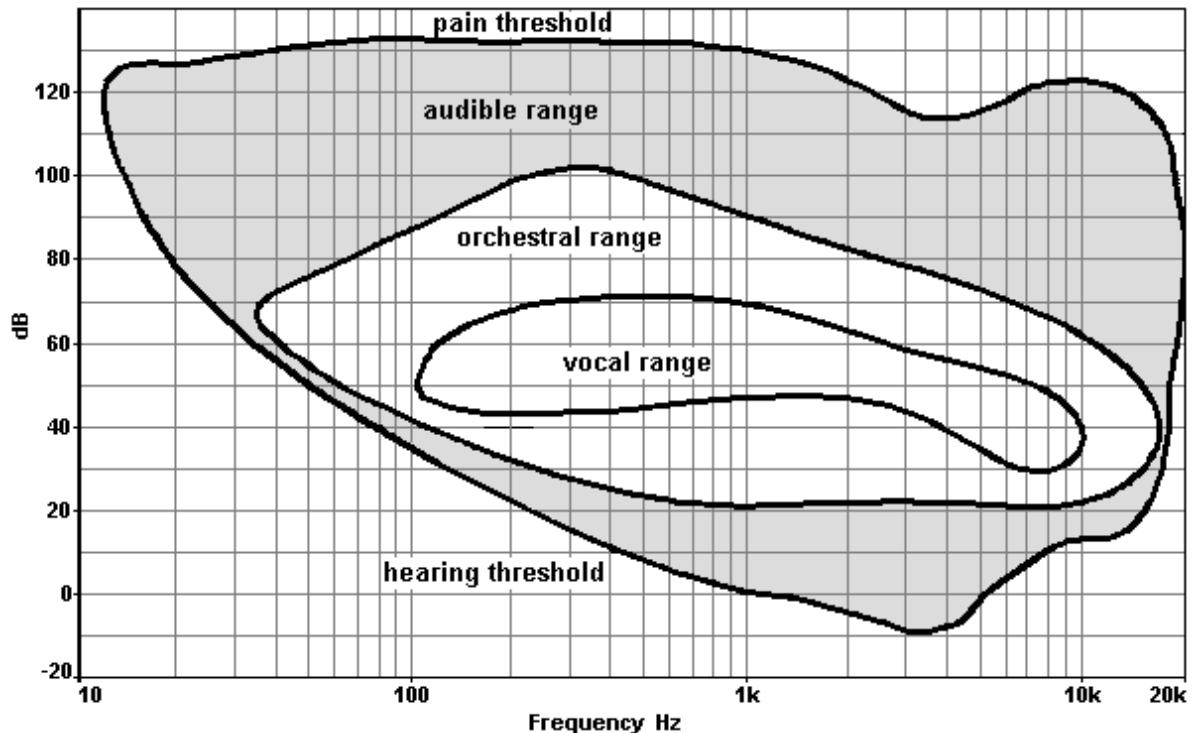


Figure 3.1. Frequency and sound strength (SPL) of different ranges

3.2.3 Sensitivity of the human ear as a function of frequency

The human ear is not uniformly sensitive, as we hear only sounds between roughly 20 Hz and 20,000 Hz. Because of this the earlier description is valid only to a sound of a given frequency. Usually, corresponding to the official standards, the frequency of 1000 Hz is taken. Thus two 40 dB sounds can generate a different perception, if their frequencies are different. This is illustrated in Figure 3.1.

Measure of sound level (SPL) is thus decibel (dB). During one measurement, however the audible sounds are spread over a large range of frequencies, and, as shown in Figure 3.1, lower sounds are less loud than high ones. Thus when performing a measurement, strength of sounds is weighted to make sounds that are more perceived “count” more than those ones barely heard. The standard weighting filter that models the frequency dependence of human hearing is called “A”-filter. The sound level (SPL) measured with such a filter is called A-level.

Noise sources generate constant or variable noise. In noise protection variable noise is characterized by the so-called equivalent noise level. This is a time average of noise over a given time range, and it approximates subjective human noise perception. The quantity that describes equivalent noise level is L_{eq} . This is measured by continuously monitoring the noise level and then averaging it on the required time. This time is usually a second or a tenth of a second. However, the noise of transformers and other devices is such that they are very loud for a very short time only, for example for a hundredth of a second. Thus an average over a second does not reflect this very sharp noise-peak, as the time average will be low. However, this kind of noise may also be very disturbing, as the eardrum might be harmed by very short but intense impulses as well, even if they are not perceived as loud noise. Thus both the average and peak of the noise carry very important information, both have to be measured.

3.2.4 Noise control

Protection against noise can be done by reducing the number and strength of sources (more quiet cars, factories etc.), as well as by applying one or more layers of noise barriers.

Generally flat, hard surfaces (concrete or metal walls, tiles) are good sound reflectors, while porous, soft materials absorb sound, thus latter materials are good tools to reduce noise in a noisy volume or room. Thin and/or light materials are bad noise barriers, while heavy, tense structures block sound from going from one side to the other, thus these are soundproof materials.

3.3 Our noise measuring device

Our device is a Brüel & Kjær 2239 A type Integrating Sound Level Meter. Important parts of it are the following: a microphone as sound sensor; a display that shows average sound level and other, below listed values and actual settings of the device.

With this device, sounds between 30 and 140 dB can be measured with an uncertainty of a couple dB. Measured values can be saved, it is however more convenient (in the present lab course) to write the results down to our notebook.

After turning the device on, it pursues a short self-test and starts to measure immediately. It performs the measurement for a preset time interval, e.g. for a couple of seconds or minutes. During the measurement it displays the actual sound level averaged over short intervals, and the peak sound level during these short intervals (usually seconds). At the end of the measurement, the global average and peak sound level are displayed. These values can then be stored and a new measurement can be started. The displayed values are summarized in Table 3.1.

Measurement results	
Peak	Maximal peak sound level in the last second (or the preset averaging interval)
Inst	Random sample sound level during the last second
SPL	Average sound level during the last second
Leq	Average sound level of the last (or the ongoing) measurement
MaxP	Maximal peak sound level of the last/ongoing measurement
MaxL	Maximal sound level of the last/ongoing measurement
MinL	Minimal sound level of the last/ongoing measurement

Table 3.1. Explanation of the measured quantities

Important settings of the device are the following:

- Time of the measurement, i.e. how long one measurement shall be. This value may go from a couple of seconds to a couple of minutes.
- Averaging time of the measurement, i.e. the short interval for which average and peak sound levels shall be displayed. This is typically part of a second.
- Limits of the measurement, i.e. what the maximum and minimum SPL should be. Outside this range the device will not be able to measure sound levels; it overflows for loud sounds and does not display anything for quiet sounds. There are three ranges available:: 30-100 dB, 50-120 dB and 70-140dB

The information below are displayed by the device:

- Measurement range.
- Measured sound level on a quasi-analog display.
- Status of the measurement (STOP/RUN).
- Elapsed time in the ongoing measurement.
- Peak sound level (by several standards).
- Average sound level (by several standards).
- Other general settings



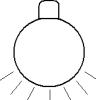




Buttons available on the device	
	ON/OFF
	START/STOP; choosing a setting ("YES" button)
	Display illumination on/off (automatically turned off after 30 sec)
	Change currently displayed measurement result (average, max, etc.)
	Store measurement results, view and delete older stored results.
	Up/down movement in the displayed options.
	Setting the measurement range

Table 3.2. Buttons of the device

3.4 Lab course tasks

3.4.1 Background noise

As a first task, let us measure the average background noise in the room. This should be done by setting the device to a half minute measurement time, then we measure noise while being in complete silence. The important quantity to write down is the equivalent sound level, the L_{eq} value displayed by our device.

- In order to be able to neglect the background noise, the measured results of the next tasks should be significantly (at least 20 dB) louder than the background.
- It would be interesting to compare our results to the European Union standard for quiet bedrooms during the night, the 32 dB noise threshold.
- Also note the maximal peak sound level marked by MaxP. This corresponds to the maximal value of the instantaneous noise. It might be not so disturbing for our ears as an equivalently high average sound level would be, however, it still might cause physical damage on the long run.
- In the further measurements the above two values should be noted (L_{eq} and MaxP). The calculation exercises shall only be done with the equivalent (average) sound level.

3.4.2 Distance dependence of sound level and intensity

The next task is to measure the sound level of a constant power source of sound or noise, as a function of distance from the source. For a constant source it is most convenient to use a mobile phone, with a ringtone or mp3 file that is homogeneous in time as far as possible, to reduce ambiguities as much as possible.

The measurement should be performed as follows. We chose 5 or 10 seconds as measurement interval on the device, then we search for a ringtone on the mobile phone that is homogeneous in time at least for 5 or 10 seconds, depending on the previous setting. The homogeneity has to be investigated by starting the measurement at different times and checking the difference in sound level resulting from not measuring the same interval of the ringtone.

Then we measure the sound level of the ringtone for several distances, which may be for example 5 cm, 10 cm, 20 cm, 30 cm, 0.5m, 1m, 2m, 3m and 4m. We plot a graph of sound level versus distance. Compare the largest value to the generally accepted value for pain threshold, 120 dB.

3.4.3 Estimating the power of the source

With the previous graph we can calculate (estimate) the power of the source (the speaker of the mobile phone), as explained below.

Assume that sound is not absorbed (converted into heat) in the medium, and draw imaginary spheres around the source. The energy flowing through these spheres per unit time is the same for all spheres, because no energy was lost in the medium. Imagine a hose that sprinkles water all around in the air. If one liter of water is sprinkled out in a second, then on all imaginary spheres (drawn around the sprinkler as source) one liter water is flowing through per second, independently of the radius, whether it is 10 cm or 1 m – because the water did not get lost during the sprinkling (we neglected the possibility of the water falling down on the ground). This is illustrated in [Figure 3.2](#).

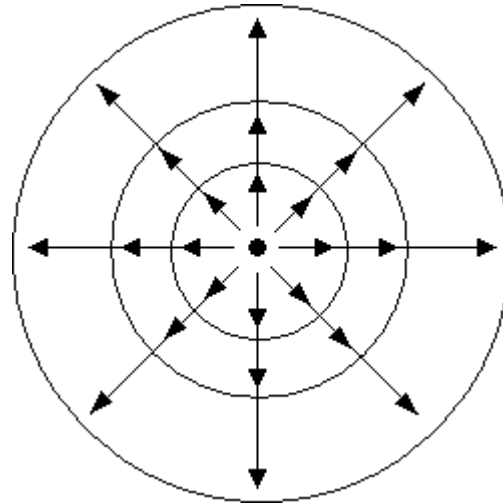


Figure 3.2. Illustration to the propagation of sound from a point-like source

Thus the energy per unit time flowing through the spheres of different radii is constant. However, the intensity is defined as energy flowing through unit area per unit time, i.e.

$$I = \frac{E}{At} = \frac{P}{A},$$

where P is the power of the source. This way we get a simple estimate how the sound intensity depends on the distance.. Recall that the area of a sphere with radius R , i.e. being in a distance of R from the source is $A=4R^2\pi$. Using this one gets:

$$I(R) = \frac{P}{4R^2\pi}.$$

From this the sound level (d , measured in dB) depends on distance as:

$$d(R) = 10 \log_{10} \left(\frac{P}{I_0 4R^2\pi} \right) = 10 \log_{10} \left(\frac{P \cdot R_0^2}{I_0 4R^2 R_0^2 \pi} \right) = 10 \log_{10} \left(\frac{P}{I_0 4R_0^2 \pi} \right) - 20 \log_{10} \frac{R}{R_0}.$$

Here I_0 is the audibility limit, as mentioned earlier, while R_0 is an arbitrary unit of distance (e.g. 1 meter), which had to be introduced to have a dimensionless measure of distance (one can take the logarithm only of dimensionless quantities). Clearly, the first part of the sum is a constant quantity, while the second one shows the linear dependence on the logarithm of distance.

From the above one can deduce that if the previously measured sound level versus distance graph is plotted on a logarithmic distance scale, then one gets points fitting on a linear curve. Let us investigate this linear curve and the above relationship. Clearly,

$$d(R_0) = 10 \log_{10} \left(\frac{P}{I_0 4R_0^2 \pi} \right),$$

because $\log_{10}(1)=0$. Let us then take the constant in the linear curve of the logarithmic plot above, this is the intersection of the curve with the vertical (y) axis. Let us call this value d_0 . Then from the earlier equality, one gets:

$$d_0 = 10 \log_{10} \left(\frac{P}{I_0 4R_0^2 \pi} \right),$$

i.e.

$$P = I_0 4\pi R_0^2 10^{d_0/10},$$

thus by measuring d_0 , one can calculate the power of the sound source.

Our task in this exercise is the following: replot the previous graph, on a logarithmic distance scale (by choosing a distance unit R_0), fit the data points by a linear curve, determine the axis intersection d_0 , and determine the power of the source, based on the above equations.

3.4.4 Testing of noise barriers

At this point we will investigate the sound insulation of different materials. We will use our sound source applied earlier. Choose a suitable distance (e.g. 40 cm) and measure the sound level at this distance. Then put different materials between the source and the measuring device.

First use a regular sheet of paper. Then use two sheets of paper placed right behind each other, then leave a 10 cm distance between the two layers. This illustrates the layered soundproof materials. Measure the sound level with the following materials as well: a newspaper, polystyrene foam (e.g. Styrofoam or “Hungarocell” in Hungary), glass and a door.

Let us make a simple calculation here as well! If the sound intensity is reduced by ΔI due to the material, this difference was absorbed (in form of thermal energy or heat) or reflected by the material. Using the area of the material and the time range of the measurement we may give an upper bound on the energy absorbed by the material. The required formula is

$$\Delta I = \frac{\Delta E}{At}.$$

The intensity drop can be calculated as follows:

$$d_1 = 10 \lg \frac{I_1}{I_0} \Rightarrow I_1 = I_0 10^{\frac{d_1}{10}}$$

$$d_2 = 10 \lg \frac{I_2}{I_0} \Rightarrow I_2 = I_0 10^{\frac{d_2}{10}}$$

$$\Delta I = I_2 - I_1 = I_0 \left(10^{\frac{d_2}{10}} - 10^{\frac{d_1}{10}} \right)$$

Try to estimate the absorbed energy of the different materials in the above approximation! Compare this energy, absorbed in a couple of seconds, to the energy needed to heat up a glass of water by one degree (this couple hundred Joules). Estimate, how much time would be needed to gather this energy by sound absorption (neglecting the transmitted heat by the water to the air).

3.4.5 Testing of everyday noise sources

In this last exercise we will investigate noises from our everyday environment. Adventurous students may measure the sound level of their shouts or screams – in this case, however, the upper limit of the measurement has to be set to 120 or even to 140 dB.

Then measure the noise on different, frequently visited or maybe isolated locations of the University Campus. The measurement interval should be set to half a minute in this case, to get a more realistic picture.

Last but not least, investigate the noise of the surroundings of the University Campus (tram station of trams no. 4 and 6, wharf underneath the Petőfi bridge). Measure the noise here, and also the decrease of noise with distance. Nowadays houses may be built at least 30-50 meters from very busy streets, measure the noise reduction of the Irinyi József street from this distance.

If there is time left, choose some noise sources and measure they average and peak noise. These sources may be the following: arrival or departure of a tram, engine noise of a car, arrival or departure of an elevator, a computer running on high power, or any other noise that is part of our everyday life.

Organize the measured sound levels at your discretion; compare them to each other and to the pain threshold of 120 dB. Investigate connection between the personal disturbing effect of a noise to its objective sound level.

3.5 Test questions

1. What is the difference and the connection between sound and noise?
2. What is the intensity and sound level of noise? What are their definitions?
3. What is the audibility limit? What numbers and units are connected to it?
4. Which sound is perceived as louder, a 50 dB musical A sound or an equally 50 dB, but 40 Hz frequency sound?
5. What is the quantity denoted by L_{eq} ? What is measured by it and why?
6. Which materials are good sound absorbers? And which are good as noise barriers? What is the difference?
7. What is the definition of sound pressure level? And how are the quantities therein defined?
8. How much energy is flowing through the 1 cm^2 eardrum during one second, if a continuous sound of 50 dB is reaching it?
9. Make a sketch of the frequency dependence of the sensitivity of the human ear! Where is the audibility threshold and the pain threshold?

4. Infrasound wave detection (INF)

4.1 Introduction

An infrasound microphone device is capable of detecting low frequency sound waves that are not audible. We use a device like this during the laboratory practice which is a home made detector. The original design of the detector was developed at the Los Alamos National Laboratory, and has been recently further developed by the Eötvös Gravity Research Group (EGRG) for exploiting its capabilities in the LIGO experiment. LIGO is an international collaboration of almost 50 research institutes and universities to detect gravitational waves. In the experiment MIT and CalTech scientists developed a classical Michelson interferometer to a precision, where it can detect a 10-18 meter change in the 4 km long interferometer arm (this distance is equivalent to the thousandth of the proton diameter). Since the detectors are extremely sensitive to any environmental noise, it is a serious challenge to decouple the system from the outside world. If decoupling for one particular environmental effect is not feasible, somehow we have to monitor the effect itself. Our microphone is designed to continuously monitor the low-frequency pressure wave background.

Basically sources of infrasound can be separated to two main types. The first type of the sources emits low-frequency sound waves impulsively, causing the ambient air pressure to raise very fast, and then drop below the original value, quickly after the original pressure is restored. This is called an N-wave, because of its typical N-shape. These sources can be volcanic eruptions, meteor explosions, rocket and spacecrafts, supersonic aircrafts, artificial explosions, etc..

The second category of infrasound sources emit continuous pressure waves that last for at least several full periods of the wave. Typically continuous infrasound waves are generated by earthquakes, microbaroms (standing waves on the oceanic surface), mountains, storms, wind-farms, tornadoes, etc.

Pressure waves in air with a frequency less than 20 Hz are called infrasound, generally considered inaudible, however, note that sound remains audible even under 20 Hz, if the pressure amplitude is high enough. Sound is attenuated by air exponentially with its frequency, meaning that low-frequency waves can propagate for longer distances (>20000 km), while high-frequency sound waves are absorbed by air in a shorter range. This important property of sound waves is due to the inelastic collisions of the particles in the medium of propagation, the air molecules. Another important property of low frequency pressure waves is that they cannot be easily mitigated. According to the mass law, attenuation of a sound wave in a material will decrease 6 dB with each halving of the frequency. The extremely long wavelength of these mechanical waves makes them also able to travel even further, than shorter wavelength waves. So, we can conclude that there are numerous sources of infrasound waves that produce signals with a very high propagation range. These waves are also hard to be mitigated with conventional methods; therefore infrasound effects have to be taken into consideration when a state of the art, high precision instrument is designed.

Our infrasonic microphone will be a part of the Advanced LIGO (starting 2014 fall) environmental monitoring system. One prototype of the instrument is already gathering data at the LIGO Hanford Observatory, USA; the data will be used for correlation analysis. Thus, this lab course is different from other courses somewhat. We do not investigate one single

phenomenon with one single experiment. Instead of this, we will investigate a real experiment, where the understanding of several phenomena is needed.

4.2 Basic physics of infrasound

Infrasound, like any other type of sound, is a longitudinal pressure wave in the air (in gases there are no shear forces, thus transversal waves are impossible to form). Sound velocity is constant relative to the medium (but of course it depends on the medium itself, on temperature, humidity etc.). The pressure field can be described through pressure, so we will use the $p(\mathbf{x}, t)$ function, where \mathbf{x} is the position vector, and t is elapsed time.

In case of a plane wave, the pressure looks like $p(\mathbf{x}, t) = p_0 \sin(2\pi \mathbf{x} \mathbf{n} / \lambda + t\omega + \varphi_0)$, where \mathbf{n} is the unity vector pointing towards the wave propagation direction, λ is the wavelength, ω is the angular frequency and φ_0 the initial phase.

If the plane wave propagates in the x direction, then the pressure is a function of x and t only:

$$p(x, t) = p_0 \sin(2\pi x / \lambda + t\omega + \varphi_0).$$

Based on frequency, sound waves are put into three categories. Above 20 kHz it is ultrasound (used by bats and dolphins for example), between 20 Hz and 20 kHz we talk about audible sound, and below 20 Hz is infrasound (used by elephants for example).

4.2.1 An estimation on the effects of infrasound

Imagine infrasound as a very large wavelength pressure wave propagating in the air, with amplitude p_0 (wavelength λ can be calculated from frequency and velocity, i.e. $\lambda = c/f$). Let us check how big the movement is which is caused by the infrasound, on a body with l length and A cross-section area that is orthogonal to the wave fronts.

Pressure difference between the two ends of the body can be written as follows:

$$\delta p = p_0 \sin(\omega t + \omega l / 2c) - p_0 \sin(\omega t - \omega l / 2c) = 2 \cos(\omega t) \sin(\omega l / 2c) \approx p_0 \frac{\omega l}{c} \cos(\omega t)$$

Let us then write down the equation of motion:

$$ma = A \rho a = \delta p A$$

$$a \approx \frac{p_0 \omega}{c \rho} \cos(\omega t)$$

$$v(t) = \int a dt \approx -\frac{p_0}{c \rho} \sin(\omega t)$$

$$x(t) = \int v dt \approx -\frac{p_0}{c \rho \omega} \cos(\omega t)$$

If 1 mPa pressure, 2500 kg/m³ density (typically glass) and 10 mHz frequency is used, a movement of ~20 nm comes out as a result, which is comparable of laser wavelengths, thus it has a large effect on LIGO. This means that if infrasound enters the system, they will cause a large frequency systematic error in the gravitational wave channel signal. Due to nonlinearity of the system, upper harmonics will also appear.

Strength of a sound wave of course can be given with the amplitude of pressure fluctuations (sound pressure). It is well known that the sensitivity of the human ear is logarithmic, thus sound volume shall be given rather on the decibel (dB) scale. On this scale, 20 dB increase

means a tenfold increase in the pressure amplitude. The scale is normalized (this can be arbitrarily chosen) such as 0 dB will correspond to the lowest sound a human can hear. This (by definition) corresponds to pressure amplitude of 20 μPa , this is a sound of a mosquito from roughly 3 meters. Sound pressure can then be calculated the following way:

$$L_p = 20 \cdot \log_{10} \left(\frac{p_{rms}}{p_{ref}} \right) \text{dB},$$

where p_{ref} is the previously mentioned amplitude of 20 μPa . On this scale typical speech is roughly 40-60 dB, while 120 dB causes permanent damage, and 130 dB is the pain-threshold. As previously calculated, amplitude of velocity oscillation is $p_0/c\rho$ (where a little air volume can be regarded as a test-object). Power propagated by the sound wave is clearly pressure times cross-section (which is force) times velocity (movement per unit time), i.e. $(p_0)^2 A/c\rho$. Here air density is roughly 1.2 kg/m^3 , sound velocity is 331 m/s.

Please find more about sound pressure in http://en.wikipedia.org/wiki/Sound_pressure

Source	Sound pressure (Pa)	Sound pressure level (dB)
Krakatau explosion from 160km	20000	180
.30 gun from 1m	5023	168
Jet engine from 30 m	632	150
Pain threshold	63.2	130
Pneumatic hammer from 1m	2	100
Busy traffic from 10 m	$2 \times 10^{-1} - 6.32 \times 10^{-1}$	80-90
Car from 10 m	$2 \times 10^{-2} - 2 \times 10^{-1}$	60-80
Regular speech from 1 m	$2 \times 10^{-3} - 2 \times 10^{-2}$	40-60
A very quiet room	$2 \times 10^{-4} - 6.32 \times 10^{-4}$	20-30
Absolute threshold of hearing (1kHz)	2×10^{-5}	0

Table 4.1. Sound pressure of some common sources

4.3 Microphone basics

There were several methods developed for measuring sound. In old telephones there were carbon microphones. Here coal grains were placed in a closed volume. One wall of the volume could move (according the pressure on it), which would compress the particles,

causing a change in resistance, which in turn can be converted into an electric signal. This works as an absolute pressure measure, and produces a very bad quality sound, is however heavy-duty, cheap and requires minimal electronics.

In so-called dynamic microphones sound (pressure) is moving a membrane, to which a coil is secured, placed in a constant magnetic field. Movement of the membrane induces electric field (voltage) in the coil.

In condenser microphones there is also a membrane (which is moved under pressure), and together with a fixed plane they form a capacitor. The capacitor is charged with a fixed amount of electric charge. Movement of the membrane causes a change in capacity, thus voltage on the capacitor changes ($Q=C \times U$). This construction will not work on very low frequency, because the charge is fixed by electronics, which, below a given frequency, automatically compensates movements of the membrane. Capacity of a given capacitor can be calculated according to the following formula:

$$C = \frac{\epsilon A}{d},$$

where A is the area of the capacitor's conducting planes, their separation is d , while ϵ is the electric permittivity ($\sim 8 \times 10^{-12}$ F/m).

Electret microphones are similar to condenser microphones, main difference is that their conducting planes are covered with ferroelectric material (with permanently embedded static charge), thus active polarizing electronics is not needed.

In dynamic and condenser microphones two main technical solutions are used. The membrane can be open from both sides, in which case it follows the pressure gradient basically. If the membrane is closed from one side, then it measures absolute pressure.

4.3.1 Low frequency condenser microphone

The principle of the measurement is very simple. Let us take a small volume sink, roughly a couple hundred cubic centimeters; this will be used as pressure reference. A small capillary tube is put between the reference volume and its surroundings, ensuring slow pressure equalization. Size of the capillary determines the typical time of the equalization, thus with the capillary, low frequency cut-off can be directly adjusted. Infrasound can then be measured with this setup by making a differential pressure measurement between the two volumes.

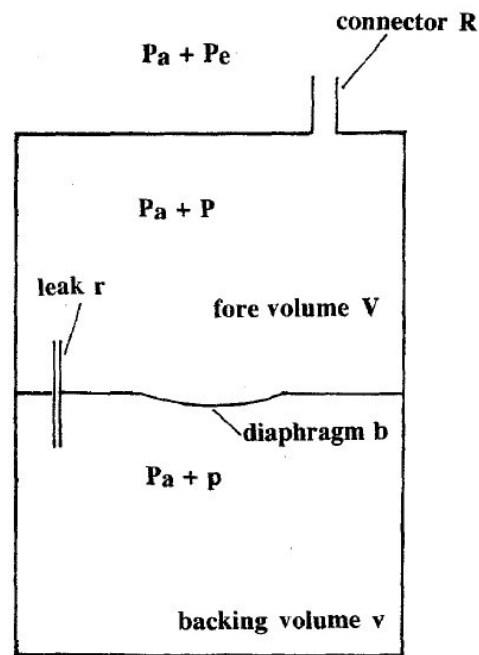


Figure 4.1. Sketch of an infrasound microphone, based on Los Alamos design

In the given microphone configuration the membrane (“diaphragm” in Figure 4.1) is the elastic plane of a capacitor, which will thus change its capacity according to the pressure difference – this is called a condenser (or capacitor) microphone. In the EGRG collaboration we developed a new condenser microphone readout electronics, which works theoretically on any required low frequency.

4.4 Basics of signal processing

Our microphone, similarly to most environmental monitoring equipment, generates a time-series as an output. This time-series is the actual signal size measured as a function time. One method of investigating the time-series, which we will also use in the measurement, is plotting the time-series graphically. This way we get visual information on the signal size. With this method, we simply get an overall picture of the characteristics of the measured signal, we can measure momentary amplitudes, can determine changes versus time. The frequency of some signals can be measured, if we know time resolution. If we count the number of periods in a second, we get the frequency in Hertz ($1/s=Hz$). Such a time-series is shown on [Figure 4.2](#).

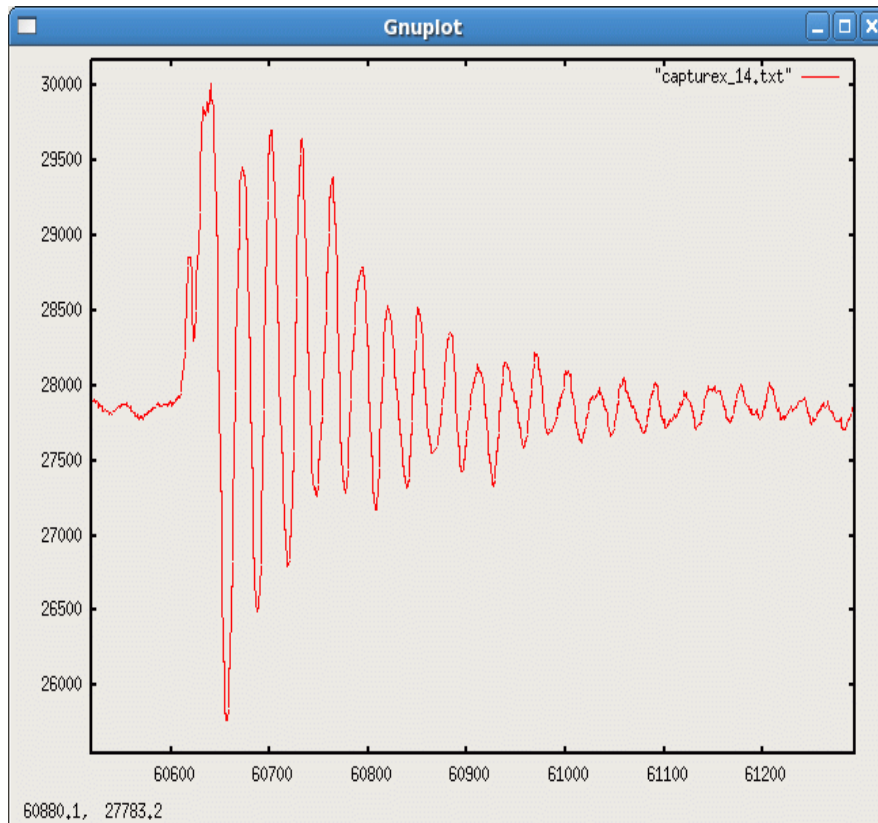


Figure 4.2. A typical time-series example. Units of the horizontal axis are 1/1024 seconds, while digital units are shown on the vertical axis.

The easiest way to plot such a time-series is to use the free, open source Gnuplot software. Its use is very easy, it has a command line interface, and there are numerous how-to's and descriptions available on the Internet. The software can be downloaded to Linux, MacOS and Windows as well. A comprehensive English description is available under <http://www.gnuplot.info/documentation.html>.

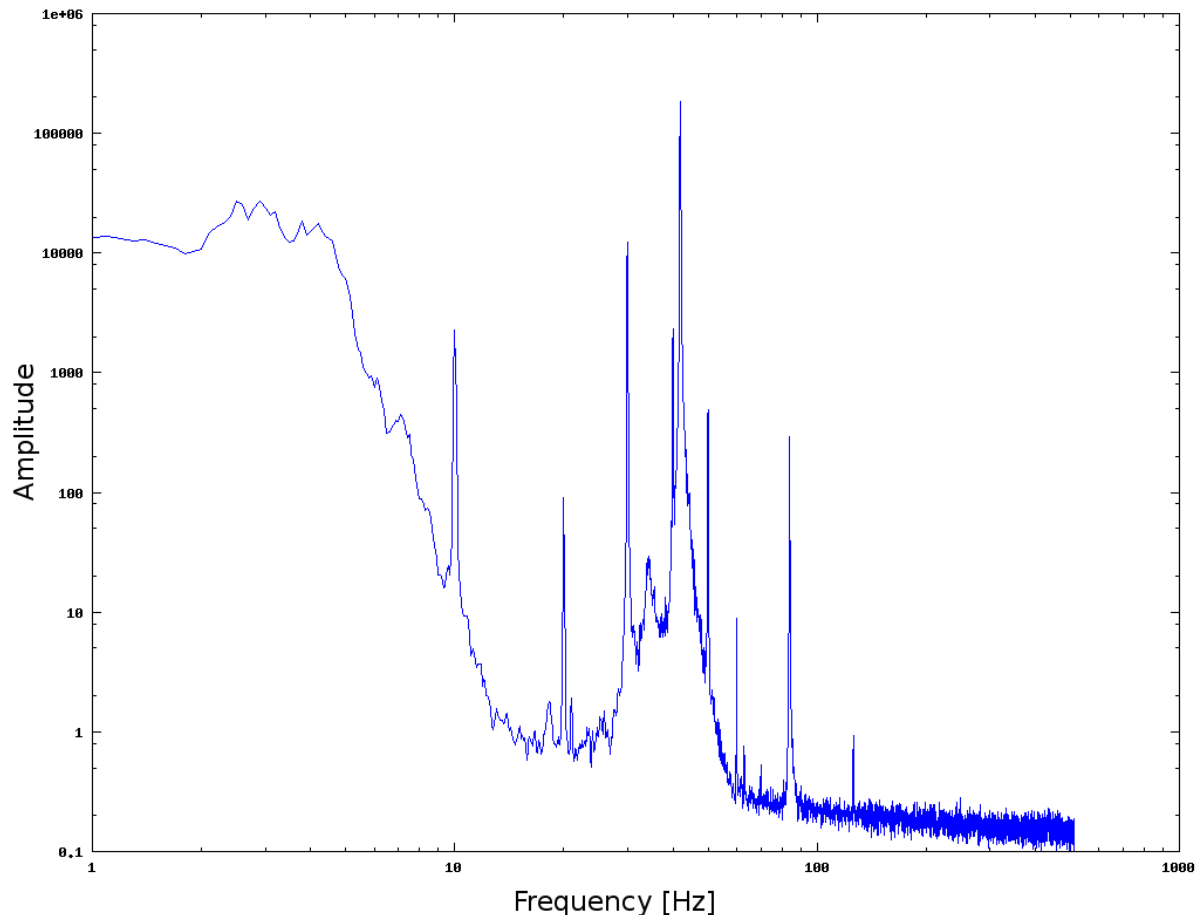


Figure 4.3. Spectrum of a signal. The axes are logarithmic, on the horizontal one frequency is plotted, while the amplitude is on the vertical axis.

Any arbitrary signal (time-series) can be decomposed to a superposition of sine and cosine functions with different amplitudes and frequencies. During this lab course we will measure the superposed signal and noise of many sources; these sources will have to be separated to tell, on which frequencies these sources generate signal or noise. A solution to this problem is the Fourier transformation, with which one is able to decompose a complicated signal to different frequency components and calculate their amplitudes. This we call the spectrum of the signal (see Figure 4.3). Thus besides being able to find all the frequencies in a complicated signal, we get an overall picture about the amplitude ratio of the given frequency components; this makes it a very useful and important tool. There are many softwares available to perform a Fourier transformation, and several of them are available freely for any platform. During the lab course we will prepare such a Fourier spectrum of the signal.

A plot with time on the horizontal axis and frequency on the vertical one is called spectrogram or sonogram. On this spectrogram the change of the different frequency components versus time can be monitored simply, and their connection might be quite obvious. A spectrogram is then basically nothing else than a series of many spectra, made right after each other with a fixed time delay. Thus changes versus time can be discovered using such a spectrogram. There might be a time-overlap between the spectra. The amplitude of individual components can be marked by the color of a given point, as the two axes are already “taken”. A spectrogram is shown in [Figure 4.4](#).

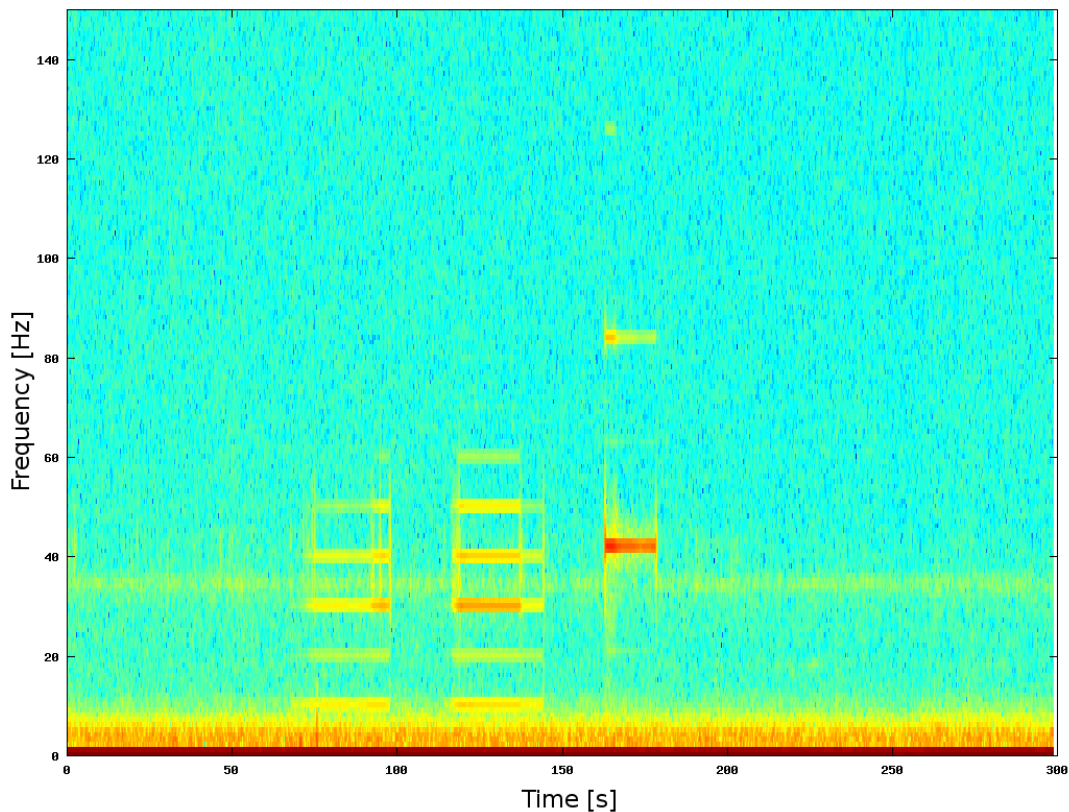


Figure 4.4. A spectrogram is shown, where the amplitude of the different frequency components are shown with different colors (blue is small amplitude, red is large).

4.5 Operation of the microphone

The microphone can be used very easily. If it is connected to the power supply, the equipment starts and after roughly a minute, it automatically starts to gather data. After the measurement is done (approximately 5 minutes) the device has to be disconnected, and the results can be downloaded from the SD card. During the measurement a blue LED is lit, but it cannot be seen, as it is inside the device.

4.6 Lab course tasks

The exact technical details of the measurement (e.g. the Fourier transformation) will be given during the course, but please be prepared to ask your questions!

1. Measure the background noise. From the Fourier transformed of the sample determine the frequency of the components of the noise.
2. Generate a deep but audible sound with a PC; the 50-100 Hz range is suggested. Take a five minute sample. Plot the sample and measure the amplitude of the sound pressure. To what dB value does this correspond to?
3. Measure the frequency of the sound with Fourier transformation.
4. Locate infrasound sources inside and near to the building. Every time, carry out a five minute measurement, and save them each to a different file. Write down the location and other characteristics of each measurement. Recommended sources are elevators, cars and air-conditioners.

4.7 Test questions

1. What is a plane wave?
2. Why is it possible for the carbon microphone to work on very high frequencies?
3. Why can't dynamic microphones work on very low frequencies?
4. Let us take a capacitor with plate radius of 33 mm and distance of 35 μm . How big is its capacity?
5. How is the capacity of the previous capacitor changing, if the membrane is moved by 1 nm?
6. What processes are the sources of infrasound in our environment?
7. How does the infrasound detector work?
8. What variable does the Fourier-transformation transform to what variable?

Chapter II.

NONIONIZING ELECTROMAGNETIC RADIATION

This chapter collects some laboratory practices involving phenomena that are important in environmental physics, and correspond to that part of the electromagnetic radiation which has no ionizing effect. This part of the spectrum is called “nonionizing electromagnetic radiation”. As the frequency grows, this regime starts from radio waves, across the microwaves, infrared waves to the visible light and the ultraviolet light. Many aspects of these radiations are important for the sustainable environment and for the good living standards. Our three laboratory practices are in the topics of electromagnetic radiation hazards and solar radiation. The electromagnetic radiation hazards covers two frequency ranges the 50 Hz low frequency radio waves and the microwaves in the GHz domain. Both are often used in our civilized environment. The solar radiation is especially important from the point of view of the energy, as it is one of the most promising source of the renewable energy sources. However, the technology to transfer it efficiently for a large portion of everyday demands has not been solved, yet. One of the highest efficiency tools is the heat collector. The practice will investigate a cheap test collector made of cans and the biologically interesting polarized light pollution that have innovative results nowadays. First an overview of the electromagnetic radiation is discussed as a framework of all these radiations. It is also the framework of the ionizing electromagnetic waves that is part of the next chapter.

5. An overview of the electromagnetic waves

Technology is being developed in huge steps these days. Vast amount of electric devices are used in our everyday life. Some of them operate on the global power network, using 50 Hz alternating current. Some of them communicate wirelessly, using electromagnetic waves, as in case of radio, mobile phones, Wi-Fi, Bluetooth or GPS devices. The microwave radiation inside a microwave oven is used to heat up food. Clearly, radiation, electric and magnetic fields are not localized in the given device, they might enter into the human body. They deposit their energy in our tissues, heat up the cells there, disturb the electric pulses of the human body or the nervous system; they can even interfere with the hormone production.

In this part we will review the theory of electromagnetism i.e. Maxwell's equations, and derive the phenomenon of electromagnetic waves from it. First we will explain the basic laws of electric and magnetic fields. Then we discuss the mathematical tools that are needed to convert these simple laws into Maxwell's equations.

5.1 Electric and magnetic fields and fluxes

When describing electricity, the quantity of electric field strength is used. Field strength is a vector quantity, equal to the force that would act upon a unit charge placed at the given spatial location. It can be defined via the relationship: $F=qE$, where F is the force, q is the electric charge and E is the field strength. Sources of electric fields are for example point-like charges (e.g. an electron), but also time-varying magnetic fields, as we will see later.

Basic signal of a magnetic field is that it turns a compass to its own direction. From the torque of the rotation the strength of the magnetic field (the magnetic induction) can be calculated. Magnetic field is denoted by B . This is also a vector quantity, defined through the torque acting on the compass; a more convenient definition is through the force acting on a charge q moving with velocity v : $F=qv \times B$. Sources of magnetic fields are moving charges or time-varying electric fields. Magnetic field is generated for example around a conductor with current flowing in it.

For demonstrating the strength of both fields we define so-called field lines. The tangent of these lines is always directed parallel to the field strength in the given point, while the density of the lines corresponds to the extent of field strength in the given point. The amount of field lines passing through a given surface is called the flux (Φ) of that surface. From this definition, for a surface of area A , the electric and magnetic fluxes are in case of constant fields:

$$\Phi_E = EA,$$

$$\Phi_B = BA.$$

What happens in case of locally varying fields, how can the flux be calculated? Of course in this case the area has to be subdivided into a large number of small surfaces where the field is almost constant, and the flux for each little surface has to be summed up. If the number of surfaces is large (i.e. the surfaces are very small) then the sum will be quite accurate. Then we take the infinitesimal limit, where the number of surfaces is infinity, and each surface is infinitely small. This is the usual method of integral calculation. The result in this case is

$$\Phi_E = \int_A E dA,$$

$$\Phi_B = \int_A B dA.$$

5.2 Basics of vector calculus

We will need the definition and rules of some basic operators in vector calculus. These are the divergence, curl and gradient operators. All are based on the Del operator denoted by the Nabla symbol. This is a differential vector operator, defined as

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$$

i.e. its components are the partial derivative. This vector may act on a scalar function $f(x, y, z)$ as

$$\text{grad } f = \nabla f = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right).$$

This is called the gradient of the given function. What does this mean? Calculate the directional derivative of $f(x, y, z)$, the derivative along a direction, like it would be a simple one-dimensional function $f(s \cdot e)$ where e is a unit vector in the given direction, and we take the derivative with respect to s then. Some directional derivatives will be small, some will be large. It can be proven that in this case, the direction of the largest directional derivative corresponds to the direction of the gradient vector, and the same is true for its size.

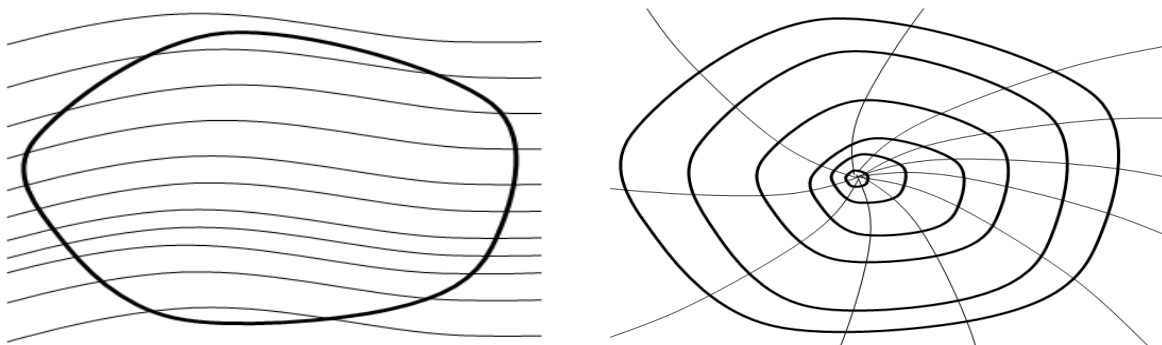


Figure 5.1. Illustration of a source-free vector field, having zero divergence (on the left), and of a vector field with a point-like source, i.e. non-zero divergence (on the right). The thick lines represent the boundaries of 3D objects, i.e. these figures are 2D cross-sections.

Let us now take a vector function, $E(x, y, z)$. We mentioned earlier (in 5.1) the definition of the flux of this arbitrary vector quantity on a given surface. Let this surface be a closed surface (e.g. a sphere). Let us now divide the flux by the volume. This is a well-defined quantity for a given volume V with surface A . If the vector field has no sources, i.e. the field lines do not end or start in any point (but are closed or infinite lines), then the flux will be zero, as any field line that crosses the surface towards the inside, it will cross the surface towards the outside as well (it does not end inside the volume). However, if the field has a source, then the lines start from there, thus the flux will be non-zero if the volume contains the source. Figure 5.1 illustrates this. What happens if we start to shrink the volume infinitely, towards one point, as shown in Figure 5.1? It can be proven that this limit has a well-defined

limiting value, which we call divergence, or it might also be called local flux density. The definition is:

$$\operatorname{div} E = \lim_{V \rightarrow 0} \frac{\Phi}{V} = \lim_{V \rightarrow 0} \frac{1}{V} \oint_A E dA.$$

It can be proven mathematically that divergences can also be calculated via the Del operator:

$$\operatorname{div} E = \nabla \cdot E = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \cdot (E_x, E_y, E_z) = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}$$

The square of this operator is called the Laplace operator, denoted by Δ (already mentioned in [Section 1.5.1](#)). It is a scalar operator, acting on scalar functions. Its effect on a function f is:

$$\Delta f = \nabla^2 f = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

Let us now take a line integral of our function $E(x, y, z)$ along a given closed curve. The value of this integral may be called the whirl or vortex of the field along the closed curve:

$$C = \oint_l E dl,$$

where the curve is denoted by l . In a simplified picture let us think of the curve as it is a circle. In a detailed mathematical picture of course it is not true, but this simplified derivation of the meaning of the curl gives the most important part of it. Let us divide the value of the integral by the area of the circle. Let us start to shrink the area and take the limit when it closes it in on a given point. It can be proven that this limit has a well-defined limiting value, and it is called the curl of the given field. The definition is (note that curl is often called rotation, shortened by rot):

$$\operatorname{rot} E = \operatorname{curl} E = \lim_{A \rightarrow 0} \frac{C}{A} = \lim_{A \rightarrow 0} \frac{1}{A} \oint_l E dl.$$

It can be proven that the curl can also be calculated via the Del (or nabla, ∇) operator:

$$\operatorname{curl} E = \nabla \times E = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \times (E_x, E_y, E_z),$$

where the vector product or cross product is used, denoted by \times .

There are two important mathematical theorems connected to the curl and div operators. One is the Gauss-Ostrogradsky theorem, which states that

$$\int_V \operatorname{div} E dV = \oint_A E dA$$

if A is the boundary surface of volume V . This clearly follows from the definition of divergence. The other theorem is called the Stokes-theorem, stating the following:

$$\int_A \operatorname{curl} E dA = \oint_l E dl$$

if l is the boundary curve of area A . This also clearly follows from the definition of curl.

5.3 Basic equations of electromagnetism

5.3.1 Gauss's law

Now let us see what kind of physical consequences the above laws lead to? In case of divergence, the question is, what the sources of electric fields are. The answer of course is: the sources charges. Gauss's law states that the flux of a closed surface will be proportional to the charges closed in. If the total charge in the given volume is Q , the law says:

$$\Phi_E = \frac{1}{\varepsilon_0} Q, \text{ or alternatively } \oint_A E dA = \frac{1}{\varepsilon_0} \int \rho dV,$$

where ρ , the charge density has been introduced, and ε_0 is the vacuum electric permittivity, a basic natural constant of value 8.85×10^{-12} As/Vm. This law can easily be checked in case of a point-like charge, the electric field of which is well known, thus it can easily be integrated on a given surface, for example on a sphere drawn around the charge as a center.

For the magnetic fields, one may draw a similar consequence. There are no magnetic charges in Nature, i.e. magnetic fields are source-less. Thus the law for magnetic fields is:

$$\Phi_B = 0, \text{ or alternatively } \oint_A B dA = 0.$$

Both laws may be transformed using the above mentioned Gauss-Ostrogadsky theorem:

$$\int \operatorname{div} E dV = \frac{1}{\varepsilon_0} \int \rho dV \quad \text{and} \quad \int \operatorname{div} B dV = 0.$$

These equations hold for any volume, thus the quantities inside the integrals have to be equal, too. Consequently, Gauss's law's for electric and magnetic fields look like:

$$\operatorname{div} E = \frac{1}{\varepsilon_0} \rho,$$

$$\operatorname{div} B = 0.$$

5.3.2 Faraday's law, Ampere's law

According to Faraday's law of induction, potential difference or voltage is induced in a closed loop shaped conductor, if the magnetic flux of the area closed in by the loop is varying in time. The induced potential difference will be the time derivative of the magnetic flux going through the given area:

$$U_{ind} = - \frac{\partial \Phi_B}{\partial t}.$$

where the negative sign is due to Lenz's law, which states that an induced potential is always in such a direction as to oppose the motion or change causing it. However, electric potential difference along a line l is defined as $U = \int_l E dl$, thus (note that the line is a closed loop here):

$$\oint_l E dl = - \int_A \frac{\partial B}{\partial t} dA.$$

From this, according to Stokes's law (converting the closed line integral to a surface integral):

$$\int_A \text{curl } E \, dA = - \int_A \frac{\partial B}{\partial t} \, dA.$$

As in case of Gauss's law, this integral equation also holds for any surface, thus the integrands have to be equal as well. From this we get the differential form of Faraday's law:

$$\text{curl } E = - \frac{\partial B}{\partial t}.$$

Ampere's circuital law relates the magnetic field to its electric current source. It says that if we integrate the magnetic field along a closed line, the result will be proportional to the current I flowing through the surface closed in by the line:

$$\oint_l B \, dl = \mu_0 I = \mu_0 \int_A j \, dA,$$

where μ_0 is a natural constant of the value $4\pi \times 10^{-7}$ Vs/Am, called the vacuum magnetic permeability, and l is the borderline of surface A . Again according to Stokes's law:

$$\int_A \text{curl } B \, dA = \mu_0 \int_A j \, dA,$$

where we converted the closed line integral to a surface integral. Maxwell discovered a correction to this equation, namely, that a term has to be added to the current density, as time-varying electric fields also generate magnetism. This additional term is called the displacement current, and its value is ϵ_0 times the time derivative of the electric field. Taking into account this correction, the differential version of Ampere's law will be (because again the integral equation holds for any surface, thus the arguments have to be equal):

$$\text{curl } B = \mu_0 j + \mu_0 \epsilon_0 \frac{\partial E}{\partial t}.$$

5.3.3 Maxwell's equations

The four equations listed above are the so-called Maxwell's equations: Gauss's law for electric fields, Gauss's law for magnetic fields, Faraday's law of induction, Ampere's circuital law. The differential form of the four equations is:

$$\text{div } E = \frac{1}{\epsilon_0} \rho$$

$$\text{div } B = 0$$

$$\text{curl } E = - \frac{\partial B}{\partial t}$$

$$\text{curl } B = \mu_0 j + \mu_0 \epsilon_0 \frac{\partial E}{\partial t}.$$

These equations contain the electric and magnetic fields, and two functions describing the spatial localization and movement of charges: charge density ρ and current density j . The most important consequence of these equations is that there is an interconnection between electric and magnetic fields, the time variation of any of the two induces the other. Even without charges present (i.e. in vacuum), there might be electric and magnetic fields. Let us discuss the vacuum case! In vacuum, there are no charges and currents, thus the above set of equations transforms to:

$$\begin{aligned} \operatorname{div} E &= 0 & \operatorname{curl} E &= -\frac{\partial B}{\partial t} \\ \operatorname{div} B &= 0 & \operatorname{curl} B &= \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} \end{aligned}$$

There is one more mathematical theorem about these vector operators we did not mention before. This says that for any vector field E

$$\operatorname{curl} \operatorname{curl} E = \operatorname{grad} \operatorname{div} E - \Delta E$$

where ΔE means the vector, the components of which are ΔE_x , ΔE_y , ΔE_z . Using this identity, the above set of equations for E and B may be transformed into:

$$\begin{aligned} \frac{\partial^2 E}{\partial t^2} &= c^2 \Delta E \\ \frac{\partial^2 B}{\partial t^2} &= c^2 \Delta B \end{aligned}$$

where $c^2 = 1/(\varepsilon_0 \mu_0)$ was introduced, and the equations are scalar equations for each component of both the magnetic and the electric fields. These equations each represent a well-known wave equation (see Section 5.4).

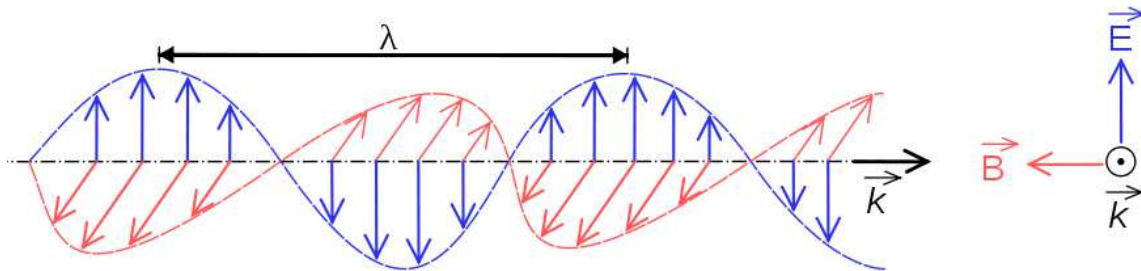


Figure 5.2. A linearly polarized electromagnetic wave. Each of the vectors k , B and E are perpendicular to each other. Image is from Wikipedia.

5.4 Wave equation of the electromagnetic fields

5.4.1 Solutions of the EM wave equations

As shown before, there is a wave equation solution to Maxwell's equations. In this case each component of both the electric and the magnetic fields fulfills a wave equation. However, these are only six equations, whereas Maxwell's equations consist of eight ones (two scalar and two vector equations). Thus some information is still in there. This information is the following. First, the wave number vector k has to be the same for each component of each vector. Second, more importantly, the following equation holds:

$$k \times E = \omega B$$

thus E and B are perpendicular to each other, and also to the wave propagation direction. Consequently both fields represent a transversal vector wave. If, similarly to [1.5.1](#), we introduce the coordinate system where the x direction is parallel to k , the solution is:

$$E_x = E_{x0} \sin(kx - \omega t + \varphi_{Ex}), \quad B_x = B_{x0} \sin(kx - \omega t + \varphi_{Ex}),$$

$$E_y = E_{y0} \sin(kx - \omega t + \varphi_{Ey}), \quad B_y = B_{y0} \sin(kx - \omega t + \varphi_{Ey}),$$

$$E_z = E_{z0} \sin(kx - \omega t + \varphi_{Ez}), \quad B_z = B_{z0} \sin(kx - \omega t + \varphi_{Ez}).$$

These are two perpendicular vector waves. Polarization of the waves is defined by the phases of the components. Circularly polarized waves were discussed in 1.2.3, see details there. One has a linearly polarized wave when the phases are all zero (or equal), in this case the direction of E and B is constant, as the sine term is always the same for each component. Figure 5.2 shows such a wave. Note furthermore that the wave propagation velocity is

$$c = 1/\sqrt{\mu_0 \epsilon_0} = 3 \times 10^8 \text{ m/s},$$

which is called the speed of light in vacuum, as light is also an electromagnetic wave. The frequency ($\omega/2\pi$) or the wavelength ($2\pi/k$) is however arbitrary. As they change, we refer to different electromagnetic radiation types. All radiations are part of the electromagnetic spectrum, detailed in the next section.

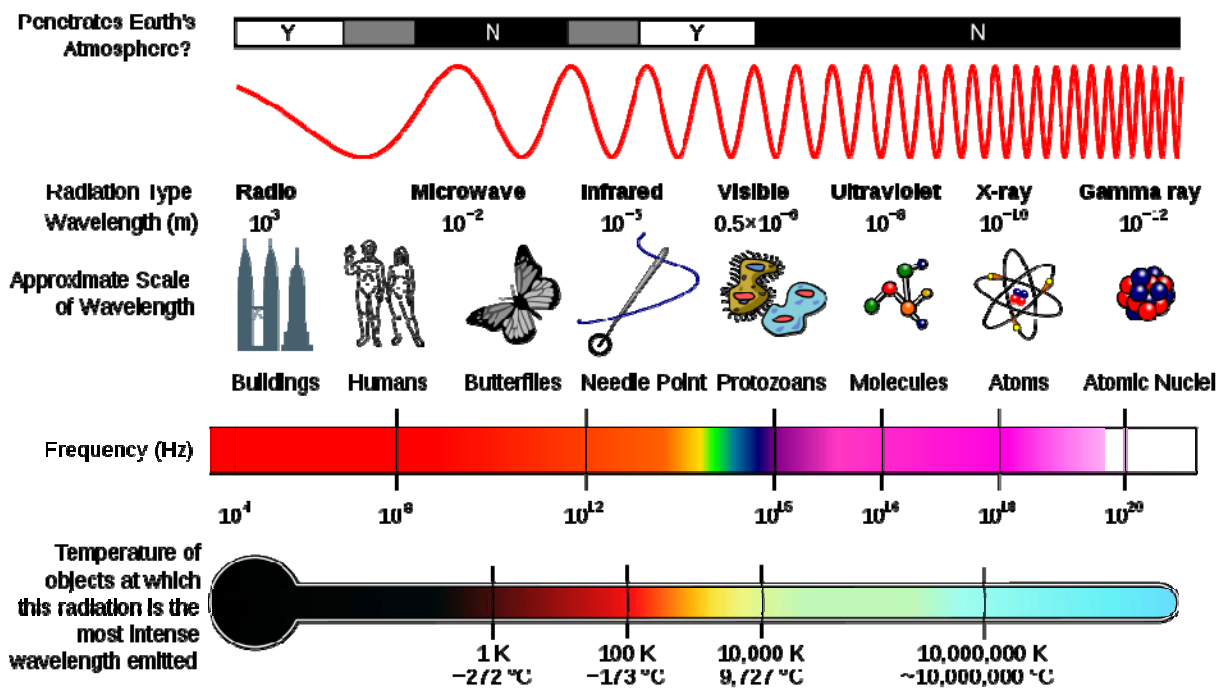


Figure 5.3. The electromagnetic spectrum, showing the type, wavelength, frequency and corresponding radiation temperature. Image is from Wikipedia.

Type of radiation	Frequency range	Wavelength range
Ionizing radiation (gamma, X-ray)	>3 PHz	< 100 nm
Ultra-violet radiation	3-0,75 PHz	100-400 nm
Visible light	750-350 THz	400-800 nm
Infrared (thermal) radiation	350-0,3 THz	0,8-1000 μm
Extremely high frequency (EHF)	300-30 GHz	1-10 mm
Super high frequency (SHF)	30-3 GHz	1-10 cm
Ultra high frequency (UHF) <i>Most communication is done in this range: TV, GSM, 3G, WiFi, GPS etc.</i>	3-0,3 GHz	10-100 cm
Very high frequency (VHF) <i>This is the range of FM radio stations</i>	300-30 MHz	1-10 m
High frequency (HF)	30-3 MHz	10-100 m
Medium frequency (MF)	3-0,3 MHz	100-1000 m
Low frequency (LF)	300-30 kHz	1-10 km
Very low frequency (VLF)	30-0,3 kHz	10-1000 km
Extremely low frequency (ELF)	3-300 Hz	> 1000 km
Static fields	0 Hz	Infinite

Table 5.1. Types of electromagnetic radiation

5.5 The electromagnetic spectrum

Wavelength of electromagnetic radiation ranges from the size of atoms to the radius of Earth. Each type of radiation is characteristically different. At one end, we have gamma radiation, X-rays and visible light, at the other end we have extremely low frequency fields generated by the 50 Hz global power network. [Figure 5.3](#) and Table 5.1 show the spectrum of electromagnetic radiations. The most important from this spectrum is visible light, while for communication, the UHF and VHF channels are also very important.

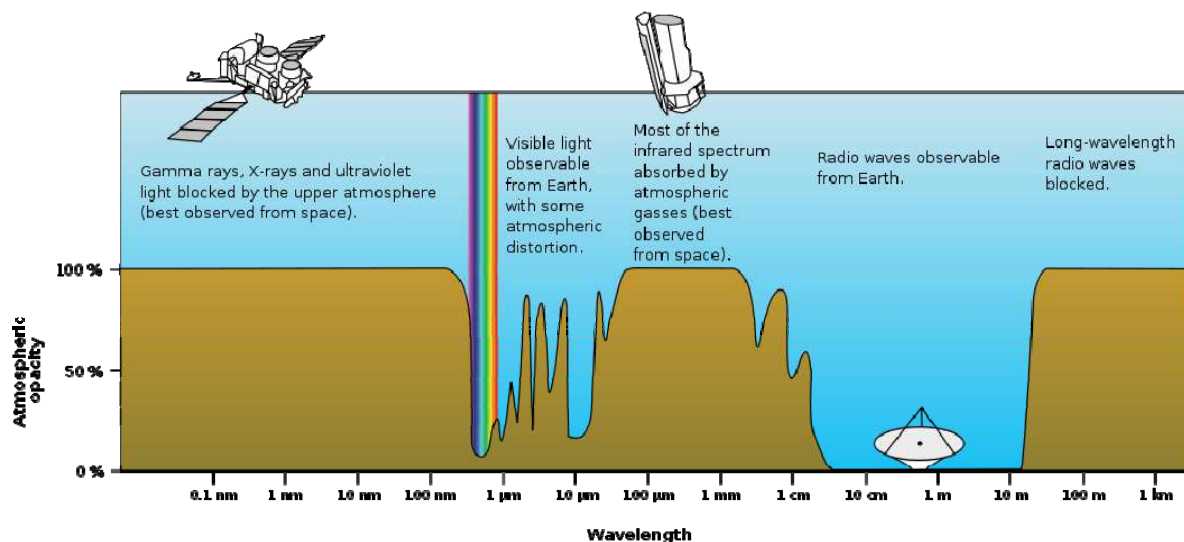


Figure 5.4. Electromagnetic transmittance of the Earth's atmosphere. Image by NASA

Figure 5.4. shows the transmittance of the Earth's atmosphere as a function of radiation frequency. The atmosphere, similarly to radio waves, does not block Visible light and part of the infrared spectrum. Other radiations are not transmitted through the thick layer of air. For example this blocking of ultraviolet-, X- and gamma-rays enables life to be evolved on Earth.

For more details about electromagnetic fields and radiations, see descriptions of lab courses “Microwave electromagnetic radiation” and “Low-frequency electromagnetic fields”.

5.6 Intensity of electromagnetic radiations

There is one more topic to discuss: the intensity of electromagnetic radiations. We did not acquire all information available from Maxwell’s equations. If we calculate the vector product of the curl equations multiplied by E and B , the following equation comes out:

$$-\nabla \frac{1}{\mu_0} E \times B - \frac{\partial}{\partial t} \left(\frac{\varepsilon_0 E^2}{2} + \frac{B^2}{2\mu_0} \right) = 0$$

This is a continuity equation, and it can be interpreted as the continuity of energy density with, if energy density e and intensity vector S (called the Poynting-vector) are expressed as:

$$e = \frac{\varepsilon_0 E^2}{2} + \frac{B^2}{2\mu_0},$$

$$S = \frac{1}{\mu_0} E \times B,$$

where the relationship $S=ce$ holds. Note furthermore that according to quantum theory, the energy of the electromagnetic waves are carried by quanta, the so-called photons. The energy of one quantum, according to quantum theory, is $E=hf$, where h is Planck’s constant with the value $h=6.63 \times 10^{-34}$ Js, and f is the frequency of the wave. An alternatively valid expression is $E=\hbar\omega$, where $\hbar=1.05 \times 10^{-34}$ Js= 6.58×10^{-16} eV·s. One quantum of visible light carries thus the energy of 5 eV (expressed in electronvolt units, which is the energy of one electron accelerated by one volt potential difference), while gamma particles have more than twenty thousand times higher energy, as the minimum energy of a gamma photon is at least 100 keV.

6. Low-frequency electromagnetic radiation of common household devices (ESZ)

6.1 Introduction

In the 20th century science and technology was being developed with huge steps. The output of the industry of the developed countries was getting larger and larger, and parallel to that the amount of energy used was also growing faster and faster. Electric energy (electricity) was produced in centralized power plants, and the distribution system needed to transport and distribute this energy was also developed. At the same time, devices started to work in households, which use this energy. Nowadays the average electric power needed by a general household is approximately a couple hundred watts. For energy transportation electric networks were developed with alternating current, and these work nowadays under well finished standards. Usage of alternating current has many advantages, for example it can be transformed to high voltage before transportation, and large voltage means small current, so energy loss can be much decreased this way. The small voltage direct current networks are intensely heating up the wires, while this heating effect is negligible in high voltage networks. The frequency of alternating current networks is standardized; it is $f=50$ Hz in Europe, while it is 60 Hz in the USA. This means that electrons oscillate in the wire and the electric and magnetic field carries the energy towards the customer. Periodicity of the oscillation is (in Europe) thus $T=1/f=20$ ms

Electric systems have a large significance in environmental physics: the electric and magnetic fields are mostly not constrained or localized to the given device, and they might enter into the human body. They can deposit their energy, heat up the cells of the given tissue, disturb the electric pulses of the human body or the nervous system; they can even interfere with the hormone production. It is well known that brain waves and heart functions are both connected to an electromagnetic activity. Do electromagnetic fields appear outside our devices and interfere with our body? Definitely they do, if there are electromagnetic radiations. In the vicinity of low frequency devices there is no real radiation, but rather (alternating) electric and magnetic fields, even if electromagnetic shielding is present. Near 50 Hz overhead power transmission lines the electromagnetic field is quite large, but radiation power is very small. Devices that emit electromagnetic radiation are radio stations or mobile phones and their transmission towers. In almost any household device has a 50 Hz alternating field in it, the high frequency fields are however mostly produced by the devices themselves. Such devices are: hair-dryer, television, microwave oven, etc. The magnetic effect of the large current in them, or their electromagnetic field extends to outside the device. The effect of these electromagnetic fields on humans is by no means clear; it is a subject of research even nowadays.

In this lab course we will acquaint ourselves with electromagnetic (EM) fields and the measurement of these; by measuring the EM field emitted by some common household devices.

6.2 Basic notions of electrodynamics

6.2.1 *Electric and magnetic field*

When describing electricity, the quantity of electric field strength is used. This is a vector quantity, equal to the force that would act upon a unit charge placed at the given spatial location. It can be defined as $\vec{E} = \vec{F}/q$. Electrostatic field is generated around a point-like

charge (e.g. an electron). If the charges are moving, then the field will be changing in time. Magnetic field is also generated around such moving charges or time-varying electric fields, e.g. around a conductor with current flowing in it. We may know that there is a static magnetic field in a given point that it turns a little compass (magnetic dipole) to its own direction. From the torque of the rotation the strength of the static magnetic field (the magnetic induction) can be calculated. This is also a vector quantity, we will denote it by \vec{B} . In the lab course however, we will be dealing with time-varying magnetic fields.

6.2.2 Magnetic field of current

Moving charges create magnetic field also, besides electric field. The strength of the created magnetic field is described by the Biot–Savart law which we do not detail here. However, from this law the so-called Ampère’s law can be derived, which is part of the Maxwell-equations. This we also do not detail here, only one of its consequences, the magnetic field of a long straight conductor. This consequence states that if current is flowing in a long straight conductor (for example an overhead power transmission line), then magnetic field is generated around it, and its strength is:

$$B = \frac{\mu_0 I}{2\pi r}$$

where r is the distance from the conductor, I is the current flowing in the conductor, while μ_0 is a natural constant of the value $4\pi \times 10^{-7}$ Vs/Am, called the vacuum magnetic permeability. This means that if 100 A current is flowing in a conductor and we are 10 m below it, then the magnetic field there will be $B = 2 \mu\text{T}$. However, in overhead power transmission lines the current is always an alternating current, i.e. the time dependence is $I = I_0 \sin(2\pi f t)$, so the magnetic field will also be alternating. In this case the magnetic field will be

$$B = \frac{\mu_0 I_0 \sin(2\pi f t)}{2\pi r},$$

i.e. for the amplitude of the magnetic field one gets

$$B_0 = \frac{\mu_0 I_0}{2\pi r}.$$

The direction of the magnetic field (B) is pointing always towards the tangent of a circle going through the given point, with the conductor being its center (this is called the right-hand law, as if the thumb of the right hand is directed towards the direction of the current, then the curved palm shows the direction of the magnetic field).

6.2.3 Electromagnetic induction

Since the experiments of Michael Faraday it is well known that time-varying magnetic field induces a curly electric field around it, and also vice versa, time-varying electric field generates magnetic field. Thus electricity and magnetism are interlocked, hence the name electromagnetism. According to Faraday’s law of induction tension (voltage) is induced in a ring-shaped (or any shape that forms a closed loop) conductor, if the number of magnetic induction lines (field lines) is varying in time. The number of field lines (otherwise called magnetic flux) can be calculated as $\Phi = BA$, where A is the area of the circle (or the area closed in by the loop), and B is the component of the magnetic induction that is perpendicular to the area mentioned before. A magnetic field parallel to the area cannot generate induction! In case of a closed loop, the induced tension, according to Faraday’s law, will be the time derivative of the flux (number of field lines) going through the given area:

$$U_{ind} = -\frac{\partial\Phi}{\partial t}$$

This equation is the third of the four Maxwell equations, which completely describe electromagnetic phenomena. If we not only have one circular conductor, but we have N loops reeled around a pipe, then we are talking about a solenoid (a straight coil of wire). In this case the induced electromotive force (induced voltage) is N times the one given before. Given the definition of magnetic flux, we get:

$$U_{ind} = -N \frac{\partial\Phi}{\partial t} = -N \frac{\partial(BA)}{\partial t}$$

Based on this and the laws mentioned in the previous parts, if we put a solenoid near an overhead transmission line, then voltage is induced on the solenoid. This is caused by the magnetic field of the overhead conductor, which is an alternating magnetic field, as discussed above. If a magnetic field with the time dependence of $B=B_0\sin(2\pi f t)$ is assumed, then one gets

$$U = -N \frac{\partial(BA)}{\partial t} = -NAB_0 \frac{\partial}{\partial t} \sin(2\pi f t) = -NAB_0 \omega \cos(2\pi f t) \equiv -U_0 \cos(2\pi f t)$$

thus the amplitude of the current will be

$$U_0 = NAB_0 2\pi f$$

which can be rearranged to get the strength of the magnetic field:

$$B_0 = \frac{U_0}{NA \cdot 2\pi f}$$

It is important that this is only true for a harmonically varying magnetic field (with sine/cosine time dependence). It is important to use the real frequency of the given conductor or household device. One also should use SI units in the above formula (i.e. units of V, m² and Hz), this way the result will be in T (Tesla) units. Note that 1 T is a very large magnetic field, so in practical results the unit of μT should be used.

As in the previous example, if we have a transmission line with a current amplitude of 100 A, a frequency of 50 Hz, and we are in a distance of 10 m (recall that the amplitude of the magnetic field is $B_0 = 2 \mu\text{T}$ here), then (if we know all properties of the solenoid) the amplitude of the induced voltage can be calculated. If the number of loops is $N=1000$, the area is $A=10 \text{ cm}^2$, while the frequency and the magnetic field is as given before, then according the previous formula, the voltage amplitude will be 0.63 (double check this result, based on the previous formulas!). In the above formula, $1/NA2\pi f$ may be regarded as a conversion constant, as if one multiplies the measured voltage amplitude with it, one simply gets the amplitude of the magnetic field.

6.2.4 Self-inductance

If current is conducted into a solenoid, then magnetic field is generated in it. The strength of this magnetic field is described by the fourth Maxwell-equation, and the result is (if the number of loops is N , the length is ℓ , and the current is I):

$$B = \frac{\mu_0 N}{\ell} I$$

with $\mu_0=4\pi\times 10^{-7}$ Vs/Am being the vacuum magnetic permeability. If, however, the current is time-varying, then the magnetic field is also time-varying, thus the changing magnetic flux in the coil is generating an induced voltage in the solenoid. This is called self-induction, and the solenoid might be characterized by its self-inductance. This means that if the current in the solenoid is changing, then only because of this, voltage is generated in the solenoid (no outer magnetic field is needed).

The magnetic field is pointed towards the axis direction of the solenoid, and is homogeneous with a good approximation inside the solenoid. Note that the magnetic field lines cannot be discontinued or ended anywhere however, so there is a small scattered magnetic field outside the solenoid as well – where the field lines turn back outside the loops. If the current in the solenoid is time-varying with the function $I(t)$, then the induced voltage is, according to Faraday's law:

$$U_{ind} = -N \frac{d\Phi(t)}{dt} = -NA \frac{dB(t)}{dt} = -NA \frac{\mu_0 N}{\ell} \frac{dI(t)}{dt} = -L \frac{dI(t)}{dt},$$

where L is the self-inductance of the solenoid, it has the unit Henry (H, equal to Vs/A). The minus sign means that the induced voltage has such a direction as to oppose the cause that is generating it – as stated by Lenz's law. The value of the self-inductance can be read off from the previous formula:

$$L = \frac{A\mu_0 N^2}{l}$$

6.3 Non-ionizing radiation and its effects

There are a lot of research results available in this field, in particular about the heat effect of high frequency electromagnetic radiation and its absorption in the human body. It is important to investigate the biological effects of radar- and radio-technology, of household devices or medical applications (e.g. magnetic resonance imaging, MRI).

Absorption of radiation in human tissue is determined by the electric permittivity and magnetic permeability of the given tissue. Energy is absorbed through dielectric polarization. If the time period of the oscillation of the outer electromagnetic field is approximately the same as the typical movement period (vibration, rotation) of the small dipoles (e.g. water molecules), then maximal absorption can be observed. This is the method of heating up water molecules in microwave ovens. Electric permittivity of biologically important materials depends strongly on frequency, and it differs largely from that of the air. Thus the amount of absorbed radiation (and so its biological effect) depends strongly on frequency. For example below 100 kHz the cell membrane screens the outer electric field, only higher frequency waves can enter the cell. The cell membrane, macromolecules, proteins, amino acids, peptides, water molecules all absorb electromagnetic radiation in different frequency ranges (frequency is growing parallel to the ordering of the above list). This absorption may have importance in medical diagnostics, too.

6.3.1 Units of dosimetry

In order to study the biological effects of radiofrequency and microwave radiation we use the standardized notions of dosimetry. Unit of electric field strength is V/m, that of magnetic induction is Tesla (T), while power per unit area (sometimes called surface power density or specific power) is proportional to the product of the two, and its unit is W/m². As in case of point-like sources with constant power, the surface power density is inversely proportional to the square of the distance (as the area is squared proportional to the radius). This is only true

however in the radiation zone, when being far enough from the source of radiation (when the distance is several orders of magnitude bigger than the radiation wavelength).

However, in case of low frequency fields (below 10 kHz) the absorption in the human body is described by the current density in the tissue, using the unit A/m^2 . For example a horizontal magnetic field of $1 \mu T$ with a frequency of 50 Hz generates a current density of almost $5 \mu A/cm^2$.

6.3.2 Radiation load of the general population

Frequency of regular CRT (cathode ray tube) computer monitors is between 15 and 60 kHz, this is the frequency with which the electron beam is deflected (by a magnetic field). If sitting in front of such a monitor, one may experience a 10 V/m electric field or a $0.2 \mu T$ magnetic field approximately. In case of very low frequency fields, 50 Hz fields are the most important because they are present in almost every household device (connected to the power network). The static magnetic field of the Earth is roughly $50 \mu T$ in Budapest (this is a constant value, so this does not appear in the 50 Hz frequency range), its natural oscillations are smaller than a couple times $0.01 \mu T$. The natural low frequency background radiation around 50 Hz is roughly $0.0005 \mu T$. The artificial sources in this frequency range, present in households, are much bigger, approximately $0.2-0.3 \mu T$. Near a 756 kV overhead power transmission line (standing on the ground) the amplitude of the magnetic field may be as large as $30 \mu T$. In the absolute proximity of an electric shaver this value might be bigger by a couple orders of magnitude, as big as $3000 \mu T$. In electric power plants, average magnetic fields are around $40 \mu T$, with rare maxima of roughly $300 \mu T$. Exposition of welding workers may be as big as $130000 \mu T$.

In case of low frequency fields, the biological effect of magnetic field turns out to be more important than that of the electric field, so during the lab course we will measure magnetic fields.

6.3.3 Radiation protection standards

According to recommendations from the International Commission on Non-Ionizing Radiation Protection, limits of exposure due to health reasons are defined. In case of a 50 Hz frequency magnetic field and constant exposure, residential threshold is $100 \mu T$, while the threshold for a professional environment is $500 \mu T$.

6.4 Lab course tasks

The measurements (MARADJON, minden egyes "task"-ot measurementnek hívunk) are done in groups of four. We will have two solenoids available, so two subgroups of two have to be formed, and each will work with one solenoid during the measurement. The tasks are as follows:

1. Determine the conversion constant of the solenoid as described in the previous sections. In order to do that, you will need the number of loops and the cross-section area of the solenoid, and also the field frequency. Calculate from the induction law that if 1 mV voltage is induced in the solenoid, how much μT magnetic field it corresponds to (assume a sine formed time dependence in each case), for a given frequency. What is the uncertainty of this conversion constant? Remember that if you measure the induced voltage with a voltmeter, the displayed value will not be the amplitude U_0 , but the effective voltage, corresponding to $U_0 / \sqrt{2}$!

2. Measure the magnetic field of several devices: a CRT monitor, an electric shaver, a hair dryer. Compare the resulting values to health thresholds and other values given above!
3. There is an underground power transmission line near the Danube side elevators of the North Building. Measure their magnetic field! First you have to determine the direction of the lines (or that of the magnetic field) by measuring the induced voltage in different directions. If the magnetic field is orthogonal to the cross-section of the solenoid, the voltage is maximal. If the direction is determined, try to find the location of the line by moving the solenoid around. How big is the biggest amplitude (in μT)? How big is this value, compared to natural backgrounds, to the static magnetic field of the Earth or health thresholds? Try to shield the magnetic field with your hand or other objects. What do you experience?
4. Measure the amplitude of the magnetic field as a function of distance from the floor, in steps of couple of centimeters. To be able to perform an uncertainty calculation, everyone shall do the measurement twice (at each location). Make a table from the results; tabularize the distance r versus measured voltage amplitude U_0 and magnetic field amplitude B_0 . Make also a graph plot of B_0 versus r !
5. The next task is to determine the current flowing in the transmission line; and the depth of the line under the floor. The distance dependence of magnetic field, as described in the previous sections, is

$$B_0 = \frac{\mu_0 I_0}{2\pi r},$$

where r is the real distance from the conductor, and this we do not know. Assume however, that the conductor line is in a depth of d , and let r be the distance we measure (from the floor). In this case our formula is modified as

$$B_0 = \frac{\mu_0 I_0}{2\pi(r+d)}.$$

Let us take the inverse of this:

$$\frac{1}{B_0} = \frac{2\pi}{\mu_0 I_0}(r+d).$$

This is the equation of a linear curve, if $1/B_0$ is plotted versus r . Hence make a graph where $1/B_0$ is plotted on the vertical axes, while the measured distance r is plotted on the horizontal axis. Our data points will follow a straight line (a linear curve). Make a fit with a $y=ax+b$ function! From the equation of the linear curve ($y=ax+b$), the amplitude can be calculated ($a=2\pi/\mu_0 I_0$), and the depth of the conductor, d , can also be derived.

6. Estimate the uncertainties of the calculations. Do the above exercise for both measured series and estimate the difference of the results. The average will be the final result, while the difference of the values from the average will be the uncertainty of the result.
7. Try to determine, how far one should be from the transmission line to be exposed to a magnetic field of 10, 100 or 1000 μT . Do the same for the field of one of the earlier investigated devices, e.g. the hair dryer.

6.5 Comments, test questions

In the lab report, note the properties of the solenoid you used. Prepare a transparent table of the measured and calculated quantities. In the table, one might use the convenient units of cm, mV or μT . In the calculations however, SI units should be used. Do not give any value with a pointlessly high number of decimal digits, pay attention to a rounding corresponding to the uncertainty of the given value. Also watch out to set the scales of the plots in such a way that each measured data points are on the plot, but there is not too much free space, either. If you use Microsoft Excel, use the plot type XY (scatter plot). When calculating the final result, check if the values are meaningful. Suspect an error in your calculations if the results are absolutely not real (i.e. a current of $I=10000$ A, a depth of $d=100$ m or a magnetic field of $B=100$ T). Work separately, do not copy the errors of your course-mates.

1. How do you calculate the induced voltage in a solenoid, if the magnetic field (B) and its time dependence is known?
2. What is the unit of magnetic field (magnetic induction)? What are the health thresholds associated with it?
3. What are the residential health thresholds for 50 Hz frequency magnetic fields in case of constant exposure?
4. What will be our main device to measure the magnetic field, and what are its properties that we need to know in order to perform our tasks?
5. How does the self-inductivity of a solenoid depend on its number of loops or geometrical sizes?
6. How do you calculate the amplitude of the magnetic field from the measured amplitude of the induced voltage, if the time dependence is sine-like?
7. How big is the static magnetic field of the Earth? And how big is the typical natural background radiation in the low-frequency range?
8. On the order of magnitude, how big of a magnetic field could we measure underneath an overhead power transmission line?
9. What determines the magnetic field below an overhead transmission line, its voltage, its current or the transmitted power?
10. How can you determine the direction of the magnetic field, with the help of a solenoid?
11. Assume you have a solenoid with 2000 loops and an area of 3 cm^2 . What was the magnetic field, if the induced voltage is as high as 10 mV?
12. How big is the magnetic field in a distance of 3 meters from a straight conductor carrying 1 A current?
13. We plot the measured magnetic field of an overhead transmission line as a function of distance (recall Ampère's law given in the previous sections). What kind of curve do you expect on a B versus r graph?

7. Microwave radiation of household devices (MIK)

7.1 Introduction

Technology is being developed in large steps nowadays. In every household or workplace, a vast amount of devices are used that are working with electric power. Some of them are communicating with each other or some central servers through wireless communication channel. Wireless communication happens with the help of electromagnetic radiation; i.e. it is present when using radio, mobile phones, WiFi, Bluetooth or GPS devices. Other devices use electromagnetic radiation for other purposes, like the electromagnetic radiation inside a microwave oven is used to heat up food. These systems have a large significance in environmental physics, as the electric and magnetic fields are mostly not constrained or localized to the given device, and they might enter into the human body. They can deposit their energy, heat up the cells of the given tissue, disturb the electric pulses of the human body or the nervous system; they can even interfere with the hormone production. In this lab course we will measure the electromagnetic (EM) radiation of several common household devices.

7.2 Electromagnetic radiation

In radio communication, radiation is generated in a straight conductor (an antenna or a transmitter) where a lot of electrons are moving upwards and downwards. The direction of current is alternating in time like a sine function. It is well known that current generates magnetic field, as described by the Ampère-law:

$$B = \frac{\mu_0 I}{2\pi r},$$

where r is the distance from the conductor, I is the current flowing in the conductor, while μ_0 is a natural constant of the value $4\pi \times 10^{-7}$ Vs/Am, called the vacuum magnetic permeability. Because of this, if the current is alternating, the magnetic field is alternating as well. The alternating magnetic field generates an electric field in any closed loop, according to Faraday's law (where instead of the electric field, electric tension is given):

$$U_{ind} = -\frac{\partial BA}{\partial t},$$

where A is the area closed in by the loop, and B is the component of the magnetic induction that is perpendicular to the area mentioned before. However, this electric field will also be alternating in time, thus again a magnetic field is generated. This way the fields are detached from the original conductor (the antenna) and will be (almost) freely propagating in space. This is called electromagnetic radiation, and is described by the wave-solution of the vacuum Maxwell's equations.

The radiation field can be detected far away from the transmitter also, its intensity decreases with distance. If being very far from the source (in a much bigger distance than the wavelength), the electric and magnetic fields have a well-known structure (as resulting from Maxwell's equations). Let us denote the direction of the wave propagation k . The electric (E) and magnetic fields (B) are orthogonal to this vector, and also orthogonal to each other, according to the right hand rule (thumb finger is k , index finger is E and middle finger is B). The vector of the electric field, E , is parallel to the direction of the original antenna. The size of both field vectors is then oscillating in their fixed direction.

The oscillating electric and magnetic field vectors thus form a transversal electromagnetic wave (transversal, as the quantities oscillate orthogonal to the direction of propagation). The propagation velocity is the speed of light (c) in the given medium; also the general $c = \lambda f$

relationship for waves holds, with f being the frequency of the radiation and λ its wavelength. Thus the wavelength of a 50 Hz radiation (generated by devices connected to the regular electric power network) is:

$$\lambda = \frac{3 \cdot 10^8 \text{ m/s}}{50 \text{ Hz}} = 6 \cdot 10^6 \text{ m} = 6000 \text{ km},$$

which is on the same order of magnitude than the radius of the Earth. The radiation power of low frequency household devices is however very small, due to the small frequency; they produce an alternating electric and magnetic field around themselves instead. These devices are far closer to us than the wavelength. Mobile phones and similar devices use frequencies around 1 GHz however. The wavelength associated with this frequency is

$$\lambda = \frac{3 \cdot 10^8 \text{ m/s}}{10^9 \text{ Hz}} = 0.3 \text{ m},$$

i.e. in this case the radiation field idea can be safely used. Wavelength of radio transmitters falls in between, their typical frequencies are around 500 kHz – 200 MHz.

7.3 Categorization of electromagnetic radiation

Non-ionizing radiations are electromagnetic radiations above a wavelength of 100 nm. The quanta of these radiations (photons) carry energies less than 10 eV, they are thus unable to change the electron structure of atoms or molecules. Thermal infrared radiation of room temperature (or temperatures on the same order of magnitude, in Kelvin of course) objects corresponds to a frequency of 1000-10000 GHz. [Table 5.1](#) contains the names of radiations in different frequency or wavelength ranges.

Type of radiation	Frequency range	Wavelength range
Ionizing radiation (gamma, X-ray)	>3 PHz	< 100 nm
Ultra-violet radiation	3-0,75 PHz	100-400 nm
Visible light	750-350 THz	400-800 nm
Infrared (thermal) radiation	350-0,3 THz	0,8-1000 μm
Extremely high frequency (EHF)	300-30 GHz	1-10 mm
Super high frequency (SHF)	30-3 GHz	1-10 cm
Ultra high frequency (UHF) <i>Most communication is done in this range: TV, GSM, 3G, WiFi, GPS etc.</i>	3-0,3 GHz	10-100 cm
Very high frequency (VHF) <i>This is the range of FM radio stations</i>	300-30 MHz	1-10 m
High frequency (HF)	30-3 MHz	10-100 m
Medium frequency (MF)	3-0,3 MHz	100-1000 m
Low frequency (LF)	300-30 kHz	1-10 km
Very low frequency (VLF)	30-0,3 kHz	10-1000 km
Extremely low frequency (ELF)	100-300 Hz	> 1000 km
Static fields	0 Hz	Infinite

Table 7.1. Types of electromagnetic radiation

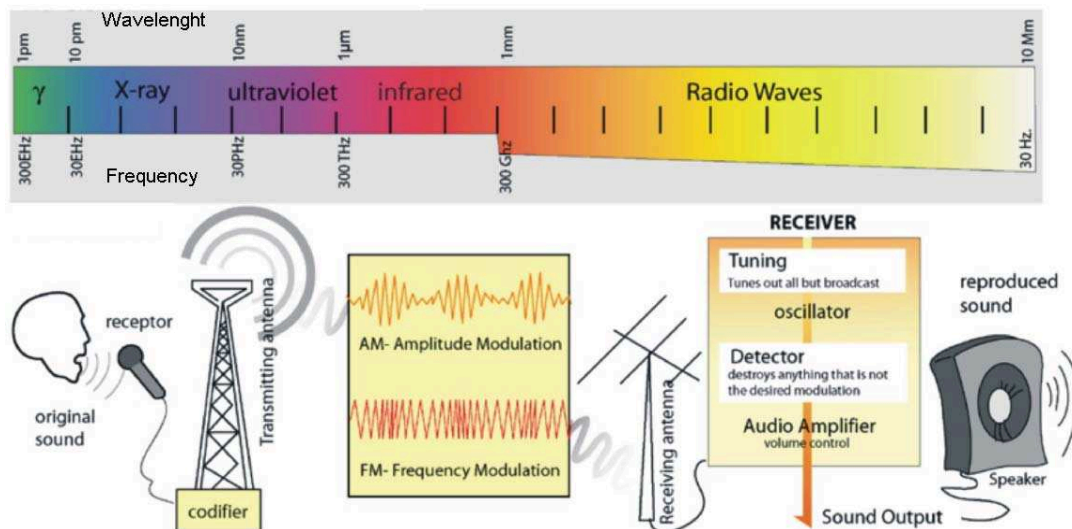


Figure 7.1. The electromagnetic spectrum and the principle of radio communication

7.4 Radiation power of electromagnetic sources

Assume that radiation is not absorbed (converted into heat) in the medium, and draw imaginary spheres around the source. The energy flowing through these spheres per unit time is the same for all spheres, because no energy was lost in the medium. Imagine a hose that sprinkles water all around in the air. If one liter of water is sprinkled out in a second, then on all imaginary spheres (drawn around the sprinkler as source) one liter water is flowing through per second, independently of the radius, whether it is 10 cm or 1 m – because the water did not get lost during the sprinkling (we neglected the possibility of the water falling down on the ground). This is illustrated in Figure 7.2.

Thus the energy per unit time flowing through the spheres of different radii is constant. However, the intensity is defined as energy flowing through unit area per unit time, i.e.

$$I = \frac{E}{At} = \frac{P}{A},$$

where P is the power of the source. This way we get a simple estimate of the distance dependence of the radiation intensity. Recall that the area of a sphere with radius R , i.e. being in a distance of R from the source is $A=4R^2\pi$. Using this one gets:

$$I(R) = \frac{P}{4R^2\pi}.$$

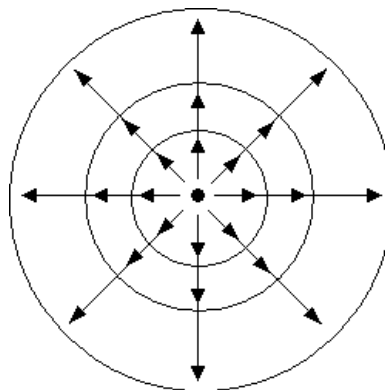


Figure 7.2. Illustration to the propagation of radiation from a point-like source

7.5 Effects of non ionizing radiation

There are a lot of research results available in this field, in particular about the heat effect of high frequency electromagnetic radiation and its absorption in the human body. It is important to investigate the biological effects of radio-technology, of household devices or medical applications.

Absorption of radiation in human tissue is determined by the electric permittivity and magnetic permeability of the given tissue. Energy is absorbed through dielectric polarization. If the time period of the oscillation of the outer electromagnetic field is approximately the same as the typical movement period (vibration, rotation) of the small dipoles (e.g. water molecules), then maximal absorption can be observed. This is the method of heating up water molecules in microwave ovens. Electric permittivity of biologically important materials depends strongly on frequency, and it differs largely from that of the air. Thus the amount of absorbed radiation (and so its biological effect) depends strongly on frequency. For example below 100 kHz the cell membrane screens the outer electric field, only higher frequency waves can enter the cell. The cell membrane, macromolecules, proteins, amino acids, peptides, water molecules all absorb electromagnetic radiation in different frequency ranges (frequency is growing parallel to the ordering of the above list). This absorption may have importance in medical diagnostics, too.

7.5.1 Units of dosimetry

In order to study the biological effects of radiofrequency and microwave radiation we use the standardized notions of dosimetry. Unit of power per unit area (sometimes called surface power density or specific power) is proportional to the product of the two, and its unit is W/m^2 . In case of point-like sources with constant power, as described above, the surface power density in the radiation zone is inversely proportional to the square of the distance.

Absorbed dose of a human body or any other biological system is characterized by the specific absorption rate (SAR). This gives the absorbed power per unit mass in units of W/kg . The quantity specific absorption (SA) is the time integrated version of the SAR, i.e. it is the total absorbed energy per unit mass, measured in J/kg . In the radiofrequency and microwave range the absorption rate is mainly characterized by the frequency and the water content of the given tissue.

Penetration depth of the electromagnetic radiation is the distance from the body surface (into the body) where the intensity of the radiation falls to $1/e$ (roughly 36.8%) of its original value (at the surface). For example, at a frequency of 915 MHz (microwave oven) penetration depth is round 3 cm in tissues with high water content, and 18 cm in tissues with very low water content. Penetration depth increases with the decrease of frequency, at 10 MHz it is 10 cm even in case of water rich tissues.

Tissues rich in water are for example muscles, skin, brain tissue, internal organs; tissues with low water content are for example bones and adipose tissue (fat). The SAR of a radiation absorbed in the human body is done via different calculation models, the results are usually given as a function of incoming surface power density and of frequency. For example, at 1 GHz frequency and $1 mW/cm^2$ surface power density a SAR of $10 mW/kg$ can be assumed. This value is strongly decreasing for frequencies below 100 MHz.

7.5.2 Radiation exposure of the population

Natural environmental background radiation in the radiofrequency range is smaller than $0.0014 \mu W/m^2$. Far away from a radio transmitter, at a distance of r , the surface power density is roughly $0.13 P/r^2$, where P is the effective radiating power of the radio source. Typical

radiation exposure values for the population (from artificial sources) are, in the FM, VHF and UHF range: $50 \mu\text{W}/\text{m}^2$ (70's, USA), $20 \mu\text{W}/\text{m}^2$ (Sweden, a metropolis office building), $<1000 \mu\text{W}/\text{m}^2$ (Hungary, any typical residential or working area). In a distance of 30-40 meters from a 100 W antenna, surface power density is roughly $10,000 \mu\text{W}/\text{m}^2$. According to national and international standards, in a 5 cm distance from a microwave oven the maximal surface power density allowed is $5 \times 10^7 \mu\text{W}/\text{m}^2$.

7.5.3 Standards of radiation protection

Recommendations of the International Commission on Non-Ionizing Radiation Protection contain health thresholds of non-ionizing electromagnetic radiations. There are different thresholds for residential and occupational radiations. In the 10 MHz – 10 GHz range the residential SAR threshold for the total body is 0.08 W/kg, for the head 2 W/kg, while it is 4 W/kg for the limbs. Occupational thresholds are five times as large. In case of low frequency fields the values are given in units of magnetic field, but we do not detail those thresholds here.

7.6 Lab course tasks

The measurements shall be done in groups of four. There are two devices available for investigation: mobile phones and a microwave oven.

1. Investigate our measuring device. What physical quantity does it display? What are its measurement limits (or range)? Try to measure the radiation of a Wi-Fi router or a mobile phone! What are typical displayed values?
2. Measure the intensity of the mobile phone and the microwave oven at different distances. You will see that below a given distance (roughly 2-3 meters) the radiation is so strong that it is outside the range of the device. Make sure the measurement is consistent, try to repeat the measurements exactly the same way, just at different distances. In order to be able to perform a uncertainty calculation, everyone shall have at least two series of data points.
3. Make a table about the measurements, with columns like the measured distance (r , in meters), intensity (I , in W/m^2). Plot the intensity values as a function of distance! How does it look like, what curve it is similar to? What did you expect?
4. The relationship between intensity and distance is, as expected, $I=P/4\pi r^2$. Thus replot the graph as follows: the vertical axis shall be unchanged, but on the horizontal axis, plot $1/4\pi r^2$ instead of the distance. What curve do you expect? It is a linear curve, and its gradient is nothing else than P , the power of the radiation source. Determine this value by fitting a linear function on the data points!
5. Perform an uncertainty calculation by repeating the above exercise for both datasets. The average of the fit parameters will be the final result, while the difference will be our uncertainty.

Prepare a transparent table of the measured and calculated quantities. In the table, one might use the convenient units of cm or mW/m^2 . In the calculations however, SI units should be used. Do not give any value with a pointlessly high number of decimal digits, pay attention to a rounding corresponding to the uncertainty of the given value. Also watch out to set the scales of the plots in such a way that each measured data points is on the plot, but there is not too much free space, either. If you use Microsoft Excel, use the plot type XY (scatter plot). When calculating the final result, check if the values are meaningful. Suspect an error in your

calculations if the results are absolutely not real (i.e. a power $P=100$ W, etc.). Work separately, do not copy the errors of your course-mates.

7.7 Test questions

1. Can the electromagnetic field or radiation enter the human body? How does the penetration depth depend on frequency?
2. What is the wavelength of visible light and that of radio waves? What is the wavelength of your favorite radio station?
3. What is the unit of absorbed dose? What are health thresholds connected to it?
4. How can you calculate the frequency of a wave from its periodic time?
5. How can you calculate the wavelength of a wave from its velocity and frequency?
6. In case of constant power radiation sources, how does the intensity depend on distance?
7. The frequency of a radi station is 100 MHz. What is its wavelength?
8. How big is the measured intensity in 10 m distance from a radiation source with the power of 10 mW?
9. We plot the distance dependence of a point-like electromagnetic radiation source (e.g. a mobile phone). What kind of curve do you expect on an I versus r graph?
10. How do you modify the above plot (what should be on the axes) to get a linear curve?
11. What is the wavelength of radiation of mobile phones and microwave ovens?

8. Solar air heat collector (NKO)

According to a large survey 70% of the energy consumption of the population is used for heating. Although the price of traditional fossil energy sources are increasing and are important factors in pollution, the usage of renewable energies in heating is very limited at the moment. The main reason of this is the energy transformer instruments – like solar cells, solar collectors, wind turbines as examples – are too expensive for a general family and the payback times are too long. During this laboratory we are working with such an alternative solution that converts the energy of the solar radiation into heat of air and it can be built up at home, therefore it is relatively cheap. On the other hand the efficiency will be not very high. This is called can-collector. We need empty used aluminum cans only and easy purchasing wood goods. The main concept of the instrument is that we ventilate the air of a room across the cans tight together. The energy of the sunshine will heat up the cans meanwhile and the cans pass the kinetic energy to the air inside that will return to the room on a higher temperature. We will investigate whether this instrument can be used for heating extension for a family house. We will investigate also the efficiency of it and we will compare different technical solutions. We will overview the cost effectiveness of this solution as well. Using a large can-solar collector the whole amount of heating energy needed for a house will not be covered. But it is worth to check if it is worth to build up, and how long the payback time is. How much will it reduce our heating costs? How much will it reduce the amount of pollution materials to the air?

8.1 The overview of solar energy

8.1.1 *The origin of the solar energy*

The Sun is the largest member of the Solar System. The Sun is in gas and plasma phase. In the core of the sun there is hydrogen, helium and heavier elements moving at 15 million K temperatures. The thermonuclear fusion is taking place here that produce the energy of the Sun. During this process helium nuclei are produced out of 4 protons (hydrogen nuclei). The energy produced in the fusion nuclear reaction is converted to photons, neutrinos and kinetic energy. This energy will leave the core and will be transferred through the convective zone of the Sun to the surface during a very long time. Except the neutrinos, those do not react with the matter of the Sun can penetrate easily across the outer layers and carry the energy rapidly out. At the end the surface of the Sun is heated up to about 5800 K. This surface is a few times hundred kilometres wide region is called photosphere. This is the source of the solar electromagnetic radiation that we can see from the Earth.

The frequency distribution of the energy from the Sun is peaked in the visible light, but it has parts in the ultra violet and infrared ranges. The radiative power of the Sun is the energy emitted by the whole star in a second. This is $3,86 \times 10^{23}$ kW. Maximum 10^{-10} part can reach the Earth. On the top of the atmosphere there are 1370 W incoming power on 1 m^2 , if it is perpendicular to the direction of the solar radiation. This value is of course smaller at the surface, moreover if the 1 m^2 is parallel with the Earth's surface. If we make an average of Earth's total surface the 1370 W is reduced to less than 400 W. Solar energy is a geographically largely varying phenomenon. Besides the geographical latitude and the morphology of the surface meteorological conditions are also very important factors. The quantity that summarizes these meteorological features is the total of sunny hours in a year. Hungary is a good place from this point of view. But generally the deserts are the best. In Hungary the total hours of sunshine in a year ranges from 1900 – 2200 hours/year. [Figure 8.1](#)

shows the geographical distribution of this quantity in Hungary. The Great Plain is the best part of the country from this point, but the other areas are also usable for solar instruments.

To demonstrate the amount of the solar energy arriving in Hungary we can investigate the average energy density that is $1760 \text{ kWh/m}^2/\text{year}$. According to the nations about 37 TWh energy need, it can be covered to use the total energy arriving on a 21 km^2 area. But that large energy plant cannot be real, and the energy conversion efficiency is also a big factor. We can say that the energy density of the solar radiation is low.

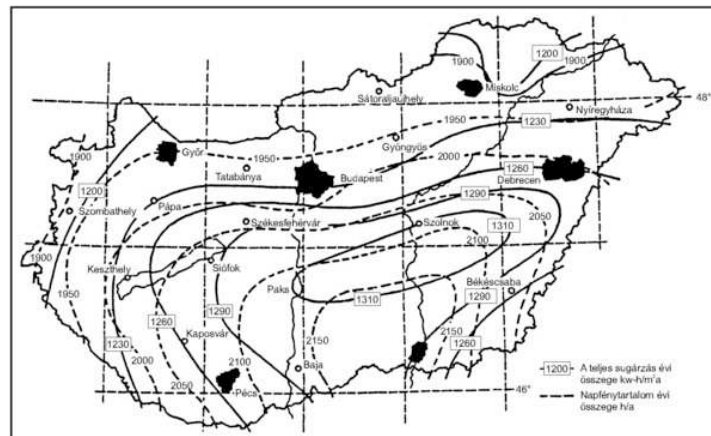


Figure 8.1. The distribution of the total sunny hours per year across Hungary

8.1.2 The usage of solar energy

There are three ways to convert the solar energy for a usable form. One is the solar cell when the energy of the radiation will produce electric current. The second way is the solar collector in which the radiation is directly transferred to heat in an instrument using air or liquid as a transfer media. These two are considered active usage, since in both case an instrument is required. The third way is the passive usage when architecture of a house is designed in a special way to collect the solar energy into the indoor air.

We should mention that almost all the renewable energy types basically use the solar energy. If we seek out what the origin of the energy of the biomass is, we conclude that it is the solar radiation. In the biomass the energy is stored in the form of chemical energy but it is converted during the photosynthesis process from the sunshine. The origin of the wind energy is the pressure difference due to the different heating of the surface of the Earth that results in the wind.

The use of solar units is feasible to Latitude 35. In deserts there are already several solar power plants, but the non-centralized use spreads very hardly.

8.1.3 Solar collectors

We focus on the second active-type usage of the solar energy: the solar collectors. These instruments can be built up onto the roof, and they should be adjusted to south for the better conversion efficiency. In the collector the heat absorber material (generally aluminum or other metals that are painted black) passes the energy of the solar radiation to a heat exchanger material that will carry the heat to raise the temperature of water in a tank. The heat exchanger material can be liquid or air. The liquid is generally antifreeze since it should stand the

meteorological environment on the roof in winter nights when nothing heats it up. See Figure 8.2 for the schematics of the collector. The collector box contains the pipes for the heat exchanger material and in some solution the pipe is put in a glass tube where there is vacuum. This is for the better efficiency, since less heat leakage. In case of the exchanger is air we call it solar air heat collector. That is the example in our laboratory practice. The collector in addition consists of an electronic controlling system that organize the operation of a pump or a fan in the air case. If the temperature of the heat exchanger material in the collector is higher than an adjusted value (threshold temperature) and the water temperature in the tank is lower than the required the pump will start to operate and the water tank will be heated up. The water tank will serve for the people in the house as warm water for bathing, or heating. There are gravitational systems where there is no pump. The density difference of the warm exchanger and the cold exchanger material will drive the flow. For solar thermal power plant there is a solution called parabolic trough, parabolic dishes or there are solar towers. The operation principles of these are always the same as of the collectors described above, but the usage of the heat is different. In these cases it drives a heat power plant. Of course the scale is much larger as well.

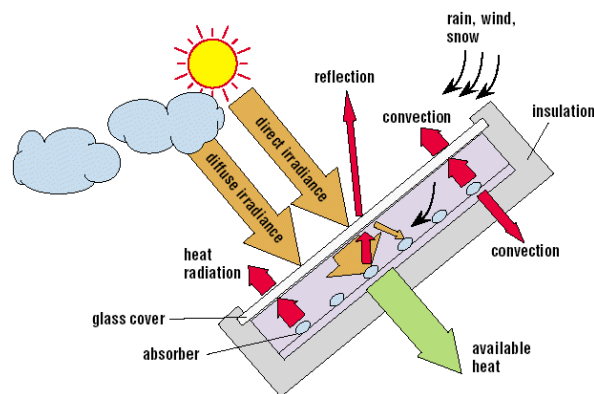


Figure 8.2. The schematics of a solar collector. The orange arrow shows that the energy from the Sun enters to the box, the red arrows represent the energy losses and the green arrow shows the used solar energy that was passed to the absorber and the heat exchanger material that is flowing in the absorber pipes.

Nowadays the time for payback the money invested to the solar collectors is generally 10-15 years at the level of actual energy prices. The tendency of these prices makes the solar collector a more attractive solution in the future. The 10-15 years are relative long compared to the lifetime of the collectors. After about 20 years the collector needs some maintenance and additional costs.

The renewable energy resources are environmental friendly solutions but we can pay attention to the aspects of conservation. These are the most important issues (these include the cases of photovoltaic solar cells as well):

- the passive solar energy usage when planning the house should be considered,
- those technologies should be used when the production phase requires the less environmental harm and which has the less waste,
- in case of storage of the energy produced with solar instruments the use of accumulators should be avoided and considered before installation.

There is one big advantage of using solar energy. It can be used in a decentralized way. There is no need for the transportation of the energy. The resources are always local; the sunshine is

available at every place. In this case the dependence on energy import will decrease and the money expended on these technologies will contribute to the local economy.

8.2 The can-solar collector

This is a solar heat collector in which the heat exchanger material is air, the absorber pipes are made of generally used cans. This is cheap and also uses waste material. The best time period for this instrument is March – April and September – October when the houses still need heating and the solar irradiation is already high enough.

8.2.1 The structure of the collector

The structure of this collector can be seen in Figure 8.3. The absorber pipes are in a wooden box that is insulated from the bottom side by rock wool. There are two cylindrical holes in the bottom of the box. That is for connecting the air pipes into the room. There is a need to drill the wall at the positions of these holes. The absorber pipes are the black tubes. These were made from cans. The bottom of the cans should be drilled and the air will move across the the drinking hole. Then they should be glued together tightly. A plain glass or plastic window should cover the whole box. We use polycarbonate window. That is a cheap and easy to use solution. The window will close the air into the box, which will be also heated up, and the material of the window itself is an insulator that hinders the heat leakage to open air. The polycarbonate is a good insulator since there are air cells in this sheet. The air in the pipes composes a closed system with the indoor air and it is separated from the inside air of the box.



Figure 8.3. The structural view of the solar air heat collector (Photo by Edina Juhász)

The collector used for our experiments consists of 4 pipes and 6 cans in every pipe. This is a relatively small size solution. The glue was Silicoplast type that is heatproof up to 200°C.

The materials and costs

The collector is cost effective since we use waste material. The materials were selected to be cheap but still the efficiency should be high enough. We list the materials and their prices

(valid around 2012) for our collector in Table 8.1. Its sizes are $138 \times 33 \times 17$ cm. The area of the surface of the collector $A = 0,4554 \text{ m}^2$.

Material	Amount	Price
Polycarbonate air cell plain (80% transparency)	$1,58 \text{ m}^2$	1 556 Ft
aluminum sheet (1mm thick)	$0,84 \text{ m}^2$	2 400 Ft
ventilation pipe (\varnothing 125 mm)	1	1 800 Ft
Silicoplast glue	1	1 200 Ft
Tool for the glueing process	1	340 Ft
joint pipes	2	2 000 Ft
Black paint 400 ml/piece	3	4 200 Ft
interface for the fan	1	300 Ft
rock-wool	1 m^2	576 Ft
wallboard	$0,9 \text{ m}^2$	432 Ft
wood for the box	$0,6 \text{ m}^2$	3 500 Ft
Computer fan 15/14 W 50/60 Hz		12000 Ft
	altogether:	30 304 Ft

Table 8.1

8.2.2 The physical background of operation

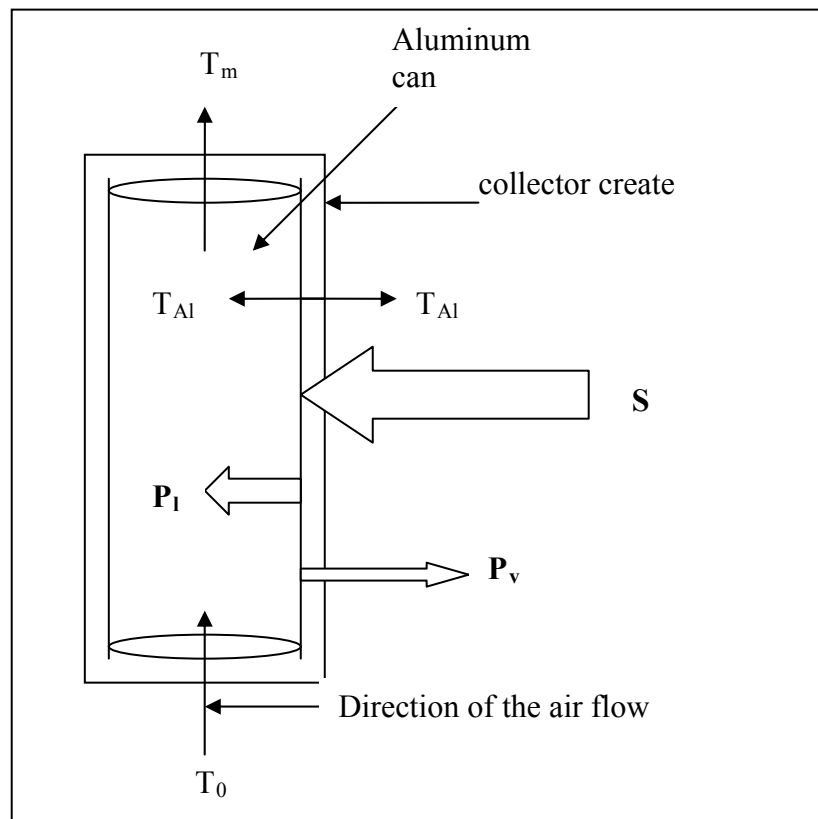


Figure 8.4. The schematic overview of the physical processes in the can-solar collector
Energy transfer in the collector

In [Figure 8.4](#) we can see the schematic overview of the heat transfers. The air going into the collector from the bottom, and its temperature is T_0 , the temperature of the outgoing air is called T_m , like T -measured, since at this point we will measure the temperature during our laboratory practice. The cylindrical shape tube demonstrates the absorber cans. The total mass of the cans is signed by m_{Al} , its specific heat is c_{Al} . S denotes the incoming solar power on the total area of the collector $[S]=W$, that reaches the absorber. P_l means that part of S which will be used for heating the absorber. On the other hand P_v is the leakage heat, the energy loss.

We will apply the energy conservation rule for the material of the cans. During a given time Δt , the absorber absorbs $S \cdot \Delta t = \Delta Q$ energy, and this amount will heat the cans up. $\Delta Q/\Delta t = P_l$ is the actual energy power that raises the temperature of the inside air. The energy loss can happen two ways. One way is useful when the air inside the cans will take the energy from the absorber by heat conduction. The other way it passes some energy (called ΔW) to the closed air and then to the open air, this part will be the loss. The actual energy power loss: $\Delta W/\Delta t = P_v$. In the Δt time interval the absorbed energy by the cans $\Delta E = (S - P_l - P_v) \Delta t$. This energy will raise the temperature of the cans: ΔT_{Al} . The relation between the absorbed energy and the temperature change is $\Delta Q = cm \Delta T$. Therefore.

$$c_{Al} m_{Al} \Delta T_{Al} = \Delta E = (S - P_l - P_v) \Delta t$$

From this conservation rule we can derive a differential equation for the time dependence of the temperature of the absorber.

$$c_{Al} m_{Al} dT_{Al}/dt = S - P_l - P_v \quad (8.1)$$

The dimension of both sides are W (watt), therefore S is the product of the intensity of the sunshine and the effective area of the collector. When we solve this equation we will apply several simplifications. To solve the detailed and real situation is quite difficult and goes much further beyond our possibilities in this lab practice. During our measurement we will keep these in mind, and select those time intervals and circumstances when the simplifications are reliable. For example S is not constant in time generally, but we will assume that. So we will use S as an average for a given time interval.

The heating of the air in the cans. This (8.1) equation is about the temperature of the cans, but we do not measure this parameter, only the temperature of the inside air. In fact these two quantities are related. If the cans are warmer the air will also be warmer. But the relationship is complex, since the air temperature is position dependent in the absorber pipe, it can depend on the ventilation speed etc. If the air flows faster that affects the mass of the air that absorbed the heat of a given time. The faster is the speed the less the temperature rises. We are not going to go into the details of these processes, and we assume instead that the temperature raise $T_m - T_0$ is proportional to the $T_{Al} - T_0$, which is the temperature difference between the cans and the incoming air (room temperature). $T_m - T_0 = p \cdot (T_{Al} - T_0)$, and the p parameter is the energy transfer parameter, depends on the ventilation speed, length of the can-pipe, the material of the cans. In this approach $T_m = T_0 + p \cdot (T_{Al} - T_0)$. Therefore $dT_m/dt = p dT_{Al}/dt$, from which we have a differential equation for T_m :

$$dT_m/dt = (S - P_l - P_v) p / c_{Al} m_{Al} \quad (8.2)$$

We can estimate the P_l and P_v parameters in this equation, and then it is ready to solve.

Determination of the useful energy transfer P_l

P_l is the energy that is absorbed by the air from the cans in a unit time on the total L length of the pipe: $P_l = \Delta Q/\Delta t$. The ΔQ heat is the total heat transferred to the air. But since the air has different temperature along the pipe, the different part of the air gets different heat at the same moment. Lets divide the pipe into N equal $l_l = L/N$ long pieces, and examine the system during

Δt time that needs the air to displace just one piece (l_1 length). The air in one piece is then called small air cell, its mass is $\Delta m = \rho A l_1 / N$. The sum of these small partial heat transfers to each small air cells along the pipe gives ΔQ . $\Delta Q = \sum_{i=1}^N \Delta Q_i$.

After Δt time the whole air will be shifted the i . air cell will be at the location of the air cell that was previously considered $i+1$. air cell. But the process is assumed to be a stationary (time independent) as a whole, therefore the temperature of the air cell shifted gets the temperature of the $i+1$. air cell after the shift. Since the stationary feature the Q_1 partial heat of the air cell at the first location equals to that partial heat transfer that any of the air cells has got when they were at the first position. All the same for the other positions, and therefore the sum of the partial heat transfers at one shift is the same as the sum of the partial heats of one given air cell during the time that went through the whole pipe. But we can calculate Q_1 , as $Q_1 = c \Delta m (T_m - T_0)$, since we know its final and initial temperature. We only have to calculate the Δm in this case. $\Delta m = \rho A l_1 / N = \rho A l_1 = \rho A v \Delta t$.

$$P_i = \frac{\sum_{i=1}^N \Delta Q_i}{\Delta t} = \frac{Q_1}{\Delta t} = \frac{c \Delta m (T_m - T_0)}{\Delta t} = \frac{c \cdot \rho A v \Delta t \cdot (T_m - T_0)}{\Delta t} = c \rho A v (T_m - T_0)$$

Determination of the energy loss P_v

We consider that the energy loss of the system to the open air is due to heat conduction. As the temperature of the cans (T_{Al}) is raising the energy loss will also be larger and larger. Therefore we will use

$$P_v = R(T_{Al} - T_{oa}) = R(T_m - T_{oa})/p.$$

In this formula T_{oa} is the temperature of open air, R is a constant that is proportional to the surface of the collector, and depends on the heat transfer constant of the covering window (e.g. of the polycarbonate). When a can-solar collector is used for real T_0 is the room temperature and T_{oa} is the open air temperature, they obviously are different. But for every case for our test collector, where no walls will be drilled out, T_0 and T_{oa} will be the same, independently of it is used inside or outside. Therefore we use $T_0 = T_{oa}$ for this laboratory practice.

Summarizing the terms discussed above equation 2 will be:

$$\frac{dT_m}{dt} = \frac{p(S - P_l - P_v)}{c_{Al} m_{Al}} = \frac{p(S - c \rho A v (T_m - T_0) - R(T_m - T_0)/p)}{c_{Al} m_{Al}} = Z - K(T_m - T_0) \quad (3)$$

where K and Z are time independent constants.

$$Z = \frac{pS}{c_{Al} m_{Al}} \quad \text{and} \quad K = \frac{pc \rho A v + R}{c_{Al} m_{Al}}$$

The initial condition for this equation is $T_m = T_0$ at the $t=0$ moment, that is when we start to irradiate the collector. Using $y(t) = T_m - T_0$ the equation simplifies to $\frac{dy(t)}{dt} = -K \cdot y(t) + Z$ type differential equation with $y(t=0) = 0$ initial condition. The solution of the (3) differential equation is

$$y(t) = -\frac{Z}{K} e^{-Kt} + \frac{Z}{K}.$$

Let's check the initial condition: $y(0) = (-Z/K) \cdot 1 + Z/K = 0$, and then the validity of our solution:

$$y'(t) = (-Z/K)(-K) \exp(-Kt) = Z \exp(-Kt)$$

and

$$Z - Ky(t) = Z - K[(-Z/K)\exp(-Kt) + Z/K] = Z + Z\exp(-Kt) - Z = Z \exp(-Kt)$$

The two sides are equal therefore the solution is verified.

Let's put the meaning of y back:

$$T_m(t) = -\frac{Z}{K}e^{-Kt} + \frac{Z}{K} + T_0$$

Here Z depends on the magnitude of the irradiation and the parameter of the collector, but K depends only on the collectors parameters, including the velocity of the air flow.

We can make this equation more usable if we investigate the meaning of the Z/K . In the $t \rightarrow \infty$ limit $T_m(\infty) = 0 + Z/K + T_0 = T_{max}$. Therefore $Z/K = T_{max} - T_0$.

$$T_m(t) = (T_0 - T_{max})e^{-Kt} + T_{max} \quad (4)$$

Where T_{max} depends on the irradiation and the characteristic time constant (K) does not. This equation describes the time dependence that can be measured with the test collector. The K constant can be determined at each air flow velocity in experiments.

8.2.3 The efficiency of the collector

In our measurements we can easily specify time periods when the S irradiation and the T_m outgoing air temperature is time independent. When the weather is quickly changing it is not easy to find a stationary interval to investigate time independent operation of the heater. But if we use a lamp irradiation in a room, it is much easier. During the practice we have a chance to use a 1 kW lamp, put close to the collector. That will simulate the situation like a direct sunshine without clouds. (Even if the weather does not allow it for us.)

The used energy per unit time that can be used for our purposes in a real use is P_l , and the incoming energy power is S . Therefore the definition of the efficiency of the system is $\eta = \frac{P_l}{S}$. The P_l should be determined when we already reached the stationary state. In this circumstances the $dT/dt = 0$, P_l is time independent, and the efficiency can be measured.

$$\eta = \frac{P_l}{S} = \frac{c\rho v A (T_{max} - T_0)}{I A_{coll}}$$

Where $S = I A_{coll}$, where I is the intensity of the sunshine (W/m^2) and A_{coll} is the surface of the collector A is the cross section of the pipes, c is the specific heat of the air (which temperature dependence can be neglected for now), ρ is the density of the air, v is the air flow velocity at a point where the cross section is A , T_0 is the ambient temperature (room temperature if we measure inside and open air temperature if we measure outside), T_{max} is the maximum temperature in equilibrium.

When we measure the air flow velocity we cannot put an instrument into the middle of the system. In place of that we create a cone to focus the air that will result in a faster air flow and it is easier to measure. Assuming that the air is incompressible material $Av = \text{constant}$, so the measured value can be applied to the inside area of the collector.

8.3 Lab course tasks

1. Measure the time dependence of the temperature of the outgoing air using a constant intensity lamp! Fit a proper function to this time dependent dataset. Determine the K value of the raising!
2. Determine the cooling curve of the collector in room temperature environment! Fit the proper K value for this case too!
3. Determine the warming curve for a case using natural sunshine! (If the meteorology allows it.)
4. Determine the cooling curve outside!
5. Determine the maximum temperature of the outgoing air depending on the velocity of the air flow!
6. Determine the warming curve at the same irradiation but different air flow rates! Evaluate the determined K values!
7. Calculate the efficiency of the collector at some cases! Evaluate the differences!
8. Determine the efficiency of the collector outside, using the irradiation data from the web! (www.naplopo.hu) The data will be available with a delay of some days.
9. Calculate the payback time based on your measurement!

8.4 Test questions

1. What is the energy source for heating the air in our collector?
2. Describe the steps of energy conversion from a media to another!
3. Why does a collector has energy loss?
4. What are the processes that decrease the efficiency of a collector?
5. How will the air warm up in time using a constant intensity lamp for heating?
6. How can the efficiency of a collector be improved?
7. Why do we paint the absorbers black?
8. Why does the efficiency depend on the outside air temperature?
9. What is the effect of using higher velocity on the efficiency?
10. What can we do with our collector on a sunny summer afternoon?
11. Is the air in the collector has homogeneous temperature?
12. Is the collector useful during a cooling curve, after sunset?

9. Polarized light pollution (POL)

The aim of the laboratory practice is to inform the students about (i) the term of *polarized light pollution* (PLP) and its physical and biological bases, (ii) the natural and artificial (anthropogenic) reflection-polarization characteristics, and (iii) the method of imaging polarimetry that is able to measure polarization patterns. Students have to measure and analyze the reflection-polarization patterns of some typical sources of polarized light pollution in the immediate vicinity of the physical and biological buildings of the Eötvös University.

9.1 Introduction

In this chapter we introduce the term 'polarized light pollution' (PLP), meaning all adverse effects on polarotactic aquatic insects attracted by exactly or nearly horizontally polarized light reflected from smooth and dark artificial surfaces. PLP is a new kind of ecological photopollution, it is global and novel in an evolutionary sense. We review the experimental evidences of PLP, such as (i) trapping of aquatic insects by dark oil surfaces; (ii) dehydration of polarotactic insects attracted to black plastic sheets used in agriculture; (iii) egg-laying of polarotactic mayflies onto dry asphalt roads; (iv) attraction of aquatic insects to black, red or dark-coloured car paintwork; (v) deception of polarotactic dragonflies by shiny black gravestones; (vi) attraction of mass-swarmed polarotactic caddis flies to glass surfaces; (vii) attraction of aquatic insects to photovoltaic solar cells and sun collectors. All such highly and horizontally polarizing artificial surfaces may act as 'polarized ecological traps' for polarotactic insects, because these surfaces are inappropriate for the development of eggs laid by the deceived aquatic insects. We discuss possible benefits and/or disadvantages of predators (spiders, birds, bats) feeding on the polarotactic insects attracted to different sources of PLP. Finally, some remedies of PLP are suggested. Conservation planners should pay much more attention to aquatic insects because of their positive polarotaxis and their demonstrated vulnerability due to PLP.

9.2 Polarized Light Pollution and Polarized Ecological Trap: Definitions

It is a well-known fact that the degradation of human views of the night sky in and near cities makes practically all astronomical observations impossible. The stars and other celestial bodies are washed out by artificial citylights that are either directed or reflected upward. This phenomenon is called the 'astronomical light pollution' (Rich and Longcore, 2006, p. 3). On the other hand, the term 'ecological light pollution' (ELP) is used to describe all kinds of 'photopollution' (a synonym of 'light pollution') which disrupts ecosystems (Rich and Longcore, 2006). ELP has been defined as the "degradation of the photic habitat by artificial light". ELP includes direct glare, chronically increased illumination, and temporary, unexpected fluctuations in lighting. Sources of ELP include sky glow, lighted structures (e.g., buildings, towers, bridges), street lights, security lights, lights of vehicles and fishing boats, flares on offshore hydrocarbon platforms, and lights on undersea research vessels (Rich and Longcore, 2006, pp. 3-4). Both the documented and possible ecological consequences of all these artificial night lighting were comprehensively summarized in the monograph edited by Rich and Longcore (2006).

The first step of all these effects of artificial night lights is the attraction, or repulsion of animals by the spatiotemporally enhanced intensity of light relative to the dark environment. This phenomenon, called positive or negative phototaxis (Nowinszky, 2003), is elicited by the intensity and/or colour of artificial light. Until now only this phototaxis has been considered

as the major visual phenomenon underlying the ELP. Here we introduce the term 'polarized light pollution' (PLP) as a new kind of ecological photopollution (Horváth *et al.*, 2009). Under PLP we mean highly and horizontally polarized light reflected from smooth (shiny) artificial surfaces (see Figures 9.1 and 9.2) having adverse effects on polarotactic aquatic insects (see Figure 9.3), including all insects, whose larvae live in water (e.g., aquatic beetles and bugs, dragonflies, mayflies, caddis flies and tabanid flies).

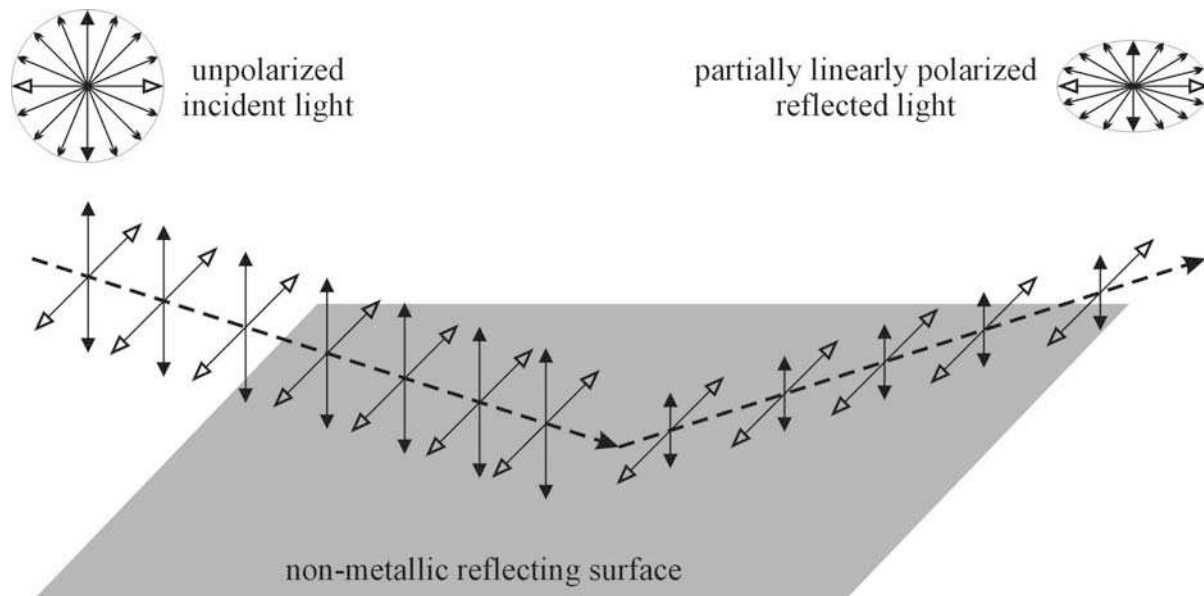


Figure 9.1. After reflection from a non-metallic surface unpolarized light becomes partially linearly polarized, whose electric field vector is smaller in the plane of reflection (double-headed arrows with black head) than that perpendicular to this plane (double-headed arrows with empty head).

In 1985 Rudolf Schwind has discovered that the backswimmer, *Notonecta glauca* detects water by means of the horizontally polarized light reflected from the water surface, rather than by the intensity or colour of water-reflected light, or by the glittering, or mirroring of the water surface. In the ventral eye region of *Notonecta* there are ultraviolet-sensitive photoreceptors with horizontal and vertical microvilli being highly sensitive to horizontally and vertically polarized light. These orthogonally polarization-sensitive photoreceptors are able to determine whether the direction of polarization of light from the optical environment is horizontal or not. In *Notonecta* an exactly or nearly horizontally polarized light stimulus elicits a typical plunge reaction. This attraction of *Notonecta* to horizontally polarized light is called 'positive polarotaxis'.

Most of the females of aquatic insects (e.g., Ephemeroptera, Odonata, Plecoptera, Trichoptera) must return to water to lay their eggs. Water bodies also often serve as rendezvous for both sexes of aquatic insects. As orientation in aquatic insects is predominantly visual, we may ask for the optical cues by which specific water bodies are recognized. The eye in many aquatic insects is sensitive to the polarization of light in the visible or ultraviolet spectral ranges (Schwind, 1991). These insects find their habitat on the basis of the horizontally polarized light reflected from the water surface. Aquatic insects detect polarization in that region of the spectrum, which is characteristic of their preferred habitat (Schwind, 1991). Depth, turbidity, transparency, colour, surface roughness of the water and substratum composition as well as the illumination strongly influence the reflection-polarization characteristics of water bod-

ies (Horváth and Varjú, 2004). Polarized light reflected by water provides important information on the quality of freshwater habitats for polarotactic insects and can aid the orientation of these insects from a distance when other cues (e.g., atmospheric humidity, dimension and shape of the water body, undulation of the water surface, water plants on the surface and the shore, temperature and odour) are still ineffective (see Figure 9.2A).

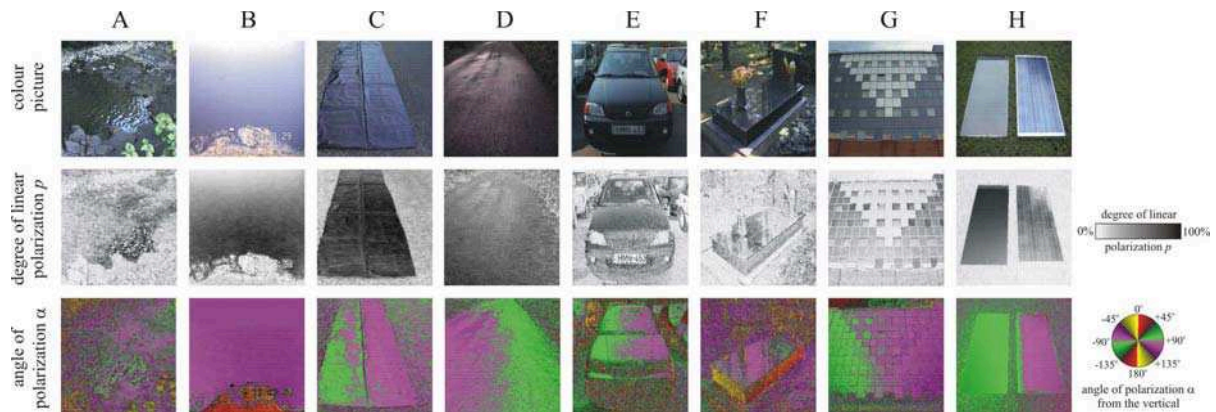


Figure 9.2. Colour pictures (row 1) and patterns of the degree of linear polarization p (row 2) and angle of polarization α from the vertical (row 3) of light reflected from a dark water surface (A) and different artificial surfaces (B-H) causing polarized light pollution. The p - and α -patterns were measured by imaging polarimetry in the blue (450 nm) part of the spectrum. A: A dark water body illuminated by the light from a clear sky after sunset. B: A sunlit crude oil lake in the desert of Kuwait (on the bottom of the picture the sandy shore can be seen). C: A dry shiny black plastic sheet laid on a dry asphalt road after sunset under a clear sky. D: A dry asphalt road lit by the setting sun. E: A sunlit black car under a clear sky. F: A shady polished black gravestone under a clear sky. G: Vertical wall of a building under a clear sky after sunset. The wall between the windows is covered by light grey and black glass surfaces as ornaments. H: Two horizontal photovoltaic solar panels. Left: homogeneous black solar panel. Right: heterogeneous polycrystalline solar panel.

In a series of observations Schwind (1991) showed that several species of aquatic bugs and beetles are polarotactic. Later studies (Kriska *et al.*, 1998, Csabai *et al.*, 2006) found that beside aquatic bugs and beetles, also many other aquatic insects like dragonflies, mayflies, tabanid flies, stone flies and caddis flies exhibit positive polarotaxis when searching for water. Until now more than 300 polarotactic aquatic insect species are known that recognize their aquatic habitat by positive polarotaxis.



Figure 9.3. Polarotactic aquatic insects and insects associated with water deceived by and attracted to different sources of polarized light pollution (photographs taken by G. Kriska). Row 1: Some typical representatives of insects trapped by a crude oil lake in the desert of Kuwait (A), and a waste oil lake in Budapest, Hungary (B-D). A: An Aeschnid dragonfly (courtesy of Dr. Jochen Zeil). B: A dragonfly (*Sympetrum vulgatum*) and scavenger beetles (*Hydrophilidae* sp.). C: A mayfly (*Cloeon dipterum*). D: A great silver diving beetle (*Hydrophilus piceus*). Row 2: Water insects landed on horizontal shiny black dry plastic sheets used in agriculture. E: Copulating mayflies (*Rhithrogena semicolorata*). F: A female mayfly (*Ephemera danica*) laying her eggs on the black plastic sheet. G: A female large stonefly (*Perla burmeisteriana*). H: A tabanid fly (*Tabanidae*). Row 3: Aquatic insects landed on dry asphalt roads. I: A male mayfly (*Epeorus silvicola*). J: Copulating mayflies (*Rhithrogena semicolorata*). K: Oviposition by a female large stonefly (*Perla burmeisteriana*), whose black egg-batch at the tip of her abdomen is shown by the tip of the white arrow. L: A great silver diving beetle (*Hydrophilus piceus*). Row 4: Insects associated with water on the dry roof of a red car. M: A mayfly (*Ecdyonuridae* sp.). N: A water bug (*Sigara striata*). O: A water beetle (*Hydrochara caraboides*). P: A tabanid fly (*Tabanidae*). Row 5: Q: A male dragonfly (*Sympetrum* sp.) perching near a polished black tombstone. R: Mass swarming of *Hydropsyche pellucidula* caddis flies (white spots) in front of the vertical glass surfaces of a building on the bank of the river Danube in Budapest. S: A *H. pellucidula* landed on a pane of glass.

T: A copulating pair of *H. pellucidula* on a glass pane.

Thus, it is quite understandable that polarotactic insects can be deceived by and attracted to every artificial surface that reflects highly and horizontally polarized light: such a man-made surface mimicks an exaggerated water surface to water-seeking aquatic insects by the highly and horizontally polarized reflected light functioning as a supernormal stimulus (see [Figures 9.1](#) and [9.2](#)). This visual ecological phenomenon is the major reason for the PLP. The physical (1), behavioural (2) and ecological (3) bases of PLP are the following (Horváth *et al.*, 2009):

According to the so-called Umow rule, the darker a surface in a given part of the spectrum, the higher the degree of linear polarization p of light reflected from it. Since diffuse reflection from rough (matt) surfaces results in depolarization, the smoother (the shinier) a surface, the higher the p of reflected light. Since the direction of polarization of light reflected from smooth dielectric materials is always perpendicular to the plane of reflection, if this plane is exactly or nearly vertical, the reflected light is exactly or approximately horizontally polarized. From these it follows:

- (1) Smooth and black artificial surfaces with exactly/nearly vertical plane of reflection mirror highly and exactly/nearly horizontally polarized light.

The higher the p of light and the less deviates its direction of polarization from the horizontal, the larger its attractivity to polarotactic aquatic insects (Horváth and Varjú, 2004). Consequently:

- (2) Smooth and black artificial surfaces with exactly/nearly vertical plane of reflection are strongly attractive to polarotactic insects.

Aquatic insects attracted to highly and horizontally polarizing dry artificial surfaces may perish due to dehydration, or may oviposit onto these surfaces, the consequence of which is that the eggs perish inevitably. From this we can conclude that:

- (3) Highly and horizontally polarizing artificial surfaces may act as 'polarized ecological traps' for ovipositing polarotactic aquatic insects (Schlaepfer *et al.*, 2002), because such surfaces are completely inappropriate for the development of eggs laid by the deceived and attracted aquatic insects.

On the basis of the above we can formulate the following visual ecological thesis: Smooth and dark artificial surfaces with exactly/nearly vertical plane of reflection are strongly attractive to polarotactic aquatic insects, and thus constitute polarized ecological traps for these animals, consequently induce PLP. In the next section we review the experimental evidences supporting this thesis. Using theoretical calculations and imaging polarimetry (Horváth and Varjú, 2004), the reflection-polarization characteristics of water surfaces and artificial reflectors can be compared. Imaging polarimetry is a useful technique to establish and monitor the sources of PLP. Excluding the astronomical photopollution, [Table 9.1](#) summarizes the major characteristics of the conventional ecological photopollution and the PLP.

characteristics	conventional ecological photopollution	polarized light pollution
source of photopollution	artificial night lights: sky glow, lighted structures, streetlamps, security lights, vehicle headlights, fishing boats, flares on hydrocarbon platforms, lights on undersea research vessels	highly and horizontally polarized light reflected from artificial surfaces: oil lakes; asphalt roads; black plastic sheets in agriculture; glass surfaces; black, red, dark-coloured car paintwork; black gravestones; photovoltaics panels
extent of photopollution	global	global
cue(s) eliciting the photoreaction of animals	intensity and colour of light	horizontal polarization of light
time of day of photopollution	between dusk and dawn	both day and night
directly or indirectly concerned animals	night-active animals	polarotactic aquatic insects and their predators (spiders, birds, bats)
effects of photopollution	<ul style="list-style-type: none"> - attraction/repulsion by lights - disturbances in biorhythm - disruption of physiological processes (reproductive condition, preparing to migrate or hibernate, egg laying, molt, hormone production) - increased deaths in collisions - disruption of seasonal changes in behaviour - disturbances in orientation and migration (dis/misorientation) - disruption of foraging - disturbances in intra/interspecific visual communication - disturbance in nest site choice - desynchronization of mating - disturbance of community interactions (e.g., increased competition and predation) 	<ul style="list-style-type: none"> - attraction of polarotactic aquatic insects to horizontally polarized light - perishment of attracted aquatic insects due to dehydration - perishment of the eggs laid onto the polarized-light-polluting surfaces due to dehydration - disruption of water-associated behaviour elicited by horizontally polarized light - disturbances in orientation to water and migration between waters - disturbances in oviposition site selection - attraction of predators (spiders, birds, bats) by the polarotactic insects lured to the sources of horizontally polarized light

Table 9.1. Major characteristics of conventional ecological photopollution and polarized light pollution (Horváth and Zeil, 1996; Horváth *et al.*, 2009; Kriska *et al.*, 1998; Horváth and Varjú, 2004; Csabai *et al.*, 2006; Malik *et al.*, 2008).

Modern human development has resulted in the introduction of different sources of PLP to natural habitats. In the natural optical environment only the flat water surface reflects horizontally polarized light, whose p is high if light is reflected from dark water bodies at the Brewster angle (Horváth and Varjú, 2004). PLP is mainly a byproduct of the human architectural, building, industrial and agricultural technology, and it may allow to function feeding webs composed of polarotactic insects and their predators, such as spiders, birds and bats, for example (see [Figure 9.4](#)).



Figure 9.4. Predators feeding on the polarotactic insects attracted to two different sources of polarized light pollution (photographs taken by G. Kriska). Rows 1-2: Urban birds feeding on the mass-swarmed caddis flies, *Hydropsyche pellucidula* attracted to vertical glass surfaces. A-B: A male house sparrow, *Passer domesticus* hunting caddis flies at a window. C: A great tit, *Parus major* standing on a window-frame. D: A *P. major* following caddis flies with attention. E: A European magpie, *Pica pica* on an edge of a building. F: A *P. pica* on wing capturing caddis flies from a glass pane. G: A white wagtail, *Motacilla alba* perching on a protrusion of a building. H: A hovering wagtail gathering caddis flies from a glass pane. Row 3: Spiders on the wall of a building where caddis flies (*H. pellucidula*) swarmed. I: An *Araneus umbraticus* feeding on a caddis fly captured by its cobweb. J: A *Tetragnathidae* sp. on a reddish brick. K: A *Thomisidae* sp. between two reddish bricks. L: A *Salticus zebraneus* on a reddish brick. Row 4: Carcasses of insectivorous vertebrates lured by polarotactic insects and trapped by the waste oil lake in Budapest. The insects were attracted to the polarized-light-polluting oil surface. M: A black redstart, *Phoenicurus ochruros*. N: A European goldfinch, *Carduelis carduelis*. O: A yellowhammer, *Emberiza citrinella*. P: Flock of European greenfinches. Q: European magpie. R: A bat (*Chiroptera*). S: An owl. T: A kestrel, *Falco tinnunculus*.

PLP is a well-documented cause of egg mortality among aquatic beetles and bugs, mayflies and caddis flies. The phenomenon of PLP is global and new in an evolutionary sense, having increased rapidly only over the last decades, following the spread of highly and horizontally polarizing artificial surfaces such as oil lakes and open-air oil reservoirs, asphalt roads, plas-

tic sheets, glass surfaces, cars and solar panels, for example. The mortality associated with PLP may threaten populations of endangered aquatic insect species.

PLP can occur not only in daytime, but also at night, if moonlight or citylight (e.g., skyglow, streetlamps) is reflected from polarized-light-polluting surfaces. The vulnerability to PLP (based on positive polarotaxis) could be enhanced by the synergetic interaction with conventional photopollution (based on positive phototaxis) caused by artificial night lighting. PLP could also be influenced by lunar cycles, especially in rural environment, where artificial night lighting is rare or lacking. It is important to determine and monitor the sources of PLP in order to minimize and/or replace them by artificial surfaces which are "aquatic insect friendly". Since many human developments with numerous polarized-light-polluting artificial surfaces are near water bodies, the aquatic insects living in/at lakes, rivers, ponds and streams are all subject to PLP. Because aquatic insects are critically important as members of food webs in aquatic ecosystems, adverse effects of PLP on these animals could have serious ecological consequences.

Flight to horizontally polarized light reflected from artificial surfaces could disturb the ecology of aquatic insects, and often can lead to high mortality of the adults and/or the eggs laid onto these polarized-light-polluting surfaces. Polarotactic aquatic insects frequently are not able to escape from the source of PLP. This behaviour is called the "polarization captivity effect", which culminates in the death of insects due to dehydration and exhaustion. The migration, dispersal, mating and reproduction of aquatic insects can also be disturbed by sources of PLP encountered in their long-distance flight paths. This is called the "polarization crash barrier effect" because of the interruption of movement across the landscape. Aquatic insects are also vulnerable to normal (unpolarized) artificial lights: Streetlamps, for example, have a long-distance effect for light-susceptible mayflies and caddis flies emerging from a small mountain stream. The night-time attraction of these insects to lamps is so strong that if there were a row of streetlamps along a stream, a species could become locally extinct in a short time. This extinction can only be accelerated by PLP.

9.3 Evidences of Polarized Light Pollution

9.3.1 *Aquatic Insects Trapped by Dark Oil Surfaces*

Many individuals of the dragonfly *Anax junius* have been killed as a result of mistaking an open surface of crude oil for water, and other dragonflies have been attracted to pools of petroleum. Horváth and Zeil (1996) observed a similar phenomenon in the desert of Kuwait: During the Gulf War in early 1991, Iraqi occupation forces blasted oil wells and pipelines in the desert of Kuwait, forming more than 900 oil ponds. Several years later, the majority of these oil lakes still existed and continued to trap a variety of animals, mainly insects, especially aquatic insects (Horváth and Zeil, 1996). Reductions in the oil level due to evaporation and percolation into the ground created distinct bands of insect carcasses at their edges. Bands of dead dragonflies, damselflies and ground-beetles reflected arrivals of migrating insects in autumn and spring. Mass arrivals of aeshnid dragonflies were witnessed by Jochen Zeil in October 1994 and February 1995, many females being trapped while attempting to lay eggs in the oil (see [Figure 9.3A](#)). Different species of water beetles (Dytiscidae, Coleoptera), giant water scorpions (*Belostoma* sp., Nepidae, Heteroptera), mole crickets (Gryllotalpidae, Orthoptera) as well as sphingid moths, *Vanessa* butterflies, solifugid spiders, scorpions, reptiles, birds and mammals were found at the edge of the oil ponds.

Horváth and Zeil (1996) suggested that polarotactic water insects were attracted by the high and horizontal polarization of light reflected from the oil surface. This hypothesis has been tested and supported in multiple-choice field experiments with dragonflies in Hungary. The numbers of dragonflies caught by water, crude oil and salad-oil traps with different reflection-polarization characteristics have been compared. It has been demonstrated that positive polarotaxis is the most important mechanism which guides dragonflies during their habitat choice and oviposition site selection, and this is the reason why dragonflies can be deceived by and attracted to crude and waste oil, tar or asphalt. Using horizontally aligned test surfaces with different reflection-polarization characteristics in multiple-choice field experiments with dragonflies, it has been obtained the same result in Switzerland.

Horváth and Zeil (1996) also compared the reflection-polarization characteristics of crude oil and transparent/translucent water surfaces by imaging polarimetry. They showed that a dark oil surface reflects always horizontally polarized light, whose degree of linear polarization p is very high (near 100%) near the Brewster angle [$\theta_{\text{Brewster}} = \arctan(n) \approx 58^\circ$ from the vertical for the refractive index $n = 1.6$ of oil] (see [Figure 9.2B](#)), while water bodies, especially the bright ones reflect light usually with lower p and not always with horizontal polarization. The larger the p and the smaller the deviation of the direction of polarization of reflected light from the horizontal, the greater is the attractiveness of a reflecting surface to polarotactic water insects (Horváth and Varjú, 2004). The consequence is that a dark oil pond can be even more attractive to polarotactic aquatic insects than a brighter water body. Thus, for polarotactic water insects oil lakes appear as an exaggerated water surface acting as a supernormal optical stimulus.

Unfortunately, in many countries plenty of temporary inland oil spills exist as a by-product of the oil industry (exploitation, transport, storage and refinery of the oil). At warm weather the surface of these oil lakes is flat, shiny and acts as an efficiently polarizing reflector of sunlight and skylight, like a water surface. Since 1951 there has existed an open-air waste oil reservoir in a suburb of Budapest, Hungary. It has been observed that this open-air oil reservoir deceived, attracted and trapped water insects in large numbers. They also measured the reflection-polarization characteristics of the waste oil surface versus time. This oil reservoir acted as a disastrous polarized insect trap for 50 years.

It has been observed that dragonflies, mayflies, caddis flies, water bugs and water beetles were trapped *en masse* by the waste oil in spring, summer and autumn at the time of their swarming and migration (see [Figure 9.3B,C,D](#)). Usually, the insects landed or plunged directly on the sticky oil surface and became immediately entrapped. Pairs of insects, e.g., dragonflies and mayflies, were trapped frequently by the oil during copulation and/or egg-laying.

Horváth and Zeil (1996) observed the behaviour of larger dragonflies above the crude oil lakes in Kuwait and the waste oil reservoir in Budapest. Male dragonflies frequently patrolled above the flat oil surface and protected their territory against all intruders. They tried to attack all flying insects. Male dragonflies often sat guard on the tip of perches at the shore. Copulating pairs of dragonflies were frequently observed flying above the oil surface or trying to lay eggs into the oil. They became trapped during water-touching manoeuvres or egg-laying. In the latter case sometimes only the female became entrapped when the tip of her abdomen was dipped into the oil. In many cases, however, the male was also carried along with the female into the oil. Touching the surface by dragonflies observed often at oil lakes is a reaction, which is typical only above water surfaces when dragonflies inspect the surface to select the optimal habitat or oviposition site. The most frequently observed behaviour types of dragon-

flies above the oil surface were the air fight, hovering and protection, which again are typical only above water surfaces.

9.3.2 Polarotactic Insects Deceived by Black Plastic Sheets on the Ground

Kriska *et al.* (1998) and Csabai *et al.* (2006) performed choice experiments with polarotactic aquatic insects in the field using white and black plastic (polyethylene) sheets laid on the ground in different wetlands. Such plastic sheets are commonly used in agriculture against weeds, and/or to keep the soil warm in order to speed up the sprouting, or simply to cover produce and to protect it against rain. A horizontal black plastic sheet reflects always horizontally polarized light with high degrees of linear polarization p near the Brewster angle [$\theta_{\text{Brewster}} = \arctan(n) \approx 56.3^\circ$ from the vertical for the refractive index $n = 1.5$ of plastic] (see [Figure 9.2C](#)), while a white plastic sheet reflects vertically or obliquely polarized light with very low p . Hence a shiny black plastic sheet is a more effective polarizer than a white plastic sheet. Thus, the light reflected from a horizontal shiny black plastic sheet acts as a supernormally polarized stimulus for polarotactic water-seeking insects.

Kriska *et al.* (1998) and Csabai *et al.* (2006) found that only the horizontal black plastic sheet attracted insects associated with water, while the white plastic sheet were unattractive to them. All these aquatic insects attracted showed similar behavioural elements on and above the black plastic sheet: landing (see [Figure 9.3G,H](#)), flying up, touching, crawling, egg-laying (see [Figure 9.3F](#)), copulating (see [Figure 9.3E](#)), reproductive activity. Finally, some of them (e.g., aquatic bugs and beetles) dried out and perished within some hours. During these field experiments almost at every sunset a rattling sound has been heard from the black plastic sheet like the pattering of raindrops. The reason for this was thousands of water insects (e.g., Corixidae water bugs or swarming mayflies) landing on and crashing into the black plastic sheet, then jumping repeatedly up and down. They did not leave the optical trap, and did not fly away from the visually attractive black plastic sheet; they remained on it. At the white plastic sheet similar effect was not observed. All these experiments and observations show that horizontal black shiny plastic sheets can deceive and attract polarotactic water insects due to the highly and horizontally polarized reflected light, while white plastic sheets are unattractive to these insects, because the reflected light is weakly and/or not horizontally polarized.

Kriska *et al.* (1998) showed that the mayfly species *Ephemera danica*, *Ecdyonurus venosus*, *Epeorus silvicola*, *Baetis rhodani*, *Rhithrogena semicolorata*, *Haproleptoides confusa* and *Palingenia longicauda* detect water by means of the horizontal polarization of light reflected from the water surface, and then move towards it, thus showing positive polarotaxis. This is the reason why these mayflies can be deceived by and attracted to horizontal shiny black plastic sheets, above which they swarm, copulate, and onto which they lay their eggs *en masse* (see [Figure 9.3E,F](#)).

It has been discovered the attraction to horizontally polarized light (positive polarotaxis) in both males and females of the tabanid fly species *Haematopota pluvialis*, *Heptatoma pellucens*, *Hybomitra ciureai*, *H. solstitialis*, *H. ucrainica*, *Tabanus bovinus*, *T. bromius*, *T. sudeticus*, *T. tergestinus*. All these tabanids are terrestrial, but lay their eggs onto the lower side of the leaves of marsh-plants leaning over the water surface, because after egg hatching the larvae must drop into the water where they develop. It has been proposed that in these tabanids the first step in the remote search for potential terrestrial rendezvous and egg-laying sites happens indirectly by means of the detection of horizontally polarized light reflected from the surface of waters, on the shore of which appropriate plants for oviposition may occur. Due to

their positive polarotaxis all the mentioned tabanid fly species can be attracted to highly and horizontally polarizing shiny black plastic sheets laid on the ground as has been demonstrated in choice experiments (see [Figure 9.3H](#)).

9.3.3 Attraction of Polarotactic Aquatic Insects to Dry Asphalt Roads

It has been observed that during their swarming above the river Tisza in Hungary *Palingenia longicauda* mayflies were attracted to a wet asphalt road running on the bank parallel to the river. It has been reported that different dragonfly species patrolled along dry asphalt roads instead of rivers and showed a typical water-touching behaviour above the asphalt surface. Kriska *et al.* (1998) observed that near sunset the individuals of the mayfly species *Ephemera danica*, *Ecdyonurus venosus*, *Epeorus silvicola*, *Baetis rhodani*, *Rhithrogena semicolorata* and *Haproleptoides confusa* swarmed, mated above and landed on dry asphalt roads (see [Figure 9.3I,J](#)), shiny black plastic sheets and windscreens and roofs of cars in the immediate vicinity of their emergence sites (mountain streamlets). After copulation the female mayflies laid their eggs on the dry asphalt surface instead of laying them on the water surface. The mayflies showed the same behaviour above the asphalt road as at water surfaces. In spring near to the mountain creeks female stoneflies with their egg-batches could also often be seen on the asphalt roads (see [Figure 9.3K](#)). These observations, especially the egg-laying by females, show that the mayflies and stoneflies were apparently deceived by and attracted to the dry asphalt surface.

Kriska *et al.* (1998) gave a convincing explanation of this behaviour of mayflies: Using imaging polarimetry, they showed that asphalt surfaces lit by skylight near sunset, when mayflies swarm, mimic a highly and horizontally polarizing water surface to water-seeking mayflies (see [Figure 9.2D](#)). The direction of polarization of sunlight reflected from sunlit asphalt roads is always perpendicular to the plane of reflection determined by the observer, the sun and the point observed. Thus, the direction of polarization of asphalt-reflected sunlight is usually tilted relative to the horizon, but it is always horizontal, if the observer views toward the solar or antisolar meridian, because in these cases the plane of reflection is vertical. If the asphalt road is diffusely illuminated by the downwelling skylight only, the reflected light is always horizontally polarized due to the extended sky, which illuminates the road from all possible directions. In multiple-choice experiments with swarming mayflies Kriska *et al.* (1998) showed that the above-mentioned six mayfly species detect water by means of the horizontal polarization of reflected light, and possess positive polarotaxis, because they are attracted to horizontally polarized light. Kriska *et al.* (1998) also showed that the darker and smoother the horizontal asphalt surface, the greater is its attractiveness to water-seeking polarotactic mayflies.

The highly and at sunset always horizontally polarized asphalt roads with a relatively homogeneous distribution of the degree p and angle α of linear polarization of reflected light can be much more attractive to polarotactic mayflies than the water surface. An asphalt road can reflect and polarize the incident light in such a way that the reflected light becomes a supernormal stimulus for water-seeking mayflies in comparison to the light reflected from water. This was also observed in the multiple-choice experiments of Kriska *et al.* (1998), when mayflies swarming above the asphalt road were attracted to the highly polarized shiny black plastic sheet after it was laid onto the road. A relatively small black plastic sheet (a few m²) attracted all the mayflies swarming above the asphalt road within several tens metres.

Positive phototaxis (Nowinszky, 2003) and polarotaxis can synergetically govern the behaviour of water-seeking polarotactic aquatic insects: We observed (2007, unpublished data) that

night-migrating scavenger beetles (*Hydrophilus piceus*) were attracted by phototaxis from large distances, and then trapped by polarotaxis by a lamplit dry asphalt surface. Hence, from a distance flying water seeking insects can be attracted to the street lamps above the littoral asphalt road by positive phototaxis. Then polarotaxis is not in action yet, because the lamps emit unpolarized light, and the horizontally polarized light reflected from the asphalt road is not visible yet. Approaching the lamplit asphalt road, the observed scavenger beetles could already perceive the horizontally polarized light reflected from the dry asphalt, thus they were attracted to the lamplit asphalt area by positive polarotaxis. Due to this strong polarization signal they remained temporarily in the circular light spot on the asphalt, because then there was no other horizontally polarized light signal in their optical environment (see [Figure 9.3L](#)).

9.3.4 Aquatic Insects Attracted by Black, Red or Dark-coloured Car Paintwork

Aquatic beetles (Coleoptera) and bugs (Heteroptera) have been frequently observed to land on the roofs, bonnets and boots of black or red cars (Kriska *et al.*, 1998), and Ephemeroptera and Odonata females often lay their eggs on these car surfaces. To reveal the reasons for this phenomenon, the numbers of aquatic beetles and bugs attracted to horizontal shiny red, yellow, white and black plastic sheets have been monitored. It has been found that horizontal red and black reflectors are equally highly attractive to aquatic beetles and bugs, while yellow and white reflectors are unattractive. The reflection-polarization characteristics of red, yellow, white and black cars have also been measured in the red, green and blue parts of the spectrum. It has been found that in the blue and green parts of the spectrum p of light reflected from red and black cars is high, and the direction of polarization of light reflected from red and black car roofs, bonnets and boots is exactly or nearly horizontal (see [Figure 9.2E](#)). Thus, the horizontal surfaces of red and black cars are highly attractive to red-blind polarotactic aquatic insects. On the other hand, p of light reflected from the horizontal surfaces of yellow and white cars is low and its direction of polarization is usually not horizontal. The same was true for the reflection-polarization characteristics of black, white, red and yellow horizontal plastic sheets used in field experiments monitoring water insects. Hence, yellow and white cars and plastic sheets are unattractive to polarotactic aquatic insects. These results show that the visual deception of aquatic insects by red, black and dark-coloured cars can be explained solely by the reflection-polarization patterns of the car paintwork and the positive polarotaxis of these insects.

An egg-batch of a female mayfly, for example, contains 6000-9000 eggs (Kriska *et al.*, 1998). All the eggs laid onto car surfaces perish. This also often occurs in the case of water insect imagoes due to dehydration on hot car surfaces. However, not only the highly and horizontally polarizing car paintworks can be dangerous to aquatic insects and/or their eggs, but the eggs laid onto car-bodies can also damage the resin of the clearcoat as does acid rain. It has been shown that the eggs of *Miathyria*, *Tauriphila* and *Erythemis* dragonflies at temperatures between 50 and 92°C produce sulfonic acids that destroy the clearcoat.

9.3.5 Polarotactic Dragonflies Deceived by Shiny Black Gravestones and Other Black Anthropogenic Products

It has been observed that the females of *Orthetrum* dragonflies laid their eggs on a shiny cement floor and *Copera marginipes* made repeated egg-laying movements in a dirty seam on a shiny black bench. Dragonflies have been attracted to shiny roofs of automobiles. It has been reported that *Libellula depressa* dragonflies laid their eggs onto a glass pane of a greenhouse. It has been experienced that mature individuals of *Pantala flavescens* performed sexual be-

haviour and oviposition movements over shiny roofs of tents. It has been found that the flight activity of dragonflies above shiny plastic sheets laid on *Sphagnum* bog was significantly higher than above control plots without plastic. Although these authors experienced that the dragonflies performed sexual behaviour and oviposition movements over the shiny surfaces mentioned, they did not recognize the important role of polarotaxis in the habitat choice of dragonflies and in their deceiving by different artificial shiny black surfaces.

It has been observed that the dragonfly species *Sympetrum flaveolum*, *S. striolatum*, *S. sanguineum*, *S. meridionale* and *S. danae* were attracted by polished black gravestones in a Hungarian cemetery without lacking any water body. The dragonflies showed the same behaviour as at water: (i) they perched persistently in the immediate vicinity of the chosen gravestones and defended their perch against other dragonflies (see [Figure 9.3Q](#)); (ii) flying individuals repeatedly touched the horizontal surface of the shiny black tombstones with the ventral side of their body; (iii) pairs in tandem position frequently circled above black gravestones. Tombstones preferred by the dragonflies had an area of at least 0.5 m² with an almost horizontal, polished, black surface, the sky was open above them, and there was at least one perch in their immediate vicinity.

Using imaging polarimetry, it has been found that the black gravestones, like smooth water surfaces, reflect highly and horizontally polarized light (see [Figure 9.2F](#)). In double-choice field experiments it has been showed that the dragonflies attracted to shiny black tombstones possess positive polarotaxis and therefore, under natural conditions, detect water by means of the horizontally polarized reflected light. Positive polarotaxis in several other dragonfly species has been demonstrated in other field experiments, in which it was also shown that these insects detect water by the horizontal polarization rather than by the intensity and/or colour of reflected light. The positive polarotaxis and the reflection-polarization characteristics of black gravestones explain why the observed *Sympetrum* dragonflies were attracted to black tombstones.

It has been observed that tandems of *Sympetrum* species oviposited onto nearly horizontal glass panes in a botanic garden. *S. striolatum* has been seen to oviposit on an old plastic wind-screen thrown into a chalk pit and on glass panels of a tractor door that was left laid flat on grass. When the door was placed in an upright position the tandems lost their interest in it. It has been reported that a tandem of *S. vulgatum* laid 50 eggs on the metallic-green bonnet of a car. Black gravestones can also elicit oviposition in dragonflies. Actual and attempted egg-laying by dragonflies onto horizontal shiny dark surfaces such as car bodies, oil pools or asphalt roads has repeatedly been observed (Horváth and Varjú, 2004). All these man-made substrata can constitute ecological traps *sensu* Schlaepfer *et al.* (2002), thus reducing the dragonflies' individual fitness. The same is true for certain gravestones. The deception of *Sympetrum* species observed in a Hungarian cemetery could be a global phenomenon. In fact, in 2007 it has also been noticed in a graveyard near Lake Neusiedl, Austria.

9.3.6 Polarotactic Caddis Flies Lured to Glass Surfaces

The caddis flies *Hydropsyche pellucidula* emerge near sunset from the river Danube and swarm around trees and bushes on the river bank. It has been observed that these aquatic insects can also be attracted *en masse* to the vertical glass surfaces of buildings on the river bank in Budapest, Hungary (see [Figure 9.3R](#)). The mass-swarmed *H. pellucidula* lured to dark vertical glass panes landed, copulated and remained on the glass for hours (see [Figure 9.3S,T](#)). Many of them became trapped by the partly open tiltable windows. In laboratory

choice experiments it has been showed that ovipositing *H. pellucidula* are attracted to highly and horizontally polarized light stimulating their ventral eye region, and thus have positive polarotaxis, the function of which is to detect water, from which they emerge, and to which they must return to oviposit. In the field it has been documented that highly polarizing vertical black glass surfaces are significantly more attractive to both female and male *H. pellucidula* than weakly polarizing white ones.

Using imaging polarimetry, Malik *et al.* (2008) measured the reflection-polarization characteristics of shady and sunlit, black and white vertical glass surfaces of buildings (where caddis flies swarmed) from different directions of view under clear and overcast skies in the red, green and blue parts of the spectrum. Using these polarization patterns, they determined those areas of the investigated glass surfaces (see [Figure 9.2G](#)), which are sensed as water by polarotactic insects facing and flying toward, or landed on a vertical pane of glass. Malik *et al.* (2008) showed that the attraction of flying caddis flies to vertical glass surfaces and that these insects remain on vertical panes of glass after landing can be explained by the reflection-polarization characteristics of vertical glass surfaces and the polarotactic behaviour of these aquatic insects. They proposed that after its emergence from the river, *H. pellucidula* is attracted to buildings by their dark silhouettes and the glass-reflected horizontally polarized light. After sunset this attraction may be strengthened by positive phototaxis elicited by the buildings' lights. Since vertical glass panes of buildings are abundant in the man-made optical environment, and polarotactic aquatic insects are spread world-wide, these results are of general interest in the visual and behavioural ecology of aquatic insects.

9.3.7 Polarized Light Pollution of Solar Panels and How it can be Reduced by White Grid-Patterns

Solar panels (including solar collectors and photovoltaic solar cells) are rapidly spreading in the world (see [Figure 9.5](#)). They usually have large, smooth, black surfaces maximizing the absorbed sunlight. Using imaging polarimetry, the reflection-polarization characteristics of solar panels have been measured (see [Figures 9.2H](#) and [9.6, 9.7](#)). It has been showed that solar panels can reflect highly and horizontally polarized light, and thus are strong sources of polarized light pollution. Therefore they can be ecological traps for aquatic insects, which are attracted to horizontally polarized light, because they detect water by the horizontal polarization of light reflected from the surface of waters, where their larvae develop. In a Hungarian nature reserve park it has been observed that stoneflies (Trichoptera), mayflies (Ephemeroptera), dolichopodid dipterans and tabanid flies (Tabanidae) were attracted in a large number to photovoltaic solar cells, where they swarmed, displayed reproductive behaviour, and the females oviposited onto the shiny black solar cells (see [Figure 9.8](#)). Multiple-choice field experiments have been performed with these insects to study their attraction to solar panels. Solar panels polarized reflected light almost completely (degree of polarization $d \approx 100\%$) and substantially exceeded typical polarization values for water ($d \approx 30-70\%$).



Figure 9.5. Solar panels, including solar collectors (SCs) and photovoltaic solar cells (PVSCs), are increasingly spreading in the world. (A) A passive, closed-circuit water heating system using a multi-flow shiny black solar collector panel providing high-efficiency absorption of solar radiation. The heated-up water is stored in black cylindrical tanks (<http://www.solahart.com>). (B) A SC on the roof of a house (<http://www.weishaupt.de>). (C) A German solar electricity plant with thousands of PVSCs (<http://www.phoenixsolar.com>). (D) Horizontal PVSCs on a rooftop in Farmingdale, Long Island, NY, USA (http://www.power-technology.com/projects/long_island). (E) PVSCs on the rooftop of a building of the Szent István University in Gödöllő, Hungary (photo by G. Kriska). (F) PVSCs on the roofs of family houses in the Japanese solar town, Kiyomino (<http://www.iea-pvps-task10.org>).

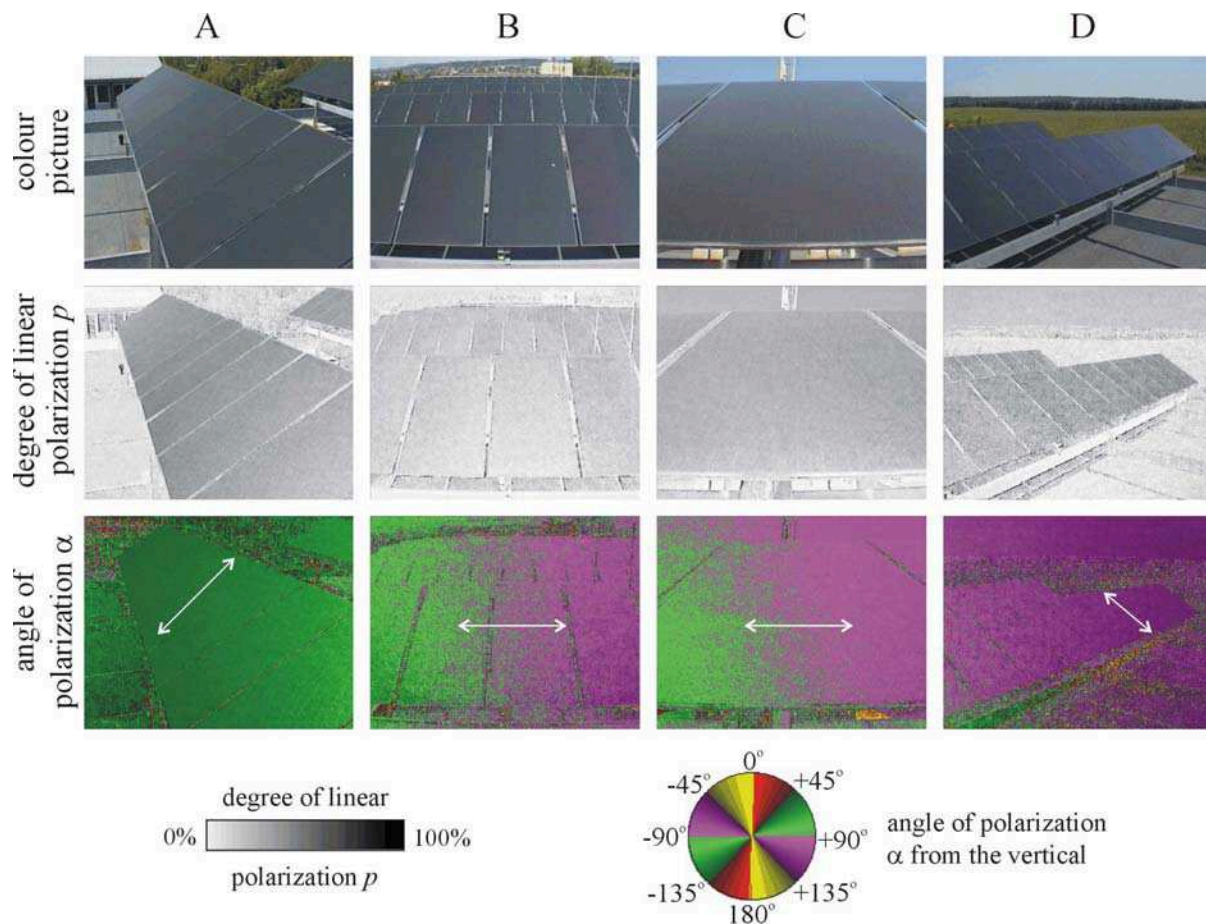


Figure 9.6. Colour pictures and patterns of the degree of linear polarization p and the angle of polarization α (clockwise from the vertical) of photovoltaic solar cells with homogeneous shiny black surface on the rooftop of the Szent István University in Gödöllő, Hungary measured in the green (550 nm) part of the spectrum from four different directions of view: from right (A), from front and above (B), from front and below (C), from left (D). In the α -patterns the double-headed arrows show the local direction of polarization of reflected light.

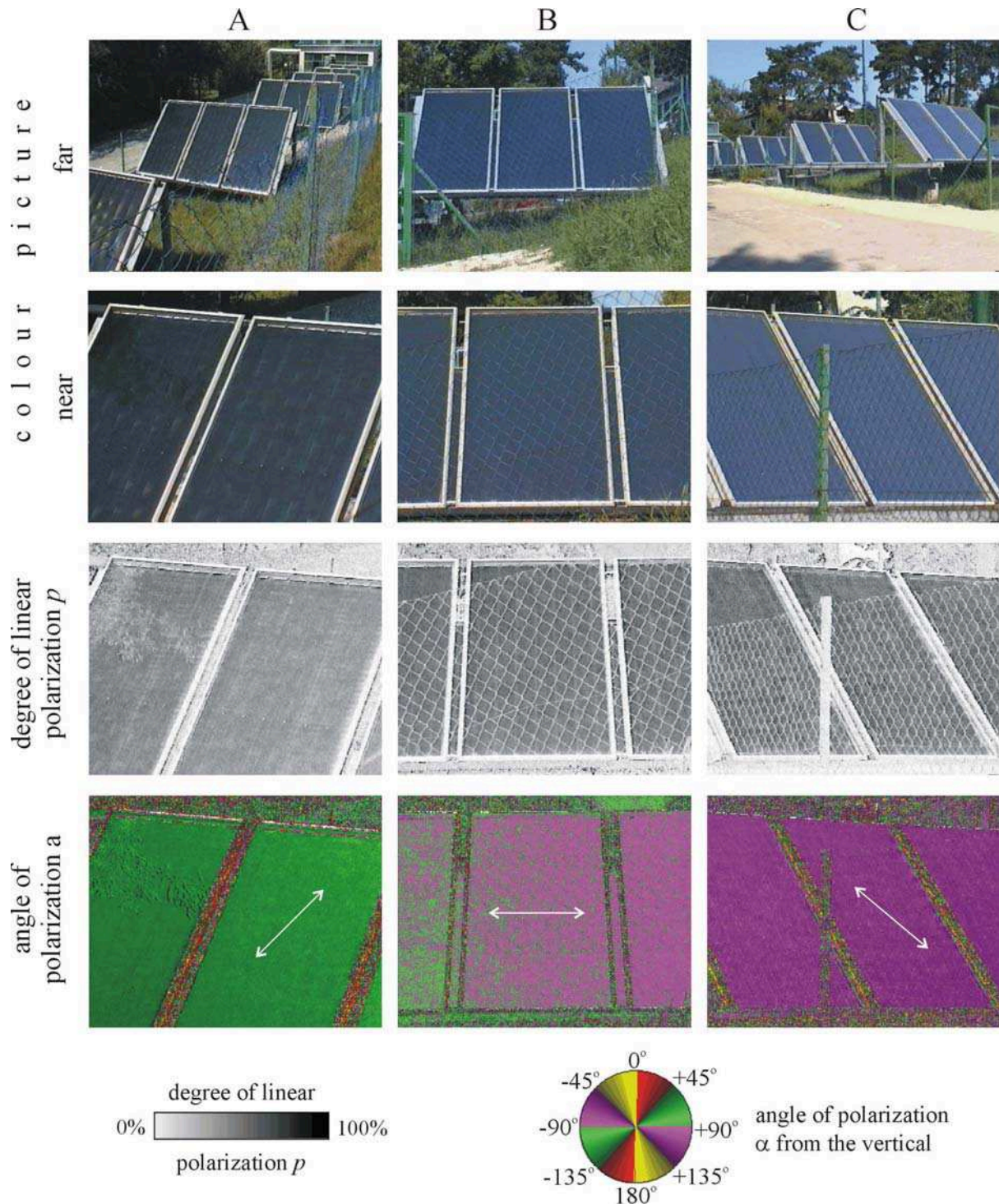


Figure 9.7. As [Figure 9.6](#) for solar collectors in the garden of the Szent István University in Gödöllő, Hungary viewed from three different directions of view: from right (A), from front (B), from left (C). In the p -patterns of columns B and C it can be seen that the wire-grid of the fence reflects practically unpolarized light.

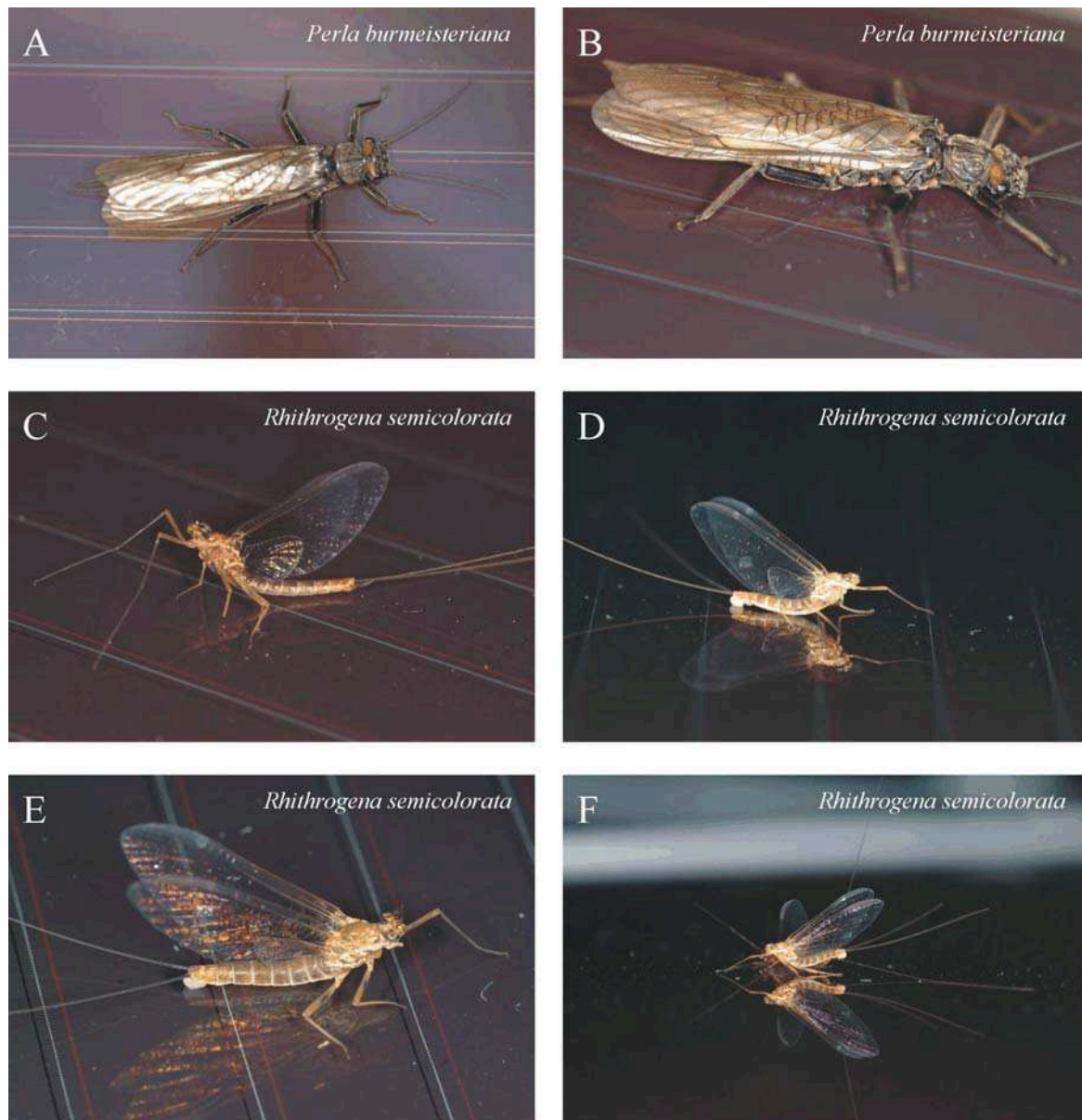


Figure 9.8. Polarotactic insects attracted to and landed on the shiny black surface of a shiny-black-framed photovoltaic solar cell on 23 May 2008 in the National Park at Dömörkapu, near Budapest, Hungary in our 3rd experiment. (A, B) Adult female stonefly (*Perla burmeisteriana*). (C) Male mayfly (*Rhithrogena semicolorata*). (D-F) Female *R. semicolorata* with a white egg batch on the end of her abdomen.

These insects were attracted to solar panels more often than to surfaces with lower degrees of polarization (including water), but in general they avoided solar panels with nonpolarizing white borders and white grates (see [Figure 9.9](#)). The highly and horizontally polarizing surfaces that had nonpolarizing, white cell borders were 10- to 26-fold less attractive to insects than the same panels without white partitions (see [Figure 9.10](#)). Although solar panels can act as ecological traps, fragmenting their solar-active area do lessen their attractiveness to polarotactic insects. This gave a new and reliable clue to reduce, or even eliminate the polarized light pollution of solar panels. These results are of general importance, because they are valid to all kinds of shiny black artificial surfaces reflecting highly and horizontally polarized light. It has been proposed that the strong polarized light pollution of such surfaces can be elimi-

nated by an appropriate depolarizing (e.g., white) grid-pattern, which diminishes their attractiveness to polarotactic aquatic insects, and thus makes them more environmentally-friendly.

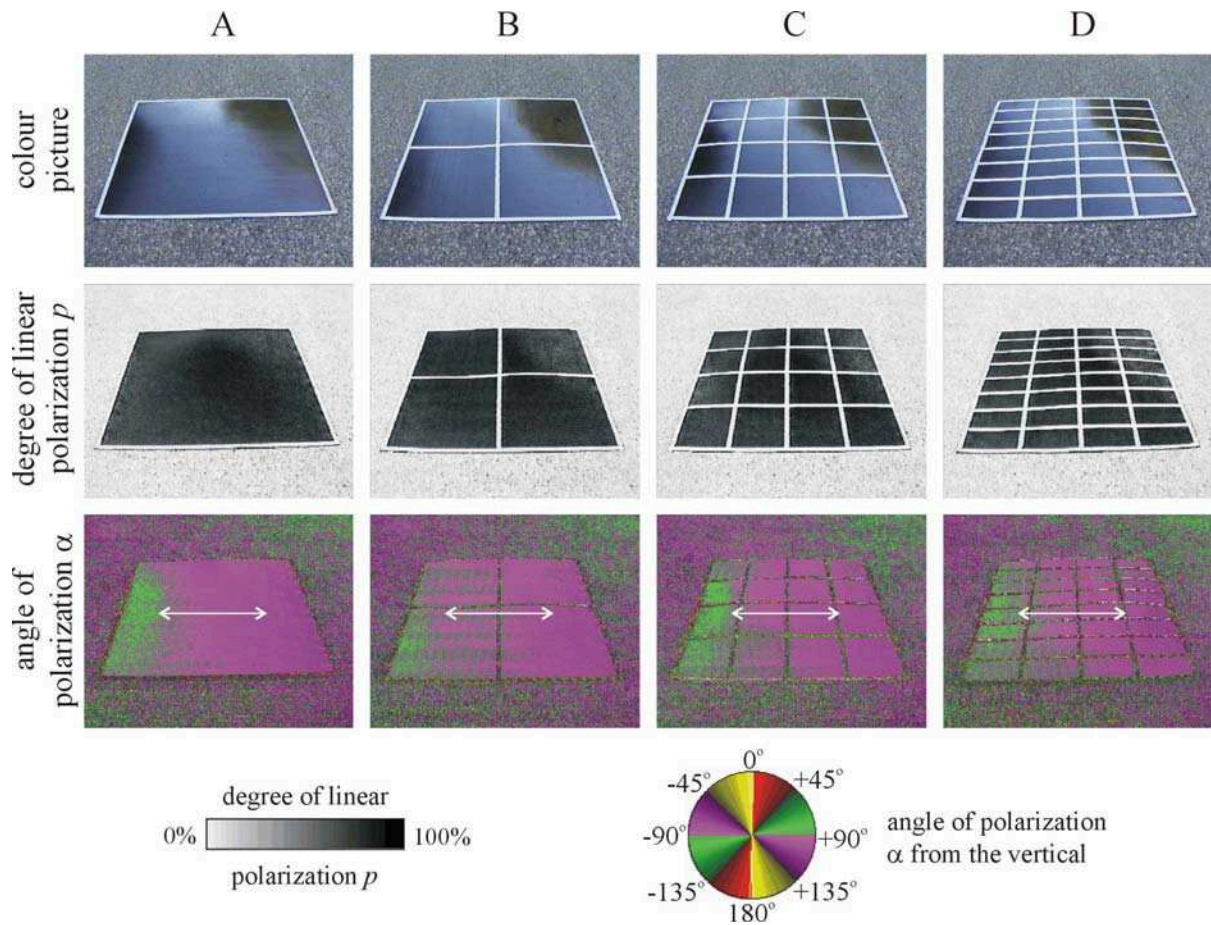


Figure 9.9. As [Fig. 9.6](#) for four different sticky test surfaces laid on a dry asphalt road in a choice experiment.

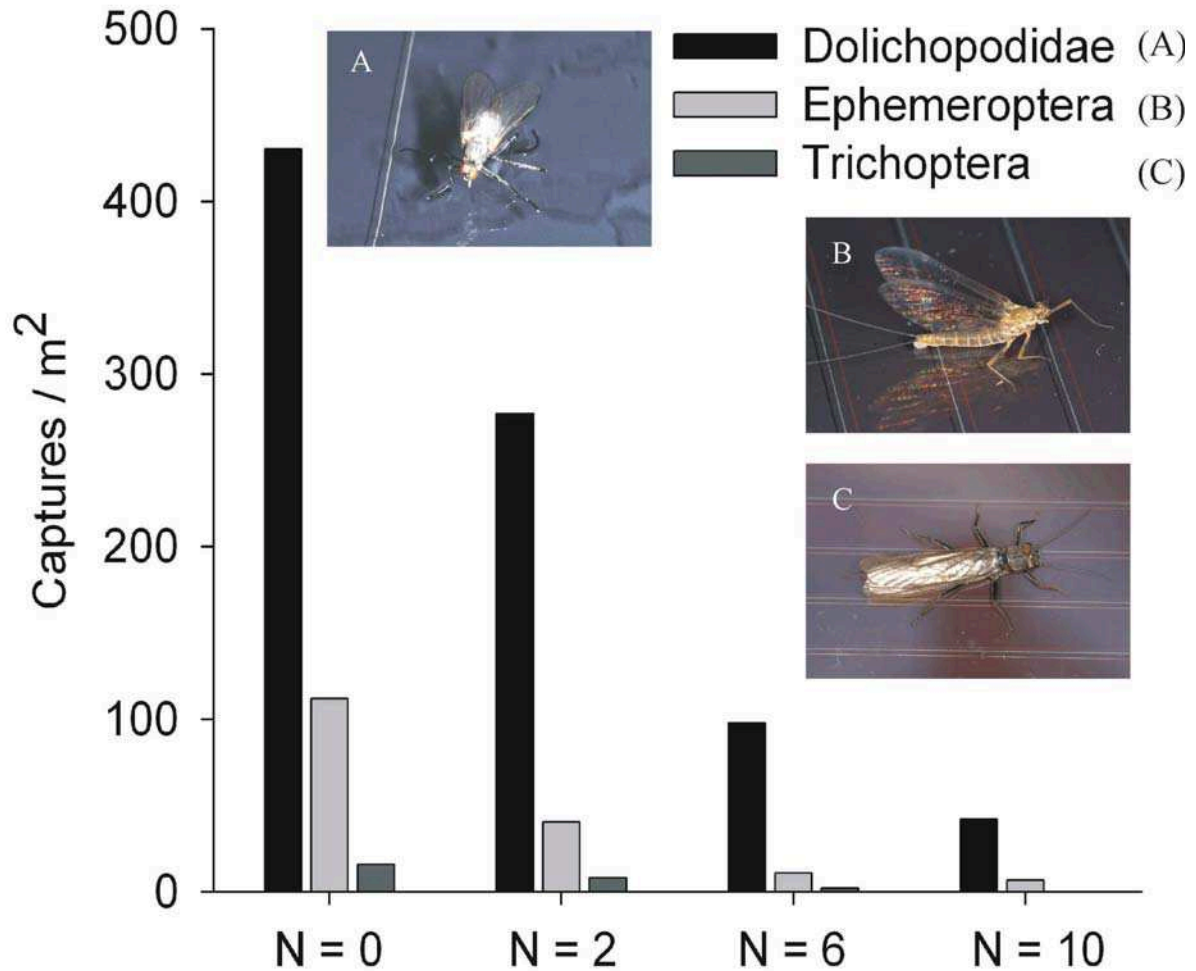


Figure. 9.10. The surface density of polarotactic dolichopodid (Diptera), mayfly (Ephemeroptera) and Philopotamus (Trichoptera) species trapped by sticky polarizing surfaces with N orthogonal white stripes in a choice experiment. Captures / m^2 are given calculated as $n(N) = m \cdot 1m^2 / A_{\text{black}}(N)$, where m is the total insect captures per test surface, $A_{\text{black}}(N)$ is the sum of the black polarizing area. Reduction in surface density associated with maximal partitioning were significant for all groups (binomial χ^2 test, $N=0 / N=10$; dolichopodids: $\chi^2 = 320$, $df = 1$, $P < 0.0001$; Ephemeroptera: $\chi^2 = 93.3$, $df = 1$, $P < 0.0001$; Trichoptera: $\chi^2 = 14.2$, $df = 1$, $P = 0.0002$). Insets: Photographs of polarotactic aquatic insects attracted to the shiny black surface of a photovoltaic solar cell in the National Park at Dömörkapu near Budapest in Hungary. (A) Aquatic dolichopodid. (B) Female mayfly (*Rhithrogena semicolorata*) with a white egg batch on the end of her abdomen. (C) Female stonefly (*Perla burmeisteriana*).

In certain cases the remedy of polarized light pollution, i.e. a depolarizing grid-pattern, is already embedded into the technology: The solar panel used in field experiments had a depolarizing white frame, and certain kinds of photovoltaic solar cells also possess such a white frame and a white grating (right panel in Figure 9.2H). These white gratings are the result of certain technical requirements, but fortunately as an unexpected byproduct, they also reduce the polarized light pollution. It is an important task of future research to determine the optimal pattern of these depolarizing gratings, which can maximally reduce the attractivity to polarotactic insects. It is probable that the optimal parameters (the width of depolarizing stripes and their distances from each other) of the depolarizing grid-pattern depends on the species. Solar

panels provided with such an optimal depolarizing grid-pattern can surely appear in the market as environmentally-friendly new products.

What could be the reasons for the phenomenon that a simple depolarizing white frame/grid-pattern can reduce/eliminate the attractivity of horizontal shiny black surfaces to positively polarotactic insects? Figure 9.11 summarizes a possible explanation of this unexpected phenomenon.

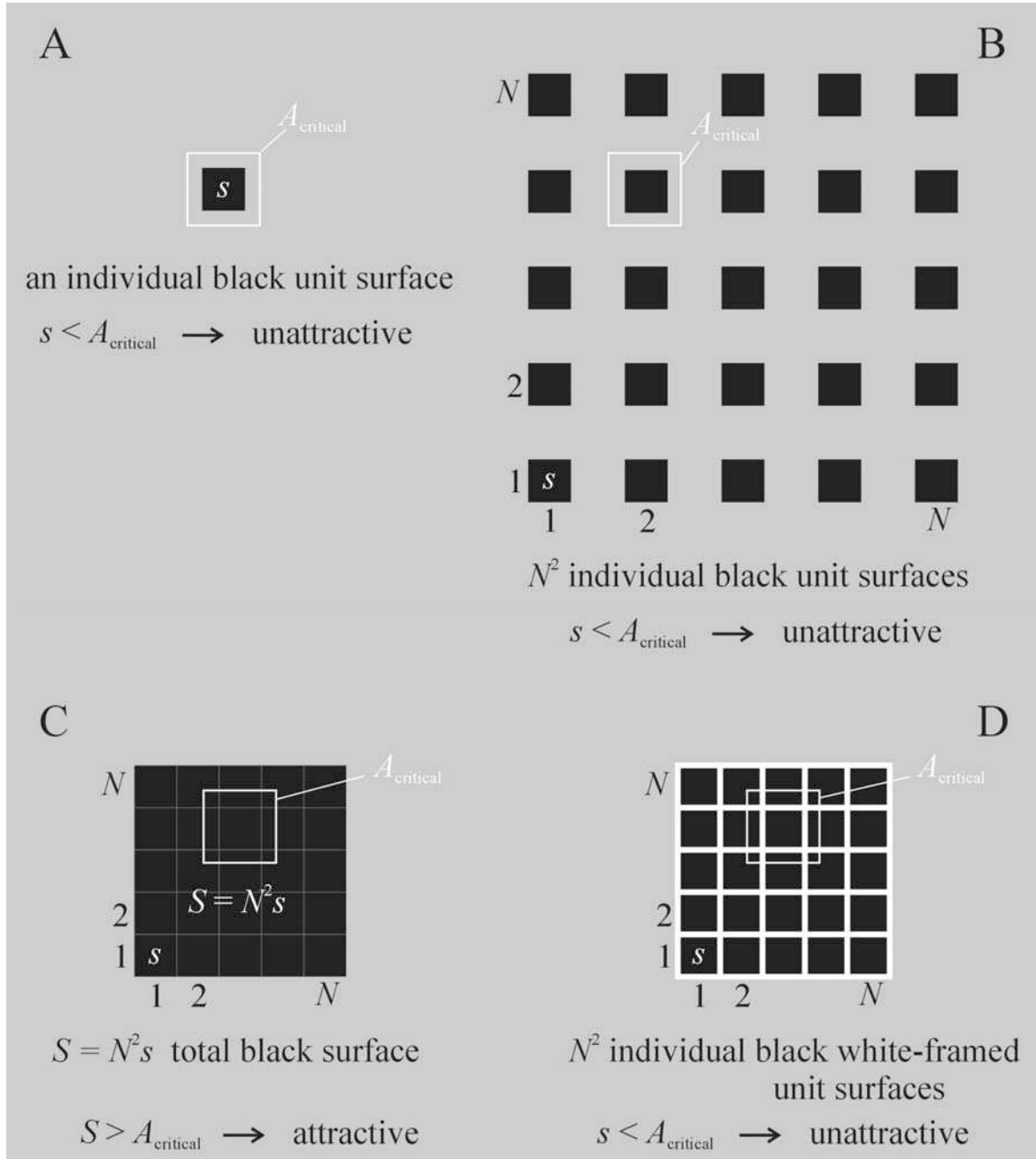


Figure 9.11. For explanation of the attractivity/unattractivity of horizontal shiny black surfaces without/with depolarizing white grid-patterns to positively polarotactic insects. A horizontal shiny black surface is attractive/unattractive to such insects, if its area is larger/smaller than a species-specific critical area A_{critical} .

A highly and horizontally polarizing shiny black surface is attractive/unattractive to polarotactic insects if its area s is larger/smaller than a critical area A_{critical} , that may be dependent on species. The biological reason for this is the following: At a given aquatic insect species there is a minimal and a maximal dimension of the bodies of water where the larvae can optimally develop. Thus the adult females of this species lay their eggs only in such water bodies, the surface of which is neither smaller, nor larger than these lower and upper limits.

- If there is only one black unit surface with $s < A_{\text{critical}}$ in a given optical environment, it is unattractive to water-seeking flying polarotactic insects (see [Figure 9.11A](#)).
- If there are $N \times N = N^2$ such black unit surfaces at large enough distances from each other, each of them functions further on as an individual unit surface s , thus remaining henceforward unattractive to polarotactic insects, because $s < A_{\text{critical}}$ (see [Figure 9.11B](#)).
- If, however, there are N^2 black unit surfaces contacting each other, their individual unit surfaces are summed up, functioning as a large surface with $S = N^2 s$ area, and if $S > A_{\text{critical}}$, they are attractive to polarotactic insects (see [Figure 9.11C](#)).
- If there are N^2 white-framed unit surfaces s contacting each other (see [Figure 9.11D](#)), they are separated from each other by a depolarizing white frame, and thus their individual unit surfaces s cannot be summed up in the visual system of the approaching polarotactic insect. Consequently each of them functions as an individual unit surface s , and if $s < A_{\text{critical}}$, they remain unattractive to polarotactic insects, in spite of the fact they contact each other (see [Figure 9.11D](#)). The prerequisites of this effect are that the depolarizing separations, i.e. the white stripes, should be wide enough, and their number has to be large enough.

9.4 Possible Benefits and Disadvantages of Insectivorous Predators from PLP

In the term PLP 'polarized light' refers to the fact that this phenomenon is elicited only by horizontally polarized light, and 'pollution' communicates the fact that the primary effects of PLP are adverse for the insects deceived by and attracted to light with horizontal polarization. Note, however, that the secondary effects of PLP could also be advantageous: If certain animals (e.g., insectivorous birds, spiders and bats) can feed on the polarotactic insects attracted to artificial horizontally polarized light, they can take advantage of PLP. The hunting of insects attracted to streetlamps at night by anuran amphibians, reptiles, birds, bats and spiders is a well known secondary effect of the conventional (non-polarized) ecological photopollution.

It has been reported that wagtails (*Motacilla alba* and *M. flava*) were lured by polarotactic insects attracted to highly and horizontally polarizing huge black dry plastic sheets laid on the ground. These wagtails systematically hunted and caught the insects above or on the plastic sheets which functioned like huge bird feeders. Kriska *et al.* (1998) observed that wagtails (*Motacilla alba*) frequently gathered the mayflies swarming and copulating above, and ovipositing on the dry asphalt roads running near creeks and rivers in suburban regions. It has also been observed that the caddis flies attracted to vertical glass surfaces of buildings on the bank of the river Danube in Budapest lured numerous different birds, such as european magpies, white wagtails, house sparrows and great tits. These birds systematically hunted and caught the caddis flies landed on the glass panes or swarming at the windows (see [Figure 9.4A-H](#)). Spiders also fed on these caddis flies on the bleak walls (see [Figure 9.4I-L](#)).

As a first approximation we can assume that the mentioned predators benefit from the abundance of caddis flies attracted to the glass surfaces as prey animals. An additional advantage of glass buildings from these predators' point of view, could be that they supply food (caddis

flies) on a temporally and spatially more predictable basis than other habitats. This may be obvious for the attracted magpies, possessing no predators around the glass buildings. On the other hand, however, the numerous magpies lured by the caddis flies mean an enhanced predatory risk for the chicks of house sparrows, white wagtails and great tits, because magpies are dangerous nest predators of other smaller birds. This situation could be an ecological trap for sparrows, wagtails and tits: (i) The abundance of caddis flies lured to the glass surfaces attracts the mentioned bird species. (ii) These birds may lay their eggs in the vicinity of the glass buildings due to the insect prey abundancy. (iii) The chicks of wagtails, sparrows and tits could be predated by the magpies, which can destroy the local wagtail, sparrow and tit populations. On the other hand, due to the temporal food abundance more wagtails could grow up, but these birds might not find enough insects for survival after the caddis fly swarming. The birds attracted by the caddis flies swarming at glass surfaces feed also on the spiders lured (see [Figure 9.4I-L](#)). Thus, these spiders are not only predators, but also prey animals in this food web.

Similar, but a more complex food web has been observed at the open-air waste oil reservoir in Budapest: The highly and horizontally polarizing black oil surface attracted different polarotactic aquatic insect species in large numbers. These insects lured various insectivorous birds and bats, which were trapped by the sticky oil (see [Figure 9.4M-P, 4R](#)). The carcasses of these entrapped birds and bats attracted different carnivorous birds (e.g., owls and hawks), which have also been trapped by the oil (see [Figure 9.4Q,S,T](#)). Finally, all members of this food web based on the PLP of the waste oil surface were killed by the oil (see [Figure 9.4M-T](#)).

We have mentioned above that tabanid flies are also polarotactic, thus they can be attracted to all highly and horizontally polarizing surfaces. This PLP of shiny black surfaces can be used to develop new optically luring tabanid traps being more efficient than the existing ones based on the attraction by the brightness and/or colour of reflected light. This is disadvantageous for the local tabanid population, but is a benefit for humans and their domestic animals, because tabanids are spread world wide, and their females are usually haematophagous. Since female tabanids suck also the blood of domestic animals and humans, they are vectors of numerous dangerous animal and human diseases and/or parasites such as anthrax, tularemia, anaplasmosis, hog cholera, equine infectious anemia, filariasis and Lyme disease.

9.5 Suggested Remedies of PLP

Not every artificial horizontal surface reflecting light with high p induces PLP. Although they are horizontal and sometimes highly polarizing, certain surfaces do not attract polarotactic aquatic insects. Such surfaces are, for example, sunlit roads and plains. On sunny days mirages may appear on these hot surfaces, when there seems to be a pool of shiny water in the distance, which dissolves on approach. The sky, landmarks and objects are mirrored in this "pool". Using imaging polarimetry, it has been measured and compared the polarization characteristics of a mirage and a water surface. It turned out that the light from the sky and the sky's mirage has the same p and α . Since the direction of polarization of skylight is usually not horizontal, the non-horizontally polarized light from mirages is unattractive to polarotactic aquatic insects. On the other hand, there are large polarization differences between the skylight and the water-reflected light, the latter being usually horizontally polarized, and thus attractive to polarotactic aquatic insects. Mirages are not usual reflections, but are formed by gradual refraction and a total reflection of light. Such gradual refractions and total reflection do not change the state of polarization of light. Mirages can imitate water surfaces only for those animals, whose visual system is polarization-blind, but sensitive to brightness and colour differences. A polarization-sensitive water-seeking insect is able to detect the polarization

characteristics of a mirage. Since these characteristics differ considerably from those of water surfaces, polarotactic insects cannot be deceived by and attracted to mirages, which thus cannot induce PLP.

Another example is a sunlit black burnt stubble-field. Due to the Umow effect (the darker a surface, the higher the degree of linear polarization of light reflected by it) p of light reflected from the black ash layer of burnt stubble-fields is very high. Numerous black burnt stubble-fields have been monitored, but aquatic insects or their carcasses have never been found in the ash, although flying polarotactic insects were abundant in the area, which was shown by attracting them to horizontal black plastic sheets in the vicinity of the investigated burnt stubble-fields. From this it was concluded that black burnt stubble-fields are unattractive to polarotactic aquatic insects. The reason for this is that the ash layer is a rough surface due to the random orientation of the charred stalks of straw. The consequences of this roughness are that the direction of polarization of light reflected from the black ash is nearly horizontal only towards the solar and antisolar meridians, and it is tilted in other directions of view, furthermore the standard deviation of both the degree p and angle α of linear polarization of reflected light is large.

On the basis of burnt stubble-fields, one of the possible remedies of PLP can be to make the reflecting surfaces inducing PLP as rough as possible: the rougher a surface, the lower the p of reflected light. If the surface roughness is so large that p of reflected light is lower than the threshold p^* of polarization sensitivity of a polarotactic insect, then the surface is unattractive to this insect, because it does not perceive the polarization of reflected light. On the other hand, the direction of polarization of light reflected from rough surfaces is usually not horizontal, thus rough surfaces are usually unattractive to polarotactic insects, which are lured only to exactly/nearly horizontally polarized light.

It has been proposed that visitors to wetland habitats should drive light-coloured (instead of black, red or dark-coloured) cars, to avoid egg loss by confused polarotactic aquatic insects. Due to depolarization by diffuse reflection, very dirty cars reflect light with much lower p than recently washed and/or waxed shiny cars. Thus, the most environmentally friendly car of all would be one that never gets washed. In other words, the "greenest" car is white and dirty. Such a car minimizes the PLP.

After the discovery of the causes of the reproductive behaviour of mayflies above dry asphalt roads (Kriska *et al.*, 1998), the experts of protection of animals and environment could take the necessary measures to prevent the egg-laying by mayflies and to reduce the amount of eggs laid and perished on asphalt surfaces: One could, for example, treat the sections of the asphalt roads running near the emergence sites of Ephemeroptera in such a way that their surface becomes relatively bright and rough to reduce reflection polarization. This could be performed by rolling down of small-sized bright gravel on the asphalt surface. This treatment of asphalt reduces significantly the p of reflected light, which abolishes its attractiveness to polarotactic mayflies.

The huge shiny black plastic sheets used in the agriculture can also deceive, attract and kill *en masse* polarotactic aquatic insects, if they are laid on the ground near the emergence sites (wet-lands) of these insects. It would be advisable to forbid the farmers to use such black plastic sheets near wet-lands, where white or light grey plastic sheets (if appropriate) should be preferred. Another possible remedy could be to develop and use such a plastic material, which would reflect light efficiently in the ultraviolet (UV) and visible (VIS) parts of the spectrum,

but absorb light strongly in the infrared (IR) spectral range. Such plastic sheets would reflect weakly polarized light in the UV and visible spectral ranges, and could keep the soil covered by them warm, which is one of the major functions of the black plastic sheets in agriculture. The UV/VIS-reflecting and IR-absorbing plastic sheets would not induce PLP in those spectral ranges (UV and VIS) where the polarization vision and positive polarotaxis of aquatic insects functions.

It has been showed that the polarotactic caddis flies *H. pellucidula* attracted to vertical glass surfaces can be trapped, if the tiltable windows are open, and thus such glass buildings can be ecological traps for mass-swarmed caddis flies *sensu* Schlaepfer *et al.* (2002). On the basis of the results of Malik *et al.* (2008) we can establish the main optical characteristics of "green", that is environment-friendly buildings considering the protection of polarotactic aquatic insects. These "green" buildings possess such features that they attract only a minimum number of polarotactic aquatic insects when standing in the vicinity of fresh waters:

- Since a smooth glass surface polarizes strongly the reflected light, a "green" building must minimize the used glass material. All unnecessary panes of glass should be avoided that would have only a decorative, ornamental function. In a building practically the only necessary glass surfaces are the windows.
- Since all smooth surfaces polarize highly the reflected light, a "green" building has to avoid bricks with shiny appearing, that is, smooth surfaces. The optimal is the use of bricks with matt surfaces.
- Since according to the Umow rule, the darker a surface, the higher the p of reflected light, a "green" building must especially avoid the use of shiny dark (black, or dark grey, or dark-colored) surfaces. A building covered by dark decorative glass surfaces functions as a gigantic highly and from certain directions of view horizontally polarizing light trap for polarotactic aquatic insects. The windows of dark rooms can also attract polarotactic insects. If the bright curtains are drawn in, the degree of linear polarization of light reflected from the window is considerably reduced, and thus the window becomes unattractive to polarotactic insects.
- Since aquatic insects usually do not perceive red light (Horváth and Varjú, 2004), and thus a red shiny surface seems to them dark and highly polarizing, a "green" building has to avoid the use of shiny red surfaces.
- The surfaces of a "green" building must not be too bright either, because near and after sunset they reflect a large amount of citylight, which can also lure insects by phototaxis. The optimal compromise is the use of medium grey and matt surfaces, which reflect light only moderately with a weak and usually non-horizontal polarization.
- If a building possesses the above-mentioned optical features, it can attract only a minimum number of polarotactic and/or phototactic insects. A further important mechanical prerequisite of the environment-friendly character is that the glass windows of a "green" building must not be tiltable around a horizontal axis of rotation. If partly open, then such tiltable windows can easily trap the insects attracted to them and got in the room. The optimal solution would be the application of windows which can be opened by rotation around a vertical axis. If a building stands near fresh water and has the mentioned unfavourable tiltable windows, it can be made easily "greener" in such a way that its windows are kept closed (if possible) during the main swarming period of the polarotactic and/or phototactic insects swarming in the surroundings.

In sum, the two major remedies of PLP are to reduce the p of reflected light by replacing the highly and horizontally polarizing dark and smooth reflecting surfaces with (1) bright and (2)

rough ones, because such surfaces reflect only weakly and not always horizontally polarized light, which is unattractive to polarotactic aquatic insects. This information should be communicated to professionals such as landscape planners, road and building designers, and policymakers, because their support is necessary to achieve these environmental measures.

The extent of PLP is global, because in the man-made environment highly and horizontally polarizing artificial surfaces (open-air oil surfaces, asphalt roads, black plastic sheets, car bodies, glass surfaces, black gravestones, etc.) are abundant and their world-wide distribution is progressive. Note that the ecologically disruptive highly and horizontally polarized reflected light can be itself the end product of anthropogenic processes that are themselves environmentally damaging: (i) the oil accumulated in oil spills and open-air waste oil reservoirs, for example, is a dangerous biological poison; or (ii) the black plastic sheets used in agriculture are usually composed of non-degradable materials, thus after their agricultural use they enhance only the plastic waste. We would like to emphasize that the measures against PLP are similarly necessary due to the protection of polarotactic aquatic insect populations as the measures against artificial night lighting to protect night-active animals (Rich and Longcore, 2006).

Populations of certain aquatic insect groups, e.g., mayflies and dragonflies are declining in countries with large human densities, which can be attributed to several different factors, including habitat change and destruction. By eliminating or controlling PLP we can reduce one of the factors responsible for this decline. If conservation of aquatic insects is a goal, we must develop and follow policies that minimize the polarized-light-polluting artificial surfaces with which insect mortality and behavioural disruption have been observed. In the urban environment with numerous water bodies or in the vicinity of wetlands an aquatic-insect-friendly building program could be developed, which is effective in reducing aquatic insect mortality by minimizing the sources of PLP.

9.6 Lab course tasks

1. Measure the reflection-polarization characteristics of some typical sources of PLP by imaging polarimetry in the Environmental Optics Laboratory and around the buildings of the Eötvös University.
2. Evaluate the measured polarization patterns by a computer program in the Environmental Optics Laboratory. Then the obtained reflection-polarization characteristics should be considering PLP.
3. Finally, answer the questions of a test about this practice.

9.7 References to this chapter

- Csabai, Z.; Boda, P.; Bernáth, B.; Kriska, G.; Horváth, G. (2006) A 'polarization sun-dial' dictates the optimal time of day for dispersal by flying aquatic insects. *Freshwater Biology* 51: 1341-1350
- Horváth, G.; Zeil, J. (1996) Kuwait oil lakes as insect traps. *Nature* 379: 303-304
- Horváth, G.; Varjú, D. (2004) *Polarized Light in Animal Vision – Polarization Patterns in Nature*. Springer Verlag, Heidelberg – Berlin – New York
- Horváth, G.; Kriska, G.; Malik, P.; Robertson, B. (2009) Polarized light pollution: a new kind of ecological photopollution. *Frontiers in Ecology and the Environment* 7: 317-325

- Kriska, G.; Horváth, G.; Andrikovics, S. (1998) Why do mayflies lay their eggs *en masse* on dry asphalt roads? Water-imitating polarized light reflected from asphalt attracts *Ephemeroptera*. *Journal of Experimental Biology* 201: 2273-2286
- Malik, P.; Hegedüs, R.; Kriska, G.; Horváth, G. (2008) Imaging polarimetry of glass buildings: Why do vertical glass surfaces attract polarotactic insects? *Applied Optics* 47: 4361-4374
- Nowinszky, L. (2003) *The Handbook of Light Trapping*. Savaria University Press, Szombathely, Hungary
- Rich, C.; Longcore, T. (eds.) (2006) *Ecological Consequences of Artificial Night Lighting*. Island Press, Washington - Covelo - London
- Schlaepfer, M. A.; Runge, M. C.; Sherman P. W. (2002) Ecological and evolutionary traps. *Trends in Ecology and Evolution* 17: 474-480
- Schwind, R. (1991) Polarization vision in water insects and insects living on a moist substrate. *Journal of Comparative Physiology A* 169: 531-540

Chapter III.

ENVIRONMENTAL RADIOACTIVITY I. (MEASUREMENTS USING X-RAYS AND GAMMA-RADIATION)

There are several topics of environmental physics that are important in the future of the civilization. Like the great ocean streams may play important role on the issues of climate, the environment-friendly materials mean a big issue of the modern technology and the question of the energy production and the energy demand of the societies is one of the most actual topics. One of the energy sources that is an alternative to the greenhouse emitter fossil fuels is nuclear energy. However, it is the leading part where the technology is developed and distributed all over the world. Of course there are controversial aspects of this kind of energy production, as the use of all the other energy sources can cause various difficulties. We can understand the shine and the shadowed parts of the nuclear energy if we know the scientific background behind it. The nature of radioactivity is one basic step in this way. But not the emission of artificial radioactivity is the only issue, since radioactivity and the ionizing radiations are built into our everyday life, like its medical use, technological applications that make living standards be higher. Moreover the natural radioactivity is an issue of the health of the population and natural radioactive isotopes are subject of scientific research as a tracer or indicator of complex processes.

This book describes experimental physics methods that are subject of the laboratory practices. In the first two chapters we overviewed the subjects of acoustic waves and non-ionizing electromagnetic radiation (EMR). If we increase the frequency of the EMR the photons of the far ultra violet region are already feasible for ionizing the outer electrons of atoms or molecules. There is no sharp limit between ionizing and non-ionizing radiation since it depends on the medium. The UV photons have 3 ranges: A, B, C. The UV B photons can already split some organic molecules in a process that has several steps but starts with ionization or excitation of an electron or these can rearrange chemical bonds. Although the atmosphere absorbs them, the more energetic UVC photons can kill living cells, but rarely ionize the air or the water as a medium. The ionizing radiation starts with the X-rays (or in other words: roentgen radiation) that are discovered by the use of cathode ray tubes, that was the basic instrument of the commercial televisions years ago. The energy of the X-ray photons starts generally around 1 keV and X-rays ionize the medium along their path.

The gamma-radiation is separated from the X-rays not based on their energy but based on the location of their creation. The X-ray photons are created in the electron shell of the atoms or by electrons slowing down or accelerated. Some particle accelerator use very high-energy electrons and lead them on a circular path. That has a radial acceleration, so electrons moving fast on a circle will radiate EMR. That is called the synchrotron radiation, that has high energy but it is not a gamma-radiation since electrons are the emitting substances.

The gamma-radiation comes from the atomic nucleus. There are small energy gamma-radiations like 14.4 keV gamma-photons from the de-excitation of an iron nucleus. That is much lower energy than several type of X-rays, but this is a gamma ray. The gamma photons are created always when an excited nucleus loses the excitation energy or part of it, and emits the energy difference in the form of electromagnetic waves. There are protons in the nuclei having charge, and their rearrangement in a simplified picture needs the acceleration of the charged particles. This de-excitation can happen by the processes where protons change their orbits inside the nucleus or the whole nucleus rotation slows down. (There are other processes, too.) This is just the main picture to describe the creation of a gamma photon.

Environmental physics uses X-ray and gamma-photons for technological benefits on one hand, but on the other hand these radiations are part of the natural human radiation dose. Eve-

ry material can contain radioactive isotopes and can emit e.g. gamma photons. The natural ionizing radiation makes harm on living cells but the cells can react and save their normal operation. We are continuously target of the natural gamma and X-ray radiations, the Homo sapiens evolved in this environment. But there are special situations where the intensity of X-rays or gamma-radiation is higher than expected. These factors make important to know the laws concerning the ionizing electromagnetic waves. Chapter IV will focus not on the ionizing electromagnetic waves but on the ionizing particle radiations. Both categories have similar health effects since the ionizing processes are similar.

In summary, Chapter III contains experiments with X-rays and gamma-radioactivity, but in Chapter IV there are experiments where alpha- and beta-radioactivity occurs.

10. The basic physical principles of radioactivity

Although this chapter is about the experimental methods that use gamma-radiation or X-rays the general knowledge on radioactivity is summarized here at the beginning and this will be applicable later in Chapter IV, as well.

10.1 The basic laws of radioactive decay

The activity A is the decay (or disintegration) per second. In a *simple decay* if the number of decaying nucleus is $N(t)$:

$$A = -\frac{dN}{dt} .$$

In this expression the minus sign shows that the activity is positive but the number of decaying nuclei are decreasing. In a decay chain for a daughter isotope the change in the number of these nuclei can be due to the decay but also due to their production from the mother nucleus, as well. Therefore the above formula is valid only for simple decays.

The activity is proportional to the number of decaying atoms in a general case. $A = \lambda N$. This gives the statistical average of the decayed atoms or nuclei in a second. Generally this number has a statistical distribution. This is called the statistical feature of the radioactivity.

The radioactive decay is described by a differential equation:

$$\frac{dN}{dt} = -\lambda N ,$$

where N is the number of radioactive nuclei ready to decay, and λ is its the decay constant.

In this case (so called “simple decay”) the solution of this differential equation is

$$N(t) = N_0 e^{-\lambda t} .$$

This shows the number of radioactive atoms decreasing exponentially. This is called the *exponential decay law*. If we have N_0 atoms at the beginning ($t=0$), then after t time there will be only $N(t)$ atoms. Here λ is the decay constant; its dimension is 1/s. It means how much the probability of the decay of one atom is in one second.

Using this formula we can determine how many atoms are there in a sample after waiting some time, or we can calculate how many isotopes were in the sample some time before it was measured. The half-life ($T_{1/2}$) is the time during the number of particles will decrease to their half:

$$N_0 / 2 = N_0 \exp(-\lambda T_{1/2}) .$$

Taking the logarithm of both sides, we get the well-known formula:

$$\lambda = \frac{\ln 2}{T_{1/2}} .$$

If we calculate the derivative of both sides of the exponential decay law we get the following expression:

$$A(t) = -N'(t) = -(-\lambda)N_0 \exp(-\lambda t) = \lambda N(t)$$

Therefore: $A = \lambda N$. This was already mentioned above, but here is already proved. This expression is the most commonly used formula in nuclear analytics. This is the basic formula that we can use for calculating the number of a given isotope in a sample from the measurement of its activity.

The heavy radioactive elements that occur in nature (^{238}U , $T_{1/2} = 4.468 \times 10^9$ years, ^{232}Th , $T_{1/2} = 1.40 \times 10^{10}$ years and ^{235}U , $T_{1/2} = 0.72 \times 10^9$ years) each have long decay chains that end around the mass number of 210, reaching stable isotopes. During the process there are many α and β decays, and after each of those, dozens of γ photons can be emitted.

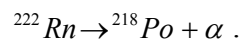
10.2 The qualitative description of radioactivity

The radioactivity or radioactive decay is the process when an atomic nucleus changes spontaneously. Its atomic number and mass number can be changed but it is not necessary. Generally at radioactive decays a fast particle is created and travels into one direction of space. This fast particle or electromagnetic radiation is called radioactive radiation.

A little more general concept is the ionizing radiation. All the charged radioactive radiations ionize the media in that they travel. But not only these radiations can ionize. A good example for the difference is the protons in the cosmic rays. The protons are mostly from the Sun in the cosmic radiation. Their origin is not a radioactive decay. Close to the surface of the Sun there are time dependent magnetic fields that can accelerate the protons. Therefore fast charged protons are coming from the Sun without radioactivity and these particles will ionize the Earth's atmosphere when they arrive. Then in the atmosphere nuclear reactions can occur. This is again not a radioactivity since these processes are not spontaneous. But they can produce radioactive isotopes. The three types of radioactive decays are the α , β and γ decays. Besides these there is one more spontaneous change of the nucleus: the fission. Nuclear fission can happen spontaneously and also can be induced in the nuclear technology by slow neutrons. However, nuclear reactors are a very important part of environmental physics, in this book we do not cover this issue. The main reason is that it is hardly a subject of a laboratory practice.

10.2.1 Alpha decay

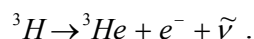
In the course of the α decay, the nucleus emits a ^4He nucleus, and its atomic number decreases by 2, its mass number by 4. (The mass number is the total number of neutrons and protons in a nuclide.) An example for the alpha decay is the decay of the radioactive noble gas, the radon:



The produced new nucleus is the ^{218}Po , is called daughter nucleus. The fast alpha particle is called the alpha-radiation. The nuclear binding of this polonium isotope is stronger than that is in the radon nucleus. The sum of the masses of the products is less than the mass of the radon-222 isotope. That means energy is released. This energy appears in form of kinetic energy. The recoil of the daughter isotope and the alpha particle can be converted to heat. The radioactive decay can produce more macroscopically measurable temperature raise if the activity of the source is high. The radioactive decays in the Earth's crust also produce heat that is important in the Earth's energy balance.

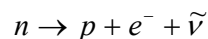
10.2.2 Beta-decay

During β -decay, only the atomic number of the nucleus is changing – increases or decreases by one, and its mass number stays unchanged. The beta-decay has three levels. As a nucleus level we can see nuclei to transform one to another. Like:

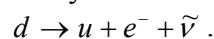


Here the daughter nucleus is the helium-3. But interestingly two fast particles are created. Besides the beta particle, that is the electron, a neutral particle of very low mass is also produced, which is very weakly interacting one. This is called antineutrino.

The second level of the beta-decay is the nucleon level. If we look at the decay above, we can realize that in the triton there are one neutron plus a neutron-proton pair. But in the helium-3 there is a proton plus the neutron-proton pair. The only change that happened is on neutron changed into a proton.



But the structures of the proton and the neutron still are similar! These two particles commonly called nucleon consist of quarks. The quarks are elementary particles to the best knowledge of the science nowadays. The nucleons contain many quark-antiquark pairs in fact, although we generally say there are only three quarks. This question leads to deep parts of particle physics. We now think of only the three quarks picture uud for the proton and udd for the neutron. Again the difference is more simple than expected. Only a d quark changed into u quark. This is the quark level of the beta-decay:



The mechanism of this decay actually is more complex. As an interesting but not very much important detail we mention this. When the d is transformed into u a new particle is created for a very short time: W^{-} that is the exchange particle of this beta decay.

More important feature of the beta-decay that it has very long half-life in a lot of cases, therefore it can produce radioactive isotopes that remain in the environment for a long time. Alpha decay also can have a very long half-life, but this has a different reason.

10.2.3 Gamma-rays

The γ decay leaves both the mass number and atomic number unaltered; the nucleus just de-excites into a state that has a lower energy level, while emitting the energy difference between its states in the form of an electrically neutral particle, the photon (the quantum of gamma-radiation).

10.3 The radioactive families and the secular equilibrium

As we have seen above, radioactive decays either decrease the mass number by 4, or do not change it. For example, the mass number of the members of the ^{238}U chain are 234, 230, 226, 222 etc., stepping down four by four. Thus, isotopes with mass numbers 237, 236 or 235 cannot be produced from a nuclide with a mass number of 238. This means that only four decay families can exist, according to the residual (0, 1, 2 or 3) from the division of the mass number by 4. From these families only those exist today where the half-life of the mother nuclide is not much smaller than the age of the Earth – these are the already mentioned three mother nuclides. The mother of the fourth family is ^{237}Np with a half-life of 2.14 million years, so it has already decayed to extinction over the long lifetime of the Earth.

Let us examine a decay chain where the half-life of the mother nuclide is much longer than the half-life of any of its daughters. In other words, the mothers' decay constant is much smaller than the decay constant of any of its daughters. In this case, after certain amount of time, the activity of all the daughters will be determined by the activity of the mother.

Let us denote the members of the decay chain by **1, 2, 3, 4, ...**, where the sequence of the decay is of the form **1 \rightarrow 2 \rightarrow 3 \rightarrow 4...** Then, the following set of differential equations will be valid for the decay chain:

$$\begin{aligned} \frac{dN_1}{dt} &= -\lambda_1 N_1 \\ \frac{dN_2}{dt} &= -\lambda_2 N_2 + \lambda_1 N_1 \\ \frac{dN_3}{dt} &= -\lambda_3 N_3 + \lambda_2 N_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ \frac{dN_i}{dt} &= -\lambda_i N_i + \lambda_{i-1} N_{i-1} \\ &\dots \end{aligned}$$

Here, the index i denotes the respective member of the $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow \dots \rightarrow i \rightarrow \dots$ decay chain. A simple decay law applies to the 1 mother nuclide. The same would be true for the daughters too, if they were by themselves; for example, the equation $\frac{dN_2}{dt} = -\lambda_2 N_2$ would apply to N_2 . Since over a unit time interval $\lambda_1 N_1$ mother nuclei will decay and become daughters, the number of nuclei 2 increases by that amount, not only decreases due to its own decay.

As an example, let us assume that the half-life of nuclide 1 is 10 years, while for nuclide 2 it is $1\mu\text{s}$. Then, after a few μs , the activity of nuclide 2 will be equal to the activity of nuclide 1 , because only so many nuclide 2 can decay per second, as many has been created, per second. Each decay of the mother is followed immediately by the decay of the daughter. If we conduct a measurement, which is very short with respect to the half-life of the mother (let's say, a few hours or days), the activity of the mother does not really change during our measurement.

This means that $\frac{dN_2}{dt} = 0$. This situation – on a different time scale – is similar to that when the mother nuclide is ^{238}U with $T_{1/2} = 4.468 \times 10^9$ years, since the half-life of its longest living daughter is only a small fraction of that (^{234}U , $T_{1/2} = 2.445 \times 10^5$ years). In that case, of course, the equilibrium does not set in after a few μs , only after a few million years. Based on the set of equations, the activity does not change if the time span of the measurement is very small compared to the half-life of the mother particle:

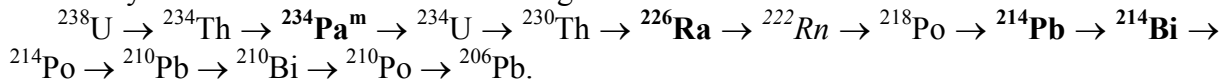
$$\frac{dN_1}{dt} = \frac{dN_2}{dt} = \frac{dN_3}{dt} = \dots = \frac{dN_i}{dt} = \dots = 0.$$

From this it follows that

$$\lambda_1 N_1 = \lambda_2 N_2 = \lambda_3 N_3 = \dots = \lambda_i N_i = \dots = \text{activity},$$

which means that the activity of each daughter is the same as the activity of the mother. This is the equilibrium between the activities, which is called *secular equilibrium*. This way we can obtain the activity of the mother particle by measuring the activity of any of its daughters. From the activity and the decay constant of the mother particle, we can then easily calculate the total number of mother nuclides. We also know that, for example, that 238 g uranium equals 1 mol, in other words, 6.022×10^{23} uranium atoms. Thus, the uranium content of the sample can be calculated from the number of uranium atoms.

We can assume that a natural piece of granite rock is at least several million years old, therefore the secular equilibrium has set in long ago between the mother and the daughter nuclides. The decay chain of the ^{238}U is the following:



The **bold** letters denote the isotopes that emit easily measurable gamma rays. In spite of the secular equilibrium, it is not sure that the activities of all isotopes are equal, because some of the ^{222}Rn (noble gas) can leave the granite by diffusion, before it decays with 3.8 days half-life. In this case, it gets dispersed in the air of the room; its daughters will decay far from the detector and remain undetected. Thus, the activity of the nuclides before and after the radon in the decay chain can differ (as opposed, for example, to the ^{214}Pb and the ^{214}Bi that must have the same activity). In case of thin samples almost the total amount of radon can escape, in which case the gamma lines of the daughters are not visible in the measured spectrum. By the comparison of the radiation of the radium and of the radon-daughters one can even estimate the porosity – or diffusion constant – of the given rock sample.

It can also happen that we find radium (^{226}Ra) in the sample, but no uranium (i.e. neither $^{234}\text{Pa}^m$, nor ^{235}U). This is possible in case of the measurement of timescale with high radium content deposited on water pipes of radioactive thermal springs, or old watch dials painted with radioluminescent radium paint. One should be careful to interpret the measured spectrum correctly.

10.4 The radioactive equilibrium

When a radioactive mother nucleus has a longer half-life than its daughter's, there is a chance to reach radioactive equilibrium. The most used type of this equilibrium is the above-mentioned secular equilibrium. This happens when the decay of the mother nucleus cannot be measured well during our characteristic time scale. For example the uranium has the half-life in the order of billion years. During a person's life cycle only a negligible amount of the uranium will decay. The total numbers of the uranium, hence its activity, remains constant in the statistical uncertainties, generally. But we still observe its decay products, but this number is so small, and will not change the total number. There are different cases, too. For example the ^{222}Rn has a half-life of 3.82 days. That is easily observable when the radon and its daughters in equilibrium change their activity. This time the mother's half-life is greater than the daughters' but not by many orders. This latter case is called the moving equilibrium. For this case we cannot state that we did above:

$$\lambda_1 N_1 = \lambda_2 N_2$$

Since the left side is the decay rate of the mother and at the same time the production rate of the daughter, the right side is the decay of the daughter. If these two are equal, the number of daughter nuclei will not change in time. But it changes simultaneously with its mother, and it is measurable already, in spite of the secular case.

A little more general case of the radioactive equilibrium is when the ratio of the mother and the daughter activities is in between small ranges around a time dependent constant.

$$R - \varepsilon < \frac{A_{\text{daughter}}(t)}{A_{\text{mother}}(t)} < R + \varepsilon$$

In a two-step serial decay: $M \rightarrow D \rightarrow S$ (mother, daughter, stable) $R = \frac{\lambda_D}{\lambda_D - \lambda_M}$. Here λ_i ($i=M, D$) are the decay constant of the mother and daughter nuclei, respectively. If $\lambda_D \gg \lambda_M$ we get the $R=1$ limit case, that is the secular equilibrium.

One important aspect of this is the time that needs to reach the equilibrium. We call it access time. In the simple two-step serial decay scheme, when at the beginning no daughter nuclei are present, the $R(t)=A_D(t)/A_M(t)$ ratio can be calculated:

$$R(t) = \frac{\lambda_D}{\lambda_D - \lambda_M} (1 - e^{-(\lambda_M - \lambda_D)t}).$$

This function will reach the $R-\varepsilon$ threshold during the access time. If we introduce the difference from the time dependent constant in relative units $R-\varepsilon=R(1-\eta)$, where ($\eta=\varepsilon/R$), we can say that $e^{-(\lambda_M - \lambda_D)t} < \eta$ is the criteria to be in equilibrium. This can be reached only if $\lambda_D > \lambda_M$, as we restrained it in the beginning. Therefore the access time in this case is:

$$t_A = \frac{\ln \eta}{\lambda_M - \lambda_D} = \frac{-\ln \eta}{\ln 2} T_D \left(1 + \frac{T_D}{T_M + T_D} \right).$$

If we set the $\eta=3\%$, the $t_A \approx 5T_D$.

10.5 The absorption of ionizing radiation

The ionizing radiation interacts with the medium in what it travels. The main process here is the ionization. This is where its name is coming from. The charged particle radiations simply interact with the Coulomb force between them and the electrons of the medium. But the neutral particles like X-ray and gamma photons or the neutrons first hit a charged particle and then that will ionize the medium.

Charged particles stop soon in solids. The distance that the charged particles travel until they slow down to room temperature (or to the temperature of the medium) is called the range. As a general approach the range of the natural alpha-radiations, whose energy is in the 5-10 MeV range, is in the order of 10 μm . But in air, whose density is 1000 times lower, the range is 1000 times longer, it is about 3-10 cm. The range of electrons from beta-decays in solids is about several cm. These numbers depend on the parameters of the medium. This radiation can be stopped with a material that contains a lot of electrons. But for an electron there is other process to lose its energy. This is the bremsstrahlung, electromagnetic radiation (X-rays) that occurs when an electron has acceleration, in most cases deceleration. In lead shields for example higher energy electrons can produce this radiation.

Our first subject is the thickness dependence of the gamma ray absorption. There are several reaction types, which can absorb the gamma photons. These are the photoelectric effect, the Compton effect and the pair production. The interaction of a certain energy gamma or X-ray photon and an electron via these processes has a probability depending on the photon's energy. The probabilities of atomic processes are described by the concept of the cross section. That gives the number of reaction in a unit time (\dot{N}_r) including some other parameters: $\dot{N}_r = \sigma I \rho dx$, where I is the intensity of the radiation, $I = \Delta N / \Delta t$, N is the number of photons going through during Δt time, ρ is the particle density of the medium, dx is the thickness of a thin layer, and σ is the cross section. The $R = \dot{N}_r \Delta t$ number of reactions will reduce the ΔN after the dx zone. So we can write for the decrease of the intensity that $\Delta N / \Delta t_{\text{before}}$

$R = \Delta N / \Delta t_{\text{after}}$, therefore $I_{\text{after}} - I_{\text{before}} = -R / \Delta t = \sigma I \rho dx$ and from this we can get the differential equation: $\frac{dI}{dx} = -\sigma \rho I$. The solution of this is

$$I(x) = I_0 e^{-\sigma \rho x},$$

the exponential weakening of the intensity. The more absorbing material we put into a given volume, or the more we increase the thickness, the intensity loss follows these exponentially.

This law is valid for gamma rays and X-rays, but for visible light, infrared and UV photons also. The main concept is that it should be neutral particle interaction. Photons do not slow down, as the alpha and beta particles but they disappear by photoeffect or pair production or they change their energy by Compton effect. There are actually other processes also with much less importance. Like gamma-photons can interact with the nuclei of the atoms.

11. Determination of heavy metal contents using x-ray fluorescence analysis (NFS)

11.1 The role of the X-ray fluorescence analysis in the investigation of environmental samples

The civilization has developed quickly in the last centuries. The industry gave tremendous advantages for the everyday life, but it also produced a significant load for the environment. Nobody would like to give up the advantages. But people need to make a balance on the advantages and the disadvantages. We cannot do this without quantitative knowledge on the loads. One of the loads with proved health effects on humans is heavy metal contamination. During this laboratory practice we learn about a method to determine heavy metal concentrations of environmental samples using X-ray fluorescence analysis.

One of these metals is the lead. The traffic used gasoline that contained lead contribution for 80 years. Since the unleaded gasoline is used worldwide the lead atoms has been moving in the soils and in other part of the environment. The contamination was not eliminated, but the growth of it was. The lead was used not only in the traffic but the pipelines were made of lead until the sixties. Many types of paints contain lead as well.

Our everyday life cannot miss the small electronic devices that work using batteries. Almost all types of the batteries contain heavy metals. Cadmium, mercury are two known examples. The selective waste collection is a good solution but still there are contaminations at industrial places.

The production of electric energy means also a large load for the environment. An average type conventional power plant emits about six tons of heavy metal salts in a day.

There are analytical chemical methods available to determine the heavy metal concentrations. But these instruments are not built for field usage. The device using X-ray fluorescence analysis (XRF) or in other words roentgenfluorescence analysis (RFA) can be used on the field. Without preparation of samples it can give quick but not precise results in situ. After homogenization of the samples and lab preparation the method reaches analytical precision. This is a nondestructive method. This feature is very advantageous in many cases.

The RFA method is applicable for investigation of mineral contents as well. Using comparative studies one can give useful information on layout and localization of ore in a mine or research site.

Food also can contain these heavy elements, since the processes, environmental chains will result in this kind of contamination. The RFA method applied for food samples is also useful since the samples can be measured afterward using other, more precise techniques.

11.2 The characteristic X-rays

The roentgen-radiation or X-rays can be created in two ways. One is the Bremsstrahlung (radiation that occurs when charged particles, mainly electrons, slow down), and the other is the characteristic X-rays. X-rays are produced when kicking out an electron from the inner shells, like from the K-shell, ionizes an atom. An electron of the same atom will fill in the vacancy shortly but that has more energy. The energy difference of the electron orbitals will be radi-

ated out by an electromagnetic wave. This energy difference is characteristic of the atomic number of the emitting atom. This is the origin of its name. The energy range of it is generally in the X-ray range if the vacancy was in the K or L shell. An interesting question is how we can kick out an electron from the deep inner part of the electron cloud of the atom. It is possible in several ways. Proton or electron beams can do this, or photoelectric effect by gamma rays can also produce a vacancy most probably on the K-shell (if it is energetically allowed). Therefore to make an atom to emit characteristic X-rays is technologically not difficult.

There is an alternate process for the energy conversion. The energy difference of the two orbitals mentioned above can be transferred to outer electrons of the atom, therefore in some cases not an X-ray will be produced, but electrons will be ejected. This is called Auger-effect. At low atomic numbers the Auger-effect, but at higher atomic numbers the emission of characteristic X-rays is the dominant process.

The energy or the frequency of the emitted X-ray photons was firstly measured by Henry Gwyn-Jeffreys Moseley, and he discovered a relationship between the frequency and the atomic number of the emitting atom in 1913. (Two years later, in 1915, at the age of 27, he died in Gallipoli in a battle of World War I.) The atomic electrons have difficult energy structure but the K-shell electrons show hydrogen-like orbitals. This is due to the screening effect and the electron-electron interaction that is not very important for the deepest energy levels. According to the Bohr-model the energy levels of the hydrogen-like atom, that has atomic number Z , can be calculated in the following way:

$$E_n = -\frac{Z^2}{n^2} hR.$$

Here R is the Rydberg-constant. $R=3.288 \cdot 10^{15} \text{s}^{-1}$, and n is the principal quantum number of the electron. Assuming that this formula is valid for that electron orbital from where the electron fills the vacancy in, we can calculate the energy of the emitted photon:

$$h\nu = \Delta E = Z^2 hR \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \quad (11.1)$$

Here $m(>n)$ is the principal quantum number of the initial electron orbital. This is why the photon energy is characteristic of the atom. Measuring the energy of the photons, the atomic number can be determined using a table. Moreover, if we count how many characteristic photons are emitted, and we precisely know the parameters of the setup, we can determine the quantity of the atoms having atomic number Z . The quantitative analysis is complicated, though, due to the secondary effects that generally occur in environmental samples. But in spite of these obstacles the method is already technologically developed and used in several instruments on the industrial and scientific market. During this laboratory practice we only take care about the qualitative analysis. We will determine what kind of atoms there are in the sample in a measurable amount, but we will not determine their concentration.

For the qualitative analysis the energy of the photons should be measured and be associated to an atomic number. This task is not so easy since the vacancy can be filled in from several electron orbitals. Further complication is that not always the K-shell electron is vacated but the L-shell or higher principal quantum number shells but these are less probable if gamma-radiation creates the vacancy. We give an overview in [Figure 11.1](#) on the energies of the emitted photons in a simplified framework. The energy of the photon is the energy difference of

the two orbitals, and there are fine effects to modify a little bit the electron energy. These small modifications can be observed in our measurement in some cases as a splitting of a peak in the energy spectrum.

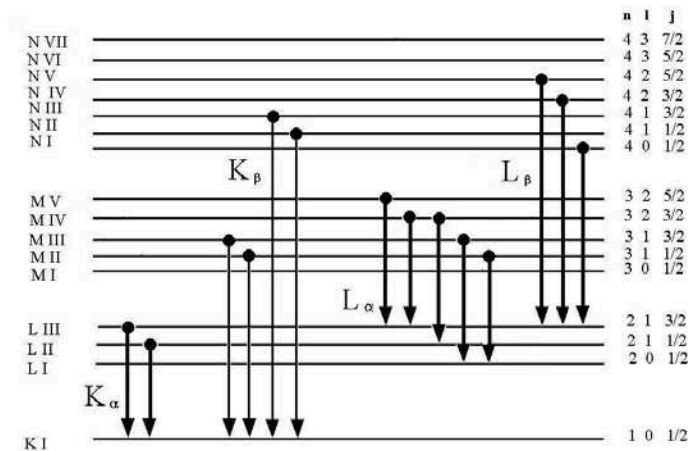


Figure 11.1. The possible electron transitions in an atom. Simplified diagram of the electron energy levels.

The notations of the electron transitions and the corresponding X-ray energies are K_α , K_β , L_α , L_β , etc. K-lines mean that the electron vacancy is on one of the orbits with $n=1$ principal quantum number, this is the K-shell. K_α photons are emitted when the initial orbit is on the L-shell, $n=2$. The radiation is called K_β radiation when the electron fills up the K-shell vacancy from the M-shell. In Figure 11.1 the LI, LII, LIII lines correspond to the orbits on the L-shell, and their energies are different a little bit, due to the fine interaction. There are two kinds of K_α photons according to the initial orbits fine energy. From one of these orbits the electron dipole transition is not allowed to the K-shell. These two photons are called $K_{\alpha 1}$, $K_{\alpha 2}$, but these are hardly separable. L_α , L_β , L_γ transitions or L_α , L_β , L_γ lines go to the L-shell from the M or N shell. In these transitions there is a lot of variation and some of them are quantum mechanically forbidden as a dipole transition. A general rule tells three intense L-lines occur and their energies are close to each other, while there are only two K-lines. In Figure 11.1 the y-axis shows the energy of the levels, but on a distorted scale! The small energy differences of a shell are magnified for better understanding.

11.3 X-ray fluorescence analysis

The Moseley-rule gives the energy of a characteristic X-ray photon:

$$E = A \cdot (Z-B)^2.$$

Here E means the energy of the photon; Z is the atomic number of the atom, A and B are constants that are different for the different transition types: K_α , K_β , L_α , L_β , L_γ . The B parameter corresponds to the screening effect. It is a simplified parameterization of the energy of the energy levels of the L, M, etc. orbits. In fact, there is a hardly calculable many-electron system. We make a simplification that the inner orbitals screen the charge of the nucleus, like its charge is less. From quantum mechanical point of view it is a confirmed assumption. The constant A depends on the n and m quantum numbers (see [equation 11.1](#)), and in fact also on the fine interaction.

This simple atomic number dependence of the characteristic X-ray photons and the development of the semiconductor detectors gave a handful and relatively cheap method to the researchers that is applicable on field measurements as well.

The simplest application of the method uses gamma or X-ray photons to kick out a K-electron from the atoms of the samples. These radiations always kick the K-electrons with the highest probability, much more than the L-shell and the interaction with the outers is mainly negligible. The most commonly used exciting sources are ^{55}Fe , ^{109}Cd , ^{125}I , ^{241}Am radioactive isotopes. The americium source has a huge advantage: its half-life is long enough, but on the other hand its energy is 60 keV, a little far from the optimal 5-20 keV. The energy of the source photons E_γ is important from the point of view of the excitation probability: the probability of generating a vacancy on the electron shell of energy E_v . This probability strongly decreases with $(E_\gamma - E_v)$. Therefore the most effective source has the gamma energy about the ionization energy of the vacated electron that is about 5-20 keV. The gamma photons from ^{125}I source are less energetic, but there are more than one. That makes the intensity of the emission of characteristic photons higher but makes the quantitative analysis much more complicated. There is another disadvantage of this source: its half-life is about 60 days. The source disappears easily in one year; one should spend money to replace it. The americium source has a 432-year half-life, and therefore usable for a long time.

Qualitative X-ray fluorescence analysis (RFA) and quantitative RFA can be executed. For the first one the energy of the emitted photons should be determined, for the second one the intensity of a given energy photon should be measured and a complex analysis should be applied to calculate the concentration. The energy spectrum of an RFA measurement can be seen in Figure 11.2. There are K_α , K_β lines, K_β always has a little larger energy, and 3-5 times less intensity. Many times a K_β line overlaps with another K_α line. That is to be solved in the quantitative analysis.

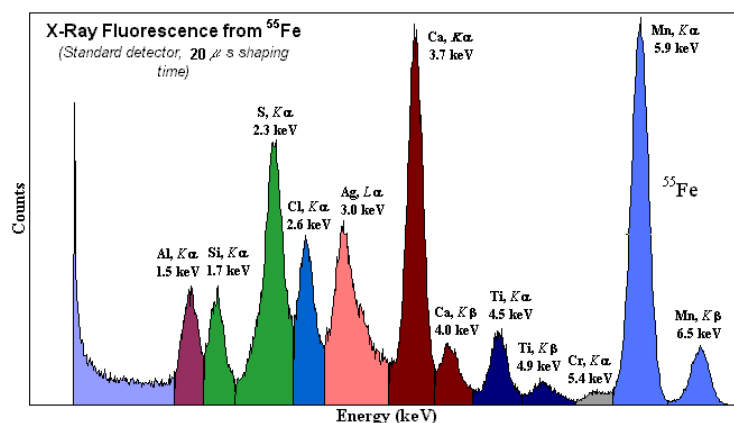


Figure 11.2. A RFA spectrum using ^{55}Fe exciting source. The energy of the source photon is little less than 7 keV, so here the elements with low atomic number can be detected.

This spectrum was detected by the PathFinder on the surface of the Mars.

The sensitivity of the method is different for different samples. The atoms in the sample with low atomic number are poorly excitable and generally their photons are not detectable by the semiconductor detectors, only by crystal spectrometers. The higher atomic number elements can be excited with higher probability. The gamma or X-ray radiation that generates the vacancy will kick out the electrons by photoelectric effect. In this process the photon is absorbed and its total energy is converted to eject the electron from the atomic shell. This process strongly depends on the strength of the electric field at the position of the vacated electron. In case of the high atomic number atoms at the position of the K-shell for example, the strength of the electric field is proportional to the Z^2 . Therefore the high atomic number atoms are much more excitable, and therefore the characteristic photons corresponding to these are

much intensely detectable. This is a favorable condition if one determines heavy metal contamination. The method is especially applicable for observing mercury, lead and cadmium in very low concentration. For example in soils where the main elements are aluminum, potassium, calcium, silicon, oxygen etc. have low atomic numbers. The characteristic photons from these atoms will not appear at all, we can investigate clearly the contamination.

For heavy metals the Auger-effect will not cause a problem either. Since its atomic number dependence also favors the low atomic number region.

The most simple type quantitative analysis is the relative measurement. This means that we have a material of known concentration of the element to be determined. We can add a little amount of this to the sample and observe the increase of the intensity of the corresponding peak in the energy spectrum. If we add one or more times this additional amount and detect the peak intensity, we determine a systematic behavior that should be a monotone function of the concentration. Generally the 0 peak intensity corresponds to 0 concentration and it gives another point for the analysis. Determining the function gives an opportunity to calculate the concentration of the selected element in the sample.

The RFA gives an average concentration of the irradiated area. The intensity of the gamma and X-ray photons going into a material will decrease exponentially. The irradiated depth is small due to the characteristic length of this exponential (penetration depth parameter). The method examines only the top some millimeter of the sample. This effective thickness depends on the atomic number of the sample. A sample with higher average atomic number gives smaller effective thickness. Therefore the same concentration of an element will produce lower intensity peak in the spectrum in a higher atomic number environment. The self-absorption of the sample further increases this effect.

In principle the RFA is a non-destroying method. Of course the simple relative measurement needs preparation and we lose this advantage. If the sample is not homogeneous it can be an effect, but mostly during the preparation we homogenize the material. At field measurements this is impossible. Of course, when the sample is not allowed to be destroyed, the preparation and addition of known amount of an element is not possible. But there are complicated procedures for the data analysis in this case as well.

The matrix effect. One of the main reasons of the complication of the quantitative analysis is the matrix effect. It means that the characteristic photon of one element can excite (make a vacancy) in another lower atomic number atom in the sample. So the excitation is not only from the outer source but can happen inside. Further complication is that the sample absorbs its characteristic photons and the absorption depends on the average atomic number of the sample.

The detection limit of this method of course depends on the actual geometry and parameters of the excitation and detection device, but it can be as low as some ppm.

11.4 The detector system

The RFA detector system consists of a source, a sample and a detector. In our measurement it has a cylindrical geometry. The source is a ring shape source shadowed into the direction of the detector that is located at the bottom of the system at central position. The sample is on the top, and the radiation of the source hits it from below. The characteristic photons from the sample go across the empty middle part of the ring source down to the detector.

The detector is a Silicon-Lithium semiconductor detector Si(Li). The X-ray photons interact with the electrons of the detector material (the silicon) mostly by photoelectric effect and accordingly by Compton effect. The useful part is the electrons hit by photoeffect, since the total energy of the photon is converted into the kinetic energy of the electron. The electron in the semiconductor will be released from its atom and becomes movable in the crystal. There is an electric field put on the crystal that will make the electrons move to the side of the detector and give an electric pulse. At every photon detection several electrons are collected. The pulse is about 1 μ s long and is to be amplified. The energy of the photon is proportional to the amplitude of the electric signal. This is the base of the energy measurement with this detector. Several electronic units (e.g. amplifier, pulse shaper) analyze the pulses; at the end an analog to digital converter will produce an integer number that is proportional to the energy of the photons. This number is called channel number, and calibration translates this value to energy in keV. At the beginning of every measurement a calibration should be done using known composition sample.

In the energy spectrum we collect the frequency distribution of the channel numbers. At the characteristic lines e.g. K_{α} , K_{β} we can detect a Gaussian type peak. The width of the peak is small compared to its average. This is a specific feature of the semiconductor detectors and enables us to separate the characteristic lines of different elements. The Full Width of Half Maximum (FWHM) of the peaks is generally some percent. This has statistical reasons. To create one delocalized electron in the silicon it needs about 1 eV. For a 10 keV energy photon it means ten thousands electrons (and the ions as their pair). The statistical uncertainty of this number, assuming Poisson distribution, is the square root of 10^4 that is 100. This means 1% is the sigma parameter of the Gaussian peak. This translates to 2.4% FWHM, which means 240 eV at 10 keV. The energy resolution of the detector is the relative value of this FWHM and the photon energy (10 keV in this example). The modern devices can have as little as 150 keV FWHM at 7 keV of ^{55}Fe .

11.5 Lab course tasks

1. Calibrate the setup using a mixed sample! Determine the energy resolution in percentage for the K_{α} peak of the iron.
2. Determine the elements that can be found in the unknown samples!
3. Determine the elemental composition of the minerals, $Z > 20$.
4. Determine the barium content of a soil sample! First make a calibrating series then determine the barium-intensity function for the soil sample.
5. Determine the lead content of a leaf sample applying the relative intensity method!

11.6 Test questions

1. What is the characteristic X-ray?
2. What is the Auger-effect?
3. What is the meaning of the transitions called: K_{α} , K_{β} ?
6. What is the Moseley-rule? What is the meaning of the B parameter?
7. What is the phenomenon of X-ray fluorescence?
8. How can the elements be identified in a sample?
9. How can you determine the concentration of an element?
10. What is the matrix-effect?
11. What is the energy resolution coming from?
12. How can we execute the energy calibration?

12. Film dosimetry (FDO)

Use of ionizing electromagnetic waves presents a lot of advantages for the society. The most known example is the medical diagnostic X-ray. Bone breaks and several other invisible details of the human body can be revealed using this radiation. However, it means health hazard for the people who absorb it. Also there are diagnostic technologies to investigate welding or soldering in a material that use X-rays. These inspections bring a lot of advantages and save other types of health hazard and lower health risk. But smaller risk will occur when these works are being done. Usage of radioactive isotopes is also helpful in several cases, but means health risk due to gamma rays. To quantitatively control the absorbed dose of these ionizing radiations several methods were developed. The most prevalent method is the film badge dosimetry. This is an official control method for the radiation workers.

12.1 Overview of the dosimetry

12.1.1 *The general concepts of the dosimetry*

The control over a work phase that uses ionizing radiation has three aspects: i) reasonability, ii) optimization of shielding, iii) restriction of radiation dose. The necessity of using ionizing radiation should be proved and explained. If an application can be executed without radiation the necessity is not valid, the radiation cannot be used. The absorbed dose during a work should be minimized by using shields. Of course more money cannot be spent on building a shield than the amount we are producing with the use of the radiation. The volume and the material of the shield should be calculated and unnecessary materials cannot be built in. The optimization of the shielding is important, and generally includes investigation of the setup from geometric point of view, the movements of the sources, and the times should be optimized with the shielding. The third aspect requires limits of radiation doses. This is a complex regulatory system. There are limits on the dose for the whole body, for specific tissues (like: eyes, skin), for a year duration and for 5 years duration, too. But always the most important rule is that the dose should be measured. The dosimeters are used for this purpose. Therefore dosimeters should detect not only counts, but also energy that is lost in the material, or some other property that is proportional or at least depends in a monotonic way on it.

Another phrasing of these concepts is the ALARA (or ALARP) principle: “As Low As Reasonably Achievable” or in Britain “As Low As Reasonably Practicable”. The reasonable is a question that should be approached from different aspects, like what we mentioned above.

12.1.2 *The principles of protection*

The protection from the radiation that is used or its remaining part that can hurt the human health has three posts: a) distance, b) time, c) shielding. The activity of the radioactive source or the intensity of the X-rays is given for a given work, it cannot be changed. Instead, the workers should be protected. Since the radiation travels to every possible direction of space its intensity is decreasing by the $1/r^2$ law. This helps to avoid higher doses. The minimum distance from the source has to be elaborated and as far as it is possible. The radiation sources, for example, should be taken by nippers. The absorbed dose is proportional to the time of absorption. The work should be planned in such a way that the minimum time should be stayed at the source. The shielding is an important part of the protection, too. Enough shielding should be used as planned.

12.1.3 Measurable doses

The United Nations Scientific Committee on the Effects of Atomic Radiation has been collecting data on several aspects of dosimetry and developed a well-established system for measuring and evaluating doses. The health hazards were investigated for several cases and the information is accumulated. These are the fact bases of the dosimetric standards, the used constants and methods.

The dosimetry is based on the concept: dose. (On those that can be measured or at least can be estimated.) The first type is the *absorbed dose*. This is the absorbed energy per unit mass of a homogeneous material. Generally the material means a tissue of the body, but it can be a test material for detector investigation or simulation.

$$D = dE/dm,$$

where D is the absorbed dose, dE is the energy absorbed in the material having mass dm . Its unit is $[D] = \text{J/kg} = \text{Gy}$ (gray).

But the biological effects of the different types of radiations (alpha, beta, gamma, neutron or proton) can be different. This is because these particles interact in different ways with the material of the cells. Mainly the ionization density can be different for electrons and alpha particles. That results in a situation where in one cell there will be much more free radicals or more proteins will lose their secondary structure. The reaction of the human processes can handle these ionizing densities in a different way. Therefore the equivalent dose will describe the intensity of the biological harms of the absorbed dose:

$$H_{T,r} = Q_r D_r$$

This depends on the r radiation and is associated to a T tissue. The dimension of the equivalent dose is called Sievert: $[H] = \text{Sv} = \text{J/kg}$. The J/kg is still valid here, but the meaning is much different than in the case of absorbed dose. The Q -values are $Q=1$ for electrons, gammas and muons. $Q=5$ for protons, $Q=5-20$ for neutrons, depending on their energy and $Q=20$ for alpha particles.

Different tissues have different complexity and sensitivity for radiation harms. The dose type that characterizes the effect for the whole human body is called the effective dose.

$$E = \sum_T w_T \cdot H_T$$

Here the weight factors w_T are effective for a T tissue. Many years of research resulted in these values. The dimension of the effective dose is the same as for H , this is Sievert. The film badge dosimeters will determine an effective dose, and this is the dose type for that the most regulations apply.

12.1.4 Dose ranges

The effective dose is calculated for one year period. During one year people absorb continuously doses from natural sources: the background gamma-radiation, the cosmic radiation, mainly the dose from muons. People contain radioactive isotopes like ^{14}C , ^3H , ^{40}K . The dose of these radioactive isotopes also continuously reaches the cells. A big fraction of the natural dose is coming from the radon daughters. As a survey of the United Kingdom reported the effective dose from the mentioned parts of the natural ionizing radiations are the following: cosmic 12%, gamma background (terrestrial, building material) 13.5%, internal dose 10%, artificial dose (medical dose mainly) 14.5% and the radon 50%. The total amount for an average citizen is 2.6 mSv. (Source: National Radiation Protection Board.) This dose per year has been the environment of people for millions of years.

The other end of the scale of the doses is the half lethal dose that is about 7 Sv. This happens only in case of accidents and for only a few people. Although in case of radiation therapy several steps of 2 Sv is used for curing patients in hospitals. There is a factor of about 3000 between the two sides of this scale that is surprisingly narrow.

The radiation dose can cause two types of biological effects. One is the stochastic effect that has a probability, and there are no necessary serious biological consequences. If there is, it evolves many years later than the irradiation happened. Above 300 mSv there is a deterministic effect of the irradiation. The effects happen within one day or some days. These are not necessary lethal illnesses.

The radiation dose presents a health risk. But all the occupational activities have health risks as well. In construction industry and chemical industry there are accidents during work that are dangerous. We can calculate a risk level that is average for the population and therefore is widely acceptable. This is the population averaged workplace risk. That radiation dose which causes this average risk is feasible for a society. If someone's work is changed and a different type of work will happen, the risk is unchanged. This dose is 20 mSv nowadays in the EU, but it is a country dependent value, the life standards affect it deeply. It is worth to mention that the natural dose is only a factor of 10 smaller than the dose limit for radiation workers. In case of an anomalous natural environment the natural dose can raise and the certain dose can exceed the limit for workers. Of course that is not applicable for the homes but points out the importance of high natural doses. Being able to avoid building new houses onto radon rich soils is for example one of the valuable results of the common effort of environmental physics and geology for the population.

12.2 Film badge dosimetry

12.2.1 *The film dosimeter and the basics of the method*

It is known that normal photographic films are sensitive to exposure of ionizing radiation. This radiation interacts with the matter of the film and sometimes hits molecules containing silver and makes changes in them. During the chemical processes of developing the film from these molecules silver atoms will spring up. These atoms then can absorb the visible light and cause higher optical density or blackening of the film. The more energy is deposited in a small part of the film (dose) the more silver atoms will arise, and the darker the film will be. In a case when the film is not overexposed this relationship is linear. The parameter, which is to be measured, therefore is the optical transparency of the films after irradiation.

The method of film dosimetry always uses calibrating films and the determination of the absorbed dose is based on the comparison of the darknesses of different films.

This lab practice corresponds only to gamma ray irradiation. Generally film badge dosimetry can be applied to beta- and rarely to alpha-radiation, too. Our calibrating films were irradiated by gamma rays only, so the irradiation during this lab means always gamma irradiation.



Figure 12.1. Film dosimeter holder (blue) and the film (white) in it.

The films will change their optical density due to the irradiation, but gamma rays with different energy interact with different probabilities in the matter, so it gives different amount of density change at a given dose. Therefore to determine the dose by evaluating the transparency of the films we need to know the energy of the irradiating gammas as well.

Of course in most cases it is not a monoenergetic gamma-radiation. Like if there are no extra gamma rays hitting the film, the blackening happens by the background irradiation, which is a mixed radiation. Even though there are characteristic energies in it. In a case of special work with radioactive sources at workplaces, however, it can be a general assumption that the work was done by a given energy gamma-source. In other cases, like dental X-ray diagnostic, a X-ray tube is the source of the X-rays (that are equivalent to gamma rays from this dosimetric point of view). This source has a maximum energy and less photon energies can occur in the radiation. For these cases an average energy is used instead of dealing with the energy distribution of the measured radiation.

The energy of the gammas has to be determined. For this purpose the film badge holder has 6 “windows” (see Figure 12.1 and [Figure 12.2](#)). 2 are real transparent windows but the other 4 “windows” have specific filters applied in the holder. In each filter the absorptions of the gamma rays are different. Therefore the ratios of these windows give information about the energy of the gammas. These filters are thin absorbers with different absorption parameters, i.e. different average atomic numbers and different thicknesses. The films are covered light-tight to avoid the blackening by visible light. [Figure 12.2](#) shows the parameters of the absorber filters.

At the evaluation of these films the optical transparency will be measured at given points of the film. The centers of the filters are measured always. We will use three absorbers (positions) in our lab: 300 mg/cm² thick plastic (plastic is the material of the holder), an Al-Mg-Si alloy called “dural” and a Sn-Pb filter that has the highest average atomic number and strongly absorbs the low energy gamma rays.

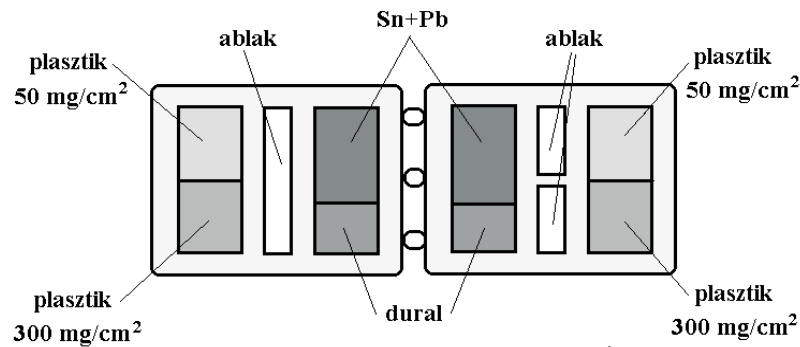


Figure 12.2. The filters in the windows of the film dosimeter holder.

12.2.2 The calibrating series

We have two calibrating series.

1. The so-called cobalt-series (Co-series). Each film of this series was irradiated by ^{60}Co gamma source that has an average gamma-energy of 1250 keV, but using different doses (different duration of the exposure). This series contains 10 films.

2. The so-called energy series. These films were irradiated by different energy gamma rays and by different doses. There are 6 films in this series. But one of the films is the same as a member of the Co-series.

12.2.3 The determination of the transparency of the films

The basis of the evaluation is the measurement of the transparencies (t) of the films. We have a given light source (lamp) that emits I_L intensity of visible light. The films are placed at a given distance from the lamp that is fixed during our measurements. We use a light intensity measurement device (lux-meter) to determine the amount of the transmitted light through the films: I , and then we calculate the transparency (transmittance) as $t=I/I_L$. I depends on the measured films but I_L is always the same unless we change the intensity of the lamp ourselves (not recommended, that would destroy the whole measurement).

The darkness of a film is defined as $S = -\ln t$. (Using the minus sign S has positive value.) When the visible photons from the lamp are absorbed on the same silver atoms there are less photons remaining to be absorbed on the other silver atoms. The absorption is proportional to the absorbing particles but also with the remaining photons. Therefore if we increase the number of silver atoms the transmitted photons will not decrease linearly, but exponentially as a function of the absorbing particles. If we assume that only the silver atoms are the absorbers, the t ratio is an exponential function of their number N : $t = e^{-\text{const} \cdot N}$. Therefore the darkness (as defined above) is proportional to the N .

The films are irradiated generally by gamma rays to be measured but always there are background gamma rays, which will be added to them. This background will produce silver atoms in the film, which is to be deduced:

$$S_{\text{irradiation}} = S_{\text{measured}} - S_{\text{bkg}}$$

We need a reference background film that was produced at the same time and was developed the same way as the others. Since the two series were applied onto the same type of film and

the unknowns, too, we have only one background film that is applicable for all the three types of films. We measure the transmitting intensity through this non-irradiated film: I_{bkg} . This will give us the $S_{bkg} = -\ln(I_{bkg}/I_L)$, which is the 0 dose case.

We measure the I_{bkg} at the positions below all the three filters of the holder three times: before the measurements of the energy series, between the two series, and after the measurements of the unknown films. The average of them gives the I_{bkg} used. The *darkness* of a film ($S_{irradiated}$) at a position below a given filter can be determined as follows:

$$S_{film,filter} = S_{meas,filter} - S_{bkg,filter} = -\ln t_{meas,filter} + \ln t_{bkg,filter} = \ln \frac{t_{bkg,filter}}{t_{meas,filter}} = \ln \frac{I_{bkg,filter}}{I_{film,filter}} \quad (12.1)$$

In a not overexposed film the number of silver atoms is proportional to the absorbed dose, as we already saw that the darkness is also proportional to the number of silver atoms, therefore the net darkness (S_{film}) is proportional to the dose, which has to be measured.

12.2.4 Relative sensitivity and D^*

The different areas of the films at different filters get different doses. We denote the dose that is absorbed at the empty window part of the film by D_0 . The doses at the locations of the filters: $D(f)$, f =plastic, dural and Sn+Pb. These are less than D_0 . $D(f) = D_0 \cdot A(f)$, where $A(f)$ is the absorption factor of the filter. Moreover, this depends on the energy of the gamma-rays, as it was discussed: $A(f, E)$.

As we mentioned the dose – darkness monotone relationship stands only at a given energy of the gammas. At different energy the photons interact with different probability with the material especially with the silver atoms as well. Therefore at a given dose different silver atoms will be produced. The quantity that shows how many silver atoms were produced by a unit dose at a given energy is the *sensitivity* $R(E)$. This is energy dependent, but does not depend on the dose, neither on the filter used.

$$R(E) = \frac{S_{film}(f, E)}{D(f, E)} = \frac{const \cdot N_{silver}(f, E)}{D(f, E)}$$

The more relevant quantity that measures the dose for the person who wears the badge is the D_0 . Altogether the factors that affect the darkness – D_0 relationship at a filter location:

$$S_{film}(f, E) = D_0 \cdot A(f, E) \cdot R(E)$$

The cobalt-series is a calibrating series where the same gamma rays (1250 keV average energy) were used for irradiation but the durations were different. We know the applied D_0 doses for each film. At these circumstances the darkness is proportional to the dose. In other words the sensitivity of the films ($R(E)$) is the same, since the same photon energy was used. The linear relationships between the darkness and the dose, however, are different for the three filter positions. The darkness at the lowest atomic number plastic filter is always the highest. This difference is due to the different absorption properties of the filters $A(f, E)$.

$$S_{plasztik}(D_0) = a_{pl} D_0, S_{dural}(D_0) = a_d D_0, S_{Sn+Pb}(D_0) = a_{SnPb} D_0.$$

We can write it in the formalism like above:

$$S_{film}(f, 1250keV, D_0) = D_0 \cdot A(f, 1250keV) \cdot R(1250keV) = a_f D_0$$

If we know the a_i constant after a calibration, we can determine the dose for any darkness assuming that the irradiation was by a cobalt source:

$$D^* = S_i/a_i.$$

If the irradiation is not by cobalt, but different energy gamma rays, this formula still can be applied but it will not give the right dose. We call this D^* value then the *cobalt-equivalent dose*. It gives the dose of photons from a cobalt source that causes the same darkness as measured.

For the cobalt-series D^* -s for all filters will be D_0 , the transformation S to D^* eliminates the effect of the different absorptions of the filters. But for the energy-series D^* will be different at the three locations.

$$D^* = \frac{S(f, E, D_0)}{A(f, 1250\text{keV})R(1250\text{keV})} = D_0 \frac{A(f, E)R(E)}{A(f, 1250\text{keV})R(1250\text{keV})}$$

This difference has two reasons. i) Since the absorption depends not only on the absorbing material but also on the gamma energy, $A(f, E) \neq A(f, 1250\text{keV})$. ii) The sensitivity is different, the gamma photons will interact with the molecules containing silver in different ways at different energies, so $R(E) \neq R(1250\text{keV})$.

Because of the reason i) for the films that had been irradiated by low energy gammas the darkness at the Sn+Pb filter is much weaker than the darkness at the plastic and at the dural filter. It can be seen just by looking at it. At the Sn+Pb location it is less dark, as if it had got less dose. But not, only the absorption factor ($A(f, E)$) is much larger for the Sn+Pb filter.

We can define the relative sensitivity as an easy measurable quantity that is useful for the energy determination.

$$N_{rel}(f, E) = \frac{D^*(f, E)}{D_0} = \frac{A(f, E)R(E)}{A(f, 1250\text{keV})R(1250\text{keV})} \quad (12.2.)$$

This quantity varies very strangely. If we plot this at the energies given in the energy-series, it can be seen that it is generally decreasing monotonously. But at about 25 keV it has an increasing part. This is due to the $R(E)$ factor. At 25.52 keV the silver atom has their K line absorption edge. At 25.52 keV the probability that the gammas interact with the silver atoms rises like a step-function.

For the energy determination we apply another parameter, the ratio of the D^* -s of a given film. In this case the $R(E)$ will cancel out and the above-mentioned difficulty with $R(E)$ will be unloaded. We call it contrast-difference Q :

$$Q_1(E) = \frac{D^*(pl, E)}{D^*(dural, E)} = \frac{A(pl, E)}{A(dural, E)} \quad \text{and} \quad Q_2(E) = \frac{D^*(dural, E)}{D^*(Sn + Pb, E)} = \frac{A(dural, E)}{A(Sn + Pb, E)} \quad (12.3)$$

The Q , the ratio of cobalt equivalent doses for the different filter pairs will give better information on the energy of the irradiating gamma photons. In a rough estimation the $Q(E)$ functions are hyperbolic ones.

The negative darkness values obtained during our evaluation can be neglected. At low gamma energy the absorption can be strong and the dose is only slightly different from the background, therefore in this case the uncertainties are important.

12.2.5 Determination of the parameters of the unknown films

First we estimate the energy of the irradiating gamma rays. Using the above mentioned ratios of D^* values for the three filter windows of an unknown film

$$\frac{D_{Pl}^*}{D_D^*} \quad \text{and} \quad \frac{D_D^*}{D_{Sn+Pb}^*},$$

and the graphs corresponding to Eq.12.6 and Eq.12.7 formulas for calibrating films the energy can be estimated. With the known energy then the relative sensitivities (N_{rel}) can be calculated using linear interpolation and (12.3, 12.4, 12.5) formulas. From these and the values of the D^* we can calculate the irradiation doses corresponding to all the 3 filter windows. The arithmetic mean of these values gives the dose that we have been looking for.

It can happen that the Q_1 and Q_2 agree within their uncertainties to one of the Q_1 and Q_2 pair of the calibrating series. In this case the comparison of the two films based on the three D^* values the dose can be estimated directly without using the other films of the calibrating series. This is because the relative sensitivities of the two films are experimentally the same.

12.2.6 Formal requirements with the lab report

Since the analysis is based on comparison it is necessary to summarize all the measured and calculated values (S , D^* , D^*/D^* , N) in one big table, where the number of the calibrating films, their energy and dose can be seen. The serial number of the unknown films should be indicated altogether with their S , D^* , D_1^*/D_2^* (Pl/Du and Du/Sn) data as well. Do not present too many meaningless decimal figures and do not use too little, either!

12.3 Lab course tasks

12.3.1 General tasks

1. Measure the transmitted light intensities for the Co-series. Determine $I_{i,f}$ values, where i =number of films, f =filter (plastic, dural, Sn-Pb). (10 films)
2. Measure the transmitted light intensities for the energy-series. Determine $I_{i,f}$ values, where i =number of films, f =filter (plastic, dural, Sn-Pb). (6 films)
3. Measure the transmitted light intensities for the unknown films. Determine $I_{i,f}$ values, where i =number of films, f =filter (plastic, dural, Sn-Pb). (4 films)
4. Determine the $I_{bkg,f}$ from its three measurements f =filter (plastic, dural, Sn-Pb).
5. Determine the darkness – dose relationship for the Co-series and for the energy-series. Calculate the S values from the above measured values.
6. Make a graph of the dose – darkness ($x - y$) data points of each filter for the Co-series:

$$S_{plasztik}(D), S_{dural}(D), S_{Sn+Pb}(D).$$

(Here the difference of the irradiation of the calibrating films is the consequence exclusively of the duration of the irradiation, therefore the dose and the darkness are proportional to each other.) The next step will be a linear fit to the data, but this step will be executed in a different way by each person, so it is in the personalized tasks part.)

12.3.2 Personalized tasks

1. Fit the appropriate function to the dose – darkness data points, and determine the dose of the unknown films from the measured darkness values! First person use points 1-5 and analyze the films I and III. The second person use data points 1-6 and analyze films II and IV. The third person of the group use data point 1-7 and analyze films II and III. The fourth person use five data points 3-7 and analyze films I and IV. Note into the write-up which configuration was used! (Hint: Calculate the D^* values and the Q values first, then determine the energy, and then the dose.)
2. Compare the dose absorbed by the workers who wore these films with the maximum permissible dose to radiation workers! Assumptions: the worker wore the film for 2 months at a radiation workplace, during the year he worked 11 months and that the irradiation was uniform during that time. Determine how many times more dose did he get than the limit!

12.4 Test questions

1. What is the ALARA principle?
2. How do we control the use of radioactivity at a workplace?
3. What are the dose quantities?
4. What dose limits do you know? For whom are these applicable?
5. Why does the film badge dosimeter contain different shielding filters?
6. What is the sensitivity of a film and what does it depend on?
7. Why can be a film overexposed?
8. What can we learn from the ratio of the D^* quantities?
9. How do we measure the darkness?
10. Why is the darkness a logarithmic quantity?

13. Thermoluminescence dosimetry (PTL)

13.1 Principles of a thermoluminescent dosimeter

Devices to measure ionizing radiation (mostly gamma rays) in a certifiable way are film and solid-state dosimeters. Important examples of the latter group are thermoluminescent crystals, with their small size, energy independence and high level of sensitivity.

Basic principle of a thermoluminescent dosimeter (TLD) is that the ionizing radiation (mostly gamma-radiation) pushes the electrons of the crystal in an excited state, and then they are captured by the doping atoms. From there they can return to their ground state only through heating. During this return to the ground state they emit visible light (or photons with a wavelength near the range of visible light). Number of emitted photons (measured by photoelectron multipliers) is proportional to the absorbed dose of the dosimeter (the crystals).

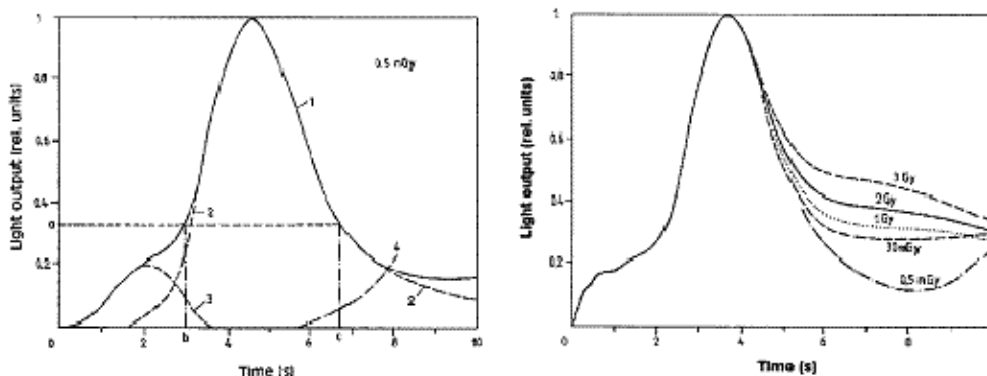


Figure 13.1. Time dependence of the light yield of a TLD during heating

Due to heating, temperature of the crystal changes more or less linearly. The light yield shows a typical curve shown in Figure 13.1 (curve 1). This is the sum of several curves. One belongs to a small temperature, and has a small, narrow peak with a quick cut-off (3). The most important one is the broad one (2), this is used for dosimetry. At the end of the heating there are no more excited electrons, thus the curve is cut off. If the heating goes on, photons coming from thermal radiation (4) are detected by the photoelectron-multiplier. In case of large doses, a peak belonging to higher temperatures appears as well, as shown on the right plot of Figure 13.1. This however, will not be important during our measurements. Processing of the measurement consists of measuring the area under the largest, broadest peak in the middle. To do this, we have to integrate the light curve numerically. One has to set the integration range in a way, so that the low temperature peak and thermal radiation have both only a small contribution, but the largest possible fraction of the middle peak should fall in the range. Then the integral is proportional to the irradiated dose. The factor of proportionality is determined by calibration via a source of known activity. This factor is determined by the amount (mass) of the crystal, its sensitivity, and the efficiency of the photoelectron multiplier, so the factor is different for each dosimeter. This and the integration limits are coded in the memory of most of the dosimeters (e.g. in our “Pille” dosimeter), but they may be overwritten. It has also to be noted that with the heating, the dosimeter is reset, there remain only very few excited electrons. There might be a small residual dose, but in case of Pille, it is very small, less than 1 nGy.

13.2 Layout of a dosimeter and its readout unit

In this section we will review a typical dosimeter, the Pille, developed by the KFKI Atomic Energy Research Institute for usage on space stations. This detector consists of two parts:

- a) Four TLDs, each consisting of thermoluminescent $\text{CaSO}_4:\text{Dy}$ crystalline grains in a vacuum glass pipe with a heating unit. The crystal will be excited by the ionizing radiation.
- b) A small, compact, portable and programmable TLD readout system with a clock. This way, the device left in measurement position, is able to determine the time profile of the dose-power even without human interaction¹. Of course it is possible to read out dosimeters irradiated away from the device.

The pen-shaped mechanical layout of TL dosimeters is shown in Figure 13.2. The TL is housed in a light-blocking piece made of two concentric cylinders, such that absolutely no outer light can reach the inside. This can be put in the readout and turned like a key, the inner cylinder is turned as well. In this position, a window opens for light to exit (in order to perform a measurement). The TLD can be withdrawn from the readout only when turned back, and the light-permitting window again closed.

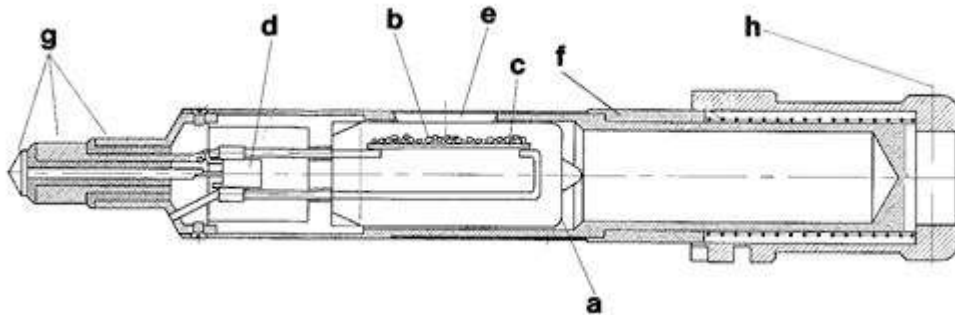


Figure 13.2. Cross section of a TLD.

Parts of the TLD are the following. (a): vacuumed glass cover. (b): TL crystal, i.e. $\text{CaSO}_4:\text{Dy}$ grains layered on a metal plate of desired specific resistance (c), and this plate can be electrically heated. (d): programmable memory chip containing the calibration parameter identifiers, fitting in the oxidized aluminum case. (e): window on the case, by default closed by a stainless steel tube. (f): the stainless steel tube that protects the inner parts from light and mechanical impact, as well as the operator from the heat after the readout. This automatically slides away when the dosimeter is put into the readout unit. (g): gilded contacts for the heating current and the memory chip. (h): the code is visible here that is stored in the memory during readout. When not in readout mode, the dosimeter is in a metal protection case.

The microprocessor (μP) controlled readout unit of the TLD ensures the preliminary evaluation of the irradiated dose. The readout heats the TL material inside the vacuum cover in a predefined way, and measures the amount of emitted light, and thus the absorbed dose is measurable, its value can be visualized and stored on a memory card. The card is capable of storing the results (dose, identifier of the readout unit and the dosimeter, date and time, error codes, parameters of the measurements and the readout, the digital heating curve) of 8000 measurements.

¹ Hourly measurements carried out for almost a week showed the excess dose coming from crossing the South Atlantic Anomaly of the Van Allen belts twice per day.

Main parts of the readout unit: microprocessor (μP), power supply for the heating, photoelectron multiplier tube (PMT), broadband I/U and A/D converter, memory card driver, high voltage power supply (HV). Logic scheme of the readout unit is shown in Figure 13.3.

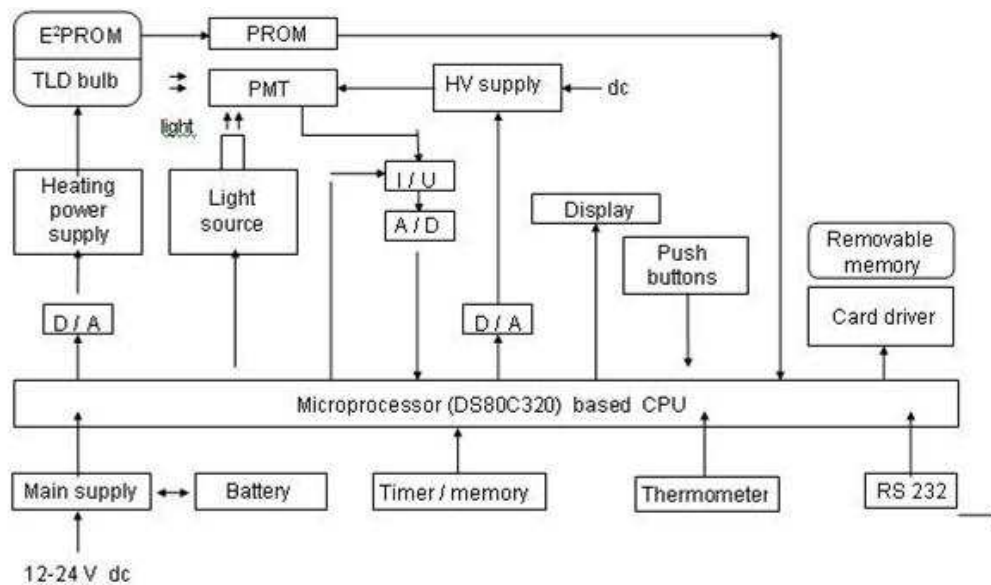


Figure 13.3. Schematic block diagram of the readout unit

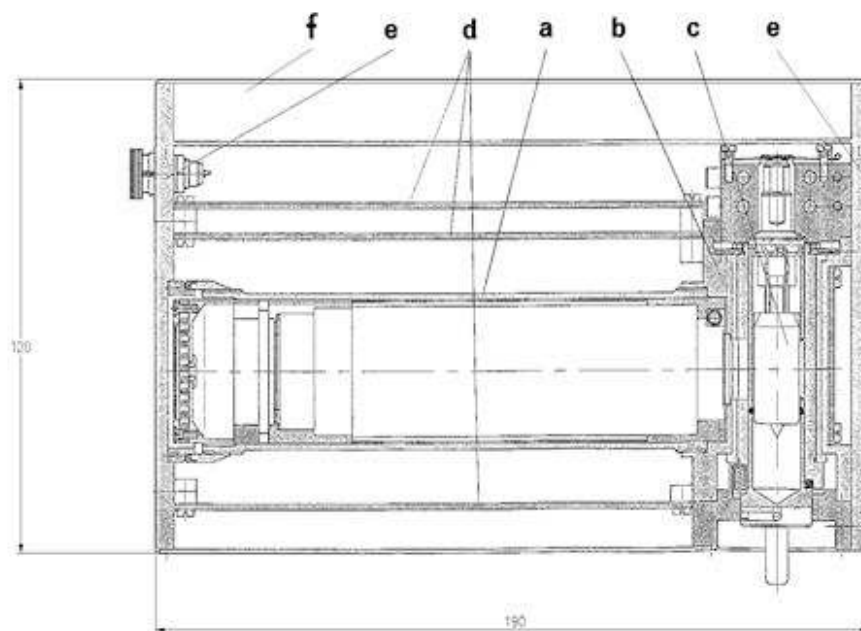


Figure 13.4. Cross section of the readout unit

Cross section of the readout unit is shown in Figure 13.4. Here (a) is the mechanical support structure of the unit, holding an aluminum cylinder, within which there is a PMT. Inside the transverse light protection case (b) one finds the dosimeter (c). The tube is surrounded by printed circuit boards (d). These are secured inside the thick aluminum wall (e) of the unit. The NiCd batteries are in the separated backside (f) of the unit.

Important warning: dosimeters should never be heated up twice within 5 minutes, because the crystal is not cold yet after the first heating, may thus be harmed. One also has to take care not

to let the dosimeters fall, as the glass covers may break easily. The replacement of the damaged dosimeters is beyond usual financial scope of a lab course.

13.3 A brief history of Pille and the PorTL

Bertalan Farkas has first used Pille in 1980, on the Salyut 6 space station, and it was also left there. Later on, soviet cosmonauts performed measurements with it. In 1983, a newer, more sensitive version was brought to Salyut 7. This was then transported to the Mir space station, where it was used for the first time in 1987 to measure the irradiated dose during a space walk. Already in 1984, according to a cooperation contract with NASA, the first American spacewoman took a modified Pille with her to the Challenger spaceship.

In 1994 a new version microprocessor was installed, and newer and newer versions flew in the Euromir framework of the ESA and some NASA missions to Mir. A lot of successful measurements were performed with it, among others during space walks. The newest version of Pille made it to the American module of the International Space Station (ISS) in March 2001, and in 2003 a modified version was brought to the Russian labs of ISS as well. Charles Simonyi used a Pille as well during his 2007 space flight.

There is a version of Pille to be used on Earth, which is also portable, it is called PorTL. This is widely used to monitor environmental radiations, for example in the nuclear reactors at Paks. These developed readout units can make corrections for the wide range fluctuations of the temperature of the surroundings as well. The PorTL consists of dosimeter cells and a portable, battery-powered readout unit, shown in Figure 13.5.



Figure 13.5. A PorTL dosimeter and readout unit.²

13.4 Lab course tasks

All measurements start with a heating, during which we also measure the average background dose at the location of the detector since the last measurement (usually a week in our case). For the measurement the dosimeter has to be taken out of its case and put into the reader. The dose has to be calculated based on the actual irradiation time, however, some devices use the

² From the PorTL website, <http://portl.kfki.hu/>

time between two measurements for their calculations, which only agrees to the irradiation time in case of the background measurement.

After measuring the background, we perform measurements above the opened lead case of a ^{22}Na source (for 30 minutes, 1275 and 511 keV photons are radiated here), as well as a ^{241}Am source (for 10 minutes, with 60 keV photons).

In the lab course proceedings, one has to answer the following questions.

- a) What is the dose rate of the natural background radiation at the entrance of the building and in the P11 lab? What is the gamma dose rate of the day measured by the National Meteorology Service (www.met.hu)? Which one is the biggest? Use the unit of nGy/h)
- b) What is the dose rate of the ^{22}Na and the ^{241}Am source, with and without shielding? Measure the dose rate at the locations given by the lab course leader, and subtract the background from it. How big are these values?
- c) How much bigger is the dose rate at the sources than the background? For how long may one be there in their vicinity if the dose from the source shall not exceed that from the background (with and without shielding)?
- d) Each participant of the lab course shall make a measurement of each source with and without shielding, and make the calculations based on that. The participants should then compare their results in a documented way.

The calculated and measured quantities should always be given with an uncertainty. The relative systematic error of the calibration is 20%, the relative stochastic error is decreasing with the measured dose: $(1+(33 \mu\text{Gy/D})^2)^{1/2}$ in percent. Calculate the quadratic sum of the two different errors as measurement uncertainty. Use the units of μGy , $\mu\text{Gy/h}$ or nGy/h.

Data of the used sources (activity, dates and gamma energies) are all written in the P11 lab.

13.5 Test questions

1. What are the working principles of a TLD?
What components does the light-curve of a TLD have?
2. What do we measure to calculate the dose in case of a TLD?
3. What is the most important rule when reheating a TLD?
4. How shall one set the integration range for the TLD measurement? What do we integrate?
5. What are the units of dose rate?
6. What is the remaining dose in case of a TLD?
7. What important part is there in the readout of a TLD, and what is its role?
8. What does one compare to the measured background radiation, and where can one get it?
9. What is the mechanical setup of TL dosimeters, and how do they work?
10. What are the occupational exposure limits for ionizing radiation?
11. What are the public exposure limits for ionizing radiation?

14. Gamma spectroscopy with scintillation detectors (NAI)

Our mother nature is radioactive. The human body itself contains radioactive isotopes like ^{14}C , ^3H , ^{40}K etc, but there are uranium and thorium on the Earth surface due to volcanic activities and other geological processes. The uranium and the thorium have alpha-radioactive isotopes, but their radioactive chain contains many isotopes that emit gamma rays. The natural gamma-radiation can penetrate through soil layers, brick walls therefore all humans are target of this radiation. Besides the natural gamma-photons if radioactive isotopes are released by industrial activity we can absorb gamma rays from artificial source, like medical diagnostic processes as well.

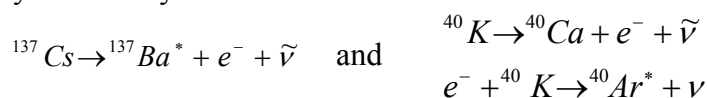
14.1 Gamma-spectroscopy with scintillation detector

The gamma-spectroscopy is a method to determine the radioactivity of given isotopes in a sample. Generally from this quantity the concentration can be calculated or in other cases the radiation dose can be determined. The activity of an isotope in the sample is proportional to the net area of the corresponding full energy gamma peak (we will describe this in detail later). But some other parameters affect the relationship of activity with the number of detected photons. For example the sample itself absorbs the photons, the measurement time is also proportional to the counts and there is a fundamental constant for each isotopes that tells how many gamma photons of the given energy are released of 100 decays. This latter is the relative intensity.

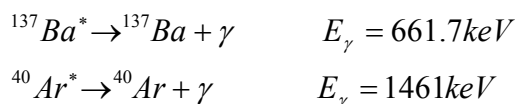
14.1.1 The gamma sources: ^{137}Cs and ^{40}K

In this laboratory practice we will investigate the gamma-radiation from ^{137}Cs and ^{40}K sources. Both isotopes have long half-life. The ^{137}Cs has 30 years and this is an artificial isotope. It occurred on the surface of the Earth after the nuclear weapon tests and the Chernobyl accident. ^{40}K is a natural isotope of potassium. Its half-life is 1.25 billion years (10^9 y). The 1.18×10^{-4} part of potassium atoms is this radioactive isotope. This counts as a large fraction, since the number of potassium atoms are in the order of 10^{23} . ^{40}K frequently appears in building materials, in food and in the material of the human body, too. As a rough estimate a person of 70 kg has about 3 000 decays of ^{40}K in every second. These isotopes decay by beta decay, but the emitted electrons are stopped in the case of the NaI detector or even already in the sample that contains the isotopes. Therefore only the gamma-photons can be detected generally.

Both isotopes decay by beta-decay:



The star in the upper index means that those nuclei are created in excited states. That is general in beta-decay. This excitation energy will be emitted then using gamma-radiation during the deexcitation process.



In fact these gamma rays are coming from the daughter nuclei, not from the beta decaying isotopes after which the sources are named.

The relative intensity means the fraction of decays that results in a given energy gamma-photon. It is expressed generally in percentages. The ^{40}K can decay in two channels as it is written above: electron capture and negative beta-decay. Only the electron capture process that leads to an excited argon nucleus emits gamma rays, this is why the relative intensity is quite small: 10.66%. For the ^{137}Cs source the relative intensity is much higher: 85.1%.

14.1.2 Interaction of gamma photons with the detector material

During the detection of a single gamma-photon we can measure the energy remained inside the detector active area. The energy loss is determined by the radiation – matter interaction. In our case the gamma-photons interact with the electrons of the detector material that is a sodium iodine crystal.

For a gamma-photon there are three possible interactions:

i) Photoeffect, where the photon loses its total energy that is transferred to a bound electron, which is ejected from its location. This electron then ionizes many other electrons on a short distance until the whole energy loss is transferred to many electrons.

ii) Compton-effect, where the gamma hits a quasi free electron but only a part of its energy will be transferred: a lower energy photon goes out from this scattering process and might leave the detector material or can cause another interaction within the detector.

iii) Pair production, if the gamma-photon has more energy than 1.02 MeV the creation of an electron – positron pair is possible in the electric field of the atomic nucleus. In the pair production the photon perishes, the electron and the positron lose their kinetic energy via ionization of the medium and at the end the positron will annihilate.

The probability of occurring one of these processes is energy dependent and it also depends on the atomic number of the detector material. At low energies photoeffect is always more probable than Compton effect. Below certain energy the photoeffect dominates and above it the Compton effect will be the most probable process, but at higher than 1 MeV the pair production is also becoming important. In case of a high atomic number material that certain energy goes to higher values. This is because the probability of the photoeffect is proportional to the Z^5 , where Z is the atomic number of the media, while the probability of the Compton-effect is proportional to Z . High atomic number materials are a good candidate for detecting gamma rays via photoeffect and hence the total energy of the photon can be measured. In the sodium-iodine crystal the I has a high atomic number, $Z=53$. In organic materials for example the average atomic number is below 8, that is Z for oxygen. In these materials the Compton effect dominates the interaction at around 1 MeV and the total energy of the gamma-photon is hardly measurable.

14.1.3 Scintillation counters

The scintillation counters or in other words scintillation detectors consist of two main parts: scintillator and the photomultiplier tube (PMT). The scintillator material interacts with the gamma-photon and produce visible or UV light flash at each detection event. These scintillation photons go to the window of the PMT and this device will convert them to an electric signal. The schematic view of the total setup can be seen in [Figure 14.1](#). In the scintillator the

energy loss of the radiation is proportional to the number of scintillation photons (that is called light output, L). The PMT will produce an electric pulse, which has a given shape, its time dependence is governed by the electric units (resistors and capacitors) and it is the same for all events, independently of the amplitude of the pulse. Therefore the amplitude is proportional to the light output, and the energy loss in the scintillator.

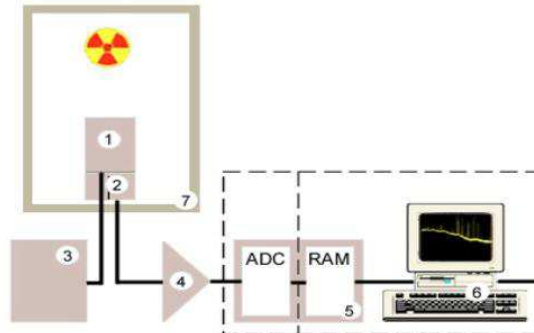


Figure 14.1. The schematic layout of the experimental setup. 1. NaI(Tl) scintillator crystal, 2. Photomultiplier tube (PMT), 3. Power supply, 4. Spectroscopic amplifier, 5. Analog Digital Converter and spectrum analyzer, 6. personal computer, 7. shielding

The investigation of the energy spectrum of the gamma-photons will be in fact the investigation of the amplitude of their pulses. In gamma-spectroscopy we collect and count those events where the total energy of the gamma-photon equals to the energy loss in the detector.

Big crystals and high atomic numbers increase the probability of this requirement. There are scintillation detectors for alpha or beta-radiation, not only for gamma rays. These types of scintillators are thinner according to the smaller range of these radiations in matter than the gammas have. The atomic number of the material is also not so important in these latter cases, since we do not have to take care of photoelectric effect of gamma photons, the alphas and beta-electrons are charged particles and ionize permanently along their path. The big problem, however, is how these particles will get into the scintillator that should be light-tight, so their surface are covered with saving material. These scintillators have a special coating that stops the room light but its thickness is thin enough to let the alphas go through. These detectors are called scintillators with alpha-window or end-window.

The scintillation photons created in the scintillator travel into the whole solid angle. We have to ensure that as many photons can reach the photocathode window of the PMT, as it is possible. Therefore the scintillator generally covered by reflective material to avoid the loss of the scintillation photons. In many cases the scintillator and the PMT are connected using a light guide that connects the two parts with the minimal loss of scintillation photons. When the photons reach the photocathode window of the PMT they will cause a photoeffect on the thin layer. These photons have about 3 eV energy and the probability of kicking out an electron from the photocathode is about 10%. These electrons first are accelerated by the voltage applied on the PMT and focused into the first dynode. There are about 10-12 dynodes in a PMT and on each the number of the electrons is multiplied by about 3. One electron starting at the beginning will result in 3^{12} (about half million) electrons at the last dynode that is called anode. The more electrons will be produced on the photocathode the more electrons will arrive on the anode and that is proportional to the current impulse that is the output of the PMT. The PMT uses 3000 – 4000 V high voltage that is distributed among the dynodes using a resistor ladder. When we use the PMT and switch the high voltage on it should be in dark. Otherwise

the room light will generate so many electrons that wear down the PMT quickly. The output signal generally goes through electronic units to be amplified and shape it before the amplitude selection.

14.1.4 Operation of the amplitude analyzer

After the electronic units the PMT output signal goes into an analog digital converter unit. The amplitude of the incoming analog pulse will be measured here and as the output an integer number is calculated from 0-1023, that is proportional to the height of the electric signal. This integer number is called the channel number. The analog digital conversion is done at this point, and the process continues in a multichannel analyzer (MCA). This is a digital device that collects the integer numbers (channel numbers). It is an integrated circuit that has memory in it. There are 1024 units (channels) in its memory that is cleared at the beginning of the measurement, but after each event the value of one channel is increased by 1, that one which corresponds to the height of the electric pulse.

In this way the MCA will determine the frequency distribution of the heights of the electric pulses out of the PMT. That is in fact the distribution of the energy loss in the scintillator, after an energy calibration procedure.

The energy resolution of this system is about 10%. That means if we detect the same energy gamma photons many times we will not get the same channel number. The events will be spread out into several channels. The energy resolution of the scintillation detectors are not the best among the gamma-detectors. But on the other hand their efficiency can be very satisfactory since the high atomic number of I, and the available big size of the crystal. The competing detector type is the semiconductor detector. It is made of germanium, and to grow an intact single crystal of it costs much more than that from NaI. But the semiconductor detectors have very nice sharp energy resolution.

14.1.5 The distribution of the detected energy of a ^{137}Cs source detected by a NaI(Tl) scintillation detector.

In [Figure 14.2](#) we can see a spectrum that is a good example for gamma-spectra. On the y-axis there are the counts per channel divided by the measurement time. The x-axis shows the channel number (that is proportional to the energy loss in the scintillator). The NaI detector in this case was irradiated by the monoenergetic 661.7 keV gammas from a ^{137}Cs source. According to the analysis the center of the full energy peak is at channel number = 280, and its width is 32 channels.

On this spectrum we can see not only one Gaussian shape peak but also a few other interesting objects. From right to left: the first is the big peak at 280 that corresponds to the full energy of the photons. Below about the 180-th channel there is a constant distribution that ends at about 180. This end is called an edge: Compton-edge. The events in this range correspond to a Compton-scattering in the scintillator when the scattered gamma did not interact more in the crystal and the gamma lost only part of their energy. The Compton-edge is at the maximum energy that can be transferred to an electron by Compton effect. There is another object in the spectrum at the 100-th channel. This is called the backscattered peak. The photons detected in this peak travelled through the scintillator without interaction but they interacted

with the material around and scattered by an electron. Generally the geometry of the setup is so that these scatterings correspond to a 180° angle, that is why we call it backscattering.

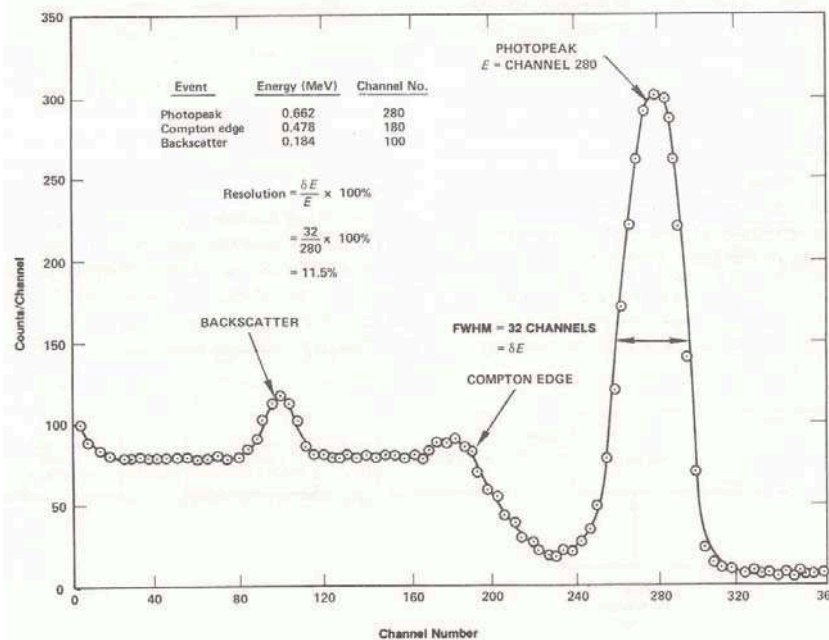


Figure 14.2. The distribution of the energy loss in a scintillator by the 661.7 keV gammas of a ^{137}Cs source.

(http://www.physics.uoguelph.ca/~detong/phys3510_4500/low-res-gamma-ray%20fall07.pdf)

14.1.6 The energy calibration

We assume that the PMT works in a linear way. The light output of the scintillation event is proportional to the height of the electric pulse out of the PMT. Also we assume that the scintillator's response is linear, proportional to the energy loss. This is true for electrons. In this case the centers of the full energy peak (if there are more) correspond to a linear function of the channel number. The determination of this function is the energy calibration. After this process we can identify unknown peaks in the spectra.

We collect a spectrum with a ^{137}Cs and a ^{40}K source simultaneously. This spectrum contains two full energy peaks, and these can be realized easily. After determining the centers and using the known energies (661.7 keV, 1461 keV) we can apply it into our measuring software and get the energy calibration.

14.2 Lab course tasks

During the measurement:

1. Calibrate the gamma spectrometer using ^{137}Cs and ^{40}K isotopes!
2. Measure the unknown sample, take their gamma-spectra using 5 minutes collecting time.

During measurement mark the Regions of Interests (ROI) where there are the total energy peaks of the gamma-photons. Save the results into a file!

Issues are to be worked out in the lab report:

3. Make graphs of the spectra that you took. Denote the full energy peaks and the Compton regions!
4. Determine the precise location of the full energy peaks. Assign energy to them; determine the brutto and the net peak areas!
5. Determine the energy resolution (FWHM) of the system at the full energy peaks of ^{137}Cs and ^{40}K !
7. Find the full energy peaks of the decay chain, if you know that uranium or thorium is in the sample. Select the peaks to given isotopes that are listed in the gamma-table.

14.3 Test questions

1. How many natural radioactive decay chains exist? What kind of radiation do they emit?
2. In the decay chains what is that isotope which has high importance in reaching the secular equilibrium?
3. What is called secular equilibrium?
4. What is scintillation?
5. Why do we use different size NaI detectors for different radiations?
6. How does the photomultiplier tube work?
7. List at least 5 radioactive isotopes that can be found in nature!
8. Tell the name of some radioactive isotopes that are artificial but can be found in nature.
9. What does the ADC do?
10. What does the calibration mean? How do we calibrate the system?
11. What is the photoeffect?
12. What is the Compton-scattering?
13. What is the pair production and what is the minimum energy this to happen?
14. What is the shape of the full energy peak?
15. What is the energy resolution?
16. How can you determine the net peak area?
17. What are the main elements of the measuring system?

15. Radioactivity of natural soil and rock samples (TAU)

15.1 The origin of natural rock and soil radioactivity

The isotopes of uranium and thorium, and their daughter products (members of their decay chains) can be found abundantly in the outer layers of the Earth: ^{232}Th , ^{238}U , ^{235}U and their daughters. We are going to examine a granite sample with high uranium content, and estimate its ^{238}U -concentration. Besides, the radioactivity and composition of several other natural samples will be studied.

The average uranium content of the soil is a few mg/kg (or g/t), in other words, a few ppm (parts per million). Of course, the observed values can vary widely. It is notable that the uranium concentration can be quite high close to volcanic mountain regions. There are certain locations famous for their surface rock layers with high uranium content, like in Kerala, India. Soils with more than average radioactivity can be found also in Hungary. Those are located primarily in the vicinity of our andesite- and granite-based mountains, like the Velencei-hegység or the Mecsek.

Another source of the natural surface radioactivity is the potassium isotope with the mass number 40, whose relative abundance among all potassium isotopes is **0,0117%**. The ^{40}K isotope decays to ^{40}Ca with β -decay, or (with 10,7% probability) to ^{40}Ar . (During the β -decay the nucleus emits an electron or a positron, while its atomic number increases or decreases respectively, by one, but its mass number does not change.) In the latter case each decay is followed by a radiation of a γ -photon with energy of 1460 keV^3 (the γ radiation consists of photons, the quanta of the electromagnetic field). The presence of the radioactive potassium, with a half-life of 1.248 billion years, can be detected in virtually all soil samples, since the ^{40}K appears in the energy spectrum with a single, well separable γ -energy. Thus, it is easy to tell it apart from the radiation of the uranium and thorium daughters. In the latter cases, the decay chain is long and more than a hundred gamma-energy shows up in the spectrum, originating from the many daughter products. The radiating isotopes can be identified via the determination of the γ -energies.

Among the artificial radioactive isotopes, the ^{137}Cs (with a half-life of 30 years) can be detected most easily in soil samples, which originate from the atmospheric nuclear bomb experiments and from the fallout that followed the nuclear power plant accident in Chernobyl.

The method to be demonstrated on this laboratory exercise is rather general; it is also used to determine the radioactive levels of the building materials at construction projects, or to examine the soil at a construction site. The decay chain of the uranium contains a radioactive radon isotope, the ^{222}Rn , which is a noble gas, and can escape from the soil or from the rocks by diffusion. Since the half-life of the ^{222}Rn is 3.8 days, it can exit the soil or building materials and mix into the air, and may accumulate in the air of the rooms in buildings. The decay products of the radon easily stick to dust particles, and can enter the respiratory system (lungs) by breathing, and emit significant amount of α -radiation.

In the course of the α -radiation the decaying nucleus emits a He nucleus (called α -particle) with an energy of a few MeV. The α -particle is doubly charged (contains two protons), therefore it ionizes strongly the materials it passes through, and decelerates quickly. While external

³ We are using the *eV* (*electron volt*) as the unit of energy: the amount of kinetic energy gained by an electron if accelerated through a voltage of 1 V. With SI units, $1\text{ eV}=1.6\times 10^{-19}\text{ J}$. The thousand- and millionfold of this quantity are also often used, that are the *keV* and the *MeV*.

α -radiation is already absorbed in the human skin, the total energy of the above described internal α -exposure is deposited inside the human body, and strains the sensitive inner organs.

In order to avoid high radon concentration in buildings, it is advised to examine the soil at the construction site already before the work starts. Dross from forges was often used as a building or insulating material. Dross often contains – depending on its origin– an elevated level of ^{238}U due to the burning process. (For example uranium level of dross from Tatabánya in Hungary, has been shown to be elevated.) It is known that the radioactivity of these building materials can be of concern. Besides the higher gamma dose it produces in the building, the radon concentration can be higher than average there. It is because the higher uranium content means higher radium content (radioactive equilibrium) and the bricks, blocks of this material having high gas permeability, so radon can escape out of these.

The same measurement method can be used to examine the uranium and thorium content of the rocks below surface, for mining exploration purposes.

We emphasized the gamma-radiation of the mentioned radioactive materials, since that is the radiation that is important for us at this laboratory exercise. However, the alpha- and beta-decaying isotopes dominate among the daughters of the uranium and thorium. These charged particle radiations get almost immediately absorbed in the soil sample itself, already after a short distance. Large volume samples thus emit almost exclusively gamma-radiation, which can be detected with our experimental setup.

The aim of this laboratory exercise is to get familiar with the reasons for natural rocks and soils to show radioactivity, and to learn an experimental method that is widely used in examinations of environmental samples. During this process, we will discuss those – sometimes complex – problems that have to be solved by the experimenter in order to apply the method in a reasonable and reliable way.

15.2 The determination of uranium content

The uranium content of our natural samples will be determined using gamma spectroscopy. As we will see during the measurement, only the gamma lines of the daughters of the ^{238}U will appear in case of the granite sample with significant intensity during the 10-15 minutes of the measurement. Since the ^{235}U is only present in the uranium with 0.7% fraction, its daughters can be analyzed only with a longer measurement. Furthermore, in case of this granite sample and short measuring time, the lines of the ^{232}Th are hardly visible. In our case, thus, almost all of the gamma-radiation dose originates from ^{238}U daughters.

15.2.1 The HPGe semiconductor detector

During the detection of a radiation quantum (particle) we can only measure the amount of energy deposited in the volume of the detector. This amount of energy is determined by the interaction between the particle to be measured (in our case the γ -quantum) and the material of the detector. These interactions (as it was already mentioned) are the photoelectric effect, the Compton-scattering and the pair production. The photoelectric effect is most probable if the photon has a small energy (less than a few hundred keV). Nearly all of the photon's energy is transferred to the electron in the atom of the material. The atom gets ionized and the electron starts to move with high velocity in the material, but it loses its energy quickly, since it interacts with the other electrons in the material by their electric repulsion. In the course of a Compton effect, the photon scatters on an electron, which gets free from the atomic shell,

similarly to the photoelectric effect. In this case, though, the photon also survives and passes on with smaller energy (longer wavelength) than it originally had. The electron and the photon share the original photon energy randomly: therefore the energy of the electron is *not* a well-defined value. The scattered photon later escapes from the detector or it can be scattered again. As its energy decreases at each scattering (but not between them!), it can end up finally in a photoelectric effect. In such a case the total energy of the photon gets finally deposited in the detector, in the form of kinetic energy of the electrons. During the pair production the photon polarizes an electron and a positron (the antiparticle of the electron) out of the vacuum. This electron was not present initially in the material. For that to happen, the photon energy has to be larger than the total rest energy of the electron and the proton, which is $511+511 \text{ keV} = 1022 \text{ keV}$. The produced electron and positron slows down in the material quickly. The positron approaches another electron of the material after it has lost most of its kinetic energy, and annihilates with the electron. In the annihilation, two (rarely three) photons are radiated out. Both photons have 511 keV energy and they propagate in opposite directions (to satisfy the energy and momentum conservation laws). These photons are not energetic enough any more to produce another electron-positron pair, but they can participate in the Compton and photoelectric effects. If one, or both photons escape from the material without these interactions, we will measure less energy deposited in the detector than the original photon energy was, by precisely 511 or 1022 keV, respectively. This will lead to the appearance of the single (SE) and double (DE) escape peaks in the measured energy spectrum. If, however, the two photons from the annihilation also deposit all their energy in the detector, we will measure an energy deposit that does correspond to the total energy of the original photon.

To summarize, all of the above described processes, elementary photon-material interactions lead to the liberation or creation of one or two highly energetic *charged* particles. The kinetic energy of these particles is much larger than the binding energy of the electrons in the atoms of the material. These energetic particles create *many electron-hole pairs* (the hole is the vacancy at the place of the electron that is kicked out) as they travel in the *semiconductor* material, by kicking electrons into the conduction band. To create such an electron-hole pair only a few *eV* energy transfer is sufficient. Thus, an energetic electron (or positron) can create as much as 10^5-10^7 charge carrier pair in semiconductors, their number being proportional to the kinetic energy of the particle. This total amount of charge is measurable, the charges are collected for a few microseconds. The precision of the energy measurement (the energy resolution) is on the order of 0.1%, thanks to the large number of electron-hole pairs created: large numbers fluctuate statistically less (in a relative sense) than small numbers.

Our detector is a high purity germanium (HPGe) semiconductor detector. The total energy of the γ -photon can be deposited by photoelectric effect, multiple Compton-scattering, or pair production followed by the absorption of the two 511 *keV* photons emitted in the annihilation. In case of the detector we use (with linear dimensions about 5 cm) the multiple photon interactions take place within a few times 10^{-9} s (a few *ns*). With this detector size, γ -s above 200 *keV* that are fully absorbed usually go through multiple Compton-scattering, and not a single photoelectric effect. Therefore, it is more correct to use the term „total energy peak” instead of the „photopeak” for the sharp energy values in the measured spectrum. The energy resolution of semiconductor detectors at 1 *MeV* is about 1-2 *keV*.

A high voltage (3000-4000 *V*) is connected to our HPGe detector, to collect the electrons and the holes on the positive and the negative electrodes, respectively, thereby creating an electric pulse, instead of letting them recombine with each other quickly after the irradiation. It is also

necessary to keep the detector at a low temperature, because under the high voltage there would be already some current flowing, without irradiation. Low temperature decreases the chance that the kinetic vibrations in the crystal push electrons into the conduction band. The detector is placed on the top of a copper rod, while the bottom end of the rod reaches into a liquid nitrogen tank filled with LN₂, keeping the rod at –196 °C (the boiling point of the nitrogen). Copper is a good heat conductor, thus not only its lower end, but also its upper end cools down, keeping the HPGe at low temperatures. The nitrogen evaporates continuously, and should it boil away completely, the power supply would sense the change in the current flowing through, and would turn off automatically. Nevertheless, failing to care about the periodic nitrogen refill ultimately damages the detector.

The signals of the detector are amplified and shaped by a spectroscopic amplifier unit, producing a short pulse with a few volts amplitude. Each photon that deposited any amount of energy in the detector produces such an electronic signal. This chain of events (photon arriving, interacting, creating a pulse) is called briefly a *hit*.

15.2.2 The method of the measurement

During the measurement, the gamma-energy spectrum of the samples should be collected for a known duration of time. We will use a portable spectroscopy analyzer module, which connects to a personal computer via the USB port. The information will be evaluated by software that handles the analyzer data.

The operation of the amplitude analyzer

The analyzer holds a vector with 8192 ($=2^{13}$) elements, and each element is cleared in the beginning of the measurement. The analyzer measures the height of the peak of the incoming analog signal and digitizes it: characterizes its amplitude by an integer between 0 and 8192, proportionally to the peak voltage. The larger the energy deposit was in the detector, the higher this value is. This is called the channel (number). We increase the element of the vector that corresponds to this channel by 1. For example, if the energy of a given hit – in this arbitrary unit – is 536, then we add 1 to the content of the 536th channel. At the end of the measurement the content of this channel will be the number of those photons that deposited precisely that much energy that corresponds to the channel 536. This way the probability of occurrence of various energies is measured. In case there is a well-defined energy that was deposited in the detector many times, a peak appears in this histogram. The resolution of the detector is finite (not infinitely good), so even in case our photons had the same very sharp, well-defined energy value every time, the electric pulses will not be exactly the same, and the integers associated to them may also differ. Therefore, the hits corresponding to the very same energy will not fall into one single channel, but will be distributed over 5-10 channels, forming a peak. If we want to determine the total number of photons at a given energy, we have to take the total *area* of the peak, that is, we have to sum up the contents of all the channels that form the peak.

15.2.3 Energy calibration

Assuming that the location of the peaks is in simple linear relation with the energy deposited in the detector, we can determine this energy-channel number relation using a nuclide that radiates several gamma photons with known energies. After this energy calibration, other unknown isotopes can be easily identified, based on their γ -energies.

The energy calibration can be carried out with a natural ^{232}Th source, collecting its spectrum for a few minutes. It is recommended to use the peaks that appear at 238.6 keV and 2614.7 keV energies (these are the highest intensity peak at low energy, and the highest energy peak, respectively)! For a more precise result, let us repeat the calibration also taking into account the peaks at 338.3 keV and at 721.2 keV. For the calibration, one has to highlight the peaks with the two cursor lines, and add the peak to the list of ROI's (Region of Interest). We can type in the known energies of the peaks in keV units. After the calibration, placing the unknown sample on the detector and measuring for a known duration, we get a spectrum that completely differs from the thorium spectrum. Obviously, the area of the peaks at various energies will be proportional to the duration of the measurement and to the activity of the sample. The total and net area of the peaks as well as the statistical error of the peaks can be determined with the help of the control software of the analyzer.

15.2.4 The identification of the nuclides in the sample, based on the observed γ -energies

The gamma-radiation that arrives from the sample follows an alpha- or beta-decay, when the daughter nucleus is initially in an excited state. Usually there are several possible energy levels involved, and not all of the possible gamma-energies appear in each decay. Also, the emission of several different gamma photons can follow an alpha- or beta-decay. It has to be kept in mind that only natural radioactivity is expected from a granite rock sample. In these natural decay chains the larger mass number that occurs is 238 (uranium), and the decay chain ends around the mass number of 210, where the decay daughter is already a stable isotope.

The software called DECAY is also available, in which several thousand gamma energies are listed together with the emitting nuclides. Based on the measured energies, and the fact that the mass number of the possible nuclei should be between 210 and 238, we can easily identify the emitting nuclides. The fraction of decays in which a given gamma energy occurs can be found for each nucleus. This fraction is called intensity, or probability. Obviously, if the measured gamma photon is emitted only in 19% of the decays ($I=0.19$), then the number of detected gammas should be divided by 0.19 to estimate the activity.

15.2.5 Determination of the detector efficiency

A further problem is to determine the chance that a given photon energy entirely remains in the detector. It is quite difficult to estimate this *efficiency*, especially in case of bulky, large samples. In our case a simulation method, so called Monte-Carlo method is used to determine the probability of detection.

The basis of the method is the following:

We can assume that the emission of gamma photons from all the volume elements of the sample to all directions is equally probable. The computer generates photons emitted to random directions, and we count them. In case a photon started to the direction of the detector, we examine if it interacted with the material of the detector. We know that there can be three types of interactions: photoelectric effect, Compton-scattering and pair production. The software contains the cross section (the probability) of these processes for germanium as a function of photon energy. We follow the photon until it loses all of its energy or it leaves the detector. We count the cases where the total energy was deposited. The ratio of the number of these events and the number of all generated photons gives the *efficiency* (η) we wanted to determine. If there was a photoelectric effect, the total energy of the photon remained in the

detector we do not have to follow up that event any more. In case of the Compton-scattering, however, we have to follow up the scattered photon, since it can scatter again in the detector. Since this process completes in less than a nanosecond, and the charge collection time is about a microsecond, the total energy can apparently remain in the detector also with multiple Compton scattering. The situation is similar in case of the pair production, too, since the positron slows down, encounters an electron and annihilates, and the created two photons have to be followed further, separately.

When applying this method, several million events have to be generated to estimate the efficiency precisely. The statistical accuracy of the calculation improves inversely to the generated event number. Thus, to decrease the statistical error by half, we have to generate four times more events.

15.2.6 The problem of self-absorption of the sample

A further problem is that in case of extended samples the photons emitted close to the detector reach the detector with a high probability, but photons emitted from the far side of the sample „see” the detector at a much smaller solid angle. In addition, the photons have to traverse the material of the sample without interaction! To estimate this chance, again the three types of interactions have to be considered: the photoelectric effect, the Compton effect and the pair production. In case any of these interactions happen, a photon with a completely different energy will arrive at the detector, or it does not even reach it. These cases should not be counted into the efficiency. This phenomenon is called self-absorption.

- To estimate the magnitude of the self-absorption, we have to know the approximate chemical composition of the sample, and the atomic number- and energy-dependence of all the three interactions for the whole periodic system of elements. It is important to know that the cross section of the photoelectric effect is proportional to the fifth power of the atomic number. Although this probability is decreasing with energy quickly, at smaller gamma energies (100-400 keV) already the light elements cause a significant self-absorption. Unfortunately, the atomic number and energy dependence of the interaction probabilities for all elements is programmed only via approximate formulas in the software, and the precision can be further degraded by the lack of information about the chemical composition of the sample. Due to these facts, the systematic error on the efficiency is about 6-8%, thus we cannot ultimately measure the ^{238}U content of the sample.
- For precise measurement we could measure the self-absorption of the sample using radioactive isotopes radiating with known energies, then we could correct the efficiencies calculated without self-absorption based on these results. This method, however, would require much longer time to carry out, and thus it is out of the scope of the laboratory exercise.

15.2.7 Measuring the granite sample

We can see sharp, Gaussian-shaped peaks in the energy spectrum at certain characteristic energies. The area of those peaks (without the Compton-background) is related to the activity of the decaying nuclide, and also proportional to the probability (intensity) of that gamma energy occurring at the decay of the given isotope. For example, if we have 1000 ^{214}Pb decaying nuclei, then – because of the competing de-excitation channels – in only 192 cases is there a photon with 295.2 keV energy emitted. We attempt to determine the number of ^{214}Pb atoms from the number of measured γ -photons, therefore we have to divide the measured number of photons by the intensity factor, in this case by 0.192.

The concentration is thus proportional to the area of the peak, but it is important to take the energy dependent self-absorption and the energy dependent detector efficiency into account. Both factors are measurable or calculable. Since the self-absorption depends on the composition of the sample, it has to be recalculated for each sample separately.

The half-life of the ^{238}U and some of its daughters ($T_{1/2}$, from which the decay constant $\lambda = \ln 2/T_{1/2}$ can be calculated): ^{238}U : 4.468 billion years, ^{234}U : 244.5 thousand years, ^{230}Th : 77 thousand years, ^{226}Ra : 1600 years, ^{214}Pb : 26.8 minutes, ^{214}Bi : 19.9 minutes. The half-life has to be converted to seconds to calculate the number of atoms correctly. The way to calculate the activity is thus:

$$A = \frac{N}{\eta \cdot I \cdot t},$$

where A is the activity to be obtained, N is the measured net peak area, η is the efficiency, I is the intensity of the gamma-photon and t is the duration of the measurement. The activity calculated in this way has to be divided by the decay constant of the ^{238}U to get the number of uranium nuclei, then it is easy to calculate the uranium content of the sample – that is, how many grams of uranium is contained in a ton (1000 kg) of rock sample.

15.3 Lab course tasks

1. Calibrate our equipment using the Th-sample with a few minutes of measurement!
2. Place the granite sample on the detector. Conduct a measurement for 10-15 minutes! Determine the net area of the peaks at 186 keV, 295 keV, 352 keV, 609 keV, 1001 keV energy, and the net area of 5-6 further peaks!
3. Using the DECAY program, find the nuclides and the line intensities that possibly correspond to these lines!
4. Using the Monte-Carlo program, determine the detection efficiency for the above energies! For that, assume that the granite is essentially composed of SiO_2 , in which the mass number of the Si is 28, its atomic number is 14, the mass number of the oxygen is 16, its atomic number is 8, and the sample has a 2.5 cm radius, 5 cm height and a mass of 280.85 g.
5. Calculate the activity for each measured lines using the net peak areas, the duration of the experiment, the peak intensities and the efficiencies. Determine the experimental uncertainty of those activities, and calculate the average activity! (Here, the uncertainty of each obtained activity can be quite different, therefore we have to use weighted average!)
6. Based on the average activity, calculate the uranium content of the sample!
7. Measure and analyze the salt mixture that can be found in the laboratory in a similar way!
8. Measure and analyze one or two unknown samples found in the laboratory!

15.4 Calculation of experimental uncertainties of the measured and calculated quantities

It is not sufficient to provide the measured quantities (and their units!), their experimental uncertainties („errors”) are also needed. (The unit of the error is always the same as the unit of the corresponding quantity.) Without knowing the errors, it is not possible, for example, to compare two quantities meaningfully. The experimental error originates from the error of the net peak area on one hand, and from the systematic error of the efficiency (including self-

absorption) on the other hand. Compared to those, the measured time duration and the intensities are very precise values, we can neglect their uncertainties. It is difficult to estimate the error of the efficiency, and it is out of the scope of this laboratory exercise, therefore let us take it to be 7% in case of the granite sample, and 20% for the other samples. We can deal with the errors of the peak areas, however, and the errors propagated from those. We summarize in the following the important knowledge we need to calculate these statistical errors.

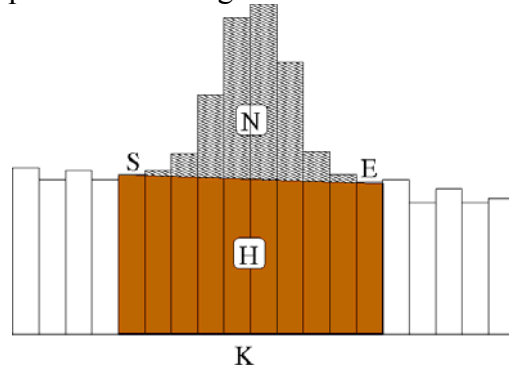


Figure 15.1. The uncertainties of the counts in the selected areas in a gamma-spectrum

The error of the total peak area (T) is the square root of the total area (total number of hits) in the peak: $\sigma_T = \sqrt{T}$. The net peak area (N) is the area obtained after subtracting the Compton-background (H) from the total area: $N = T - H$. If two quantities are statistically *independent*, then the error of their sum (or difference) is the square root of the sum of the square of their error. Therefore, the error of N can be calculated like this: $\sigma_N = \sqrt{\sigma_T^2 + \sigma_H^2} = \sqrt{T + \sigma_H^2}$. The background is subtracted by calculating the area of the trapezoid determined by the left and right edges of the peak we have highlighted. The area of this trapezoid is $H = K(S + E)/2$, where K is the number of channels that compose the peak, S is the content of the first, E is the content of the last channel. Since K is fixed (does not change from measurement to measurement), and the error of S and E is \sqrt{S} and \sqrt{E} , the error of H becomes: $\sigma_H = K\sqrt{S + E}/2$. Then, as discussed above, the absolute error of the net peak area:

$$\sigma_N = \sqrt{T + K^2(S + E)/4} = \sqrt{T + H \cdot K/2} = \sqrt{N + H \cdot (K/2 + 1)}.$$

We have to note, however, that the total peak area T is not strictly independent of S and E , since the first and last channel is also contained in the peak. Due to that, the above formula has to be corrected slightly. This more exact formula is (we do not provide its derivation, but a bonus is due for it in the lab report):

$$\sigma_N = \sqrt{N + H \cdot (K/2 - 1)}.$$

As we can see, we need the number of channels (K) that compose the peak (ROI) besides its total area T and net area N . K is easily readable from the computer display. We have to write down this σ_N statistical error next to each net area value! Now it is easy to get the *relative error* of the net peak area: $\sigma_N^{rel} = \sigma_N / N$ (or hundred times that if expressed in percent).

The *relative error* of the product or ratio of two independent quantities is the square root of the sum of the square of the *relative errors* of the two quantities. Since the net peak area and the efficiency has to be divided when calculating the activity, the *relative error* of the activity, A , is obtained as:

$$\sigma_A^{rel} = \sqrt{(\sigma_N^{rel})^2 + (\sigma_\eta^{rel})^2}.$$

If we assumed the efficiency to be very precise, then the above would become simply $\sigma_A^{rel} = \sigma_N^{rel}$. Of course, the *absolute* error of the activity is obtained as: $\sigma_A = A \cdot \sigma_A^{rel}$. When we calculate the average of the activities, we have to take into account that the activities are not equally accurate, their errors differ. At this averaging, do not consider the error of the efficiency yet, only the statistical errors! The correct method is the *weighted averaging*, where the weights are given by the reciprocals of the squares of the individual errors:

$\langle A \rangle = \frac{\sum A_i / \sigma_i^2}{\sum 1 / \sigma_i^2}$, where A_i is the activity of the i^{th} peak, and σ_i is its absolute error. The statistical error of the average obtained in this way is given by the formula $\frac{1}{\sigma_{\langle A \rangle}^2} = \sum \frac{1}{\sigma_i^2}$, that

is: $\sigma_{\langle A \rangle} = \frac{1}{\sqrt{\sum \frac{1}{\sigma_i^2}}}$. We still have to add – using the square rule – to that (more precisely to

the $\sigma_{\langle A \rangle} / \langle A \rangle$ *relative error*) the *relative* error of the efficiency to get the *relative* error of the average activity. The *absolute* error of the activity from that is: $\sigma_{\langle A \rangle} = A \cdot \sigma_{\langle A \rangle}^{rel}$.

We always have to give the absolute statistical error of the net peak areas and the statistical and total absolute errors of the average activities in the laboratory report.

It is recommended to use many decimal digits in the intermediate calculations, but the final results should always be given to only as many decimal digits as reasonable and meaningful. Do not write those decimals that are completely undeterminable due to the larger experimental error! The errors have to be given up to *two or three* (nonzero) digits only. For example, „32,9 ± 1,7 Bq” is correct, because the quantity and its error are given up to the same decimal place, but „32,9235 ± 1,6854945 Bq” and „32,9235 ± 1,68 Bq” are incorrect. Similarly incorrect is to round the values excessively, like „30 ± 2 Bq”.

15.5 Personal exercises

Each member of the group is supposed to briefly describe the method of the measurement and the samples and tools they have seen in the laboratory. Besides, everyone has to note the measured quantities in the report. Parts of this laboratory text should not be copied into the report. Finally, each member of the group should choose *one* of the problems below, and work it out in detail in his/her laboratory report.

A) Calculate the activity of the ^{214}Pb and the ^{214}Bi based on their measured peaks (no other peaks should be considered for this problem)! First, calculate the activities that correspond to the lead, and take their weighted average (where the different errors of the activities are taken into account). Let us do the same for the bismuth peaks! Determine whether the activity of the lead is the same, or not the same as the activity of the bismuth within experimental errors! Then calculate the weighted average of the average activity of the lead and of the bismuth. Assume that none of the radon could exit the sample, and the secular equilibrium is valid. Calculate from that the total uranium content of the sample in grams, and the uranium concentration in „grams of uranium / ton of rock” units!

B) Calculate the activity of the $^{234}\text{Pa}^m$ and its experimental error! This isotope is certainly in secular equilibrium with the ^{238}U . Based on that, calculate the amount and concentration of ^{238}U in the granite! Give the experimental errors! Then calculate the amount of ^{235}U as well, using the fact that – in natural uranium – the ^{235}U content is always 0.72% of the total uranium content. How many ^{235}U nuclei are there in the sample? Calculate, what the contribution of ^{235}U to the net area of the peak at 185 keV is (that is, what would be the peak area if only ^{235}U would be present?). Subtract this ^{235}U -contribution from the measured net peak area and give the remaining area, together with its error! This remaining area is due to the radiation of the ^{226}Ra . Based on this, calculate the activity of the ^{226}Ra and its error! Does this activity agree with the activity of the $^{234}\text{Pa}^m$ within errors? Calculate the weighted average of the activity of the ^{226}Ra and the $^{234}\text{Pa}^m$ and the error of this average! If we got a larger value than the person who calculated exercise A), which means that part of the radon could escape from the sample before it decayed. If so, what percentage of the radon escapes?

C) Measure the gamma spectrum of the „Horváth Rozi” salt mixture (mass: 231 g) found in the laboratory! This salt only contains 60% NaCl for health reasons; the rest is KCl. Measure the amount of ^{40}K in the sample (and its error)! Is there any other radioactive isotope in the salt? Calculate, based on the measured values, the total amount of potassium in the sample (given in grams), using the fact that the fraction of ^{40}K in natural potassium is always 0,0117%! Always give the experimental errors! Calculate the total K-content of the salt from the fact that 40% of the total mass of the mixture is KCl! Do the two values (the measured and the calculated K-content) agree within the experimental error?

The laboratory instructor sometimes gives a further simple exercise connected to another unknown sample.

D) Let us choose an unknown sample from the many interesting samples in the laboratory! We can also bring in a soil sample from our own garden in a glass food container. Based on the energy spectrum of this unknown sample, let us identify as many different radioactive isotopes as we can! Explain the observed peaks, as well as the eventually missing peaks (compared to the granite)! Estimate (calculate) the amount of uranium, thorium, potassium, cesium, and eventually other isotopes, similarly to exercise A)! Find out if there is uranium in the sample, or, eventually, only radium and its daughters! Give the experimental errors of all measured quantities! If we *do not find* the characteristic peaks that correspond to certain isotopes (like ^{137}Cs or ^{40}K), then let us give an *upper bound* for the amount that can be in the sample at most, without producing a visible peak during the measurement. This can be done in two ways. If, in the given range, there is not a single hit, and not even Compton-background, then we have to calculate the number of nuclei, for which we should have seen at least *one hit* at the given energy, if it was present in the sample. If there is a nonzero Compton-background, then we can define a peak area (ROI) around the expected energy value, similarly to the case of the other visible peaks. Then, we can add the net „peak area” (which will be of course a small number) and two times its error, obtaining this way a fictive net peak area. We can state with high confidence, which we have measured *less than* this fictive peak area. The amount of isotope that would correspond to this area is the *upper bound* we were looking for. (Note that upper bounds *do not have* experimental errors, so do not give any uncertainty for upper bounds).

15.6 Test questions

- 1.) What kind of natural radioactive decay series are known?
- 2.) How do we determine the uranium content of the granite?
- 3.) How does our detector work, and what kind of detector it is?
- 4.) Would it be possible to measure alpha or beta radiation with our detector? Why?
- 5.) If 1 kg soil contains 0,01% uranium, how large is its activity?
- 6.) Which quantities do we need in our experiment to calculate the activity (formula)?
- 7.) Why do we need high voltage when we use the germanium detector?
- 8.) Why do we need to keep the detector at low temperature?
- 9.) What is the activity of 0,119 g clean ^{238}U ? Its decay constant is about $6 \cdot 10^{-18} \text{ s}^{-1}$.
- 10.) What is secular equilibrium, and what is the necessary condition for it to set in?
- 11.) How does our amplitude analyzer work, and what is its role?
- 12.) How do we calibrate our equipment (energy calibration)?
- 13.) How does the incoming gamma-photon interact with the detector? How does all this depend on the energy of the photon?
- 14.) We find two peaks in the spectrum, corresponding to the same nuclide. We calculate the activity for both. For the first we get $100 \pm 10 \text{ Bq}$, for the second $112 \pm 5 \text{ Bq}$. What is the weighted average of the two, and the error of this average?
- 15.) Our result for the net peak area is 200 ± 10 , and the efficiency is $0.02 \pm 10\%$. The intensity for the given line is 0.5, and we have measured for 20 seconds. What is the activity and its error?
- 16.) What is the self-absorption, and what kind of difficulty does it cause at the measurement?
- 17.) Why is it necessary to know the chemical composition of the sample to determine its uranium content?
- 18.) Let us say, the activity of the ^{214}Bi in our sample is $1000 \pm 55 \text{ Bq}$, while the activity of the ^{226}Ra is $1500 \pm 75 \text{ Bq}$. How can we explain the difference?
- 19.) In case of one of our samples only the gamma-radiation of the ^{235}U can be detected, there is no sign of the ^{214}Pb nuclide. What can be the reason for that?
- 20.) Which natural and artificial isotopes can be usually observed in soil samples with gamma spectroscopy?

16. Annihilation radiation from positrons and positron emission tomography (PET)

16.1 Introduction

During this lab course we will investigate the annihilation of positrons; we will acquaint ourselves with the basic principles of positron emission tomography (PET). Our task will be to perform a PET investigation on a test marker, and during this investigation we will have to find the location of an idealized tumor with the largest possible accuracy. The quantities to determine are: number of tumors, their location on the (x,y) plane, the accuracy of the measurement and the activity in the tumors. First we will review the knowledge needed to understand the measurement, then we will get to know the measuring device.

In our everyday life, objects around us (and ourselves, too) are built up of protons, neutrons (forming nuclei) and electrons around them (thus forming atoms and molecules). Protons and neutrons are not elementary particles, they consist of quarks (quarks named “up” and “down”), but electrons are elementary, as far as our present knowledge goes. All of these particles have anti-particles, even if we do not meet them very often. There are antiquarks (anti-up and anti-down), these form anti-protons, but electrons also have anti-partners. These are called positrons. Most properties of an anti-particle are the same (mass, lifetime, etc.), but the electric charge is exactly the opposite. Thus the charge of positrons is positive; the charge of an antiproton is negative. An anti-hydrogen atom can be formed out of these two, and the properties of this will be exactly the same as the regular hydrogen atom. It is important to know that if a particle meets its anti-partner, they annihilate and all of their mass is converted into energy (in form of photons) according to Einstein’s formula, $E=mc^2$. This process is similar to that takes place in a semiconductor, when an electron and a hole (an electron-defect) meet, and the electron jumps into the hole with an effectively positive charge, and so both “charges” disappear, accompanied by energy liberation.

The existence of the positron was theoretically predicted by P. A. M. Dirac in 1928, and Carl D. Anderson was the one who, in 1932, first experimentally detected a particle (in cosmic showers) with exactly the electron mass but with an opposite charge. This particle could be identified with the hypothetic positron. In the next years, electron-positron annihilation was also observed. Both scientists received a Nobel Prize for their discoveries.

A positron can be pictured as an electron-defect in the vacuum. Vacuum in particle physics is basically a sea of particles and anti-particles bound to each other, and a large amount of energy is needed in order to separate them, to create free particles or anti-particles. But if an electron may jump into an electron-defect, a positron, and then the vacuum state is restored there, as there are no more free particles or anti-particles. As the vacuum state is the lowest energy state possible, energy is liberated by this process. However, if energy is being invested, then the opposite can happen, an electron-positron pair may be formed out of nowhere, just from energy. This can be pictured as if an electron and an electron-defect, a hole would be formed. In high energy particle colliders this happens many times in a high energy collision, particle-antiparticle pairs are formed in large amounts.

During annihilation, the total energy and momentum is also conserved. The total energy of a resting electron (when we are in its rest system) is mc^2 , where m is the mass of the electron and c is the speed of light. Thus the total energy available from one annihilation is $2mc^2$, as the mass of the electron and positron are the same. This energy is radiated in form of two photons (due to momentum-conservation), which, due to their large energy, are called gamma-photons, denoted by γ .

16.2 Antimatter in Nature

In Nature, we mostly find only matter and no antimatter, thus we do not observe these spectacular annihilations, and neither a large amount of high energy γ -radiation. If there were some antimatter available on Earth, it would immediately annihilate with regular matter, and we would only find matter after these annihilations are done. The amount of energy arising from these annihilations is very large. If 1 gram of protons annihilate with 1 gram of antiprotons, then more than 10^{14} Joules are liberated, equivalent to 50 GWh (the energy produced during one day in the Paks nuclear power plant). But 1 gram of antiprotons is a very large amount of antimatter, which has never been seen on Earth. The usual quantity for energy and mass in this case is the electronvolt (eV). This is the energy which one electron acquires while being accelerated with an accelerating voltage of 1 V. As the charge of the electron is $1.6 \cdot 10^{-19}$ Coulomb, this energy of 1 eV is equivalent to $(1.6 \cdot 10^{-19} \text{ C}) \times (1 \text{ V}) = 1.6 \cdot 10^{-19} \text{ J}$, in SI units. The mass of the proton is 938 million eV/ c^2 in these units (note that energy per speed of light squared equals to mass). Thus if a proton and an antiproton annihilate, the 1876 MeV energy is created. The mass of the electron is 511 keV/ c^2 . In the most common case of electron-positron annihilation, two photons are created, each with energy of 511 keV.

In Nature, antimatter might be created in natural or artificial radioactivity (positive β decay), or in particle showers generated by high energy cosmic radiation. In particle accelerators a large amount of antimatter might be produced, and by the complex technology of storing anti-particles in small magnetic traps, even anti-atoms like anti-hydrogen can be formed.

During natural radioactive β decay, not only a positron (or an electron) is formed, but a particle called neutrino (or an antineutrino) as well. This does not interact with any other particles almost at all, so from an experimental point of view, it is not observable. The well-defined energy arising from the decay is taken by the neutrino and the positron, but with a random ratio. Thus the energy of the positron will not be a fixed value, but is varying according to a broad probability distribution. The maximal positron energy is however well defined, and is usually on the order of a couple times 100 keV.

The β^+ -decaying isotopes used in material studies are e.g. ^{22}Na ($T_{1/2}=2.58$ years, maximal positron energy is $E_{\max} = 545$ keV), ^{58}Co ($T_{1/2}=71$ days, $E_{\max} = 470$ keV) or ^{64}Cu ($T_{1/2}=12.8$ hours, $E_{\max} = 1340$ keV). In medical sciences the isotopes of biologically important atoms are used, e.g. ^{11}C ($T_{1/2}=20$ minutes), ^{13}N ($T_{1/2}=10$ minutes), ^{15}O ($T_{1/2}=2$ minutes) or ^{18}F ($T_{1/2}=110$ minutes). These medical sources have a very short lifetime, so they do not deposit radiation in the patient's body for a very long time. However, due to their short lifetime, they do not exist in Nature, so they have to be created in nuclear reactions (in particle accelerators) not very far from the medical investigation location. These isotopes are then injected into the human body by putting them into sugar, water or ammonia molecules. In oncology, mostly ^{18}F is used, while in neurological investigations ^{15}O is common. In this lab course we will use the isotope ^{22}Na .

16.3 Positron Emission Tomography

16.3.1 Positron annihilation in details

The lifetime of the positron is infinite similarly to that of an electron. However, inside any material (even the radioactive source itself) the positron meets an electron almost immediately. Annihilation happens however only, if the relative speed of the electron and the positron is small. Thus the positron first loses speed inside the material (due to its electric charge). This slowing process lasts approximately 10^{-12} s (0.001 ns), and at the end, the energy of the

positron is smaller than 0.1 eV. During this time, the positron can cover a distance of typically 0.1 mm. After slowing down, another 0.1 μm distance may be covered, and then the annihilation happens. Figure 16.3 shows this process. The positrons emitted by the ^{22}Na source are slowed down in the material illustrated by a rectangle. During annihilation, two almost oppositely directed photons exit the material. V.

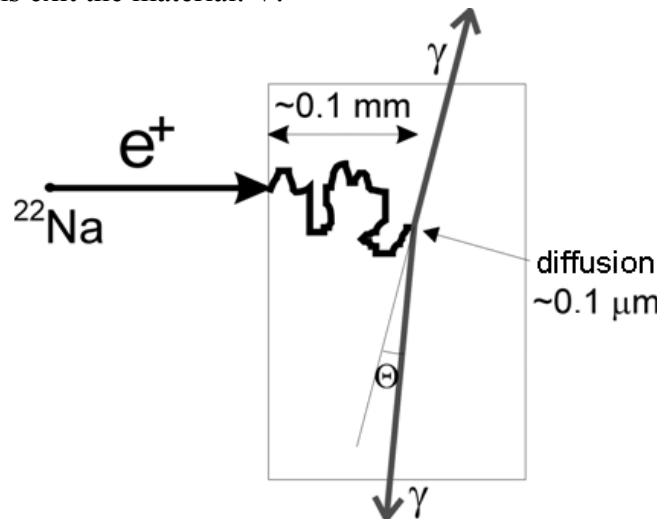


Figure 16.3. Annihilation of a positron inside a material

In an idealized case, if the momentum of the positron and the electron are almost zero, then the total energy, as discussed above, is $2 \times 511 \text{ keV} = 1022 \text{ keV}$. Due to momentum conservation, it is impossible that only one photon is created (the total mass of the two particles would be converted into momentum of the photon); two opposite photons have to be formed as then their total momentum can be zero (or very small, the same as that of the positron and the electron). Note that in real materials, the annihilation creating only one photon may also occur if the other half of the momentum is taken by a nucleus near the annihilation.

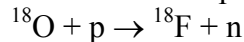
Another very important issue is that the direction of the two photons might not be exactly 180° , if the annihilation happens with a high speed electron. In this case, there is some initial momentum of the system. The kinetic energy of the positron before annihilation is approximately 0.02 eV, while that of the electron inside an atom is 10 eV. The energy of the created photons is however 511 keV, so the initial momentum is almost negligible. The difference from 180° , denoted by θ in the figure, is very small, $1\text{-}2^\circ$ at the very maximum, but usually even smaller. If this difference can be measured, it gives us information about the velocity distribution of electrons in the material. This is an important method in solid state physics, our device is however far less accurate.

16.3.2 Medical applications of positron annihilation

Positron annihilation is used in medical practice in the Positron Emission Tomography (PET) technique. Goal of these investigations is to find biologically hyperactive areas, such as tumors. With modern PET techniques, a 2- or 3-dimensional image of the body can be drawn without surgery. These images are needed in order to make an accurate diagnosis, as well as to plan further treatment of the patient. The PET imaging method was developed by Michael E. Phelps in 1975.

First step of a PET investigation is to insert a short lifetime radioactive (positron emitting) isotope to the patient's body. This isotope has to be part of a biologically active molecule, which is then absorbed by the tissues with active metabolism. In most cases, the used mole-

cule is $C_6H_{11}FO_5$, a glucose molecule having radioactive fluorine instead of the sixth oxygen. These molecules can be created in cyclotrons (low energy particle accelerators) where water molecules enriched with ^{18}O atoms are bombarded with protons. Then the



reaction takes place, and the created fluorine can be gathered and connected to glucose molecules in specialized laboratories. Then these molecules have to be transported very quickly to the hospital performing the PET investigation. As this transportation might be complicated, in newer PET devices cyclotrons and isotope labs are attached as well. In case of adults, isotopes of 200-400 MBq activity are entered into the circulatory system. The modified glucose may enter into any cell that needs an enhanced amount of sugar, e.g. brain tissue, liver and also every kind of tumor. The molecules stay in the given cell while the fluorine performs the positron emitting β -decay.

Less than an hour after the injection of the isotope, the patient is put into the PET device. The positrons are annihilating in the patients body and thus are producing two oppositely directed photons each having energy of 511 keV. These photons can exit the human body without any interaction or energy loss, so they reach the detectors placed around the patient. The photons can then be detected with silicon photo-diodes or scintillation detectors (we will use these in present lab course). Figure 16.4 shows the setup of such a PET investigation.

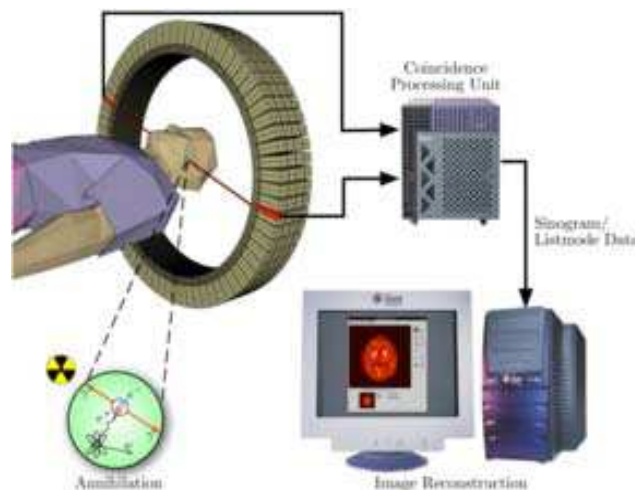


Figure 16.4. Setup of a PET investigation

16.3.3 Principles of PET measurements

In case of a PET measurement the radioactive source (injected into the patient's body) is surrounded by several dozens or even hundreds of detectors organized in a ring shape. These detectors are sensitive to the 511 keV energy oppositely directed photons coming from positron annihilations; the system can then also tell if the two opposite photons arrived at the detectors at the same time, i.e. their coincidence is investigated.

The material used also in our lab course, the NaI scintillator (see [Subchapter 14](#)) is such a special material, which generates a light scintillation if a charged particle (e.g. an electron) goes through it. The γ -particles are uncharged; however, they hit electrons of the scintillator material, which then break out of the atoms. These electrons are already charged particles, so they generate light scintillation. This light enters a photoelectron-multiplier tube (PMT) connected to the scintillator. The PMT has a transparent window where light particles generate electrons. These electrons are then manifolded by a voltage of several hundred or thousand volts. Several hundred thousand electrons are created in this way; these can be detected al-

ready with sensitive amplifiers. The amount of the signal (number of electrons) is proportional to the energy of the incoming photon (which is 511 keV in our case). Due to this proportionality one can select the photons coming from annihilations and filter out any other photons or particles reaching our detectors.

There is a very important condition to assign the signals of the detectors to a positron-annihilation: the two opposite photons have to reach our detectors at the same time (as they were also created at the same time). This simultaneity is of course approximate, but far within the accurateness of our detectors. The method of investigating the simultaneity is called coincidence, and for the exact definition one needs the coincidence window width, the time interval where two signals can be regarded as simultaneous. This is usually on the order of 10 or 100 ns, as light travels a 30 cm distance during 1 ns (if we do not want the coincidence to be sensitive to the exact location of the annihilation, at least a couple ns tolerance has to be set). In the lab course the coincidence window will be on the order of 1 μ s. The coincidence requirement is a perfect tool to filter out non-annihilation photons or any other kind of gamma-radiation. This is especially important in case of small activity sources.

Goal of a PET examination is to determine (map) the local concentration of the injected radioactive isotope. First let us see what happens in case of one single radioactive grain. The gamma-radiation exits the body from the exactly same point (or the same mm^3 sized volume). The 511 keV photons reach two of the detectors placed around the patient. It is for sure that the radioactive grain is somewhere on the straight line connecting the two detectors (note that the two photons might not be exactly opposite, and the detector cells are also not point-like, so we should rather talk about a straight tube not just a line). As also a next photon pair will be emitted, two lines can be drawn. The grain will be on the cross-section of the lines. These lines are called response lines, as shown in Figure 16.5. Usually the more response lines are drawn, the more accurately one can determine the location of the radioactive grain.

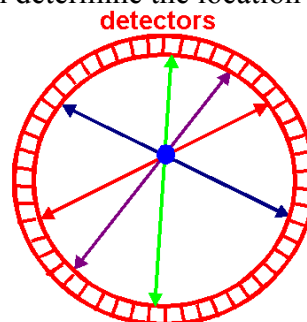


Figure 16.5. Response lines from a single radioactive grain

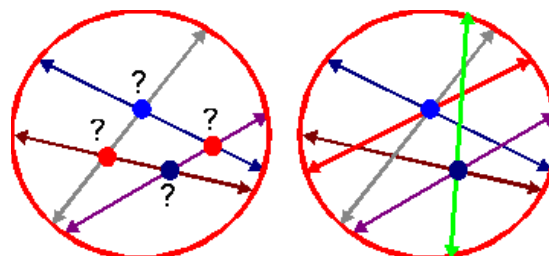


Figure 16.6. Response lines needed to determine the location of two radioactive grains

Note that, as the speed of light is finite, from the arrival time difference of the two photons the exact location of the grain could be calculated. However, this time difference is very small (< 1 ns), so this is not possible with our device, nor with regular PET machines, only with the most modern devices. Their great advantage is that only a smaller amount of radioactive material has to be injected into the patient's body, thus reducing his or her radiation load.

If we have two or more grains, two response lines are not enough. If measuring three response lines, one of the grains definitely produced zero or one line, thus the location of that grain cannot be determined. Four lines are also not enough, as shown in [Figure 16.6](#), as there will be four sections. Thus at least five measurements have to be done here. By measuring several response lines one has to find the spatial location of those two intersecting points where the most lines intersect each other. The method is similar if there are not only two but several point-like sources.

In reality however, we do not have point-like sources, but a distribution of isotope concentration developed in the patient's body. In this case the body divided into imaginary cubes or cells in each of which the isotope-concentration is unknown. Then we measure a large amount of response lines, and the number of lines going through a given cell will help to determine the concentration in a cell, with the help of a computer. The spatial resolution (the size of the cells) can be increased by increasing the measurement time or the amount of injected isotope (one could also increase the size or number of detectors, but this would heavily increase the price of the device). During one PET examination usually several million response lines are produced. Medical PET devices contain a lot of detectors, the data of which are evaluated by special, very fast computers and software. The results have to be corrected for *background* radiation, for photon interactions inside the body (cf. Compton-effect), for the dead time of the detectors (after a detection, the detector is "blind" for a very short time), etc. Early PET devices consisted of a single ring of detectors as shown above, but modern devices contain cylinders of several rings. In case of such devices 3D pictures can be produced by allowing or disallowing coincidences between different rings. The latter case is less sensitive but also faster. The end-result is the 3D image of the isotope concentration. With such an image, the doctor or the radiologist may infer valuable information on the size of the tumor or the seriousness of the disease. See such an image in [Figure 16.7](#).

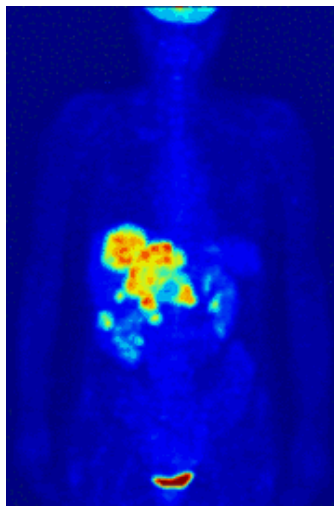


Figure 16.7. A section of a 3D image produced during a PET examination

Medical PET diagnostics is usually connected to other imaging techniques, in order to gain a better reliability. These other techniques may be simple X-ray images, X-ray computed tomography (CT), ultrasound examinations, but also with the method of magnetic nuclear resonance or magnetic resonance imaging (MRI) that has a far better spatial resolution, produces different type of information however. While the MRI image result gives an exact anatomic picture of the patient, the PET image shows the metabolism of the patient's body, e.g. a tumor with very active metabolism. The two different images can be taken at once, while the patient is not moving; from the two-folded information it can be very accurately seen, which part of which organ has been attacked by the disease. Besides diagnosing tumors, PET plays a very

important role in the exploration of brain dementia (impairment of cognitive brain functions) or the Alzheimer-disease.

Drug tests performed on pets are also often evaluated with PET examinations. This is so important for the drug industry that it has a separate name: small animal PET imaging. With its help the number of sacrificed animals can be reduced drastically, as during drug tests no autopsy has to be performed on the animal's body, and one animal can be used in several tests. For humans, PET together with MRI and CT makes it possible to recognize diseases at an early stage, as PET is sensitive to the functional changes of an organ, and these functional changes happen usually much earlier than anatomic changes examinable by MRI or CT. It is a problem however, that PET examinations are much more expensive than CT or MRI, so the accessibility of these investigations is more restricted.

16.3.4 Radiation protection at PET tomography

PET examinations do not require any surgical intervention, but loads the body with a small amount of ionizing radiation. The usual radiation load per examination is approximately 7 mSv. One may compare this to the background radiation (2 mSv/year), to a lung X-ray (0.02 mSv), to chest CT (8 mSv) or to the cosmic radiation load of airplane captains or stewardesses (2-6 mSv/year). In our lab we use a very small activity (<0.1 MBq) source, thus the radiation load during the course is roughly 0.0001 mSv. This is a very small amount of radiation, however, according to the ALARA (as low as reasonably achievable) principle, we should always keep a maximal distance from the source, touch them only with tweezers, and let the lab course leader manipulate the sources. If our distance is larger than one meter, the absorbed dose will be negligible.

16.4 The course of the measurement

During the course we will deal with a simplified model of the above-discussed PET device, a so-called positron scanner, used in medical practice before PET devices were available. We will put one or two ^{22}Na isotope samples (in form of NaCl for example) in a test dummy; this will represent the tumor (as the tumor, with its increased metabolism, captures most of the injected radioactive NaCl). Our goal will be to determine the location of the tumor(s). Touching the dummy is forbidden. It is placed in a transparent acrylic glass box. A transparent foil has to be glued on top of the box, so the response lines can be drawn on it.

There are two rotatable gamma-detectors around the dummy, with which we will measure coincidence events as described above. As we have only two detectors, we will only have coincidences if the tumor is on the line connecting them. Thus by rotating one of the detectors the location of the source can be mapped. The maximal number of coincidences will be observed if the two detectors and the sources are on one single line. If we plot the number of coincidences during a fixed amount of time we get a peak, the location of which will be at the required angle. See such a peak in [Figure 16.8](#). The measurement setup is shown in [Figure 16.9](#).

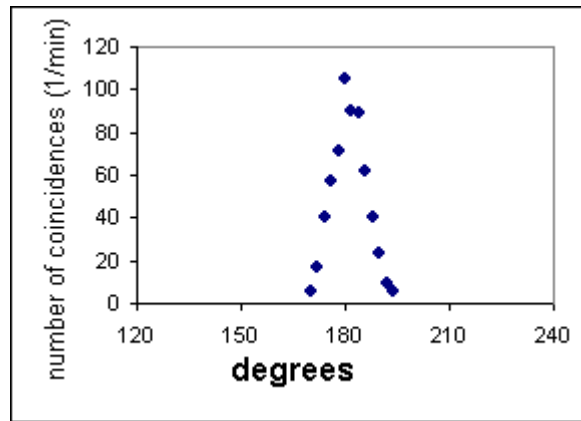


Figure 16.8. A typical angular distribution of coincidences

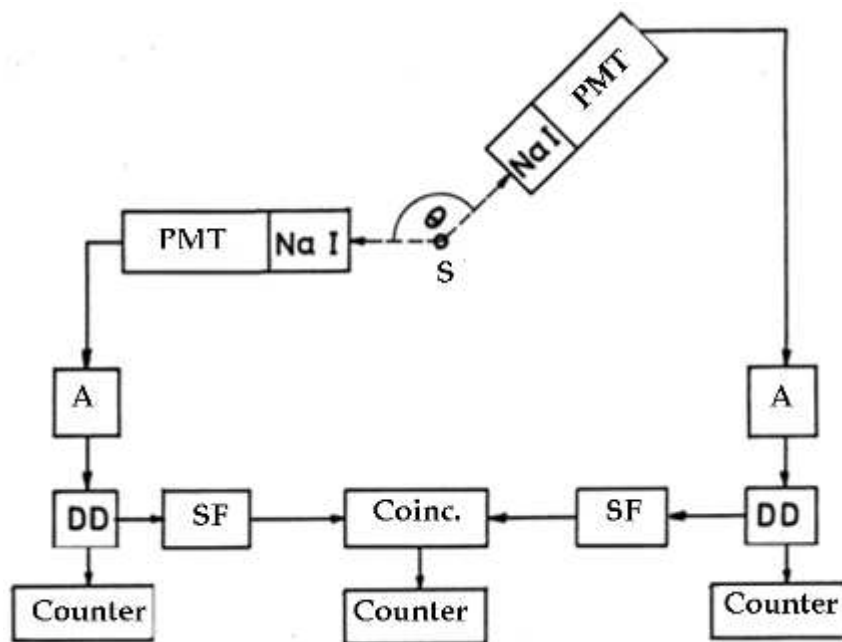


Figure 16.9. Experimental setup. S: source (the test dummy); NaI: scintillator; PMT: photoelectron-multiplier; A: amplifier; DD: differential discriminator; SF: signal forming circuit; Coinc: coincidence device; Counter: counting devices

For our measurements we will use detectors with NaI scintillator heads connected to photoelectron-multiplier tubes. One of the detectors is fixed, the other one can be rotated around the center. Their relative angle can be measured with the built-in protractor. The high voltage of the PMT's and the amplifiers is delivered by one single power supply.

The two analyzed branches are connected to amplifiers and differential discriminators (DD). The DD's will be selecting the photons with the required energy. They only pass signal through the size of which are in a preset ($V, V+dV$) range. The value V is called basis while dV is the windows width. These can be changed by a potentiometer in the 0.1-10 volts range. The signals can be investigated on an oscilloscope.

The signal coming from the differential discriminators is multiplexed to two branches. On one branch the signals are counted by simple digital counters. On the other one, the signals go directly (or, in case of the moving detector, after a preset delay) to the coincidence unit, and

then the coincidence events are also counted. The fourth counter measures time. After setting up everything, we start the measurement with the push of a button. The measurement is stopped after a preset time. The counters, amplifiers, discriminators and the high voltage are set to the right values already. If anything needs to be changed, please refer to the lab course leader.

The ^{22}Na nucleus decays into a ^{22}Ne with a half-life of 2.58 years, and emits a positron during the decay. From the annihilation of the positron, two 511 keV photons arise. Basically right after this, a third, 1280 keV photon is also created when the ^{22}Ne nucleus reaches its ground state (Figure 16.10 shows the decay scheme of ^{22}Na). Because of this, it is important to set up the detector (the differential discriminator) in such a way that it is sensitive only in an energy range around 511 keV. The required values of the differential discriminator settings (V and dV) can be determined by acquiring the energy spectrum of the ^{22}Na source. This can be done by measuring the number of photons reaching the detector as a function of energy – i.e. as a function of the $V \dots V+dV$ interval set in the differential discriminator. The dV value shall be unchanged (for example, 0.1) and V should be scanning a large interval from 0 to 2.5 for example. On the resulting plot will show two peaks, one at 511 keV and one at 1280 keV. The differential discriminator should then be set to include only the close vicinity of the 511 keV peak. This process should be done for both detectors (can be done at the same time however).

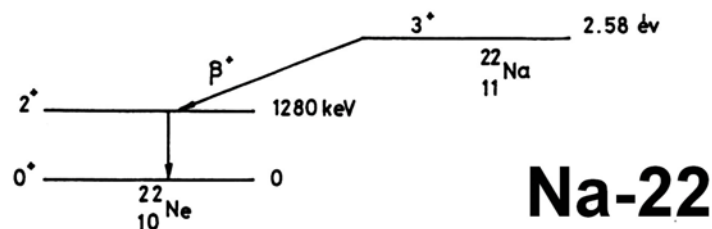


Figure 16.10. Decay scheme of the ^{22}Na isotope

16.5 Lab course exercises

1. Ask the lab course leader to put the test dummy with the “tumor” in it into the plastic box. Glue a transparent foil on it and draw the contours of the dummy on the foil (with a dashed line). The box should be put into its place then. Turn on the measurement device and the high voltage power supply.
2. Measure the energy spectrum of the ^{22}Na with the differential discriminators as described above. The measurement time should be set to 0.2 minutes (12 seconds), and the discriminator window should be $dV=0.1$. Start the base level from $V=0.1$ and go up until at least 2.5 in steps of 0.1, for both detectors. Note that the potentiometer setting shows the tenfold of the window width!
3. After acquiring the spectrum set the discriminator to include the 511 keV annihilation line of the ^{22}Na isotope (this is the first peak, the second is at 1280 keV). Set the measurement time to 1 minute and the angle between the two detectors to 140° and start a measurement.
4. Increase the angle of the two detectors by 5° until you reach 220° . Perform a 1 minute measurement at each location. Plot the number of coincidences as a function of the angle. Determine the location of the peak and its uncertainty.
5. Set the detector to angle corresponding to the location of the peak and draw a straight line on the transparent foil between the two detectors. This will be our first response line.

6. Rotate the test dummy (with the box) by roughly 60° . Repeat the exercises number 4 and 5 to get a second response line!
7. Rotate the box again and repeat exercises 4 and 5 to get a third response line! How many tumors does the dummy have according to the intersections? If possibly more than one, try to find more response lines!
8. Determine the intersection(s) of the response lines and through that the 2D location of the source on the foil. Determine the accuracy of the location by the accuracy of the intersection(s) of the response lines, and also by the accuracy of the lines (which should rather be thick bands as described in the previous parts). Compare the two errors. Determine the organ of the dummy in which the tumor is. If there are more than one tumor, try to estimate the ratio of activity gathered in the tumors, and try to tell which tumor is more active!
9. Turn of the high voltage and all the devices. Remove the foil from the box and attach its photocopy to your lab notes. By setting the bottom left corner of the foil to have coordinates (0,0), determine the (x,y) coordinates of the tumor(s), and also the uncertainty of the coordinates.

The lab report should contain the short description (documentation) of the above tests, a table and a plot of all results (with axis labels and units), the final results with uncertainty and the photocopy of the foil. Try to explain everything in a concise manner, but make sure everything you did is clear from the report. Do not quote more than a couple of sentences from this description.

16.6 Test questions

1. What is the mass and electric charge of the positron? How does this compare to the same properties of the electron?
2. What happens when an electron and a positron “meet”?
3. What kind of particles (and how many) are created in case of an annihilation?
4. How big is the energy of the created particles in case of an annihilation?
5. What angle is there between the particles created in annihilations?
6. What is the velocity of the particles created in annihilations?
7. What is the positron lifetime in vacuum or on matter?
8. Approximately how much antimatter is there on the Earth (in kg units, rounded)?
9. Where can we observe antimatter in Nature?
10. What kind of β decays do you know? What particles are created in these decays?
11. List β^+ decays used in solid state physics and ones used in medical studies!
12. What is the usual half-life of isotopes used in PET examinations?
13. Is it possible that only one photon is created during annihilation? If yes, how?
14. Is it possible that three photons are created during annihilation? If yes, how?
15. May the positron be annihilated if it is colliding into a proton?
16. What determines the angle of photons (the difference from 180°) created in annihilation?

17. What kind of isotope is injected into the patient's body in a PET examination? How big is the activity of the used isotope?
18. May the annihilation photons trespass human tissues?
19. How do our detectors detect photons?
20. How much distance is covered by γ -radiation during 1 ns?
21. What is the essence of the coincidence method?
22. What is a response line, and how thick is it?
23. How many response lines are needed to localize one point-like source?
24. How many response lines are needed to localize two point-like sources?
25. What kinds of corrections are needed when investigating the results of a real PET examination?
26. What other imaging techniques are used together with PET diagnostics?
27. From a medical point of view, what kind of information do we get from PET? And from MRI?
28. List two diseases where PET has a large significance in diagnostics!
29. Why is PET of small animals important?
30. What kind of surgery is needed in case of a PET examination?
31. How big dose does a patient get during PET? And us, during the course?
32. Why do we apply a rotating detector during the measurement?
33. Why do we have a differential discriminator, what does it do?
34. How many photons are created in one decay of a ^{22}Na ?
35. How does the energy spectrum of the ^{22}Na look like?
36. How do you determine the location of the tumor on the test dummy? And how do you calculate the accuracy of the location?

Chapter IV.

ENVIRONMENTAL RADIOACTIVITY II. (PARTICLE RADIATIONS)

This chapter focuses on the particle radiations that can be important in the environment. These particle radiations are the alpha- and the beta-radiation. We will describe a technique to determine the concentration of tritium in waters using their radioactivity. Tritium atoms are good trace elements in the hydrogeology and in those topics where water is the central target in the environmental sciences. Tritium is created in the atmosphere and mixed up in the atmosphere, but the subsurface waters are free of tritium since their long traveling time. To understand the origin of the waters the tritium concentration is useful information. The beta-decaying isotopes emit Cherenkov radiation. We will be introduced to the techniques to count this radiation, which is a general method for determination of the concentration of the beta-decaying isotopes or other chemicals with the help of the beta-decaying nuclei.

The alpha-spectroscopy part of the exercises is centered around the radon. This is an important issue in environmental physics, since high radon activity areas can harm the population. The radon can accumulate in the indoor spaces, and their daughter product can attach to the pipes of the lung, which will then irradiate the tissues directly by their alpha-radiation. This gives importance to radon. The natural radioactivity presents an effective dose for the humans around the world that is 2.4 mSv yearly. Half of it is due to the radiation of the radon and its daughter product. This is the second most frequent death reason in the USA after the lung cancer. Therefore to measure the radon level and discovering high radon activity areas is a useful impact to the society.

For these purposes three laboratory practices cover this field. The first one shows how to determine the indoor air radon concentration, the second one demonstrates how to measure the radon in water. These two cases are where the radon can come close to people. More general motivation stands behind the exhalation measurement. That will give information of a soil type that is significant for an area and helps to decide if the area should be investigated more from this point of view. Or in other cases a small area, a building site can be investigated, and the danger from the radon levels can hold off. These are basic methods and several specified research projects can count on these.

17. Tritium content of water samples (TRI)

17.1 Introduction

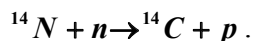
There are several natural radioactive isotopes in our environment. The most well-known ones are the uranium and the thorium that can be found in the soil everywhere, and their daughter products. The half-lives of the most abundant isotopes of the uranium and the thorium are 4.4 billion and 10 billion years. These are longer than the age of the Earth. At the creation of the material of our Earth there were several radioactive atoms but most are decayed into stable isotopes already a long time ago. The long half-life isotopes such as ^{235}U , ^{238}U , ^{232}Th , ^{40}K , to mention the most important ones, survived. These isotopes are in several types of rocks, and they appear in every kind of soil. Their concentrations are sometimes small, but sometimes can be higher, too. These are the first group of natural radioactivity. But there are other radioactive atoms in our environment, different from these ones. The reason how the others can appear is that those are produced continuously. The main source of it is the cosmic radiation. The fast protons coming from the Sun can enter the upper atmosphere and they can collide to the nuclei of nitrogen and oxygen atoms that are available in O_2 and N_2 molecules. If the energy of the colliding proton is not very high in the collision, the protons and the neutrons will be the participants. The proton can hit a neutron out of a nucleus for example, that will make a neutron flux at the upper atmosphere. Several reactions can happen and many isotopes are produced. If their half-lives are long enough, those can mix up in the whole atmosphere taking part in the great circulations, and then these isotopes can spill out from the air to the ground with the precipitation. These isotopes can accumulate homogeneously in the atmosphere. Their equilibrium concentration is determined by their production rate namely the cross section of those reactions and their half-life.

The third group of naturally occurring ionizing radiation consists of the cosmic showers. The cosmic protons have many times higher energy than a typical energy of a nuclear physics reaction. In this case particle physics will govern a process. The protons can interact to a smaller part of the colliding nucleus. The quarks inside will be the participants of the reactions. In these reactions new particles can appear. In a proton nucleus collision a new particle, for example a pion, can be created. The pions then will decay into muons. The muons are like the electrons, are elementary particles but they have about 200 times larger mass. These muons will go through the atmosphere and reach the surface even though they enter the ground and will slow down there. During these processes new photons and electrons can be produced in a cascade process. These are the cosmic showers. Their extent can be several hundred meters on the surface.

17.2 The radiocarbon and the tritium in our environment

17.2.1 *The origin of the tritium and the radiocarbon*

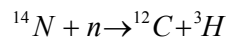
We saw that neutrons can be produced in the upper layer of the atmosphere. These neutrons hit another atomic nucleus. Mainly oxygen and nitrogen nuclei are available. If a neutron hits a ^{14}N nucleus it can hit out a proton and it can build into the place of the proton.



The short description of this reaction according to the nuclear physics tradition: $^{14}\text{N}(n,p)^{14}\text{C}$. This is the way in which the radiocarbon (the radioactive isotope of the carbon element) ap-

pears in the Earth, first in the upper layer. But its half-life is 5730 years and this time is very long compared to the time scales of the circulation and the diffusion, convection that happens in the atmosphere. This radiocarbon will be built into the live organisms, cells. For example the photosynthesis builds the CO₂ into the body of the plants, so they put ¹⁴C into the plant. The animals will eat that and in the food chain these isotopes can be mixed into every kind of living object. The radiocarbon is present in its equilibrium concentration in, for example, trees or in the human body as well. Although the radiocarbon decays in the bodies but while those are living they will take it also from the environment. At last an equilibrium concentration will be reached. This equilibrium level has not been time independent for the last 50 years due to the artificial production of the ¹⁴C as a result of the nuclear weapon tests. However, as a good approximation one can use the ratio, that is $^{14}\text{C}/^{12}\text{C}=10^{-12}$.

The neutron in the upper atmosphere can collide to the nuclei in different ways, too. They can hit the same ¹⁴N nucleus but it can pull a proton and a neutron out with itself. Several other reaction mechanisms can be imagined, but these two have the highest probability, because in these very stable nuclei will be formed. In these reactions the electric charge and the number of nucleons (proton number and the neutron number) are conserved. Before the reaction there were a ¹⁴N nucleus and a neutron present, where the sum of the electric charge is 7 times the elementary charge, and the sum of the nucleon number is 15. After the reaction the carbon has only 6 elementary charges and the tritium has one. The nucleon numbers are 3 in the tritium and therefore the A=12 isotope of the carbon will be formed, which is very stable isotope.



The atomic nucleus which consists of 1 proton + 2 neutrons is called triton (this is the nucleus of the element tritium). This is the A=3 isotope of the hydrogen. Its half-life is 12,3 years, which is long enough to participate in the environmental processes. This can be built into the live bodies similarly to the ¹⁴C, and there is an equilibrium concentration of it. The tritium can appear in the water molecules mostly, but not exclusively. It can reach the surface waters like ocean, rivers, and lakes. But the subsurface hydrogeological flows have so long time scale (more than thousands of years), so in those the tritium will decay to zero. This is a big difference, and a possible method to determine whether the water sample comes from a long hydrogeological flow or its origin is instead from a fresh precipitation. The equilibrium ³H concentration of the surface rivers and rains is less than this for the ¹⁴C, since the tritium decays faster. Its ³H/H ratio is 10⁻¹⁸.

17.2.2 Environmental activities of tritium and radiocarbon

In subsurface water flow the tritium activities are falling according to the exponential decay law. After 12.3 years the surface equilibrium concentration will be only the half of it and there is no way to supply it. Generally after t time the initial activity A_0 will fall to:

$$A(t) = A_0 \cdot 2^{-t/T} = A_0 \cdot e^{-(\ln 2/T) \cdot t},$$

where T is the half-life of the tritium. There is no supply alike the case of the subsurface flow, if we close the water into a bottle. In this way the natural equilibrium concentration that is formed in grapes and surface waters will be smaller and smaller as the time goes by. Therefore when the wines are closed in a bottle the tritium activity falls according to the rule above. If we determine the tritium activity on the other hand, we can tell the age of the wine in a time scale of 1-50 years. The older liquids generally contain so little amount of tritium that only with expensive precision methods can its concentration be successfully determined.

For the radiocarbon the equilibrium activity concentration e.g. in a tree is 0.154 Bq/kg. If the tree dies the radiocarbon activity starts to fall exponentially, since there is no supply. This is how the age of a tree can be determined besides the counting of the annual rings. In the last 60 years the nuclear tests modified these radioactivity concentrations, unfortunately.

The radiocarbon chronology is good in the time range of thousand years, until about 30 000 - 40 000 years. Therefore the radiocarbon chronology gives information about the age of historical events.

The tritium and the radiocarbon are present in the human body according to the mentioned equilibrium activity. Therefore everyone is slightly radioactive. In the radioactive decay of these isotopes small energy electrons leave the decaying nuclei, and these electrons will be stopped in some micrometer. The cells absorb ionizing dose in this process, many molecules in the cell will be ionized. This radiation will not leave the body. Since the origin is from inside the body it is called inside dose. This has been going on for millions of years since the life appeared on Earth, and it is a part of the functioning of the body. The civilizations although are capable to increase this natural level of radioactivity. Therefore there are strict rules for the allowed radioactivity concentrations in water. This is also a check for illegal nuclear activities around the world. The International Atomic Energy Agency runs a worldwide monitoring system that checks the radioactive levels continuously.

17.2.3 The decay of the tritium and the radiocarbon

The tritium and the radiocarbon as we already mentioned decays by beta-minus-decay. In both nuclei there are more neutrons than protons, and there are more neutrons than in the most stable isotope of the element. In the stable hydrogen the nucleus is a single proton. In the tritium there are 2 more neutrons. The most stable carbon has 6 neutrons, while the radiocarbon has 8. In the beta-decays of these isotopes 1 neutron will transform to a proton in a beta decay that is governed by the weak interaction. The weak interaction is one of nature's elementary interactions. In this process an electron (or its antiparticle the positron) and one kind of neutrino will participate. In the beta-minus-decay the neutron is converted to proton and the nuclear structure will be more effective, the final nucleus has larger binding energy after the neutrons and the protons realign. During the neutron \rightarrow proton conversion the mass number of the nucleus will not be changed, but the atomic number increases by one, since there will be one more proton. (In the other type of the beta-decay, in the beta-plus-decay, the case is opposite, the proton converts to neutron.)



In the neutron \rightarrow proton conversion according to the conservation of the electric charge there should be a negative charge to be released. This is the electron, and due to the main features of the weak interaction an anti-neutrino will be also created. It is interesting that in this process the energy released in the reaction will be divided into three ways. One part will go to the remaining (daughter) nucleus, one part goes to the electron, and the third goes to the antineutrino. The first is negligible, since the nucleus has large mass compared to the light particles like electron and neutrino. The electron and the anti-neutrino will share almost the total decay energy. The decay energy was determined already in the 20th century by mass spectrometers. The decay energy can be calculated from the masses:

$$Q = m({}^3\text{H}) - m({}^3\text{He}) - m(e^-) = 18.6\text{keV}$$

$$Q = m({}^{14}\text{C}) - m({}^{14}\text{N}) - m(e^-) = 156.5\text{keV}$$

The maximum electron energy (when the neutrino does not get any) is 18.6 keV and 156.5 keV, and both are less than the rest energy of the electron, that is 511 keV. This is another contribution why we call these small energy decays.

17.3 Measurements of tritium in water

17.3.1 The tritium activity

The amount of tritium can be shown according to its radioactivity. For radiocarbon it is also a well-known widely used method, but in that case the isotope separation by mass spectrometers is also an established method. If we determine the radioactivity of a sample and we know the half-life of it, we can calculate the number of radioactive atoms in the sample according to the following formula:

$$A = \lambda N = \frac{\ln 2}{T_{1/2}} N.$$

17.3.2 The energy distribution in beta decay

As we mentioned the electron and the anti-neutrino share the energy released in the decay. We assume that this share is a random process (in this study). Every velocity vector of an electron is equally probable and the corresponding direction of the anti-neutrino is also random. Therefore there is a given probability for each kinetic energy of the electron. The bases of this distribution lead to difficult physics, so we do not go into the details now. Instead we will measure this distribution. There are general features of these distributions. When the neutrino gets 0 energy, the electron brings the total. When the electron gets zero all the total decay energy will be carried out by the anti-neutrino. The distribution has its maximum at about 1/3 of the maximum energy.

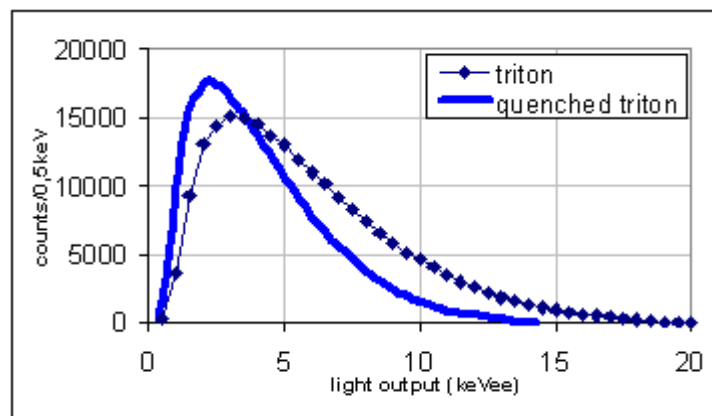


Figure 17.1. The light output distribution of the tritium beta-decay. The squares were used for the case of the unquenched (see details later) tritium standard, and the continuous line is for a real quenched sample.

Figure 17.1 shows the light output distribution in two experimental cases. The light output is a quantity that is proportional to the kinetic energy of the electrons. We will discuss it later. The shape of this distribution is characteristic for the beta-decaying isotope, and is useful to identify it if we do not know which isotopes cause the radioactivity of the sample.

17.3.3 The detecting intensity

In this laboratory practice and in the general practice, too, the experimental question is the tritium activity concentration of a sample. Generally the required unit is Bq/l. From this we can tell the number of tritium atoms in the sample using the $A=\lambda N$ formula, after determining the activity. We should determine how many atoms decay in a second, and we should identify the element itself. Generally we cannot detect every decay that happens in the sample. Therefore the detected intensity differs from the activity. The loss of event contributes to the detecting efficiency, which is $\eta=I_d/A$, where I_d is the detecting intensity, or in other words CPM, counts per minutes. The efficiency does not depend on the activity but can depend on several factors. For example it depends strongly on which isotope is used, what kind of detector is used, and its parameters are also important. If we give the activity in an unusual unit like DPM, disintegration per minutes (decay per minutes), and the efficiency gets this form:

$$CPM = \eta \cdot DPM$$

For a tritium measurement our goals are to measure the CPM and the efficiency.

17.4 The liquid scintillation counting techniques (LSC)

The ranges of the released electron in the beta-decays of the radiocarbon and the tritium are so short, that these electrons cannot penetrate even a thin aluminum sheet. But in a general case most of the detectors are covered, they are in a box. The best solution is about the aluminum box, but that is still not enough for the tritium and for the radiocarbon case to let the electrons go through. Therefore the sample cannot be outside the sensitive volume of the detector. It is better to mix up the sample with the detector material. The liquid scintillation technique gives just this possibility. The sample will be mixed up with a scintillator material, in this case it is in liquid phase.

The instrument that will be used is the TriCarb 1000A machine, designed for liquid scintillation analysis of triton and radiocarbon samples. It is also applicable for determination of other isotopes like dissolved radon, or phosphorus in water.

17.4.1 The scintillation detector

The scintillation detectors consist of two parts. One is the scintillator material and the other is the photomultiplier tube (PMT). The scintillator is a material, the molecules of which can emit visible light after excitation. In our case it will be liquid. The scintillation is the process when the scintillator material emits visible or UV photons just after one radioactive decay. The other part, the PMT is used to convert this light to electric pulse. Then an electronic system will analyze the pulses and for example will count them.

The scintillator

The electron that was created in the beta decay will lose its energy quickly in such a way that it will ionize or excite the atoms and molecules of the surrounding medium. The excited atoms or molecules will return to their ground state or lower excited states in some time via emitting the visible or UV photons. The number of excited atoms or molecules can be described with an exponential decay law, if we assume that the excitation process is very fast, and the decay is slower. The characteristic time is called the decay time of the scintillator.

The light output is the number of photons emitted and got out of the scintillator material in one radioactive decay N_{LF} . The number of the created photons can be smaller N_{CR} . But for the first case $N_{LF}=N_{CR}$. We have seen that the light output is an important parameter.

The ionizing particles (e.g. from a radioactive decay) will lose kinetic energy in the scintillator. This is called E_d , and it is proportional to the number of excited atoms or molecules in the scintillator, this will be proportional to the number of scintillation photons, which is the light output itself. Therefore the light output is proportional to the energy deposited in the detector material. Generally in beta decays E_d equals to the kinetic energy of the electron. (If the electron is formed at the very side of the scintillator volume, it might travel out of the volume and in this way some of its energy will not excite the scintillator but the wall of the detector.) If we measure the light output we can determine the maximum kinetic energy of the electrons in the beta-decay, and therefore we can identify the isotope.

If there is an electron that deposited 100keV energy in the scintillator, it cannot be from the decay of the triton, since its maximum energy is 18.6 keV. That event can be a background event or from a different isotope.

The unit of the light output: The light output is measured relatively to the number of the scintillation photons of a special case. The light output is in fact not a number, but a ratio. The special case is the case of 1 keV electrons. The light output tells how many times more photons were emitted in decay than it was emitted in the case of a 1 keV electron.

$$L=N_{LF}/N_0 .$$

Here L is the light output, it is a dimensionless quantity, but we use keVee for this, N_0 is the number of scintillation photons from 1 keV electron case. The keVee is called kiloelectronvolt electron equivalent. This unit is material-dependent. But the L itself is not! A 1 keV electron will create different number of photons in one scintillator than in another. In the case of alpha-decays L is proportional to the alpha energy but there is a conversion factor, since we defined the L for electrons. This is called the light emission efficiency. For alphas the ionizing process is different and the excitation mode of the molecules is also different. That is the reason why there will be much less visible photons after exciting the medium by fast alpha particle. As a summary we can write:

$$E_{electron} \sim E_{deposited} \sim N_{excited.mol.} \sim N_{LF} \sim L$$

In a general scintillator $E_d=1$ keV will produce 7 visible photons. The light production is a statistical effect, so it has deviation. Two exactly identical electrons will create slightly different number of photons.

For detection of the soft beta-decays we select liquid scintillator that can be solved in the sample (mainly water). In this case the detector material will cover the whole cone around the decay, and can detect very small energy electrons. The liquid scintillators consist of three components, and therefore they are called cocktail. The largest part of the cocktail is the solvent. This phase will absorb the kinetic energy of the ionizing particle and emit photons of which the wavelength is not optimal. Therefore there are two more components in the cocktail that absorb the photons from the solvent and reemit the energy by longer wavelength photons. We call them wavelength shifters. The first is the primer scintillator and the other is the secuder scintillator that emits already at the optimal wavelength range, which is determined by the photocathode features of the PMT. The primer and secuder scintillators are in the cocktail in a very low concentration.

The scintillation photons will not necessarily go out of the cocktail volume. There is a process that is called quenching, which means the photons will be absorbed in the vial, or on the wall of the vial. So the number of the outgoing photons is $N_{LF} < N_{CR}$. The part that survived the absorption is $e = N_{LF}/N_{CR}$, but the quenched part is $q = 1 - e$.

The quench can be resulted in many processes. The liquid can be colorful, it can contain contamination that absorbs the light, there can be contamination on the glass of the vial, there are several ways how a light absorber center can be involved in the detection mechanism.

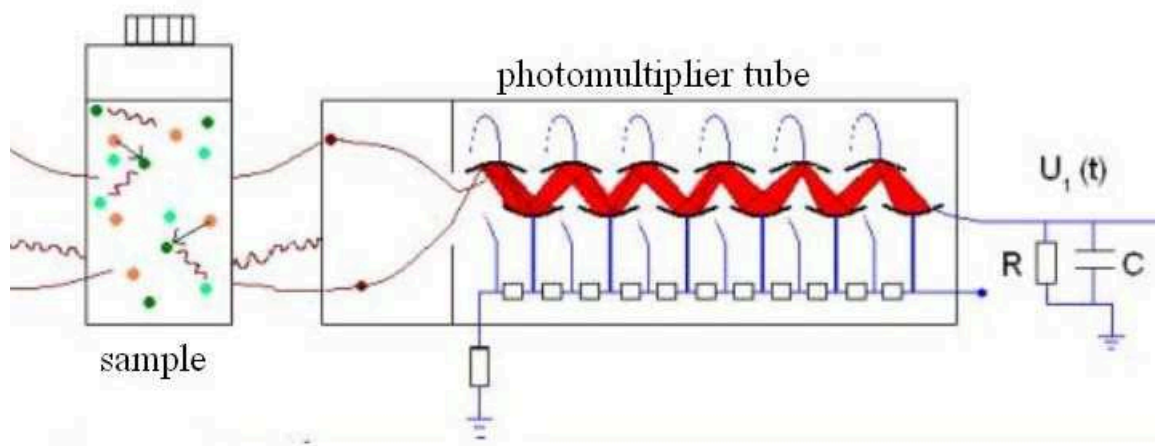


Figure 17.2 The photomultiplier-tube (PMT). On the left side there is vial containing the cocktail and the sample solved in it. The excited atoms are shown and the lines symbolize the emitted scintillation photons. In the left wall of the PMT is the photocathode, where a photon kicked out an electron. That was multiplied by the dynodes. $U(t)$ shows the electric signal.

The photomultiplier tube

The PMT will convert the light output into electric signal. One part of the scintillation photons will enter the photocathode ($N_{FC} \sim N_{LF}$) of the PMT and hit a photoelectron out in a photoelectric process. Every 5th photon will do it generally. The number of photoelectrons taking part in a single decay is N_{PE} . The photocathode is a very thin metal layer that is evaporated on the inside part of the glass of the PMT window. There is vacuum in the PMT, the photoelectrons will be focused into a metal layer, so-called dynode. The dynodes are special metals. If a “fast” electron hits its surface there will be about 3 slower electrons flying out. Then the slower electrons will be accelerated by electric voltage of about 300 V, and the multiplication starts over. In a PMT there are 10-12 dynodes, and those can multiply the number of electrons up to about 3^{12} . The last metal layer is called anode, this will not multiply more the number of electrons, but lead them out of the PMT, the number of electrons that went through the anode is N_A . These electrons will go to a ground potential point across a resistance. There is always scattered capacitance, which also should be taken into account. The short signal is about 1 μ s long, or sometimes much shorter. So the average electric current that flowed through the resistance and was caused by 1 photoelectron is $I_I = N_A \cdot e / \Delta t = 3^{12} \cdot e / 10^{-6} \text{ s} = 1,5 \cdot 10^{12} \cdot 1,6 \cdot 10^{-19} \text{ Cb/s} = 0,24 \mu\text{A}$. If there are more photoelectrons created this current is proportionally higher: $I = N_{FE} \cdot I_I$. Some photoelectrons already give measurable current. While the multiplication factor of the dynodes is constant the amplitude of the electric signal is proportional to N_{FK} ,

therefore to L as well, and so to the energy of the beta electron. To be able to accelerate the electrons in the vacuum of the PMT we have to apply about 2000-3000 V high voltage on the PMT between the anode part and the photocathode.

The scattered capacitances and the resistance, and also the decay constant of the scintillator determine the time dependence of the electric signal on the resistance. The time shape therefore does not change during the detection. Only the amplitude can be different according to the N_A , or L , which is proportional to it. The constant shape of the signal is a very important feature. This makes the proportionality possible. Taking every step into account the energy of the beta electron is proportional to the amplitude of the electric signal. Digital electronic technologies, very much developed in the last decades, are applied, and an integer number will be determined which is proportional with the electron energy. This number is called the channel number. This can be considered as particle energy but in fact it is proportional to L , the light output.

17.4.2 The measured light output spectrum

In 17.2.2 we described that the energy distribution of the electrons shows a continuous curve, since the electron and the antineutrino share the energy. This is sharply different from the energy distribution of the alpha decays. Where there is no 3rd particle created in the reaction, so the decay energy will determine the kinetic energy of the alpha particle. We get 1 sharp peak at this energy in the alpha decay case.

The light output distribution, in fact the distribution of the channel numbers, that is proportional to the light output, we also can call it amplitude distribution. This distribution will be measured using a multi-channel analyzer (MCA). Above we saw that for every decay the instrument calculates a channel number. If we have a lot of decays, a lot of events, we have a lot of channel numbers. The MCA calculates how many times each channel number happened in the measurement. So it means how many times each electron energies occurred among the beta decays of sample.

The light output distribution should be investigated in every measurement. That will allow us to identify the isotope and/or measure the quenching quantitatively. The effect of the quenching on the energy spectrum can be seen in [Figure 17.1](#). If the quenching reduces every light outputs by a factor of q , the channel numbers will be smaller, and the distribution will be like to which is pressed left along the x-axis in a linear way. The quenching (q) can be measured by a gamma source that should be put aside of the vial. The quenching will modify the known and constant light output distribution of the Compton electrons of this source. The TriCarb machine quantifies a comparison to the standard. This is the tSIE (transformed spectral index) value or quench that we can get for each sample. The value 1000 means the unquenched case and the heavy absorption results in very small tSIE values.

17.4.3 Determination of the efficiency

The investigated samples have different quenching properties. Therefore they have different counting efficiency. The higher the quench the smaller the efficiency is. There is a built-in method, the tSIE parameter, which is useful for determination of the efficiency. There was a calibration measurement when one tritium sample of known activity was mixed with solutions of different quenching properties. The tSIE values (see below) and the efficiencies $\eta = \text{CPM}/\text{DPM}$ were determined and it resulted in a calibration curve for the efficiency. We

should use this curve that is valid on the other hand a given cocktail: ULTIMA GOLD, which is the most frequently used scintillator for tritium. The background of this method using tSIE parameter is the following. We should measure the known light output distribution in two light output intervals: one at low light outputs and one at higher light outputs. In a measurement the counts in the lower light output range are called N_{low} and in the higher range are N_{high} . For an unquenched sample the ratio of the counts $R=N_{high}/N_{low}$ is a given number. But this ratio will decrease in case of quench. The counts from the higher light output interval will move to the lower light output range. The tSIE value will also be smaller if quench goes high.

The method to determine the efficiency in our case is the following: We have to find a mathematical function to describe the known tSIE – efficiency calibration points. Then tSIE should be measured, and at last efficiency will be calculated using the function and the measured value of tSIE.

17.4.4 Operating the TriCarb machine

Our liquid scintillation machine has two PMT-s in it. We detect the events if there is a coincidence between those PMTs. The vial that holds the liquid sample will be lifted down to the place just between the two PMTs. Lead sheets shadow this area to allow lower background. The machine communicates with the users by a small LCD window. There are command buttons. We need only the button called COUNT that starts the measurements. Before starting the measurement place the vial into the right position and then press COUNT. If this button will be pressed once again the machine cancels the measurement and gives the vial back (lifts it up, out of the inside area).

On the small LCD monitor we can follow how much time elapsed and what the CPM value of the measurement is. After the measurement the machine transfers each channel number to a computer to draw the light output distribution. It also sends the most important values to a printer.

The machine can differentiate 2 regions according to the light outputs. Region A and B can be programmed. The lower and upper light output limits should be given in keVee units. The general setup for tritium is: Region A – lower L=0 keVee, upper L=18.6 keVee, Region B – lower L=2 keVee, upper L=18.6 keVee. The ratio of A and B regions immediately gives information on the quench. For radiocarbon a different setup is necessary. Region A: lower L=0 keVee, upper L=145.6 keVee

17.5 Lab course tasks

1. Calculate the tritium and radiocarbon activity of an idealized human body!
2. Determine the counting efficiency of an unquenched tritium standard in the A and B regions! Use the initial activity information from the top of the vial.
3. Measure the tritium standard n times in standard ^3H protocol. $n=10-20$. Determine the standard deviation of the CPM values! Compare this with the uncertainty printed by the machine. That is calculated from the \sqrt{N} error estimation!
4. Investigate the light output curves of the tritium samples!
5. Determine the tritium content of an unknown sample!
6. How can we determine the radioactivity of tritium and radiocarbon in an unquenched sample if both isotopes are present in the sample? (We have ^3H and ^{14}C standards as well.)

17.6 Test questions

1. Why do we use liquid scintillator for detecting the decay of triton?
2. What process is called scintillation?
3. What does a liquid scintillator contain?
4. What is the light-output and how does it change in case of quenching?
5. How do the radiocarbon and the triton decay?
6. How do the radiocarbon and the triton is created in our planet?
7. Why is the beta-spectrum continuous?
8. What is the spectrum of the amplitude of the electric pulses during a triton measurement?
9. How can the scintillation light be quenched?
10. How can we determine the efficiency of a triton measurement from the measured value of the quench?
11. Why is the amplitude of the electric signal proportional to the energy deposited in the vial?
12. What are the operating purposes of the cocktail ingredients?
13. How does the photomultiplier work?
14. How can we discriminate the triton counts from the radiocarbon counts?

18. Investigation of the Cherenkov-radiation from potassium beta decay (CSE)

During this laboratory practice we will determine the concentration of such elements in water solution that have naturally occurring beta-emitting isotope. The basis of the determination is the Cherenkov-radiation of the fast electrons emitted in the beta decay, with which the decays can be detected individually. The Cherenkov-radiation is a small intensity visible and ultra violet radiation, therefore it can be absorbed easily in colorful solutions. This gives a second possibility to determine the concentration of colorful solutions using additional natural beta-emitting radionuclides. The method has several applications, in which generally ^{40}K is added to the sample.

This is a cheap way of the determination of potassium concentration if the medium is appropriate for this. Also the concentration of quenching materials can be determined using additional potassium with this technique.

18.1 The Cherenkov-radiation

18.1.1 The phenomenon of this radiation

Cherenkov-radiation is created when a particle moves with such a high velocity in a medium (e.g. water or solutions) that is faster than the velocity of the speed of light in the medium. In our case the particle is an electron. The speed of light in vacuum is $c=300\,000\text{ km/s}$, in water this is smaller according to the refraction index: $v=c/n$, that is about $225\,000\text{ km/s}$.

If an electron moves with $270\,000\text{ km/s}$ velocity in vacuum, and it enters a water body, it can not slow down immediately since it has finite rest mass and it can not have infinite acceleration. So after entering the water first it should slow down. During this fast movement of the electron we should apply the laws of the theory of special relativity. We therefore summarize these laws:

a) One of the theoretical postulates of the theory of relativity is that the velocity of a particle cannot be faster than c , the speed of light in vacuum. That is a maximum velocity for ordinary particles appearing in our real life. There is a consequence of it: the addition of velocity vectors (those are close to the speed of light) will be very unusual.

b) Another theoretical principle established a new thinking on the concept of energy and mass. Those are not two different attributes but they are strongly related to each other. If something has energy it has mass and vice versa, too. These two earlier different features of a particle are proportional to each other. $E=mc^2$.

c) The third surprising principle of special relativity is how the total energy of a particle depends on its velocity. The concept of the rest mass, and the corresponding rest energy states that the particles have energy already if they are in rest: $E_0=m_0c^2$. This quantity is a universal basic property of a particle. For example every electron has a rest energy of 511 keV . This principle states the total energy will increase with the velocity in the following form:

$$E = mc^2 = \frac{m_0c^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Here v is the velocity of a particle, c is the speed of light in vacuum, m_0 is the rest mass of the particle. The kinetic energy of a particle can be calculated from this. The kinetic energy is a part of the total energy of a free particle, which is above the rest energy:

$$E_{kinetic} = mc^2 - m_0c^2 = \frac{m_0c^2}{\sqrt{1-\frac{v^2}{c^2}}} - m_0c^2 = m_0c^2 \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1 \right)$$

18.1.2 The threshold energy of the Cherenkov-radiation

That electron which moves just with $v=c/n$ velocity in a medium has total energy E_{th} . E_{th} can be calculated from the formulas above:

$$E_{th} = m_0c^2 \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1 \right) = 511keV \left(\frac{1}{\sqrt{1-\frac{(c/n)^2}{c^2}}} - 1 \right) = 511keV \left(\frac{1}{\sqrt{1-\frac{1}{n^2}}} - 1 \right)$$

An electron that has more energy than this moves faster than the speed of light in the medium, therefore it can create Cherenkov-radiation. Electrons, which are slower than this, cannot emit that radiation. E_{th} is called the threshold energy of the Cherenkov-radiation. In the case of $n=1,33$ refraction index this value is 252 keV. There are several beta decays in which case the maximum energy of the emitted electrons is larger than this threshold. So if that decay happens in water, then Cherenkov-radiation will occur. The Cherenkov-radiation is a mixture of electromagnetic radiation in the range of blue visible and ultra violet light. According to the emitted photons that require energy the electron slows down, and when it reaches the speed of light of the certain medium, the Cherenkov-radiation falls dead.

18.1.3 Examples for Cherenkov-radiation

The Cherenkov-radiation was discovered in the 1920-s by Pavel Cherenkov. He examined gamma emitters in solutions, and surprisingly he observed low intensity blue light. In that case the fast electrons arose in Compton scattering of gamma-radiation.

We can observe this radiation also when the muons of the cosmic radiation go through water. Those muons have negative charge, their speed is much higher than c , generally, in the cosmic showers. In oceans and seas sensitive light detectors can observe it. There are large experiments instead where big tanks are filled up by water to detect cosmic particles according to their Cherenkov-radiation using photomultiplier tubes.

In Research Nuclear Reactors one can see the Cherenkov-light. In these reactors the active zone is shadowed by more than 5 meters of water, which is open towards the top. This amount of water absorbs the neutrons, gamma rays and ionizing radiations but the visible light can go through. The beta electron emission of the fission products will result in Cherenkov-radiation, and it is easily observable.

Several isotopes are used for elemental analysis or biological tracing according to their Cherenkov-radiation. These kinds of isotopes are ^{56}Mn , ^{32}P , ^{40}K , ^{86}Rb . They can be used for investigation of plant transport processes or the phosphorus content can be measured using Cherenkov-method.

18.1.4 The features of the Cherenkov-radiation

The Cherenkov radiation is emitted not by the moving electron but the media! The charged particle will polarize the dipoles of the medium. The moving charged particle will rotate the dipoles to its direction. When the particle moves the dipoles direction is changing. The vectors of the dipoles are following the position of the particle. This can be the case when the particle goes slower than the speed of light in the medium. If it goes faster the dipoles cannot follow the particles, since the electromagnetic interaction propagates also with the speed of light (first principle) it cannot propagate with an infinite velocity. While the interaction reaches the location of the dipoles the electron moved ahead with faster speed than the interaction propagated. At last the dipoles will have an aligned motion, and the electron is the “conductor”. This aligned motion is in fact a motion of electric dipoles, when they have acceleration. So it is not surprising that the dipoles will emit electromagnetic radiation.

The angular distribution of the Cherenkov-radiation is very special. The Cherenkov-photons are not isotropic. They will be emitted along a cone, centered at the direction of the electron. In Figure 18.1 we can see the photons (red vectors) and the wave front which is also a cone, but on the other direction. The circles demonstrate the Cherenkov-light starting at one point at one moment. The particle goes faster than the radius of the circles. This makes the shape of the wave-front be a cone. Considering that the radii of the circles are ct/n starting from a center of them (blue line), and the displacement of the electron is vt , for the angle of the cone we can write the following formula:

$$\cos \alpha = \frac{c/n}{v} = \frac{c}{vn}$$

Here c is the speed of light in vacuum, n is the refraction index and v is the velocity of the electron, as before. In the figure the photons started from the center of the first circle, other photons started later, therefore those circles have smaller radii.

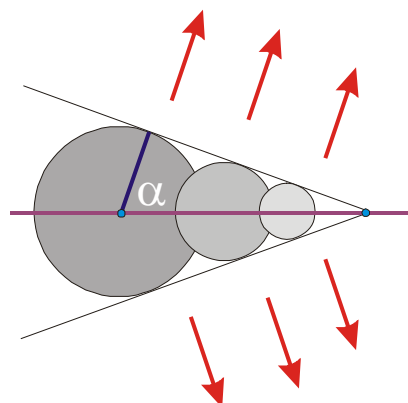


Figure 18.1. The angular distribution of the Cherenkov-radiation. The red arrows show the direction of the emitted photons, and the solid lines that show a cone are the wave-front of this radiation.

18.2 Detecting Cherenkov-radiation due to beta decay of ^{40}K

The Cherenkov-radiation can be detected well using photomultiplier tubes (PMT). Therefore the techniques for scintillation detectors are useful for this purposes as well. During our laboratory practices we will use Tricarb 1000 liquid scintillation spectrometer, which is a scintillation detector system. The instrument can measure scintillation photons or Cherenkov-photons coming out of a 20 ml vial. Two PMTs detect the light in coincidence. To measure the Cherenkov-radiation of a beta-emitting isotope we only need water as detector material. So the method is cheap.

18.2.1 The efficiency of counting Cherenkov-photons

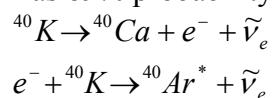
In our experiment we detect the Cherenkov-photons from the decay of ^{40}K . There are many electrons emitted in beta-minus decay that have higher kinetic energy than the threshold. But we cannot detect all of them. There are several reasons of this. The main loss is the coincidence requirement. Since the Cherenkov-photons are emitted along a cone, they will travel around 1 direction. The photons will not start towards both PMTs, which are placed opposite to each other. It is even surprising that there are coincidences from Cherenkov-radiation. We still detect them because of the photon scattering. Visible and UV light can be scattered on the material of the vial and also on the molecules or bigger aggregates in the sample. Adding scattering material will enhance the counting efficiency, but after a while it can absorb the photons as well.

Another contribution to the efficiency is that not all the electrons will have higher energy than the threshold. The beta electrons have a special distribution. At zero electron energy the probability of the emission is small, and so is at the maximum electron energy. There is always a part below the threshold. Further aspect is that the photons entering the PMT will not necessarily create a photoelectron.

There are naturally occurring beta-emitting radionuclides. The ^3H and the ^{14}C have low maximum energy 18.6 keV and 156 keV. Both are below the Cherenkov-threshold in water. For these isotopes we cannot expect Cherenkov-radiation at all.

18.2.2 The Cherenkov-radiation of the ^{40}K

The ^{40}K is a naturally occurring radionuclide that has a 1,2 billion-year half-life. The radiation of the ^{40}K is an important part of the natural radioactivity. In an average 70 kg human there are 4000 decays of ^{40}K in every second. This is because the potassium is a frequently occurring element and the relative abundance of this isotope is large: $1,17 \cdot 10^{-4}$. The ^{40}K can decay in two ways. One is the beta-minus decay when ^{40}Ca is created. The other way is the electron capture, when a K shell electron will interact with a proton in the nucleus and ^{40}Ar will be resulted. The electron emission has 89% probability, the electron capture has 11%.



The reaction in the second line describes how the argon in our atmosphere were produced. The star in the superscript denotes that the argon nucleus is in an excited state after the decay. This excitation will be soon lost by emitting a gamma-photon of 1461 keV energy. The upper line describes the emission of fast electrons. The maximum energy of the beta decay in this

case is 1320 keV, this is higher than the Cherenkov-threshold in water. Therefore in one part of the 89% there is a possibility for emission of Cherenkov-photons.

18.2.3 The light output distribution of the Cherenkov-photons from the decay of ^{40}K

The number of Cherenkov-photons according to certain decay is called: N_C . One part of these will enter the PMTs window after scattering or directly. These photons are called light output in this case. (In a general scintillation process the number of emitted photons is the light output.) In a given medium N_C is proportional to the light output and to the height of the electric pulse, which is created in the PMT for every single decay. The TriCarb instrument will digitize the amplitude of the pulse and denotes it for every decay. Therefore we can investigate the distribution of the light output as well.

The higher the velocity of the electron is the higher the energy transferred to the medium and the higher the light output is. The N_C is proportional to the light output in general, but not strictly. The light output depends namely on the direction of the electron. If the electron goes perpendicular to the line between the two PMTs, the probability to detect the photons will be different from the case when the electron started towards one PMT. The second case has less probability.

The amplitude of the light output is a relative number. The instrument determines the amplitude of the electric pulse of the PMT as it is proportional to the light output, and then compares it with a case of a reference scintillator. The unit is the number of scintillation photons that is created by an electron of 1 keV energies. This unit is called keVee. Then the light output gives the number of photons relative to this unit. Although the ^{40}K emits an electron of 1320 keV, but 1320 keVee light output would be measurable in case of a scintillation detector. Now in our case only Cherenkov-radiation makes photons, and therefore the number of photons in a single decay is much less. A large part of the energy is converted not to Cherenkov-photons. According to our experience with TriCarb 1000 the maximum light output for the Cherenkov-radiation of ^{40}K is less than 18 keV. That means the instrumental setup for ^3H isotopes (their maximum energy is 18.6 keV) is applicable for Cherenkov-radiation of ^{40}K .

18.2.4 The Cherenkov-background

If we put only distilled water into the vial and no beta-radioactive material, we still can measure Cherenkov-photons and can count a small level of background. These events are mainly also from ^{40}K isotopes, but those are outside the vial. The walls of the room, the material of the detector always contain potassium, therefore ^{40}K as well. The 1461 keV gamma photons from the decay of ^{40}K can go through thick materials and can produce a Compton effect on an electron in the vial. The electron gets some energy, which can be as high as 1243 keV (the Compton-edge of the 1461 keV gamma-photons). It is also possible that the gamma-photon hits an electron by photoeffect, when the electron gets the total 1461 keV energy. There are several ways to produce electrons of energy higher than the Cherenkov-threshold (about 252 keV). In this way the background gamma-radiation can produce events in the vial. Not only the gamma-photons from the ^{40}K but from other isotopes as well. For example the daughter isotopes of radon decay or thorium and uranium content of the building material. Our experience is that in this setup the intensity of the background is about 16 counts per minutes in 20 ml clear water.

18.3 Quantitative analysis using Cherenkov-radiation

18.3.1 Determination of KCl concentration

The intensity of Cherenkov-radiation in a given sample is proportional to the number of potassium atoms in the sample. If we measure the Cherenkov-counts per minute we can determine the potassium concentration. As an example if we know that the potassium is in KCl form we can determine the KCl concentration. The measurement uses a calibration curve. First we have to measure known concentrations in the same circumstances and then we can determine the Cherenkov-intensity of the sample and from this the potassium concentration.

18.3.2 The detection limit for ^{40}K

The general requirement for a determination now is that we want to detect twice as much counting rate than the background. If the sample has too few potassium atoms and its Cherenkov-intensity is too little, then we cannot tell the concentration with enough accuracy. The fluctuation of the background can be mixed up with our real counts.

To estimate the number of potassium atoms that will produce at least 16 counts per minute we have to know the probability of the detection of a ^{40}K electron, that is, for 100 ^{40}K atoms how many times we will be able to detect a Cherenkov-event. As an empirical number our experiences show that in ^{40}K decays the 8% of the cases will produce a coincidence. This is called light collection efficiency. In all the other cases only one PMT will detect a pulse, or neither of them.

Therefore we need 200 beta electrons per minutes. But not every ^{40}K decay will produce a beta electron, as we saw already. Only 89% of the ^{40}K decay. Therefore the minimum activity of ^{40}K is 225 decays per minute (DPM). Another factor is the electrons below the threshold. As an estimate we use 15% for this and we get 250 DPM for the minimum activity.

This is 4.2 Bq. Our aim is the potassium concentration of the minimum detectable amount in units of g/l. Therefore we calculate the number of ^{40}K atoms first, then the number of potassium atoms, and finally their mass. According to the formula: $N=A/\lambda$, we need $N=A \cdot T_{1/2}/\ln 2$ pieces of ^{40}K atoms in the 20 ml of sample. This means $N_K=A \cdot T_{1/2}/\ln 2/0,000177$ atoms of potassium. The mass of these: $m=40\text{g} \cdot A \cdot T_{1/2}/\ln 2/0,000177/N_A$

$$N=223 \cdot 10^{15}, N_K=1,9 \cdot 10^{21}, n=3,1 \cdot 10^{-3} \text{ moles}, m=0,126 \text{ g.}$$

This amount should be in 20 ml solution. Therefore the minimum detectable potassium concentration is 6,3 g/l. The potassium content of the mineral waters is generally below 500 mg/l. This method can be used only with a reduced background that is available with an advanced level of the instrument, which is more expensive. We also can improve our detection to use scintillation cocktails and normal scintillation methods, but that is also more expensive. An easier way is to improve the Cherenkov light collection efficiency by the use of a scattering material.

18.3.3 Investigation of colored solutions

We can determine the concentration of a quenching (absorbing) colorful liquid by its light-quenching feature. There are materials that are intense quenching ones. If we use mg/l concentration of them it is already possible to quantify their amount. We can use two kinds of

methods. We can have fast electrons from beta decay inside or we can put their gamma source outside the vial. For example if we put a general ^{137}Cs source just next to the sample, the 661 keV photons from the radioactive decay will hit the electrons in the sample, and we can have high enough energy electrons to create Cherenkov-light. From a known source in a controlled geometry a standard intensity of Cherenkov-light will be produced. Measuring the amount of its absorption we can determine the content of the quenching material. Of course a calibration is also necessary for these cases.

In this case the calibration curve will not be a straight line, so we have to be careful to measure several points to be able to follow the calibration curve. Then we have to parameterize the calibration curve.

18.4 Lab course tasks

18.4.1 Calculation type examples

1. Calculate the threshold energy of Cherenkov radiation for several refraction indexes, and plot it in a graph!
2. Estimate the percentage of the beta electrons that have less kinetic energy than the threshold. For $E_{max}=1320$ keV, $E_{th}=252$ keV and the electron kinetic energy distribution can be assumed to be a parabola, that is symmetric around 660 keV and its value is 0 at $E=0$ and at $E=1320$ keV, too!

18.4.2 Measurements

1. Determine the KCl content of an unknown sample!

Make a calibration series: vials with increasing but known KCl contents. Fit the data of your sample to the tendency of the calibration curve.

- a) Make solutions of KCl: Put the appropriate amount of KCl into distilled water to have 20 ml of 0.3 – 0.5 – 1 – 2 – 3 – 5 mol/l concentrations.
- b) Measure the Cherenkov-intensity of these samples 3 times in 2-minute measurements. Determine the average detected intensity and its uncertainty, then do the same for the unknown sample.
- c) Draw the calibration points on a graph and fit a line to that.
- d) Calculate the concentration of the unknown sample.
- e) Determine the detection efficiency!

2. Determine the amount of a quenching material of an unknown sample!

Our method is to have a given concentration of KCl solution that is well measurable. We can put known and different amount of quenching material to that solution. Determining the tendency of the quench we can find the concentration of the unknown quenching material from its quenching feature. The quenching material can be sodium dichromate or potassium permanganate as the most frequent cases.

- a) Make 6 pieces of known KCl solution. Into one of these do not put quenching material. Into 4 pieces put increasing amounts of quenching material. Put 20 ml of each into vials. Into the 6th solution put a given amount of the unknown sample.

- b) Measure the counting intensity of this quenching series. Make a parametrization for the data. Examine the possibilities to have linear behavior, which scale can be used for this.
- c) Calculate the unknown concentration of the sample.

3. Investigation of the light output distribution

- a) Determine the light output curve for ^{40}K solutions! Determine the maximum value!
- b) Measure the curve for a quenching material as well! Compare the results and make an interpretation!

18.5 Test questions

1. What emits the Cherenkov-light and in what case?
2. What is the relationship between the refraction index of the media and the speed of light in it?
3. What does the velocity threshold for Cherenkov-light production depend on?
4. Describe a few situations when there is Cherenkov-light produced in the nature!
5. What is the direction of the Cherenkov-radiation?
6. Is it good for us during detection?
7. We have 1000 decay of ^{40}K . How much Cherenkov-light can we detect (in average)?
8. Is a light scattering material good or bad for detecting Cherenkov-light?
9. What is the unit of the light-output?
10. What are the sources of the background of a Cherenkov-measurement?
11. What is the threshold energy of an electron to produce Cherenkov-photons in a medium that has a refraction index n ?
12. We have 1 mg potassium. How many electrons do appear in a minute?
13. How can we determine the potassium content of a mineral water using LSC technology?
14. What is the principle of the determination of the amount of a quenching material?

Reference:

Journal of Chemical Education 60, 682 (1983)

19. Radon measurements of indoor air (LEV)

The natural radioactive isotopes appear all around our living environment. Generally they are in the ground or solid materials (soils or rocks) but there is a radioactive noble gas that naturally appears in the indoor air because it can escape from the solid material and can diffuse out or get out by convection into the air of the soil then into the indoor and open air. The natural radioactivity of the air in the environment of people is mainly due to this noble gas, namely the radon.

People all around the world absorb doses of natural ionizing radiation. The world average of this is 2.4 mSv (see [Chapter IV](#) of this book). This is based on several studies in the last 30-40 years, and it was summarized in a report of the United Nation (UNSCEAR 2000). About 50% of this dose is because of the radon and its daughters! The radiation due to radon is the alpha, beta and gamma-radiation of the radon itself, but mostly of its daughters that enter the tissues of the lung. This was discovered when lung cancer appeared frequently among uranium miners. This is why the radon is important for the population. This dose is really important because there are people living in high radon environment. High radon environment can also have health effects that can lead to lung cancer as well. According to wide surveys around the USA and Europe there are about 20 000 people in every year who get lung cancer because of the high radon environment. And it always means high radon concentration of the indoor air. The rules of the radiation protection tell that small-absorbed dose can have health effect with small probability. The effect is statistical in nature. To recognize those areas where high radon concentrations can appear and find the houses that are in risk, that is the main aim of radon research.

The aim of this laboratory practice is to learn an experimental technique how to carry out radon measurements of air. We make an overview of the features of indoor radon.

19.1 Radon in houses

19.1.1 General features of radon and its decay chain

The radon is a member of the uranium decay chain. Therefore it naturally appears in the soil and rocks. The half-life of radon is 3.8 days. This is long enough time to diffuse out of the soil or from the building materials. That is why the radon is very important in environmental physics. The atomic number of radon is $Z=88$. This element has several known isotopes. However we call only the $A=222$ isotope as radon, for now. There is another very important isotope of $Z=88$. This is the ^{220}Rn nuclide that is called thoron. We will use the symbol T for this radon isotope. The other isotopes do not appear in environmental phenomena. The radon has 88 electrons in its closed electron shell. This electron shell can be deformed easily, for example in water. Therefore radon can be dissolved in water. The mass number of radon is 222, therefore it is a heavy atom. But it still can be mixed up in air. There are two processes, diffusion or convection. Both processes are important in case of the movement of radon in air or in soil gas.

The radon is created in the alpha decay of the ^{226}Ra radium isotope that has 1600 years half-life. In this decay the alpha particle has 4.78 MeV kinetic energy. The radon itself decays by alpha decay, and emits a 5.5 MeV alpha particle. That kinetic energy is enough for traveling 3.5 cm in normal air. That is the range of the alpha particle of the radon decay. In this reaction

a ^{218}Po isotope will be formed as a daughter product of the radon. The ^{218}Po has a half-life of 3 minutes. After the alpha decay of the radon there are two beta-minus decays in the radioactive chain, and then another alpha decay leads to the beta decaying ^{210}Pb isotope that has 21 years half-life. This is long enough compared to the processes that are interesting for us. During our laboratory or even a seasonal measurement the 22.2 years is considered very long. The ^{210}Pb can be accumulated in several cases and then its radioactivity is also important. The radon chain then is the sequential decay of the following 6 isotopes:

^{222}Rn	\rightarrow	^{218}Po	\rightarrow	^{214}Pb	\rightarrow	^{214}Bi	\rightarrow	^{214}Po	\rightarrow	^{210}Pb
alpha		alpha		beta		beta		alpha		beta
5.5 MeV		6.0 MeV		<1.02 MeV		<3.27 MeV		7.68 MeV		17 keV
3.823 days		3.098 min		26.8 min		19.9 min		164.3 μs		22.2 years

Table 19.1. The parameters of the radon decay chain

In this table we listed the decay mode, the decay energy and the half-lives of the members of the radon decay chain. In case of the ^{214}Pb and ^{214}Bi the beta decay can lead to several excited states of the daughter nuclei that means several different beta decays can occur all with different maximum energies. The maximum of these maximum energies is shown in the table. That corresponds to the energy difference of the ground state of the mother and the ground state of the daughter nuclei. There is an interesting feature of this radon chain. The alpha decaying isotopes rarely emit gamma rays after the decay. But the two beta decaying nuclei (^{214}Pb and ^{214}Bi) are intensive gamma emitters. Detecting these gamma rays the radon daughters can easily be determined in a sample.

19.1.2 The origin of indoor radon

The radon atoms are always born in an alpha decay of a ^{226}Ra . So the radon is coming from the radium. In most of the cases it is in a grain of a soil or rock and the radium is in radioactive equilibrium with the total uranium decay chain, but not always. There are special cases where uranium and radium are separated. This can happen if the uranium is solved into a sub-surface fluidum but the radium is not, or in the other way. Radium can appear in thermal waters and when it reaches close to the surface it can mix up with fresh water and can precipitate. This is an example for radium accumulation. Then radium will decay with its 1600 years half-life and on geological time scale it disappears. It will transform to ^{208}Pb . But a continuous accumulation can feed a high activity radium spot. The uranium can also move with a fluidum and can precipitate if the chemical environment changes. But the decay of the uranium is so long that it could not disappear in the Earth's life.

There are minerals that can bear or contain uranium or thorium atoms. They need a little larger space in the crystal. For example the rare earth elements are chemically similar to the actinides where uranium and thorium are located on the periodic table of elements. These elements also have large radius so they are good candidates for substitution of uranium. There are uranium minerals like uranite, but it is generally mixed up with other minerals. The uranium and the thorium are distributed on the Earth's surface almost homogeneously. They are accumulated in some areas but these are special cases. All soils contain uranium therefore radium as well. The world's average uranium concentration is 2.2 ppm (part per million), or we can also say it is 2.2 g/t as well.

19.1.3 The health effects due to radon and its daughters

The radon atoms in an indoor air can enter the lung of people. One breath can take 3-4 seconds. Mostly the radon atoms will go in and then come out of the lung. There is a small probability that the radon will decay just in the lung. If it happens the alpha particle will collide to the wall of the bronchia, since the diameters of these are much less than 3.5 cm. But another point is the daughters of the radon. Those elements do not have a closed electron shell. They are not noble gases. Just after the decay the daughters have lost some electrons and they are ions. It takes some time to collect the necessary number of electrons. Therefore the daughters intend to stick to aerosols or to the wall. If they collide with a dust particle the unclosed electron shell or their charge will attract the electron shell of the molecules at the side of the dust particle. This secondary attraction will glue the radon daughters to the solid phase (aerosol, wall). Therefore the radon daughters are generally on the aerosols and they are not diffusing a long time to find an aerosol. It of course depends on the properties of the air. A dusty cave environment where the miners are working does not favour for the low radiation dose due to radon. If someone breathes in an aerosol particle, the aerosol will glue to the wall of the bronchia with higher probability than radon decays in the lung. Moreover, the aerosols are accumulating in the lung. At the branch of the bronchia the air makes a band but the high mass particles will be drifted out and touch the wall and stick there. The radioactive isotopes will stick there, too. In this way the radon daughters are accumulated in the lung if the radon concentration of the air is high and there are a lot of aerosols. If people smoke that also raises the risk of radon!

The alpha decay of the accumulated ^{218}Po and ^{214}Po will emit alpha particles that penetrate into the surface cells of the bronchia and ionize the material of the cell. The range of these alphas is in the order of 10 μm . They ionize the molecules around their path. Therefore proteins can lose their secondary spatial structure, more ions can be created in the cell and other microscopic effects can happen. Recall that it has gone the same way since the life and human being appeared on Earth. The early people lived in caves which place is known for high radon concentration. This means that the average level of radon is not dangerous. The cell will react to these ionization effects and it can handle it if the frequency of the alpha particle hitting into the same cell is not very high. There are experimental hints (although not widely accepted) which tell that small radiation dose can stimulate the immune system. The problem with the radon is that between the radon dose that might be useful, and the radon dose that is harmful for the lung there are only two orders of magnitude.

19.1.4 The radon levels in houses

The radon is created by the alpha decay of the radium in the grains of the soil, rock or in the building materials. In the soil the air generally doesn't move fast, there are no turbulences, no mixture and the convection is rare. Therefore the radon concentration of the soil air is relatively high compared to the radon levels in the open air, or in indoor air. Generally it is in the 1 kBq/m^3 – 1000 kBq/m^3 range. In a general area, where there is no process that raised the uranium or radium concentration in the soil, the radon level is about 10 kBq/m^3 . According to the average uranium content of soils that is 2.2 ppm. This order of magnitude is formed in caves where the air of the cave is connected to the soil air. The radon can enter the houses as well. It of course depends on the building technology and the material used. If the house is an old type one and there is no hard insulation, the radon can diffuse into the rooms in the basement or can accumulate in other rooms that have no ventilation. The general radon level in a room in the ground level is 50-100 Bq/m^3 . It is much less than in the soil. The radon level in a

house has a seasonal variation. During the winter when the pores of the soil can freeze out and can be closed the radon level in the soil will be higher. The radon concentration in a house is always higher in winter than in summer. The radon level of the open air is also not negligible. According to lots of measurements it is between 0-30 Bq/m³. This radon concentration depends strongly on the meteorological conditions, as does the indoor radon concentration.

In a general room where we assume 100 Bq/m³ radon concentration there are several radon atoms but their number is surprisingly low, compared to the number of oxygen and nitrogen molecules. The size of a general room can be considered as 5 m × 5 m × 3 m for estimates for the order of magnitude. That is 75 m³, and the total activity is 7500 Bq. Based on the relationship between the activity and the number of radon atoms: $A = \lambda N$, we can calculate the number of radon atoms: $N = A/\lambda = AT_{1/2}/\ln 2 = 7500 \text{ Bq} \cdot 3.8 \times 24 \times 3600 \text{ s} / \ln 2 = 3.5 \times 10^9$ db. This is 14 orders less than the molar amount. The radon occurs always as a trace element, but due to the very high sensitivity of the nuclear methods, its amount can be analytically determined. On the other hand this small number is enough to make health hazards for people.

Let us calculate how many radon atoms decay in the lung during one breath. As a rough estimate the breath lasts for 4 s, during this time the radon level in the lung is the same as in the indoor air. This is a good estimate. Assume 5 l = 0.005 m³ for the volume of the lung, we get 0.5 Bq, for the radon activity inside the lung, and this means about 237000 radon atoms. Since the activity is 0.5 Bq, the time is 4 s, there will be 2 decays per one breath, out of the 237000. This does not cause a significant health hazard. But the radon daughters that are sitting out on the aerosols can attach to the pipes of the lung mostly at the junction area. This process results in much higher alpha activity in the lung.

19.2 The RAD7 radon detector

During the laboratory practice we will use a radon monitor device that contains a semiconductor detector for detection of the alpha particles. This instrument is called RAD7. This is developed for measurements of radon in any kind of air: indoor air, outdoor air, soil air or the air phase of a bubbling system. In fact, the machine detects the daughter products of the radon and not directly the radon.

19.2.1 The operating principles of the detector

The schematic drawing of the detector can be seen in [Figure 19.1](#). There are three main components: 1 – air filter, that blocks the radon daughters out from the detector chamber, 2 – air pump that makes the air circulate across the pipe system, 3 – detector chamber, the main part of the detector, there is electrostatic field in the chamber that pushes the positive ions down onto the semiconductor detector, which is placed in this chamber.

The radon atoms that decay in the chamber will produce ²¹⁸Po that is a positively charged ion. In the chamber the electric field will force these positively charged ions to be collected on the bottom of the chamber. They will drift just onto the surface of the semiconductor detector. This is an accumulation technique. Another big advantage of RAD7 and the usage of a semiconductor detector is the energy resolution feature. The detector determines the energy of the alpha particle that goes in it. The detector itself is very thin. The alpha particles stop in there but the electrons from beta decays can penetrate through and they will not deposit their total energy in the detector. On the other hand, the electron energies in the radon chain are less than 4 MeV, but the alpha energies are higher than 5 MeV. As an example 4 MeV is a good limit for separation.

If we consider the energies of ^{218}Po , ^{214}Po (see [Table 19.1](#)) their separation is 1.68 MeV and the detector's energy resolution is about 100 keV. The decays from the two isotopes can clearly be separated.

The other radon isotope, the thoron is also measurable with this detector. The ^{220}Rn alpha decays into ^{216}Po that is an alpha-decaying isotope as well. Similarly to the radon chain there is a thoron chain. The energy resolution is so good that the thoron daughters will not be mixed up with the decays in the radon chain.

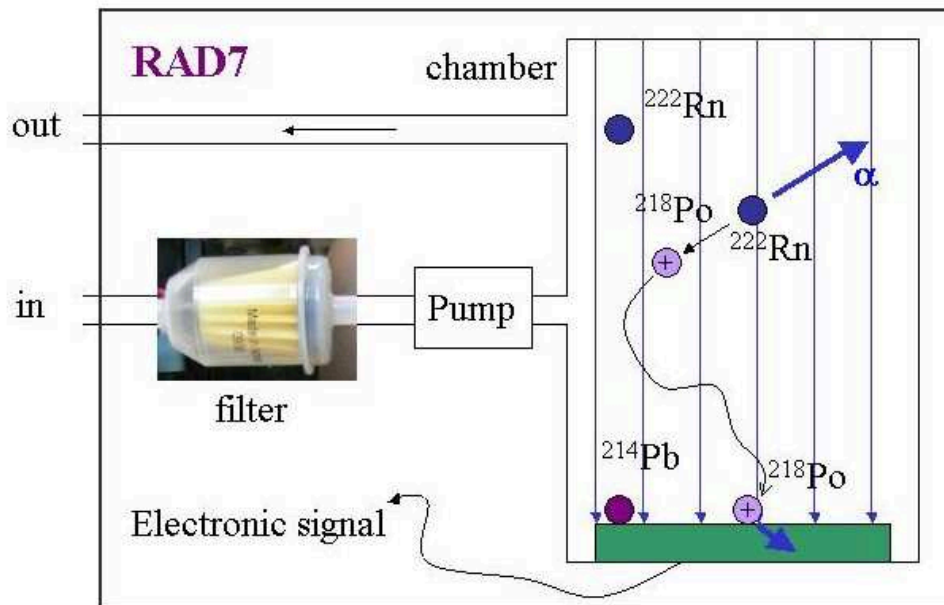


Figure 19.1. The schematic drawing of RAD7 radon monitor and the bases of the operation of the chamber

In principle the alpha decay of 4 polonium isotopes can be determined: $A=218, 216, 214, 212$. The first and the third are in the radon chain, and the others are in the thoron chain. The half-life of the ^{218}Po is about 3 minutes that means after some minutes of the radon decay the ^{218}Po will alpha decay. During that it will sit on the surface of the semiconductor detector. But for the decay of the ^{214}Po we need to wait for two beta decays in the chain, and the half-life of those are about 20 minutes. The ^{214}Po decay will happen much later. Maybe the radon went into RAD7 detector at a time and the ^{214}Po decay was 40 minutes later. So the event at the ^{214}Po energy shows that there was a radon a long time ago in the sampled air. According to these time features ^{218}Po is called “fresh radon” and ^{214}Po is called “old radon”. The same situation appears in the thoron chain. The half-life of ^{216}Po is 0.145 s but its two beta decaying daughters have half-lives of 10 hours and 60.5 minutes, those are much longer. The next isotope in this chain, the ^{212}Po sits several hours on the detector's surface before it decays. Therefore ^{216}Po is called “fresh thoron” (alpha energy = 6.9 MeV) and ^{212}Po is called “old thoron” (alpha energy = 8.9 MeV).

The energy spectrum of RAD7 shows 4 peaks corresponding to these alpha decays, the number of alpha decays are proportional to the number of those isotopes. Therefore the number of the events in a peak (in the energy range close to the alpha energy) is proportional to the number of the radon or thoron isotopes presented in the air. To select the separate isotopes RAD7 applies 4 energy windows: A, B, C and D. These are set to have the individual alphas

in the middle of the window. So the ^{218}Po corresponds to window A, ^{216}Po to the window B, ^{214}Po is in the window C and ^{212}Po 's alpha decays will be counted in window D. There is an interference although, since ^{212}Bi in the thoron chain decays by beta decay only by 64.06% probability and the other 35.96% corresponds an alpha decay that has an energy of 6.2 MeV, which goes into window A. But the computer built in the detector makes a more or less right correction of this effect. (Only if the thoron concentration is very high relative to the radon concentration, a better correction is needed.)

19.2.2 How to measure radon concentration with RAD7

RAD7 has several protocols. A protocol means a setup of measuring parameters. Like the length of the measuring time, the number of repeats that it will execute, the mode of the operation of the pump, the method how to sum the counts. The most generally used protocol for indoor air measurements is the "sniff" mode. In this case the counts in the window A will be taken only into account and the pump will operate continuously. The time can be set according to the user's choice. The other way is setting the "normal" mode, when after 3 hours the machine changes to sum the counts in two windows A and C. Window C means the old radon, therefore it is not sensitive for quick changes. That is why we generally use "sniff" mode.

Another option is the number of thoron decays. If we are interested in those we have to switch on the thoron mode.

The measured values can be seen in two ways. 1) is the computer download and 2) is the print of each measurement. RAD7 has an online printer connected by infrared communication to the detector on the top side. The printer prints out the concentration levels with their uncertainty for radon and for thoron (if thoron is set). The other way is to use a serial cable and the official software to RAD7. Using these, the data file can be downloaded and every single data can be analyzed. In this case we have the time of the measurement, the length of the measurement, the total decays detected, the percentages of the windows, the temperature and the relative humidity inside the detector, some electronic parameters like current of the pump and the calculated radon concentration and its uncertainty. The uncertainty means 2 sigma significance level (95%).

After placing the instrument and setting up the right protocol we can start the measurement. But we should know that RAD7 does not measure the alpha decays from the ^{222}Rn , only from their daughter products. This is because of the electrostatic collection in the chamber. The radon itself cannot be moved onto the detector surface since these are neutral atoms. The always-charged daughters on the other hand are easily movable by the electric field. (The alpha decay itself will eject some electrons from the radon atom as a secondary process.)

In the detector chamber there will be certain radon decay intensity, but the ^{218}Po decay intensity at the beginning is just 0 and then it rises. Time is needed while the ^{218}Po and radon will reach a radioactive equilibrium when both will decay with the same intensity (the same decay per sec) and in this stage the counts per time detected by the RAD7 will equal to the radon decays per time. This time to reach the equilibrium is about 10-15 minutes, depending on the required uncertainty. Therefore the first ten minutes cannot give the right radon concentration. If we set the measurement time to 10 minutes, and set 4 cycles, we neglect the first and make the average of the other 3 values.

On the other hand, the ^{218}Po atoms will remain on the detector surface. If we open the windows or bring the detector to the open air, to a much lower radon level, the detector will remember the radon concentration in the room, since the ^{218}Po -s from the decays still in the room will be on the surface of the silicon detector. We should wait 15 minutes after each measurement. There is an inside procedure to help this clearing process. This is the “purge”. If purge is used, the electric field will be changed and the ^{218}Po -s can be blown out of the volume of the detector chamber.

Another issue during the measurement is relative humidity (RH). We need to use desiccant material to reduce the RH below 10%.

19.3 Lab course tasks

1. Determine the number of radon atoms in the room of the measurement! First make an estimate using 100 Bq/m^3 , and later use the measured radon concentration.
2. Calculate the energy that will be lost by ionization in the cells of the lung according to one ^{218}Po attachment in the lung!
3. Determine the radon concentration of the room’s air! Calculate the percentage of the uncertainty of one measurement and that of the average.
4. Detect the decay of the remaining ^{218}Po atoms. Switch out the pump, set the measuring time to 3 minutes and watch the windows A and C as well, if their level will decrease.

19.4 Test questions

1. Why radon is dangerous for the humans?
2. What is the role of the aerosols in the mechanism of the health effect of the radon?
3. What are the average radon levels in houses and in the soil gas?
4. Are the caves interesting from the radon point of view?
5. What is the radon-series and the thoron-series?
6. Are there alpha decays in these decay chains?
7. How does the RAD7 detector work?
8. What is the time dependence of the ^{218}Po atoms in the RAD7 if we are detecting constant radon concentration?
9. What is a “old radon”?
10. What are the energy-windows in the RAD7 detector? What are these mean?
11. How does the RAD7 separate the daughter isotopes of radon?
12. Why does the RAD7 use a filter at the inlet?
13. How does the RAD7 accumulate the isotopes that it detects?
14. What detector is the main part of the RAD7 detector?

20. Measurement of radon concentration in water (RAD)

20.1 Origin of natural radioactivity

The stable isotopes of chemical elements were born in a supernova explosion in the Milky Way galaxy several billions of years ago. The radioactive, short-lived atoms were also produced there, but these are decayed out already, and cannot be found in the material of the solar system – except those radioactive atoms which have a long enough half-life, comparable to the age of the Earth. For example ^{40}K , ^{238}U and ^{232}Th are such kinds of atoms. These can be found in the deep material of Earth or on the surface as well, close to the biosphere: soil, rock, building material. The creation of our planet, therefore, resulted in a situation that our environment contains radioactivity naturally, without any civilization reasons.

The natural ionizing radiation contains two more components. There are short-lived atoms (less than million years) in our environment that have natural origin. These atoms are continuously created in the upper atmosphere due to the cosmic radiation. The third component is the cosmic radiation itself. The cosmic radiation is the particle radiation that enters the upper atmosphere. Its main component is proton from the Sun or from other sources. These protons can have very large kinetic energy, much more than man can produce in particle accelerators, but on the other hand those high energy protons are very rare.

The thorium, uranium atoms (mentioned above) are alpha decaying atoms, their atomic nucleus will spontaneously decay, emitting a helium nucleus consisting of two protons and two neutrons. After the decay the remaining nucleus is called the daughter nucleus. (From now on we talk about the nuclei only, not the whole atom.) The daughter nuclei are radioactive nuclei as well as their mothers, and so are the daughters of the daughters. We call this series of decays a radioactive series, or radioactive chain, sometimes “a radioactive family”. This latter has a bit more general meaning than the radioactive chain. Since the first mother nuclei are in nature, all of the daughters, all the elements of the chains also contribute to natural radioactivity. The chains are terminated by the decays to the stable nuclei of lead or bismuth.

Since the original creation of the elements not only the above mentioned heavy nuclei ($A > 200$) were able to survive. ^{40}K also has a long half-life: 1.227 billion years (Gy). This is a beta-decaying isotope. This isotope occurs very frequently, since its relative abundance is 0,0118%, among the potassium atoms. An interesting fact is that the average atomic weight of potassium is about 39 units, the one less atomic number argon has a weight of 40 units. This is unusual, and the reason is that all the argon on Earth is produced from the decay of the ^{40}K . There were other argon isotopes but those escaped from the atmosphere long ago. All of these radioactive nuclei have been in our environment for billion years, and earlier there were even more. The whole life evolved accompanied by radioactivity.

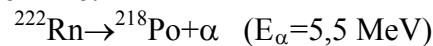
The beta-decaying nuclei of ^3H and the ^{14}C isotopes are produced nowadays in the upper atmosphere by the cosmic radiation. Their half-lives are 12.32 years and 5570 years, respectively. The high energy protons hit nucleus of oxygen or nitrogen and can produce neutrons. The neutrons then can hit a nucleus of ^{14}N and can produce a ^{14}C , or strike out a triton nucleus (^3H). These two radioactive atoms can then diffuse in air and can be built in organic materials, since the organic molecules contain a lot of H and C. Water in nature from this process can contain a small amount of radioactivity. Human beings are also radioactive from this reason, but we have to emphasize that the main component of our radioactivity is the potassium.

20.2 Some important properties of radon

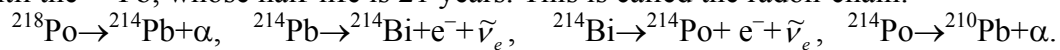
20.2.1 The decay chain of radon

There is another isotope that can enter the human body. This is the radon, which is a radioactive noble gas, and we can inhale it or drink it with the drinking water. The radon is originated from the ^{238}U . This is a member of the uranium chain. There are uranium isotopes in the rocks close to the surface of the Earth. Those are the 235 and the 238 uranium.

The half-life of ^{235}U is 0,7038 billion years, and the ^{238}U has 4,468 billion years. In the radioactive chain starting with ^{238}U (uranium chain) there is the 3,8 day half-life ^{222}Rn (radon). This time is enough to get out of the soil and can accumulate in indoor air or caves, drinking water. The radon alpha decays to ^{218}Po .



The daughter of radon has a 3 minute half-life. This is short compared even to the duration of a laboratory practice. The half-life of the daughter of the ^{218}Po is still less than an hour, and so goes until the ^{210}Pb , whose half-life is 21 years. This is called the radon-chain:



The next isotopes are beta-decaying ones with about half an hour half-lives. The ^{214}Po is again alpha emitter, but its alpha energy is higher: 7,7 MeV. Its half-life is about microseconds, and its daughter product is the “slowly” decaying ^{210}Pb . During measurement that lasts about days, it is the end of the chain. After some weeks all of the radons are converted to 210-lead.

The creation of the radon is an alpha decay of its mother nucleus: the radium. $^{226}\text{Ra} \rightarrow ^{222}\text{Rn} + \alpha$, which is the 6th element of the uranium chain.

The radon in our environment is originated always from a rock or other material that contains radium. Generally it has uranium as well, but it is not sufficient. In case of objects that have geological age, and the radium (in some geological process) is separated from the uranium the radium alone will disappear in about 8000 years, since its half-life is 1600 years. That is why one can find radium alone rarely and only in young materials.

In processes, in which radium precipitated and was separated from its preceding decay chain not more earlier than some hundred years ago, we can find natural radium alone, without uranium. Some specific thermal water are good examples for this case, where radium secession is continuously under way in geochemical processes.

20.2.2 The health effect of radon

First the uranium miners were exposed to high radon concentration. The investigation of them resulted that they had a higher risk of lung cancer. This is due to the radon daughters that can be attached on the surface of the lung pipes. The most probable place is where the pipes are split up at a junction, where the airflow bends and the heavy atoms (radon daughters) can go close to the surface. Unlike the noble gas radon, the daughters have an electron structure that is easily deformable, and can interact with the molecules of the surface. The alpha particles from radon, but much more the alphas from radon daughters sitting on the surface can ionize the cells and can make a considerable damage. In the uranium mines the exposure was long, regular and the radon concentration of the air was very high.

Nowadays it was found in Hungarian villages built on granite rocks that the health risk of radon is higher than the average. This has been well known already for decades in North European countries where granite occurs more frequently, e.g. in Finland.

There is a principal difference between short and long time exposure. In case of short time exposure the metabolism can react. New research results suggest that short-term high activity exposure can have positive health effect on the long run, since it can activate the immune system. In every thermal bath the radon concentration is high, but people are getting cured there for centuries.

Natural radioactivity has been irradiating the Earth and its population for million years. We can consider it as a non-dangerous condition. But there are cases where its intensity is anomalously high. Those cases can be dangerous and should be investigated or eliminated, avoided.

One of the important goals of the researches in environmental protection is to investigate how the radioactivity can accumulate or grow up naturally in the close neighborhood of people. Another issue is how can civilization make the radioactive levels higher.

The aim of this laboratory practice is to get known the reasons of the occurrence of the natural radioactivity in water, investigate the methods that are useful to determine the radioactivity of water.

20.3 Natural radioactivity of water

The first reason of natural radioactivity of water is the ^3H atom that is built into the water molecules. The activity built into human organs is due more to the radiocarbon (^{14}C). There is an equilibrium concentration of triton and radiocarbon in the air, and therefore in the living organs, too. This can be expressed like this: every 10^{18} -th hydrogen atom is triton, and this ratio for radiocarbon is 2×10^{12} .

The subsurface water is generally not attaching the air, so the radioactivity of this water is due to uranium (radium) or thorium content of the rocks. The elements of the decay chains can form dissolvable molecules and can go to the water. Radium-chloride or the noble gas radon can be easily dissolved. Because of these processes our natural water can be radioactive.

Measuring the water of deep drilling wells we know that generally this water is radioactive, due to radon content. The radon concentration of spring or well water is easily measurable and it is generally in the range of 10 – 500 Bq/l. For comparison, the EU suggestion for the limit for radon concentration in drinking water is 100 Bq/l. The high radon concentration water is frequent, and it is not suggested to drink it regularly, immediately after getting it from its source. On the other hand the radon can evaporate out of the water in some hours and the concentration will be much reduced after a while.

20.4 Liquid scintillation technique for radon measurement

20.4.1 Scintillator material

The radioactive radiations are ionizing radiations therefore crossing a material they will lose energy. The main process of it is ionization and excitation of the molecules and atoms of the material. The atoms and molecules will not remain excited for a long time; they will get back to the ground state shortly. During this process when the atomic electron reaches lower energy states the energy can be emitted in visible or UV photons.

Scintillation is such a process when the ionizing particles make the molecules and atoms emit visible or UV light. Many atoms and molecules will be excited, so there can be many photons emitted after one elementary decay. The number of photons is called the light output (L). These photons can be different frequencies, so there is a distribution of the frequency of photon emitted in one elementary process that is characteristic for the material which the ionizing particle went through.

The amount of the light-output is proportional to the number of excited atoms or molecules, therefore proportional to the energy deposited in the material (E_d). The unit of L is energy-like quantity: **keVee**, kiloelectronvolt-electron-equivalent. This means how many times more photons were emitted than in a case of 1 keV electron. The alpha particle with kinetic energy of 1 keV (for example) creates much less photons than 1 keVee. The reason of it is that the alpha particle ionizes the material in a more dense way, and the de-excitation can be undergoing in many different ways, like collisions and magnetic interaction. So the molecules will lose their energy not emitting a photon but colliding with another molecule.

We use scintillator material, which is liquid. It has several advantages. It can be mixed with the sample and short-range radiation can be measured. This is why the liquid scintillation is the mostly used technique to determine triton and radiocarbon. The range of radon alpha radiation in water is some microns. The liquid scintillation analysis (LSA) is also a good tool in this case.

20.4.2 The photomultiplier tube

Using scintillation detectors the issue is always to count the emitted scintillating photons. The instrument that will produce electric signal out of the scintillation light, and then we can count it, is called the photomultiplier tube (PMT).

The main principle of its operation is that the scintillation photons will hit a very thin beryllium layer (evaporated onto a glass window), and the photons will create a photoeffect in which electrons will be ejected out of the layer, they are called photoelectrons. The photoelectrons will be accelerated by some hundred volts and focused onto a special metal called dynode. If a fast electron hit the dynode it will shoot out 3 slow electrons, which will be accelerated again to be able to shoot out 3 other electrons from another dynode. In a PMT 10-12 dynodes are applied. 1 electron will be multiplied up to about half a million electrons. In this way after scintillation a measurable electric signal will be produced at the end of the PMT. The amplitude of the signal is proportional to the number of photoelectrons that were ejected from the beryllium layer (photocathode). But the number of photoelectrons is proportional to the number of photons hitting the photocathode, which is proportional to the number of photons emitted in the elementary process, and that is called light-output. This is how we can measure the energy of the particle resulted in the radioactive decay.

To determine the number of decays occurred in the sample we have to count the number of scintillation processes. Generally not all of the elementary processes will create large enough electric pulse, so we will detect fewer counts than the number of decays.

The detection intensity is the number of detected events per a given amount of time. Generally we express it as count per minutes: CPM. On the other hand the decays per minute are called: DPM. The efficiency of the detection system is $\eta = \text{CPM}/\text{DPM}$.

The activity of the sample is proportional to the detecting intensity (CPM), and a calibration process is needed to connect them.

20.5 4. The method of determination of radon concentration of water

20.5.1 Sampling

The water sample precisely is the water in its original environment, like in the porous volume of rocks. When we take a small volume of it the radon concentration can immediately change due to evaporation. So to take a sample there is generally a protocol to follow, which is called sampling process. In the sampling process we take 10 ml of the water into a syringe and put it under 10 ml of scintillator material (cocktail, in our case it is called OptiFluor-O) that is already put into a vial. If the direct sampling is not available then water should be filled into a glass bottle (plastic is not enough to avoid radon diffusion out from the bottle). The bottle should be filled totally, to avoid air bubble on the top. The radon can escape from the water to the air bubble, and we will have sampling loss of radon. The vial or the bottle should be shipped into the laboratory, and the shortest time, the better. The radon decays with a half-life of 3.8 days. After one week its activity is reduced already by more than a factor of 4. This results in higher uncertainty of detection.

20.5.2 The preparation of the water sample for the LS counting

10 ml of the sample should be already in the vial under 10 ml of OptiFluor-O, and the vial should be air tight (we use parafilm for this purpose). The scintillator OptiFluor O cannot dissolve in water, our system will be a two-phase system. The radon is first only in the lower water phase, but it cannot scintillate. We should wait for the radon diffuse up to the cocktail phase. It takes about 5 hours, but with shaking (not mixing!) we can shorten the time very well. We have to reach the diffusion equilibrium. Our cocktail is so specific for radon measurement that it dissolves the radon 13 times more than water, so the main portion of the radon atoms will be in the cocktail phase.

In the cocktail the alpha particles coming from the decay of the radon will create so many photons as an electron of about 300 keV would create. We have to set up the machine according to this. Our setup is to count an event if the light-output is in the range of: 25 – 900 keVee.

At this point we are ready to count the sample using the TriCarb 1000 type liquid scintillation spectrometer. In the machine there are two PMT-s on opposite sides of the vial and they are counting the scintillation photons in coincidence. The number of counts per minutes will be determined and will be printed onto a printer. Also a computer can evaluate the light-output spectrum.

20.5.3 Light-output spectrum of radon

In the radon chain (explained above) there are three alpha emitting isotopes. The three alpha energies are different. The alpha from the decay of ^{222}Rn has 5,5 MeV kinetic energy, the ^{218}Po and the ^{214}Po have 6,0 MeV and 7,7 MeV respectively. The alphas (from the decay of a given isotope) will have the same energy in every decay, so the accepted spectrum is a bump (e.g. Gaussian in shape) at the associated light-output. In scintillation methods this bump is quite a broad one. The two Gausses of the two lower energy alphas will overlap, because their energy difference is less than the width of the bumps. But the third alpha at higher light-output can easily be separated. The two other 214 mass number nuclei decay with beta decay. The resulting electrons have continuous energy distribution, because there is another particle in the process, antineutrino. We cannot detect this small, neutral particle, but it carries out some energy. (Sometimes the total energy.) The two beta decays in the spectra will be a broad

background like bump from 0 keVee to about 800 keVee. The Gaussians will sit on this part of the spectrum.

We detect only the counts higher light output than 25 keVee, so we cut a part of the beta-electrons, but we cut all of the background tritons and a big portion of the laboratory background.

20.5.4 The radon concentration

Our aim is to determine the dissolved radon concentration of several samples. First we measure the detection intensity of the samples, in CPM. Then we use the results of a calibration method, which was done earlier using known radium chloride solution. The calibration line will give the radon activity concentration (c) in the 10 ml sample from the measured CPM.

$$c = \frac{A}{V}$$

where A the activity of the sample (Bq), V the volume of the sample (10 ml). The unit of activity concentration is always Bq/l. The equation of the calibration line is:

$$c = \frac{CPM - 12,1}{1.98} \frac{Bq}{l}$$

This is the radon content or radon concentration. Someone can calculate the number of radon atoms from this. Using the $A = \lambda N$ expression

$$N = \frac{A}{\lambda} = \frac{A \cdot T_{1/2}}{\ln 2} = \frac{c \cdot V \cdot T_{1/2}}{\ln 2}$$

Using the counting technique we can determine the radon concentration in the vial long time after the sampling process. A lot of radon has been decayed although since the sampling. We have to determine the original concentration using the exponential decay law.

$$c(\text{counting}) = c(\text{sampling})e^{-\lambda t}$$

The symbol t denotes the time between the counting and the sampling.

20.6 Lab course tasks

1. Determine the radon concentration at the sampling time of the available samples! Generally the samples are from the Rudas-spa drinking water wells: Hungária-well, Juventus-well and Attila-well. The samples were bought there by 10 Ft/liter. These water samples are from 30-45 m deep artificial drilling, close to the Danube.
2. Personal samples are welcome. If you bring your own sample, we will determine its radon concentration. You need to bring a sample that is no more than 4 days old, and it is in a glass bottle described in the text. Beware of the radon loss at the sampling! If you appear in time you can get a prepared vial, as well. Determine the CPM and calculate the original radon concentration.
3. Determine the light output spectrum of the radon, and check if the counts are really from the decay of radon. You also can check the radioactive equilibrium between the daughters. Make a drawing of the spectrum and describe its components. The file of the spectrum will be read into excel if you set “;” as separators. The data are in columns, the first column is the

keVee values the others are the total counts per total measuring time in a 0.5 keVee light-output interval.

3.1. Draw the spectrum in that way, so the interesting parts will be well observed. The averaging of the values in excel can help in a case of low count sample. Describe the features of the spectrum in half a page.

3.2. Determine the light-output value of the center of the Gaussian associated to the 5.5 MeV and 6.0 MeV alpha particles. Calculate the light-output of a 1 keV alpha from the result.

4. How many radon atoms are there in the sample?

5. Investigate the correlation of the uncertainty of the CPM-s and the total number of radons decayed during the counting time! Use the uncertainties that are printed by the machine! How does the machine calculate it?

20.7 Test questions

1. How were the long half-life isotopes created in Nature?
2. Can we detect the alpha-particle from the radon decay?
3. How many alpha-particle scan we detect using LSC and OptiFluor O?
4. What are the advantages of using OptiFluor O scintillator?
5. There is radium in a water. Can we detects its alpha-decay using the setup in this lab?
6. Why do the subsurface water sources have dissolved radon content?
7. How does the light-output distribution look like for a radon sample?
8. What is the relationship between the radon content in a spring and that radon content which is actually measured later?
9. How can we see the beta-decays in the radon light-output spectrum?
10. What isotope is the terminator of the radon decay chain? How long its half-life?

21. Measurement of radon exhalation from rocks and soil samples (REX)

The radiological effects on people from the natural ionizing radiation consist of several parts. More than half of its biological effects comes from the radioactivity of radon gas and its daughter isotopes. The most important is the radon of the indoor air, since it is inhaled the whole night and the isotopes decay and emit alpha-radiation to the lung. The radon can enter the indoor air mainly from soil around the house and with less importance from the building material and from water sources. Firstly uranium and radium content of the soil and building material is the relevant property to control the indoor radon concentration, but secondly the process to get out of the grains of the material will determine the level of the radon actually formed. The latter is called the emanation of radon from grains to the pores of the material. There are soils that emanate only a low level of radon despite their higher radium content, since their emanation properties inhibit it. The emanation is controlled by the grain diffusion and the surface properties of the grains when the radon is formed by decay on the surface and directly moves to the pore space. The emanation coefficient tells how many percent of the radon atoms formed by decaying radium can enter the air phase. If the diffusion speed is small in the grain the radon can decay inside and the emanation coefficient will be a small number.

21.1 Radon exhalation and its measurement

21.1.1 Exhalation and emanation

Exhalation (E) is a process where radon atoms leave a macroscopic media, like soil, rock or building material. The quantity of the exhalation is the number of radon atoms leaving in one second from the sample. This depends on the size of the sample, therefore specific exhalation $M=E/m$, (where m is the mass of the sample) is worth to examine. Specific exhalation corresponds to simply the material if the size is less than the diffusion length of radon in the material. A more microscopic process is the emanation, which is where the radon atoms leave the grains of the material to the air or water phase of the pores. The exhalation can be realized in several physical processes. The two most important ones are the diffusion and the convection. For example at the soil surface both are important. The soil radon concentration is always higher (much higher) than the open-air radon concentration so the diffusion of radon will result in a net radon flux to the open air. However, this process depends on some soil parameters like water content, relative humidity of the soil temperature, and porosity. But if there are meteorological fronts or other nonequilibrium conditions, the pressure difference will make the soil air move out or in, hence the radon atoms will also move with it. This process also depends on the soil properties like permeability, humidity etc. In laboratory measurements the connection is negligible, also in case of subsurface emanation into water phase the convection does not play an important role.

The radon emanation and exhalation of the soils are important, since they are governing factors for the level of the soil gas radon concentration and therefore the radon flux to the houses. The biological effects on the inhabitants strongly depend on this.

21.1.2 Overview of the measurement

During the laboratory practice we will examine radon exhalation of soil samples using radon chambers. Our aim is to determine how many radon atoms leave the sample in a second. For

this aim we use the following setup. We close the sample into a radon chamber for 3 weeks. We wait for the equilibrium radon level in the air above the sample and then we measure the radon concentration of this air. During these 3 weeks the number of radon atoms on one hand will increase due to the exhalation but on the other hand they will decrease due to radioactive decay. In the equilibrium these two quantities are equal. The three weeks are about the 5 times the half-lives of the radon, and that is enough for the equilibrium. In this state the activity of the air is time-independent, and the same number of the radon atoms in the air will decay as will enter the air. $A=E$, where A is the radon activity of the total closed air in the radon chamber (V_{net}). We will measure A and determine E .

The measurement uses a radon monitor detector called RAD7. That is a device including a pump, pipes and semiconductor detector and analyzing electronics. The description of the detector can be found in Subchapter 19.2, and new details will be described below. This device is capable to determine exactly the energy of the alpha particle ejected in each alpha decay of the radon daughters. During this laboratory course we will apply alpha-spectroscopy, and as an advantage of this we separate the radon (^{222}Rn) and the thoron (^{220}Rn) isotopes hence the radon and the thoron fluxes of the sample. The thoron exhalation is not always important since the thoron's short half-life (55 s). But there are cases in which it has significance and fortunately our measurement method can determine both exhalations.

21.1.3 The time dependence of the radon level in a closed chamber

In our sample the source of the radon atoms are radium atoms in the soil grains. At the radium alpha decay the radon atom arises with about 100 keV kinetic energy. That destroys the vicinity (size of some micron) of the decay and the radon starts to diffuse in the grain or if the decay happens close to the surface the radon can be ejected out of the grain immediately.

The radium atom is generally in a mineral phase of the grain. There are three processes that result in radium in the grain. One is when uranium used to located in the crystal, and that decayed. The other is when radium itself is a member of the growth of the crystal, for example from a water phase during geochemical processes. The third case: the radium is a precipitation on the surface of the grain. The grain is created much before the precipitation occurred, but a change of some geochemical properties causes radium accumulation from the water phase. The third case is when radon most easily exhalates out to the pores. In these cases the exhalation coefficient is remarkably higher than generally. The first case is an example for the case when radioactive equilibrium can occur between uranium and its radium daughter isotopes. If the grain contains all the members of the uranium chain and there is no way for them to leave the grain, radioactive equilibrium will occur. In this case radium has been produced again for thousand years. The second and third cases although will not last for long time. The radium will decay out in about 7000 thousands years. So the age of these kinds of samples should be less than some thousand years. There is only one daughter of the uranium (^{238}U) that has a chance to leave the grain. This is the radon, since it is noble gas.

If the soil sample contains N_{Ra} radium atoms, the activity of them is $A_{Ra} = \lambda_{Ra}N_{Ra}$, and A_{Ra} number of radon atoms are created in every second. But not all can leave the grain. A part of it will decay inside the grain and other part (e) will escape. This e is the emanation and it is proportional to the radium activity: $e = \varepsilon_1 A_{Ra}$, where ε_1 is the emanation coefficient for the grain. The e number of radon atoms will get into the pore space and will diffuse. Some of these atoms will decay during diffusion in the soil air. If we count up all the grains, all the radon atoms in the pore air, and all of these that are able to escape from the soil we have the exhalation. That is proportional to the total radium activity of the sample. $E = \varepsilon A_{Ra}$, where ε is

the exhalation coefficient. This coefficient does not depend anymore on the radium content but only on the soil properties like grain size, radium localization, diffusion coefficient of the radon in the grains, therefore the properties of the mineral phase also play a role. At a certain site where a house is built the specific exhalation (M) is the governing property from the point of view of the radon in the house.

If we take a sample for example in a radon chamber, that has an activity A_{Ra} and E radon atoms will leave in a second during their diffusion process, but the radon back diffusion into the soil pores is negligible, that case is a constant exhalation process. The exhalation will not depend on time. In this case the radon concentration of the air has a steady fill up rate. (In many cases the exhalation for a unit area is an interesting and useful parameter. This is called radon flux.)

The number of radon atoms in the air of the chamber is $N_{air}(t)$. There is a differential equation that determines the time dependence of $N_{air}(t)$. The number of radon atoms in the air can change in two ways: decay and fill up by exhalation. Exhalation is constant and the decay has the standard term from the role of simple decay (λN_{lev} , where λ is the decay constant of the radon = $\ln 2/T_{1/2}$). Therefore the equation is:

$$\dot{N}_{air}(t) = -\lambda N_{air}(t) + E \quad (21.1)$$

Its solution if the initial condition is $N_{air}(t=0)=0$, is the following: $N_{lev}(t) = \frac{E}{\lambda}(1 - e^{-\lambda t})$, and the radon activity in the air is:

$$A_{Rn} = E(1 - e^{-\lambda t}). \quad (21.2)$$

If the time elapsed is longer than 5 times the radon half-life, $e^{-\lambda t}$ will be negligible, and the radon activity will be equal to the exhalation. The measurement is done more than 3 weeks after one closed the radon chamber, so $A_{Rn}=E$ stands. In the measurement however, we determine the activity concentration A_{Rn}/V_{air} . From this we get from $c_{air} = A_{Rn}/V = E/V$ that $E = c_{air} V_{air}$.

21.2 The experimental procedure

21.2.1 The principles of the determination of exhalation

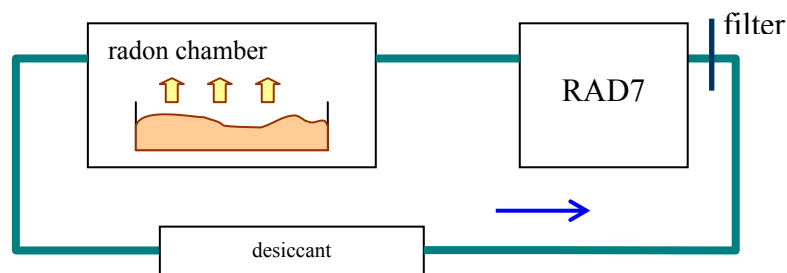


Figure 21.1. The schematic layout of the experimental setup for radon exhalation measurement

During the measurement the radon chamber will be connected to a desiccant and the radon detector (RAD7) as Figure 21.1 demonstrates. The pump inside RAD7 detector will circulate the air in this system that goes through the desiccant and loses its humidity. Keeping the relative humidity below 10% is essential for the precise determination of the radon level by RAD7. The air that enters RAD7 first goes through a filter, which adsorbs the aerosols, and

therefore the radon daughter products cannot enter the detector. Then it goes into the detector where there is a detector chamber. In the chamber a high voltage generates static electric field that will push the radon daughters (^{218}Po ions, since these generally lose a few electrons during the radon decay) onto the surface of the semiconductor detector, where the ions will glue onto. The alpha particles ejected in the decay of ^{218}Po will lose their kinetic energy in the semiconductor material, if the direction of the alpha allows this. This energy is converted to electric pulse that means the detection itself. This procedure means that this device cannot detect the decay of the radon itself but only the decay of the daughter products.

There is an effect that cannot be avoided during our measurement and it makes the detected radon concentration c_m different from c_{air} . The measured radon concentration (c_m) will be diluted since the radon in the chamber will be mixed up with the air in the detector chamber and in the pipes. The volume of the detector chamber and the pipes together are denoted by V_{det} . This is about 1 liter. The V_{air} is not completely the volume of the chamber. V_{air} equals the volume of the chamber minus the volume of the sample.

To be able to determine c_{lev} from c_m we need to know the radon concentration in the pipes and radon chamber. Therefore we execute a background measurement and determine c_h concentration. The dilution effect can be calculated if we count up all the radon atoms that are in our closed system. There are N_{air} in the chamber and N_{bck} in the background space (pipes, detector chamber). The measured radon concentration therefore is:

$$c_m = \frac{\lambda}{V_{total}} (N_{air} + N_{bck}) = \frac{c_{air} \cdot V_{air} + c_{bck} \cdot V_{bck}}{V_{air} + V_{bck}}.$$

From this formula we can get our final formula to calculate c_{air} :

$$c_{air} = c_m \left(1 + \frac{V_{bck}}{V_{air}} \right) - c_{bck} \frac{V_{bck}}{V_{air}}$$

V_{bck} should be determined by measuring the length and diameter of the pipes. Their uncertainties should also be considered!

21.2.2 The operation of RAD7 detector

The operation of the detector is described in Subchapter 19.2. Here we give an extension to that describing the alpha energy spectrum, which is essential for thoron and radon separation. The daughter products of the radon are called the radon-chain. It consists of ^{218}Po , ^{214}Pb , ^{214}Bi , ^{214}Po and ^{210}Pb . Their half-lives are 3.098 min, 26.8 min, 19.9 min, 164.3 μs , 22.2 years respectively. 22 years are long enough to treat ^{210}Pb as stable nucleus during our laboratory practice and in corresponding processes. These isotopes are beta or alpha emitters. ^{218}Po , ^{214}Po emit alpha particles with kinetic energy of 6.1 MeV and 7.7 MeV respectively.

The radioactive equilibrium is defined here as a mother isotope and its daughter reaches he same activity. In fact this is called secular equilibrium. The time needed for this is 5 times the half-life of the daughter. We can say that equilibrium between the ^{222}Rn and ^{218}Po is formed in 15 minutes, but between ^{222}Rn and ^{214}Pb , ^{214}Bi , ^{214}Po it needs 1.5 hours. ^{210}Pb never reaches the equilibrium if the original isotope is radon, since its half-life is longer than that is of radon.

The alpha energies are the properties that we use for identification of the isotopes corresponding to the detected event. The energy spectrum is divided into 4 ranges (A, B, C, D), see [Figure 21.2](#). The alphas from ^{218}Po (kinetic energy is 6.1 MeV) are placed into channel „A”,

while the alphas from the isotope ^{214}Po (7.7 MeV energy) will be collected in channel „C”. These alphas form two peaks at energies 6.1 MeV and 7.7 MeV, and the width of these peaks is thin enough for the separation. Furthermore these are so thin that channel „B” is located between these peaks.

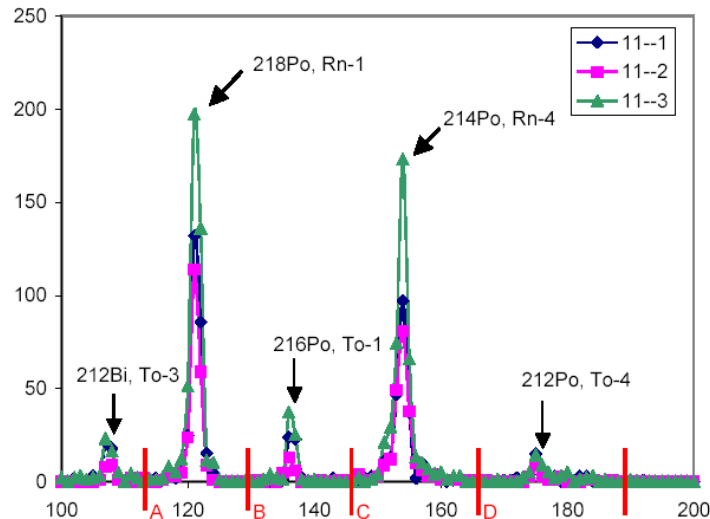


Figure 21.2. The energy spectrum of alpha particles in RAD7 radon detector. The horizontal axis shows the channel numbers that are proportional to the energy deposited in the detector by alpha particles and the y-axis shows the counting rate at a given energy bin. The red lines show the separation of the channels. The three lines correspond to three different measurements in the same detector.

The beta electrons deposit much less energy than the alphas since the detector has a thin thickness and betas go through this thin detector. On the other hand, alphas are stopped in the active area of the detector and leave their total kinetic energy there.

Counting only the events in range A, we can determine the ^{218}Po activity in the detector space. These isotopes are formed in RAD7 interior by ^{222}Rn decay, since the filter at the inlet stops the ^{218}Po -s from outside. The alpha-particles from the ^{222}Rn decay do not reach the semiconductor detector of RAD7, placed in the detector chamber, and their energy does not drop into the range A. The radon concentration is determined, hence, after only its first daughter product. Therefore one has to wait for radioactive equilibrium between ^{222}Rn and ^{218}Po . If we close the radon chamber for 3 weeks, then connect it to RAD7 and its accessories and open the radon chamber to the volume of this system we will have constant (but diluted) radon level for hours. If the radon level in the system (see [Figure 21.1](#)) is constant, the number of ^{218}Po atoms $P(t)$ in RAD7 chamber will follow the equation (A is the constant radon activity in the whole area of the detector chamber, λ_p is the decay constant of the ^{218}Po):

$$\dot{P}(t) = A - \lambda_p P(t)$$

This is a differential equation that is analogue to [Eq. 21.1](#) (see above) therefore its solution is:

$$A_{218\text{Po}} = \lambda_p P(t) = A(1 - e^{-\lambda_p t})$$

This is presented in [Figure 21.3](#). The results show that in the first 15 minutes the activity of ^{218}Po is less than that of ^{222}Rn . In this time span the detector shows less concentration than the

real radon level. After 15 minutes the equilibrium is reached, and the detector output really shows the radon level of the air. At our measurements we collect the counts during 15-minute periods 3 times after each other. The first period does not give useful data, but the others do. The orange time interval in Figure 21.3 shows when the radioactive equilibrium is not reached.

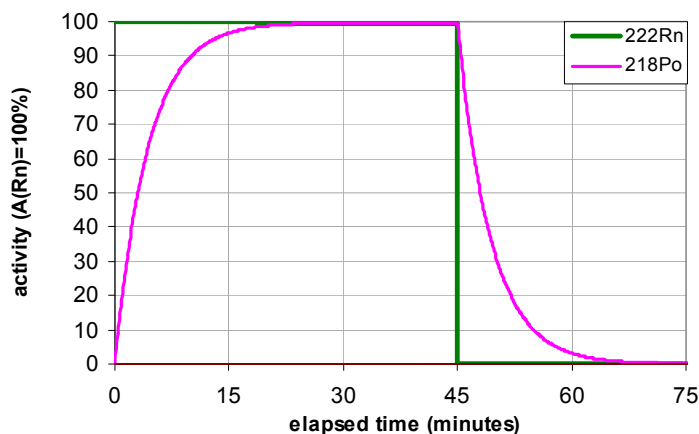


Figure 21.3. The time dependence of the ^{218}Po activity in the detector chamber of RAD7. The green line corresponds to the radon concentration to be measured. At 45 minutes the radon measurement ends, the detector is detached from the radon chamber. After this point the figure shows the decay of the ^{218}Po , and we determine the time when we can start a new measurement.

21.3. Lab Course Tasks

1. Determine the exhalation of the closed radon sample!
 - a) Choose a radon sample that is closed for more than 3 weeks and prepare it for connecting to RAD7 system.
 - b) Set up RAD7 detector to have 15 minutes measuring time.
 - c) Place a filter onto the inlet of RAD7, leave the outlet free. Start the measurement. Measure the background for 15 minutes.
 - d) Determine the geometric sizes of the connecting pipes, desiccant and determine the volumes necessary for the analysis.
 - e) Stop the background measurement, connect the radon chamber and the desiccant and the pipes.
 - f) Open the valves of the radon chamber and start the measurement! Test/Start/enter. Determine the radon concentration for 3×15 minutes, note the values.
 - g) After 3×15 minutes stop the measurement. Calculate the average radon level and the exhalation.
 - h) Investigate the statistical and systematic uncertainties of this measurement!
 - i) Give a lower limit for the radium content of the sample.

2. Investigate the decay of the radon daughters!

- a) Adjust the measuring time for 5 minutes. Setup/Cycle/enter/cursor buttons should be used then Setup/SaveUser/enter.
- b) Remove the pipes from RAD7.
- c) Start an empty measurement with RAD7. Count 7-8 cycles and listen to the displayed concentrations, when there is only room air radon. We can open the windows here for less background.
- d) At the end of these measurements stop them using Test/Stop/enter. Then press Test/Save/enter.
- e) Download the data onto a computer and analyze the concentrations. Special/ComAll/enter.
- f) The data file contains relative percentage of the counts in the A, B, C, D ranges. Determine the number of decayed particles during the 5 minutes from range A and C. Display them separately on a graph.

Test questions

1. What is the radon exhalation and emanation?
2. Why can a sample have high radon exhalation?
3. What is the definition of the exhalation coefficient?
4. What is the radon exhalation measurement protocol?
5. Why is there a dilution effect?
6. How can we register that the system is leaking?
7. What is the relationship between the soil radium content and its radon exhalation?

Summary

In this book we overviewed 17 laboratory practices in the subject of environmental physics. Our measurements mainly covered the area of environmental radiations starting from the acoustic waves, electromagnetic radiation hazard, visible light and going into the area of radioactivity: X-rays, gamma-spectroscopy, annihilation radiation, Cherenkov-radiation, alpha- and beta-spectroscopy. These exercises are good examples for those students who intend to work in laboratories using these spectroscopic or other environmental physics methods.

There are of course lots of areas in environmental physics that were not covered here, but these exercises are adjusted to the technical possibilities of the Environmental Center at Eötvös Loránd University, Budapest.

This subject is more colorful, there are many interesting areas above these. The authors hope that these practices can help to understand the complex behaviour of the processes occurring in our environment.