Maciej Marek

Field and waves...

11 physics lessons that will help you understand how modern cellular networks work

PS This book is not only for farmers and surfers

Projekt finansowany ze środków Ministerstwa Cyfryzacji

Field and waves...

11 physics lessons that will help you understand
how modern cellular networks work

**PS This book is not only for
farmers and surfers**

Project coordinator: Kinga Graczyk

Editor in Chief: dr hab. inż. Maciej Marek

Assistant editor: mgr Robert Bicki

Proofreading: Rafał Pawlak, Anna Maraszek

Graphic design, illustrations, infographics and layout: Vlada Huzieva

Typesetting: Anna Maraszek

Publisher: National Institute of Telecommunications , Szachowa 1,
04-894 Warsaw, phone: +48 22 5128 100, e-mail: info@il-pib.pl,
gov.pl/instytut-lacznosci

ISBN 978-83-916146-8-6

ISBN (Electronic edition): 978-83-916146-9-3

The material was created as part of the project "Efficient mobile telecommunications as a key to development and security" conducted by the Ministry of Digital Affairs in cooperation with The National Institute of Telecommunications – the State Research Institute. The project is implemented under the Operational Programme Digital Poland Measure 3.4. The campaign aims to increase the awareness of Poles in the scope of operation, use, security and importance of mobile telecommunications networks, and thus services based on these networks.

The translation of the physics lesson textbook was financed with funds from the Ministry of Digital Affairs.

Development and preparation: National Institute of Telecommunications.

Maciej Marek

Field and waves...

11 physics lessons that will help you understand
how modern cellular networks work

**PS This book is not only
for farmers and surfers**

Warsaw, 2025

Contents

Preface

Maciej Marek, DSc, PhD, Eng., author of the textbook

Have you ever wondered how mobile telephony works? How voice, text, photos or videos are transmitted from one phone to another?

Mobile telephony is a marvel of engineering, and a detailed explanation of how it operates would fill thick volumes and require many years of study to comprehend.

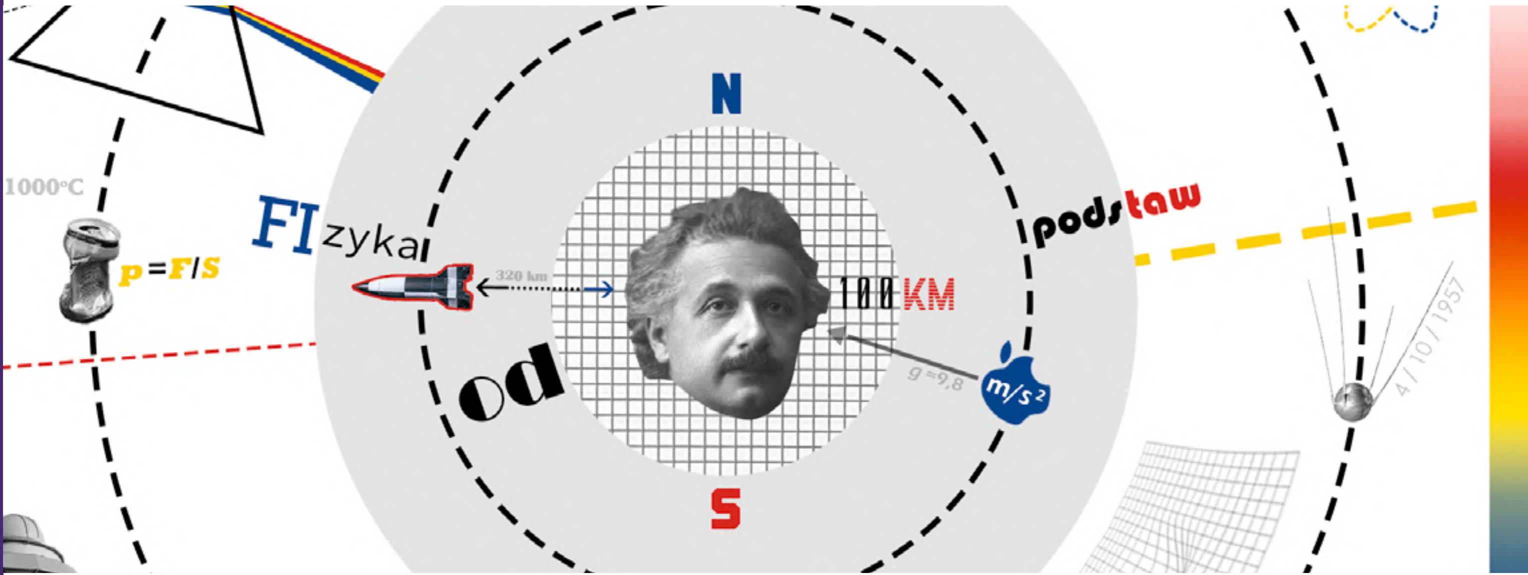
This does not mean, however, that only a specialist can master the basics of this complex system. The eleven physics lessons collected in this book are based on the middle school curriculum, only in a few places boldly going beyond it (e.g. introducing the concept of spectral analysis or signal modulation) to the extent necessary to discuss the transmission of information using the electromagnetic field.

The lessons are not independent of each other. They form one logical sequence, and the concepts introduced in a given lesson are treated as known and well-understood in subsequent lessons. Therefore, it is best to work through the lessons in order, performing experiments and solving problems that aim to consolidate and test the level of understanding of the material. The solutions to the problems can be found at the end of the book. However, it is worth looking at them only after you have faced the problem yourself and after possible tips from the teacher.

Once you have gone through all the lessons – and it will be a long and at times difficult journey – you will be able to answer questions such as:

- Why was the electromagnetic field chosen to transmit information in mobile telephony and other telecommunications systems?
- Where does the name “cellular telephony” come from? What is a "cell"?
- Why do mobile telephony use high signal frequencies? Why do subsequent generations of mobile telephony systems use increasingly higher frequencies?
- Are electromagnetic fields hazardous to health?

And much, much more. We invite you behind the scenes of the mobile telephony scene.



Robert Bicki, MSc – author of the YouTube channel "Fizyka od podstaw" (Physics from scratch)

"Hi! I'm Robert..." - with these words I began the first episode made in Paint and recorded with a microphone from my phone, which appeared on the "Fizyka od podstaw" (Physics from scratch) channel in 2017. A lot has changed since then, and thanks to thousands of hours of work, learning programs and acquired knowledge, I am able to create wonderful materials! This self-development allowed me to participate in a really difficult and engaging project. In cooperation with The National Institute of Telecommunications – the State Research Institute and the Ministry of Digital Affairs, based on the textbook you hold in your hands, I created a series of videos. In them, I discuss issues related to the transmission of information, electromagnetic waves, the interaction of these waves with the environment, the operation of the mobile network, its limitations and the impact on humans. I hope that I managed to explain these difficult issues in an accessible way, using experiments and animations - both two- and three-dimensional ones. In my opinion, they are a beautiful, and often the only visualisation on the Internet of what is usually recorded using patterns, static graphics or presented by waving hands in the air. Let these films inspire you to further discoveries and push you towards exploring exact sciences and telling others about them in an even more attractive way than I do ;).

Fizyka od podstaw

QR code for the mentioned videos on YouTube :)

Rafał Pawlak, expert, National Institute of Telecommunications

Everything that modern science knows about the electromagnetic field and waves has essentially already been said and written down. They have been described by formulas based on complex mathematical apparatus, using, for example, the nabla operator. We are aware that the concepts of divergence, rotation or gradient are not obvious or trivial to everyone. After all, we do not always remember that the rotation of a vector field rotation is the gradient of divergence minus the Laplacian.

That is why in our publication we show that the electromagnetic field is a physical phenomenon that does not necessarily need to be explained starting from Maxwell's laws that arouse fear among students of the initial semesters of polytechnic studies. It can be done much more simply, easily and most importantly - in a more accessible way. In a word, equally effectively, or even more effectively.

We are convinced that after a good reading of the lesson plans and conducting several interestingly arranged experiments, the right time will come for Maxwell's laws, which will certainly be better understood then. And maybe thanks to this, some people will even learn to live in friendship with them? Maybe we will look at our smartphone differently? Maybe we will see in it not only a source of entertainment, but also a source of electromagnetic waves? The answers to these questions belong to you, dear Readers. Interesting reading!

Łukasz Kiwicz, Counsellor, Ministry of Digital Affairs

The mobile phone is an inseparable companion in our lives. Using the opportunities it provides is widespread and taken for granted. We call, write, read, play, check, search. Even the youngest children are already great at navigating the applications available to them. We do all this practically intuitively. But apart from the ability to use it, do we know what makes us able to use our devices? How are we able to call each other? How do we connect to the Internet? Why do we sometimes have problems with network coverage? The answer should be sought in physics. It is on its principles that telecommunications is based.

We have prepared comprehensive lesson plans. Each one, in addition to the substantive part, contains exercises, experiments, a glossary of terms and suggested homework.

We hope that the publication we have prepared will help in understanding and discovering the world of telecommunications. Properly directed knowledge will make young people become conscious users of the network in the adult world.

The material was created as part of the project "Efficient mobile telecommunications as a key to development and security" conducted by the Ministry of Digital Affairs in cooperation with The National Institute of Telecommunications – the State Research Institute. The project is implemented under the Operational Programme Digital Poland Measure 3.4. The campaign aims to increase the awareness of Poles in the scope of operation, use, security and importance of mobile telecommunications networks, and thus services based on these networks.

Lesson 1

Transmission of information

Objective

Presentation of basic concepts related to various forms of information and its transmission.

Learning outcomes

- The student can list examples of different forms of information.
- The student is able to define the concept of a signal and give examples of signal forms.
- The student is able to define and state the basic characteristics of analogue and digital signals.

1. What is information?

We use the concept of information so often in our daily lives that we do not see the need to define more precisely what it actually is. On television, we encounter information programmes; in newspapers – the 'Daily Information' section; on the Internet – information portals. Every large city or tourist centre cannot do without a "Tourist Information". We eagerly ask a friend we have not seen for a long time about interesting information from the recent period.

So what is information? We will not attempt to provide a full, scientific definition of this concept here. Instead, let us try to characterise it in the most useful way for us, so that we can understand what information transfer in modern telecommunications networks is all about.

The basic feature of **information** is that it **reduces the ignorance of the recipient**.

Let us look at a simple example. Let us assume that our friend has tossed a coin in such a way that the result is known to him, but not to us. We are dealing here with the most elementary form of ignorance, because there are only two possible outcomes of a coin toss: heads (H) or tails (T). Therefore, the answer "yes" or "no" to our one question: "is it heads?", will provide us with full information about the result of the toss. If the answer is "yes", we will receive confirmation that it was heads, if "no" - it must have been tails. We call such an **elementary amount of information 1 bit**.

Fun fact. *Science is much better at precisely defining and measuring the amount of information than it is at defining what information actually is. The same is true for many other concepts, such as time and energy.*

How is it in the case of tossing two coins? There are four possible outcomes: HH, HT, TH, TT. So is the amount of information in the result as much as 4 bits, i.e. do we need four questions with a possible answer of "yes" or "no"? Fortunately - no. Let us note that two questions are really enough: "Did the first coin land heads?" and "Did the second coin land heads?" The information about the result of tossing two coins therefore contains only 2 bits.

Exercise. Let us ask a friend to think of a natural number between 1 and 8 and write it down on a piece of paper so that we can't see it. How many questions do we need to ask our friend, at worst, to answer "yes" or "no," to determine what the number is?

Discussion. Of course, if we ask in sequence: "Is this 1?", "Is this 2?", etc., in the worst case, we will need eight questions (if, by some misfortune, the friend chose the number 8). But a cleverer strategy is possible. Let us ask: "Is this a number in the range 1 to 4?" After this one question, we will be able to reject half of the possible results. If, for example, we get the answer "yes", we can ask in the second order: "Is this a number in the range 1 to 2?" Again, we will reject half of the results in this way. For example, the answer "no" will mean that the number we are looking for must be 3 or 4. We can already clarify this ambiguity in the last, third question. Moreover, two questions will certainly not be enough in the general case. Information about the selected natural number in the range 1 to 8 therefore contains 3 bits.

Our ignorance often concerns a finite set of values, as in the examples above. It is not hard to think of other examples:

- the result of a dice roll (range 1 to 6),
- the number of people present in the class (certainly not greater than the number of all students in the class),
- number of free parking spaces (limited by the number of all parking spaces),
- the name of a street in an unknown city where an archaeological museum is located (the number of streets in the city is finite),
- the number on the electoral list of the candidate who won the presidential election (the list of candidates is established by the National Electoral Commission before voting begins), etc.

In each of these cases, obtaining information involves determining which element from a finite set of possible values has actually been realised.

However, we often want to obtain information about values from an infinite set whose elements can change continuously and are not directly countable, e.g.:

- air temperature in degrees Celsius,
- the car's current speed in km/h,
- the height of the tallest building in our city in meters.

Many continuous quantities can be treated as elements of a finite set if we limit the accuracy of their specification and the range of their variability. For example, in the case of air temperature in Poland, we can limit ourselves to the range from -50 to 50 °C with an accuracy of 1 °C. The number of possible results will then be limited to 101 possible values.

Images, photos, films, and pieces of music also carry certain information, regardless of whether we consciously perceive them in this way (e.g., we read from a photo who was present at a certain historical event) or regard them solely as sources of aesthetic impression. Let us note that we can break down any photo into a set of pixels, each of which has a certain colour and brightness. The assignment of colours and brightness to pixels in a photo or image is not known to us in advance and can occur in a huge number of ways. Only when looking at a photo do we observe which of these ways has actually been realised. We receive information in exactly the same way when we observe a dice roll or read the temperature on an outdoor thermometer.

2. Forms of information. Coding

Let us go back to the example of one coin - the friend has flipped the coin, checked the result, but for us the result is still unknown. We ask: "Is it heads?" What will we consider the answer to our question, and thus the transfer of information?

Of course, we expect to hear a specific statement – the word "yes" or "no". From a physical point of view, our friend's speech organs will become the source of an acoustic wave (sound), the characteristics of which we will recognise when it reaches our ears as one of these words.

But there are many more possibilities. A friend can:

- nod or shake his head,
- write down "yes" or "no" on a piece of paper and show us the answer,
- go to the next room where there is a microphone connected to the loudspeaker in our room and transmit information using this system,
- may knock on the table once or twice.

The first way is considered as good as a direct yes/no answer. In our culture, we clearly associate these gestures with confirmation or denial, respectively. The second way requires us to be able to read, but it essentially conveys the same information. Note that in both of these ways, we perceive information as visual impressions on a physical level transmitted by light.

In the third method, the acoustic wave does not reach us directly. First, the acoustic wave is converted into electrical voltage variations by the microphone, which are transmitted via wire to the loudspeaker, where they are converted back into an acoustic wave that reaches our ears (Fig. 1).

The fourth method, in order to convey useful information for us, requires prior agreement on what one and two knocks on the table mean, because this assignment can be completely arbitrary. If we establish that one knock means "yes" and two knocks mean "no," we obtain information analogously to nodding or shaking the head.

We can see that, firstly, information can have different physical bases (acoustic wave, visual stimulus, variable electrical voltage), secondly, it can be "encoded" in different ways, and the code must be known to both the sender and the recipient for the information to be transmitted effectively. Encoding is the assignment of characters from one set of characters to another set (e.g. "yes" - nod, "no" - head shake; Morse code, etc.).

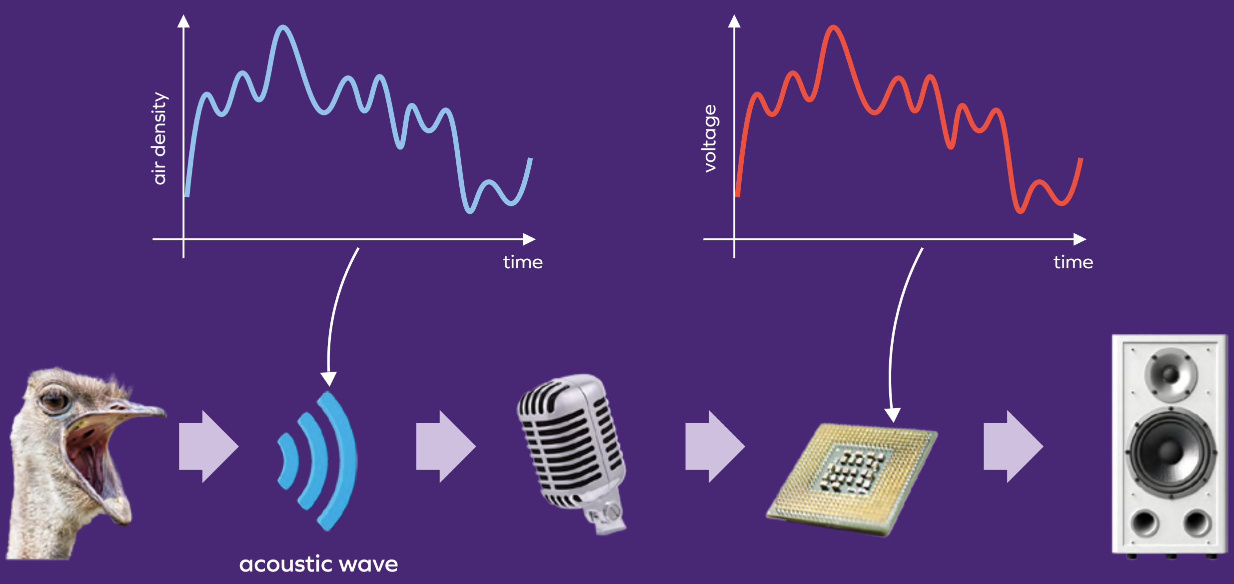


Fig. 1. Various forms of the signal carrying the same information.

A physical quantity that changes in accordance with the information carried will be called a **signal**. Between the sender and the receiver, the signal can change its form many times without actually changing the information carried.

3. Signal examples

The transmission of voice information is one of our fundamental ways of communicating. Let us take a closer look at how information is transmitted using a voice message as an example.

It all begins with a thought, which consists of neural impulses in the brain. The sender's intent is directed towards articulating a specific word, for instance, 'cat'. (Fig. 2).

- The sender's brain, via the nervous system, transmits impulses to the speech organs (electrical signal).
- These impulses lead to the appropriate positioning of the tongue, lips, jaw and tension of the vocal cords when exhaling air. Note that the characteristic positioning of the speech organs is also a kind of signal - we know that some people have acquired the ability to read from lip movement (visual signal).
- The air flow pushed out by the falling chest stimulates the vocal cords to vibrate (vibrational signal). The positioning of the oral and nasal cavities and lips creates a resonance chamber and amplifies the appropriate frequencies (vowel emission). The movements of the tongue and lips block the air flow (consonants) at the appropriate moments. A specially shaped acoustic wave is emitted from the mouth (sound signal).
- Sound spreads in the air around the sender, as compressions and rarefactions of the air. The role of the signal is therefore played by the variable density of the air.
- The sound reaches the recipient’s ear, causing the eardrum to vibrate (vibrational signal).

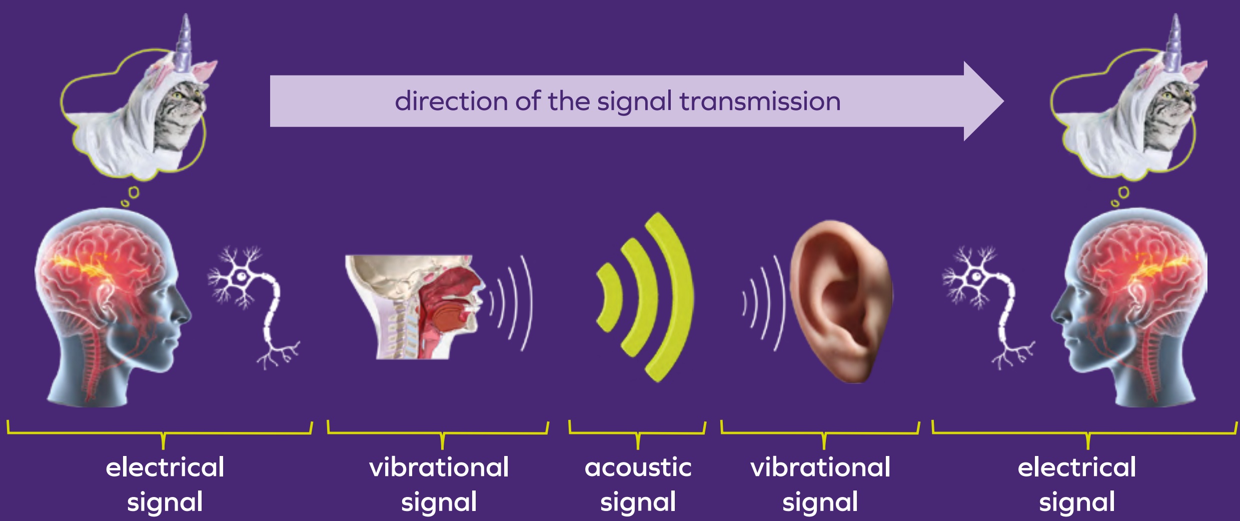


Fig. 2. Voice messaging as the transmission of information through signals of varying forms.

- The membrane vibrations are converted into electrical signals in the auditory nerve and through the nervous system reach the recipient's brain, which interprets them as the word "cat" (an electrical signal). The information has been transmitted successfully.

A conversation is a form of information exchange that takes place in turns. Telephony allows conversation to take place at virtually any distance between the interlocutors. In classic telephony, air vibrations are converted into variations in electric current in the wire. In mobile telephony, sound is converted into electromagnetic field vibrations (see Lesson 2).

Exercise. Recognise what constitutes a signal in different ways of transmitting information, i.e. what physical quantity or quantities change along the information transmission route between the sender and the recipient:

- Flashing a flashlight.
- Knocking on the wall into the next room.
- Train station announcements regarding train delays (you can consider two types of announcements: those displayed on screens and those broadcas through loudspeakers).

4. Analogue and digital signals

As we saw earlier, we can deal with signals that can take on only a few possible values, or signals that can essentially take on any value within a specified range.

Analogue signal is a signal whose value can change continuously. For example, the result of measuring the air temperature by an electronic thermometer located outside the window can be converted into an electric voltage in the wire connecting the thermometer with the equipment inside the room. If the voltage value is proportional to the temperature value, the signal in the wire is an example of an analogue signal.

On the other hand, a signal whose value is limited to a finite set of values is called a **digital signal**. Very often, digital signals take only two values – we then speak of a two-value or binary signal. For example, a signal transmitting the results of a multiple coin toss would be a **binary signal**. It could be an electrical signal with two voltage values: "heads" - 1 V, "tails" - 5 V.

In future lessons, we will see that in some cases we can convert an analogue signal to a digital signal without losing information (Lesson 9).

Both types of signals have their advantages and disadvantages. The advantage of the analogue signal is that it directly reflects the variability of information that is continuous (temperature, car speed, etc.). Continuous information can be easily converted to an analogue signal at the sender and then easily recreated at the recipient. The disadvantage of the analogue signal is its high sensitivity to interference. If noise (random interference with an unpredictable character) is added to the useful signal, the recipient is not able to easily distinguish useful information from noise, i.e. perform the so-called denoising. The variability of the signal value may result from both the information content (e.g., changes in temperature values) and random changes in noise introduced somewhere between the sender and the recipient (Fig. 3). We will talk about some possibilities of eliminating noise in analogue signals in Lesson 4.

The digital signal does not have this disadvantage, as long as the interference does not exceed half the value of the interval between the defined signal levels. Even if the signal is interfered with, we are still able to recognise what signal level was actually emitted by the sender (Fig. 3). Another advantage of the digital signal is that it can be easily processed by computers and digital electronic systems (e.g., digital filters). The disadvantage is the need to introduce appropriate converters to process continuous information into digital form (at the sender) and in the opposite direction (at the receiver). Usually, such a processed signal bears no resemblance to the waveform of the physical quantity representing the original information.

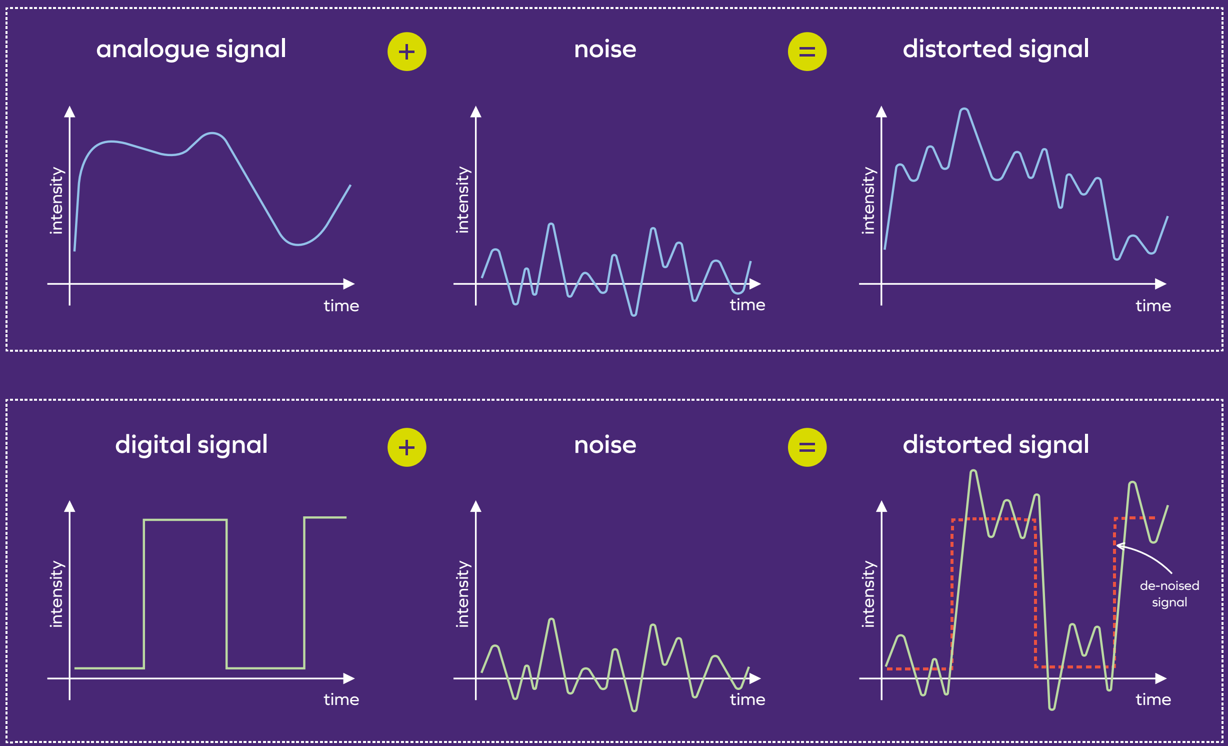


Fig. 3. Comparison of the susceptibility to interference of analogue and digital signals.

Experiments

These experiments require five participants. Two will act as the sender and receiver of information, the other three as jammers, i.e. participants who introduce some disruptions in the communication link between the sender and the receiver. The sender and the jammers receive a dozen or so blank cards and a marker. The sender and the receiver sit opposite each other, and the jammers sit between them (Fig. 4).

Fig. 4. Experiment with information transmission.

Experiment 1

The sender writes down any number (it doesn't have to be an integer) between 10 and 50 on a piece of paper so that no one else can see it, then passes the paper to the first jammer. The sender reads the number surreptitiously, then writes down the same number on his paper or increases it by 1 or decreases it by 1. Then he passes his paper to the second jammer, who can do exactly the same, then passes the paper to the third jammer, who, after possibly changing the number received from the second jammer, passes his paper to the recipient. The experiment can be repeated several times.

Try to answer the questions:

- What type of signal was transmitted here – analogue or digital?
- After reading the numbers that have been sent to him, is the recipient able to determine what numbers the sender wrote on his cards?
- If not, can he estimate it? How far off could he be in the worst case?

Experiment 2

The course of the experiment is exactly the same as in Experiment 1 with one small difference – the sender can write only one of two numbers on the card: 10 or 50. The information passes through jammers who can change the received number as before and then reaches the recipient. The experiment can be repeated several times.

Now try to answer the questions:

- What type of signal was transmitted here: analogue or digital?
- After reading the numbers that have been sent to him, is the recipient able to determine what numbers the sender wrote on his cards?
- Can it do this flawlessly? What if there were many more jammers? How many would there have to be for the recipient to lose confidence in reproducing the transmitted information?

Glossary

Bit – the amount of information contained in a "yes" or "no" answer to a question.

Information – a set of data that reduces the recipient's ignorance.

Coding – assignment of characters from one set to characters from another set.

Signal – a physical quantity that changes in accordance with the transmitted information.

Analogue signal – a signal whose value can change continuously.

Binary signal – a signal that can only take on two possible values at the time of transmission; it is a type of digital signal.

Digital signal – a signal that, at the moment of transmission, can only take on values from a finite set.

External materials

Scan QR code

1. Elżbieta Kuligowska "Everything in the Universe is information? The most important facts about apparent obviousness.".

2. Akinator app. The player thinks of a character (real or fictional), and the app's task is to guess that character in a series of questions to which the player answers yes/no. Think of any character and determine how many bits of information you had to provide for the app to guess the correct answer (you can assume that the answer "I don't know" does not convey any information, while "I think so" or "I think not" are equivalent to "yes" or "no").

Scan QR code

Homework

Problem

In a closed room, there is a person who taps a message encoded in Morse code on the wall. In a room on the other side of the wall, a second person with a flashlight listens to the message and continuously transmits the same content outside through the window by flashing the flashlight. Arrange in order the successive forms of the signal carrying information (similarly to Fig. 2) from the sender in the closed room to the emission outside and comment in detail on the individual stages.

Concepts to use:

- Electrical signal (nervous system).
- Electrical signal (flashlight electrical system).
- Electrical signal (brain).
- Acoustic signal (sound wave in air).
- Light signal (bulb).
- Vibration signal (eardrum in the ear).
- Vibration signal (wall).

Lesson 2

Electromagnetic field as an information carrier

Objective

- To familiarise students with the possible applications of the electromagnetic (EM) field in the transmission of information.

Learning outcomes

- The student is able to explain why EM waves are currently widely used as an information carriers.
- The student is able to explain the basic mechanism of generating EM waves.
- The student is able to use the law that quantitatively describes the decrease in the electric field strength, and EM wave intensity with distance from the source.

1. Long-range physical fields

In Lesson 1, we were introduced to the issue of information transmission and analysed a voice message as one of the fundamental methods of exchanging information in everyday life. Direct voice communication, however, has a major disadvantage – very limited range. Classic telephony is largely free of this disadvantage but requires a wired connection between the sender and the recipient. For mobile, wireless communication, it is necessary to use physical fields with a large range. In physics, we know two such fields: the gravitational field and the electromagnetic field.

The gravitational field as an information carrier would be ideal – as far as we know, there is no way to screen it, so it penetrates all obstacles and can reach any place in the Universe. Unfortunately, gravitational interactions between masses that we can easily control are extremely weak. According to Newton's law of gravitation, the force of attraction between two masses of mass $m = 10\,\mathrm{kg}$ located at a distance $r = 1\,\mathrm{m}$ is (G – gravitational constant):

$$F = G\frac{m \cdot m}{r^2} = 6.67 \cdot 10^{-11}\,\frac{10 \cdot 10}{1}\,\mathrm{N} = 6.67 \cdot 10^{-9}\,\mathrm{N}$$

It is therefore extremely small even at small distances and it is difficult for a force of this order to cause a detectable reaction in the information receiver. Significant gravitational effects are observed only at masses comparable to small moons. However, it is difficult to force such massive objects into a specific motion in the information transmitter. Therefore, we do not see the possibility of using the gravitational field as a practical information carrier, at least with the current state of knowledge.

The situation is different in the case of **the electromagnetic field (EM)**, which is generated by electric charges.

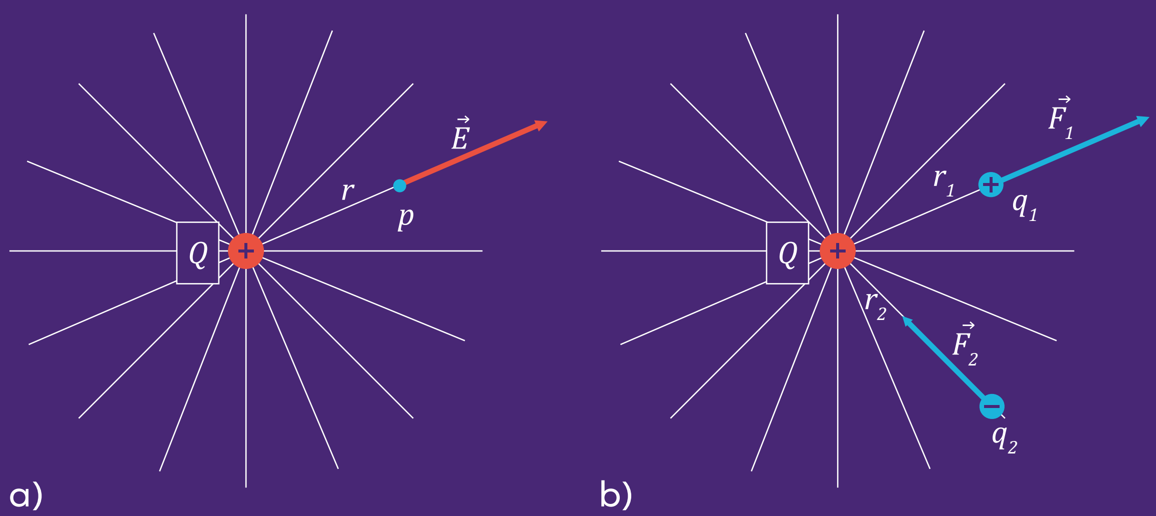


Fig. 1. Electric field around a positive charge: a) the electric field vector at the point p; b) forces exerted on the test charges q_1 and q_2.

2. Electric Field and Coulomb's Law

Each electric charge is a source of an electric field extending from the charge to infinity. At any point p of space around the charge Q we can introduce the **electric field vector** $\vec{E}$ the direction of which for a stationary charge is defined by the straight line connecting point p with charge Q, while the sense for a positive charge is chosen so that the vector is directed from the charge to infinity (Fig. 1a). The magnitude of the vector $\vec{E}$ (i.e. the field strength E) is calculated from **Coulomb's law**:

$$E = k\frac{Q}{r^2}$$

where $k = 9 \cdot 10^9\ [\mathrm{N} \cdot \mathrm{m}^2 \cdot \mathrm{C}^{-2}]$ is a physical constant for a vacuum. Electric charge is measured in SI units called *coulombs* (C). The strength of the electric field decreases inversely proportional to the square of the distance r of point p from charge Q. This means that doubling the distance from the charge reduces the field strength fourfold. On the other hand, the field strength is proportional to the value of the charge that is its source (doubling the charge doubles the field strength). If the charge is placed in a medium other than vacuum, the value of the constant k may be different (for air the differences are insignificant).

By inserting the value of the charge Q in coulombs and the distance r in meters into Coulomb's law, we obtain the value of the field strength in N/C (newtons per coulomb), but we also often use another equivalent unit - volts per meter:

$$[E] = \frac{\mathrm{N}}{\mathrm{C}} = \frac{\mathrm{V}}{\mathrm{m}}$$

The electric field is illustrated by drawing so-called **field lines**, i.e. lines starting in the charge and directed along the field vector at a given point in space. For a stationary charge, this means that the field lines are rays radiating from the source to infinity (Fig. 1). If the charge moves non-uniformly, the image of the field lines can become much more complicated, as we will see later in the lesson.

Knowledge of the field vectors allows us to find the force acting on other charges placed in the electric field (Fig. 1b). If we place a charge q, in the field, the value of the force acting on it can be calculated from the formula:

$$\vec{F} = \vec{E}q$$

$$F = Eq = k\frac{Qq}{r^2}$$

very similar to the law of gravity, but instead of the masses of interacting objects, electric charges appear.

3. EM wave generation

Just as sound information is encoded in the vibrations of a medium (like air), it can also be represented by vibrations of the electromagnetic field. Of course, in this case no medium is needed.

Let us try, at a conceptual level, to explain the formation of a propagating disturbance of the EM field, which we will also call an EM wave, analogous to an acoustic wave in air. It should be noted here that the exact analysis of this phenomenon is not simple and requires quite advanced mathematics to solve, the so-called Maxwell's equations. We will try to skip these complexities without losing the physical essence of the phenomenon itself.

Let us imagine a stationary electric charge placed at point p_1 (Fig. 2). The electric field lines generated by this charge are marked in blue. At some point, the charge moves up towards point p_2 and becomes stationary again. If it stayed there for a very long time, the field lines should run like those marked in orange.

Why do we have to specify that this will only happen after a long time? This is because no signal can travel faster than the maximum speed, which we identify with the speed of light in a vacuum, conventionally denoted by c and equal $3 \cdot 10^8$ m/s. Note that the field lines, as rays radiating from the charge to infinity, contain information about the charge's location. Even in a distant corner of the Universe, where the field lines also reach, we can take two different field lines, determine their point of intersection and expect an electric charge exactly at that point. If the displacement of a charge from the point p_1 to the point p_2 caused all the field lines to shift simultaneously to the location marked in orange, we would immediately know in a distant corner of the Universe that the charge had moved.

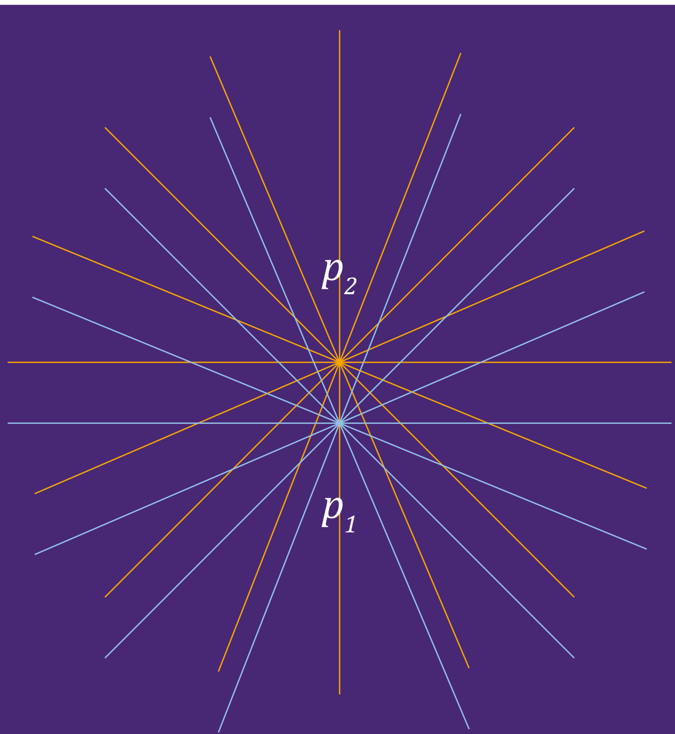


Fig. 2. Field lines of a charge placed at point p_1 and field lines of the same charge displaced to point p_2 (after a sufficiently long time).

This would therefore contradict the existence of a limited speed of information transfer. In reality, information about the movement of a charge should reach distant points with a delay, the greater the distance of the observation point from the moving charge.

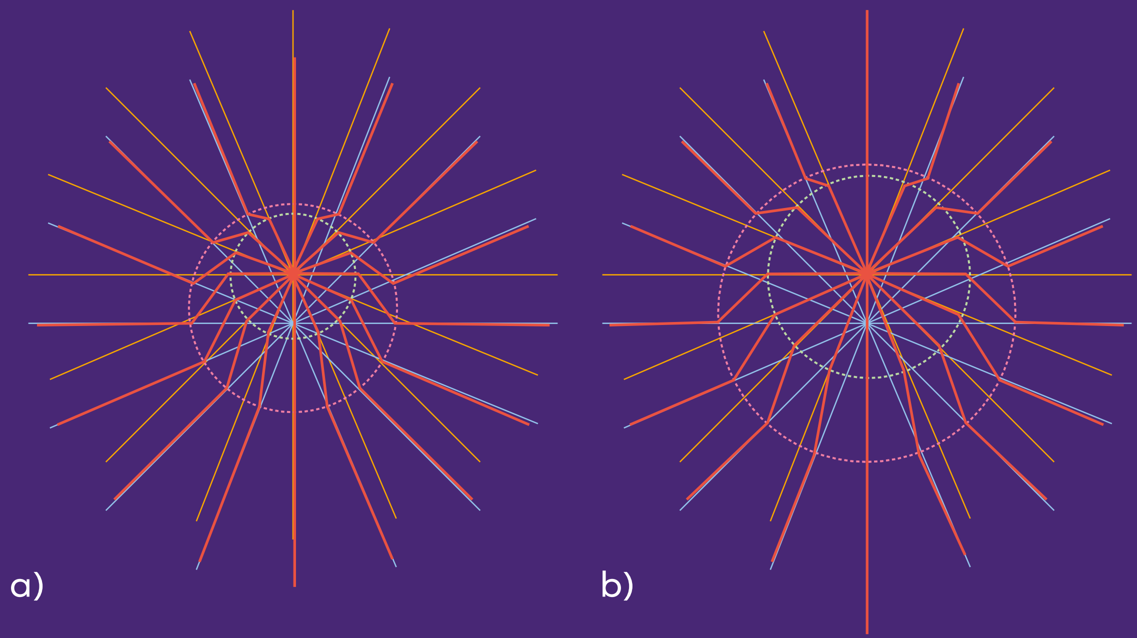


Fig. 3. Path of the electric field lines after displacing an electric charge after a certain moment (a) and after a slightly longer time (b).

So what will it really look like? Let us assume that the charge has moved to point p_2 and look at the field lines after a very short time. Information about the new position of the charge will start to spread at speed c so that it will encompass all points inside the circle marked in green in Fig. 3a. We can therefore direct the field lines inside this circle along the orange lines corresponding to the new position of the charge (red lines). However, the information about the charge's movement could not yet reach the points outside the circle marked in pink. The radius of this circle depends on the time that has passed since the charge started moving. Outside the pink circle, the field lines must therefore run along the blue (initial) lines.

What happens between the green and pink circles? Unfortunately, it is quite complicated. In this region, a rapid change of the electric field occurs, and thus a magnetic field is generated (so far we have ignored it in our discussion). This is an area that we can identify with the **EM field disturbance** – an **electromagnetic wave** propagating from the moving charge towards infinity. Since the radii of both circles increase at a speed of c, so does the speed of the EM wave. For simplicity, we will mark the electric field lines in the disturbance region by connecting the corresponding blue and orange field lines into one broken line (let us remember, however, the conventional nature of this representation).

After a while, both circles will grow larger and the field disturbance area will move further away from the charge generating the field (Fig. 3b). After a sufficiently long time, the green circle will move outside the drawing area, so the field lines will run exactly along the orange lines as in Fig. 2.

The energy of the EM wave is greatest where the field changes most rapidly. We can identify these places approximately, even based on the simplified image from Fig. 3, as the area containing the most "broken" field lines. As we can see, the field lines directed along the straight line on which the charge was moving (the vertical line) are not refracted at all. In this direction, no energy emission occurs. Conversely, in the direction perpendicular to the charge's motion, we observe the greatest refraction of the field lines – here, the energy emission is the highest. Fig. 4 illustrates the differences in the energy distribution of the EM wave in the region of the field disturbance through different shading of the area.

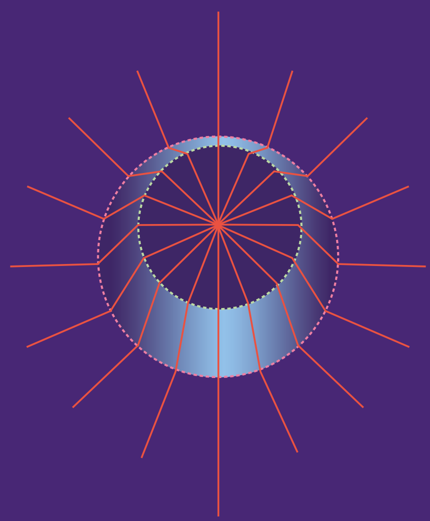


Fig. 4. The shading of the region of EM field disturbance corresponds to the wave energy (the darker the area, the greater the energy).

We can see that a charge moving in a straight line does not emit energy evenly in all directions. Although we will often refer to the analogy of waves on the surface of water or acoustic waves when analysing EM waves, remember that in this respect an EM wave differs from a circular wave propagating on the surface of water. This is a very important conclusion from the point of view of antenna design, which we will discuss later in the lesson.

4. EM wave intensity and its variation with distance from the source

We must note that the above analysis was performed for simplicity in one plane (the plane of the drawing). In reality, the field lines radiate from the charge in all directions of the three-dimensional space, and the circles marked in Fig. 3 and Fig. 4 are actually spheres with radii increasing in time. Let us see what effect this has on the variability of the EM wave energy in time and space as it moves away from the source.

As we mentioned earlier, the energy of the EM wave is concentrated between the green and pink spheres (let us call them that now). This disturbance of the EM field caused by the movement of the electric charge has somehow separated from this charge and begins its journey towards infinity. The energy of the wave cannot change in accordance with the principle of conservation of energy if it is not absorbed by the medium in which it propagates (in a vacuum, absorption does not occur, but in matter it can). However, the spheres on which this energy is distributed (unevenly!) increase with each moment in time. If we choose a point on a sphere (e.g., the green one), then as the sphere expands, wave energy at that point decreases. As a certain analogy, we can recall buttering bread. Taking a certain amount of butter, if we spread it on larger and larger slices of bread, the amount of butter in the selected place (e.g. the middle of the slice) will be smaller and smaller.

The amount of energy carried by an EM wave is related to the so-called **wave intensity** denoted with the letter I.

Warning: let us not confuse the wave intensity with the electric field strength! A thorough analysis in EM field theory shows that the wave intensity I and field strength E are related by the relation:

$$I \propto E^2$$

i.e. the wave intensity is directly proportional to the square of the field intensity.

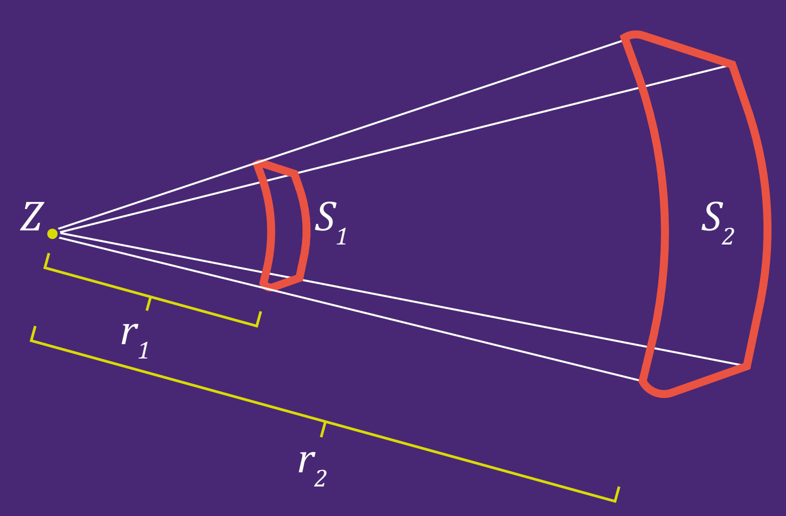


Fig. 5. Analysis of the energy distribution on expanding sections of the sphere.

The field intensities on selected sectors of the expanding sphere are inversely proportional to the areas of these sectors (Fig. 5):

$$\frac{I_2}{I_1} = \frac{S_1}{S_2}$$

On the other hand, the areas of the sector are related to each other as the squares of the radii:

$$\frac{S_1}{S_2} = \frac{r_1^2}{r_2^2}$$

Why? We can easily understand this if we approximate the sphere segments with squares. The linear dimensions of a square grow proportionally to the radius (geometric similarity of the corresponding figures). However, we know that the area of a square is equal to the square of its side – hence the proportionality of the area of the figure to the square of the sphere's radius.

Ultimately:

$$\frac{I_2}{I_1} = \frac{S_1}{S_2} = \frac{r_1^2}{r_2^2} \quad \Rightarrow \quad I_2 = I_1 r_1^2 \frac{1}{r_2^2} \propto \frac{1}{r_2^2} \quad \Rightarrow \quad I \propto \frac{1}{r^2}$$

i.e. the intensity of the EM wave decreases inversely proportional to the square of the distance from the source. Since, as previously mentioned, the intensity of the wave is proportional to the square of the field intensity, the value of the field strength in the region of the EM field disturbance decreases inversely proportional to the distance from the source:

$$E_{EM\ wave} \propto \frac{1}{r}$$

Let us emphasise – by $E_{EM\ wave}$ we understand the value of the field strength (varying in time) in the region of the EM wave, not the value of the field strength from a stationary electric charge, which (as we remember from Coulomb's law) decreases inversely proportional to the **square** of the distance from the source. The difference in this variation is huge (see "Homework"). The influence of the EM wave on distant electric charges is therefore much greater than it might seem based on Coulomb's law.

5. Transmitting and receiving EM signal

How do we construct EM wave transmitters in practice? Fig. 6 shows an example of a so-called **dipole antenna**, i.e. two segments of wire connected to an alternating current source. These segments do not form a closed circuit, so direct current flow is not possible, but our goal in this case is only to induce the movement of charges, first in one direction, then in the other. Fig. 6a shows a situation in which the current flows towards the upper section, charging it positively. At the same time, the second antenna wire is negatively charged. In the next cycle, the current flows in the opposite direction and the lower section of the antenna is positively charged (and the upper section is negatively charged). Then the cycle repeats. The movement of charges in the antenna thus becomes

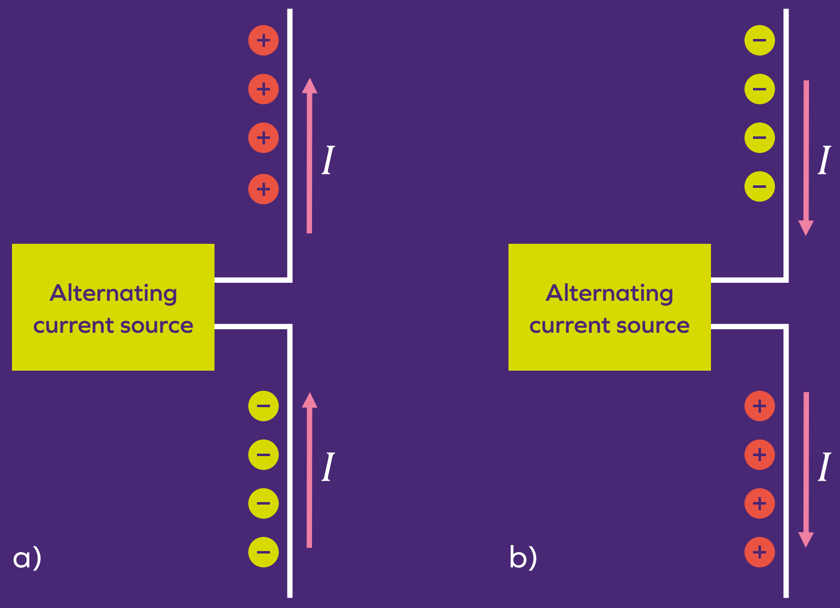


Fig. 6. **Dipole antenna as a transmitter of EM waves. Current flow toward the upper segment of the antenna (a) and in the opposite direction (b).**

a continuous source of EM waves, in contrast to the single, momentary movement of the charge as in Fig. 4, where the emission of a single, short field disturbance is shown.

We will discuss how to use the induced wave to transmit useful information in Lesson 7. Now let us just look at the design of the electromagnetic signal receiver. In Figure 7, we see on the left a transmitter in which an alternating current source connected to a dipole antenna is controlled by a specific primary signal consistent with the information we want to transmit. An alternating current will be induced in the antenna, which generates an EM wave that propagates around the antenna. As we have already noted, the main direction of energy emission is the direction perpendicular to the direction of charge movement, i.e. in this case, it is the direction perpendicular to the antenna.

The receiver is constructed very similarly to the transmitter, except that instead of a power source, a suitable circuit is placed for measuring and possibly amplifying the received signal. An EM wave reaching the dipole antenna in the receiver causes the movement of charges (electrons under the influence of $E_{EM\ wave}$) in the antenna wires to be stimulated – an alternating current is therefore induced in them. If the receiver is far from the source, the strength of the EM field reaching the antenna may be very small, but after amplification it is perfectly measurable. In this way, the signal is transmitted wirelessly to the recipient and can be further processed and interpreted.

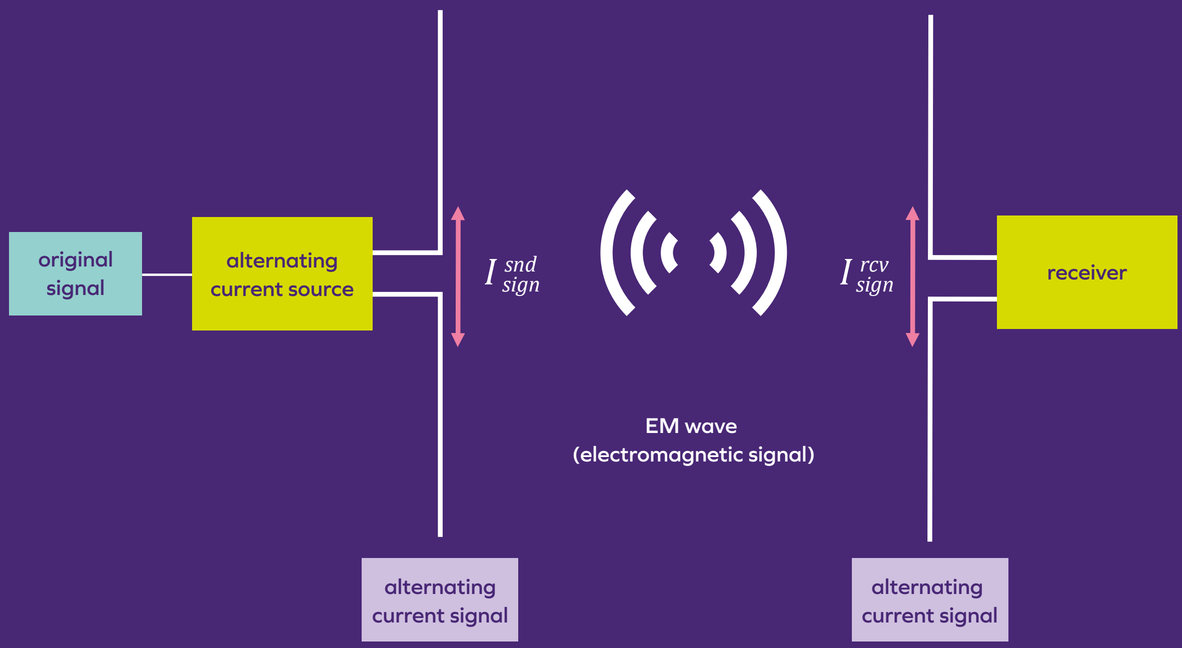


Fig. 7. Transmitter and receiver of electromagnetic signal.

We can therefore justify the widespread use of electromagnetic fields in transmitting information for the following reasons:

- ease of generating EM waves and shaping them according to information content,
- the long range of EM waves and the lack of a medium for their transport (they propagate even in a vacuum),
- the ability to easily receive EM waves even at very low intensities.

Experiment

The aim of the experiment is to confirm the quantitative dependence of the wave intensity on the distance from the source. We can perform the experiment on a desk with a lamp on a boom and a long fluorescent lamp. The lamp will act as the source of the EM wave (we will discuss the fact that light is in fact an EM wave in Lesson 6). We wrap the lamp in black cardboard, in which we cut out two square holes, one with a side $1\,\mathrm{cm}$, the second one on the side $2\,\mathrm{cm}$ (Fig. 8). This makes the surface area of the second hole four times larger than the first.

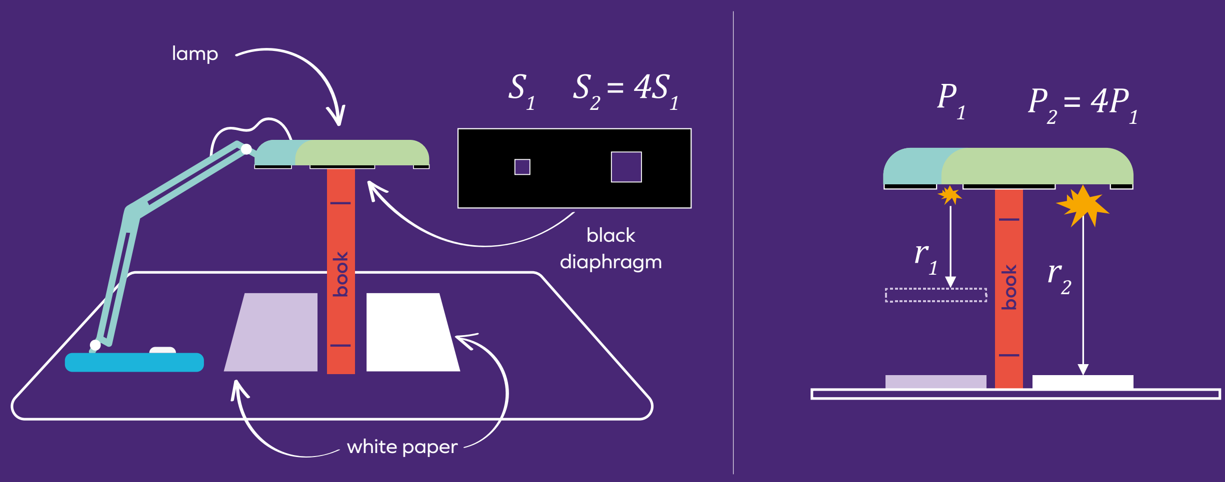


Fig. 8. **Experimental setup.**

We place a thick, large book vertically on the desk, then move the lamp close to the book so that the light emerging from both holes in the cardboard screen illuminates only the area lying on the appropriate side of the book. We place two white sheets of paper on the desk on both sides of the book, then darken the room so that the lamp is the only source of light.

We should notice that the sheet of paper on the side of the smaller hole is clearly less illuminated. The holes in the diaphragm play the role of secondary light sources, with their power being proportional to the area of the hole. Since the area of the right hole is four times larger than the first one, the power of the light source on the right is also four times larger, and since the sheets are at the same distance from the sources, the intensity of the wave incident on the sheet of paper on the left is the same number of times smaller.

Now let us stand in front of the desk so that we can easily assess the level of illumination of both pages. Let us slowly start lifting the page on the left side, constantly comparing the level of illumination of both pages. When we assume that both pages are illuminated to the same extent, let us measure the distance r_1 between the lifted page and the diaphragm and compare it with the height of the book (equal to r_2, which is the distance of the right page from the diaphragm).

Is the ratio of these distances close to ½, in other words, has the page been lifted approximately halfway up the book?

Note that according to the law of decreasing wave intensity with distance, it should decrease inversely proportional to the square of the distance. On the other hand, we expect a directly proportional dependence of the wave intensity on the source power P. It should therefore be:

$$I = a\frac{P}{r^2}$$

where a is a constant multiplier. The wave intensity (illumination) of both sheets of paper can therefore be calculated as:

$$I_1 = a\frac{P_1}{r_1^2}, \quad I_2 = a\frac{4P_1}{r_2^2}$$

and since both intensities are equal (with the left card lifted), we get:

$$I_1 = I_2 \quad \Rightarrow \quad a\frac{P_1}{r_1^2} = a\frac{4P_1}{r_2^2} \quad \Rightarrow \quad \frac{r_1^2}{r_2^2} = \frac{1}{4} \quad \Rightarrow \quad \frac{r_1}{r_2} = \frac{1}{2}$$

Glossary

Dipole antenna (dipole) – a system of two symmetrically arranged wires. When powered by alternating current, it emits an EM wave most strongly perpendicular to the wire axis. It can also function as an EM receiver.

Electromagnetic wave (EM) – disturbance of the electromagnetic field caused by accelerated electric charges.

Field lines – lines in space directed along the field vector; in the case of a stationary electric charge, these are rays attached to the charge and spreading out radially to infinity.

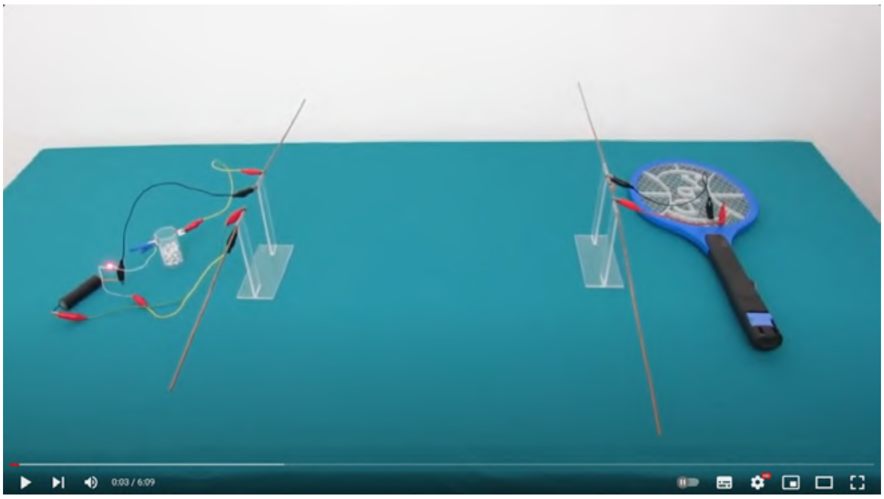
Electric field – a vector whose value depends on the force with which the field acts on a unit positive test charge; calculated using Coulomb's law; its SI unit is [V/m].

Electromagnetic wave intensity – a quantity describing the energy carried by a disturbance in the EM field. Unit – W/m^2 (watt per square meter).

Electromagnetic field (EM) – a physical field with an infinite range, generated by electric charges.

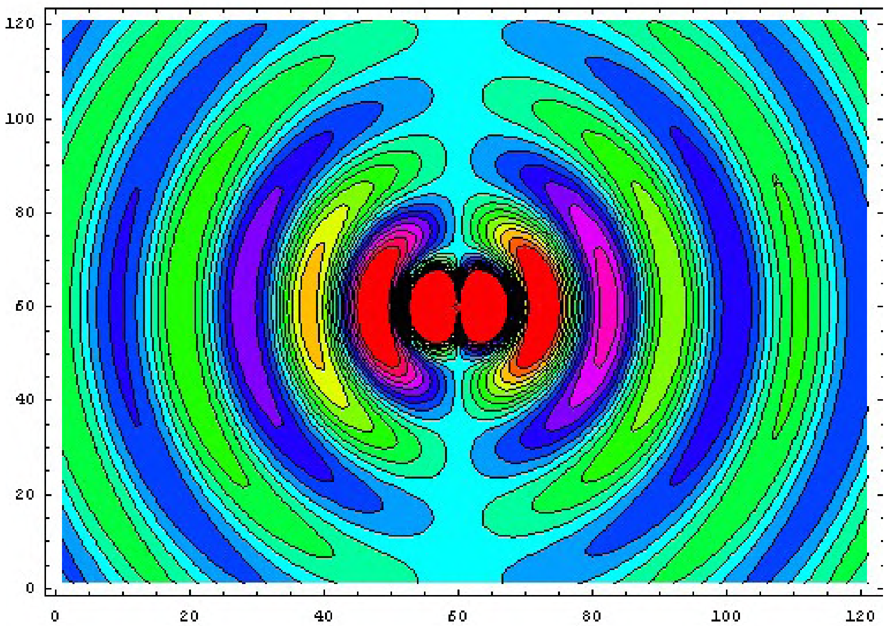
External materials

1. A simple experiment demonstrating the transmission of EM waves (*Science Project – Easy experience to detect electromagnetic waves*).



Scan QR code

2. Animation showing EM wave emission by a dipole antenna. Source: Wikipedia. File name: Dipole.gif



Scan QR code

Homework

1. Compare the $1/r$ type variation with the $1/r^2$ type variation. Calculate the values of these expressions for different r and fill in the table below. Do the obtained results allow you to compare the value of the field strength of a stationary charge (calculated from Coulomb's law) at a large distance from the charge to the value in the region of the EM wave caused by the instantaneous motion of this charge?

r	1	2	5	10	100	1000
$1/r$						
$1/r^2$						

2. Sketch the pattern of the field lines for a charge that at a certain moment starts moving from the point where the blue lines intersect to the point where the orange lines intersect. The radius of the pink circle increases with speed c from the moment the charge starts moving, and the radius of the circle increases with the same speed from the moment the charge reaches the destination point (a drawing similar to Fig. 3):

Lesson 3

Wave motion

Objective

- Introduction to basic concepts of wave motion.

Learning outcomes

- The student is able to name the main quantities describing a harmonic wave.
- The student knows the relationships between the quantities describing a harmonic wave and is able to use them in practice.
- The student is able to indicate the relationship between a harmonic wave and the motion of a point along a circle.
- The student is able to calculate the minimum antenna size at which energy-efficient transmission of an electromagnetic signal can be achieved.

1. Harmonic wave and its characteristics

By **wave motion** we mean a propagating disturbance of some elastic medium or physical field. Examples of wave motion often include water waves or sound, i.e. an acoustic wave. The electromagnetic wave we introduced in Lesson 2 does not require a material medium – it is a disturbance of an electromagnetic field that is present even in a vacuum.

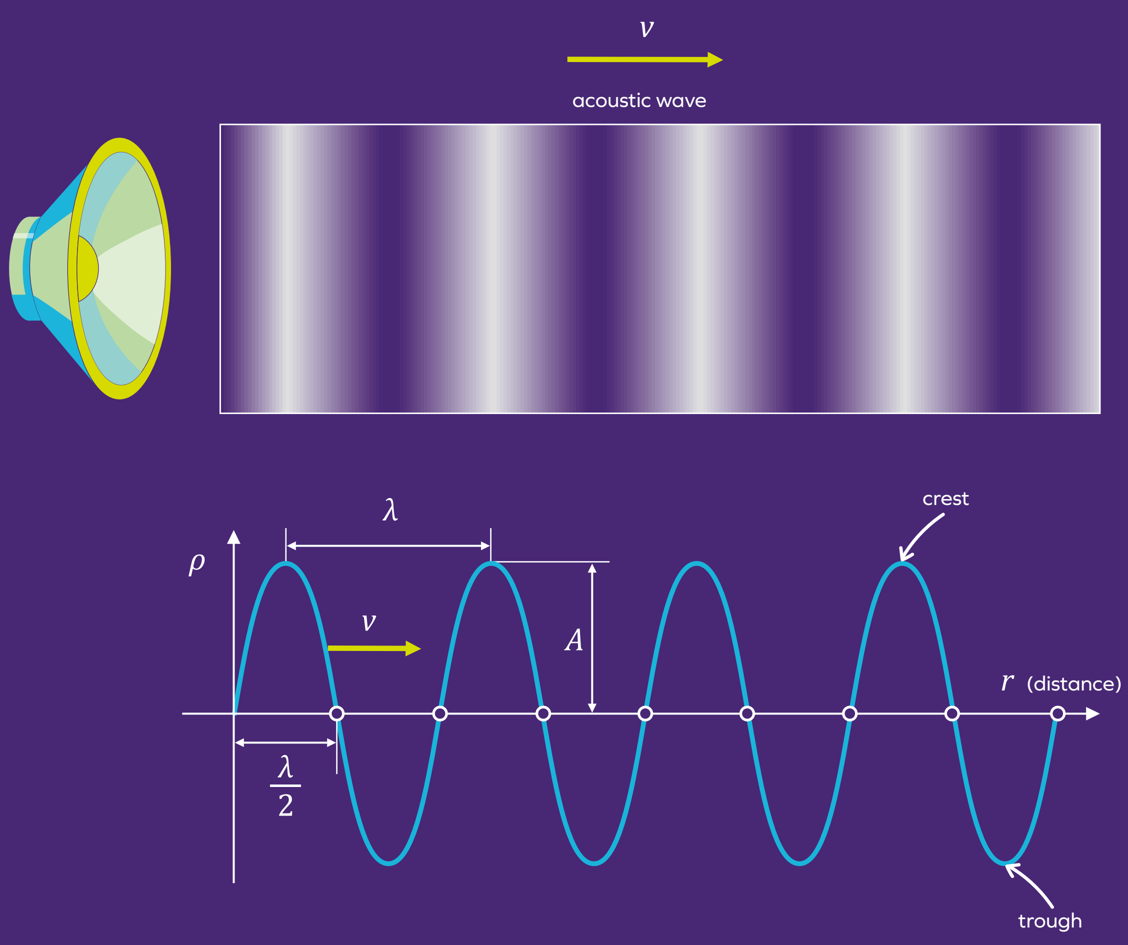


Fig. 1. Wave motion illustrated by an acoustic wave. Key parameters of a harmonic wave.

Acoustic wave is a series of air compressions and rarefactions. It can be generated by a loudspeaker membrane moving in a vibrating motion. The membrane moves so fast that the air in its immediate vicinity is not immediately pushed away (as in the case of a hand wave, for example) but is compressed – its density increases above the average air density. Only after a while does the compressed air push the further layers of air, also leading to their compression. At this time, the membrane of the loudspeaker recedes leading to the creation of a zone of rarefied air. The disturbance begins to propagate maintaining a similar pattern of density/rarefaction even far from the source. If the loudspeaker emits a so-called pure tone (of one frequency), the resulting wave has the character of a so-called harmonic wave (see Fig. 1, where different degrees of shading correspond to different air densities).

Let us assume for simplicity that it is a plane wave (i.e. surfaces of constant air density are parallel planes) and the wave energy is not dissipated. As a result, the size of the disturbances will be the same even at a distance from the source. Let us look at the characteristics of a harmonic wave.

If we plot the air density ρ against the distance r from the source, we will obtain a regular sequence of wave **crests** and **troughs** i.e. places with the highest and lowest air density, respectively (Fig. 1). In the case of general wave motion, crests and troughs are associated with extreme deviations of the medium (or field) from the equilibrium state. For example, for a wave on water, a crest is the highest position of the water surface, while a trough is the lowest.

The distance between successive crests (or troughs) of a wave is called the **wavelength** and is denoted by λ. The magnitude of the disturbance measured from the equilibrium position to the maximum deviation from equilibrium is called the **amplitude** and is denoted by A (note that this is not the distance between a crest and a trough – this distance is equal to twice the amplitude). The **speed** of a wave is denoted by v. For sound in air, it is about $340\ \mathrm{m/s}$.

As we said, a wave is a disturbance that propagates in space. However, if we choose one specific point in space and observe the behaviour of the medium (or field) as the wave passes through it, we will simply see oscillatory motion. That is, instead of making a graph as in Fig. 1, in space (depending on r), we can make an analogous graph in time (depending on t). We can then ask how long a full cycle of oscillation of the medium (or field) lasts, i.e. the time between the passage of two successive wave crests. This time is called the **period** and is denoted by T. The number of oscillation cycles in $1\ \mathrm{s}$ is called the **frequency** and is denoted by f. Frequency is measured in hertz – Hz (where 1 Hz equals one cycle per second).

There is a very simple relationship between period and frequency:

$$f = \frac{1}{T} \quad [\mathrm{Hz}]$$

$$1\ \mathrm{Hz} = \frac{1}{1\ \mathrm{s}}$$

2. Harmonic wave as an image of circular motion

Notice that we haven't really said yet what exactly a harmonic wave is. There is a very close connection between this type of wave and uniform circular motion.

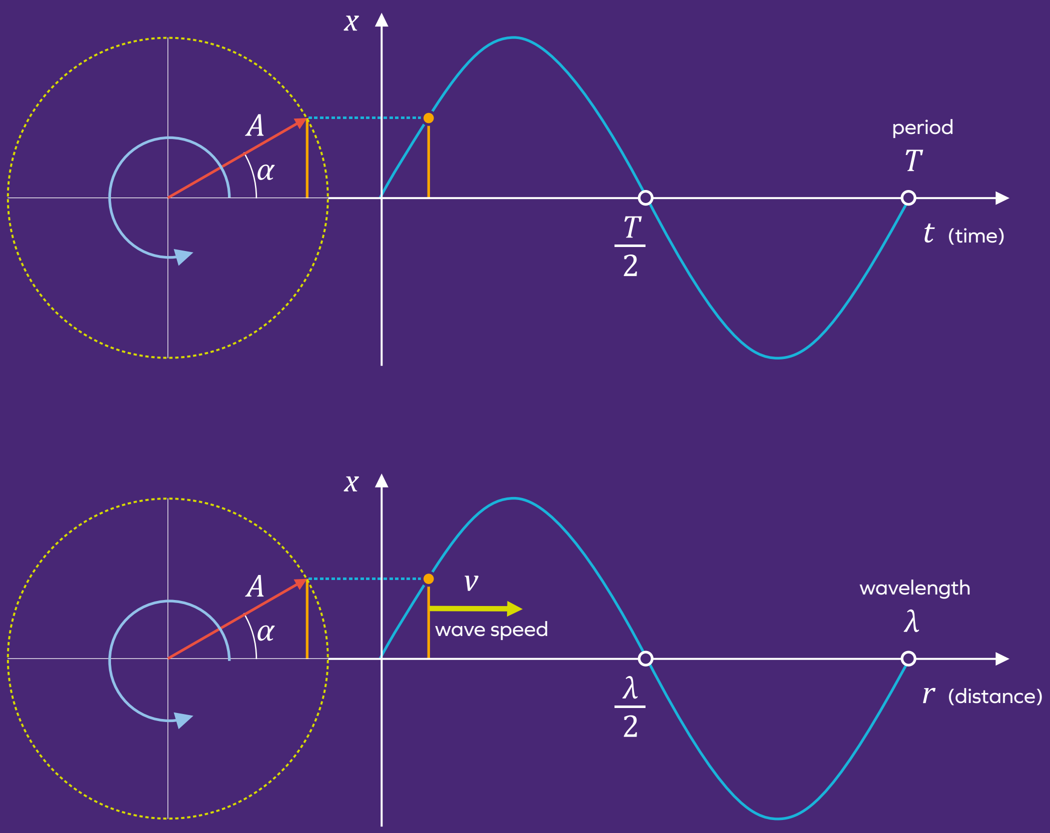


Fig. 2. Harmonic wave as a representation of a point moving along a circle.

Let us imagine a circle with a radius equal to the amplitude A of the wave and a point moving along this circle in uniform motion so that it covers the entire length of the circle in period T (Fig. 2). The height of the point above the selected circle diameter (here, the horizontal one) will be plotted on a graph as a function of time. As a result, we will recreate a waveform that corresponds exactly to one period of harmonic oscillation. Several stages of the circular motion are shown in more detail in Fig. 3.

Similarly, we can plot a graph in space. This time, on the abscissa axis, we will not put time t, but the path r covered by the wave in time t i.e., assuming a uniform wave speed, $r = vt$. In this way we will recreate one full segment of the harmonic wave, i.e. one wavelength. By continuing the circular motion, we can recreate the entire harmonic wave.

As we can see, during one period the wave travels a distance equal to the wavelength. We thus obtain a very useful relationship:

$$vT = \lambda$$

which we can write differently as: $v = \lambda f$

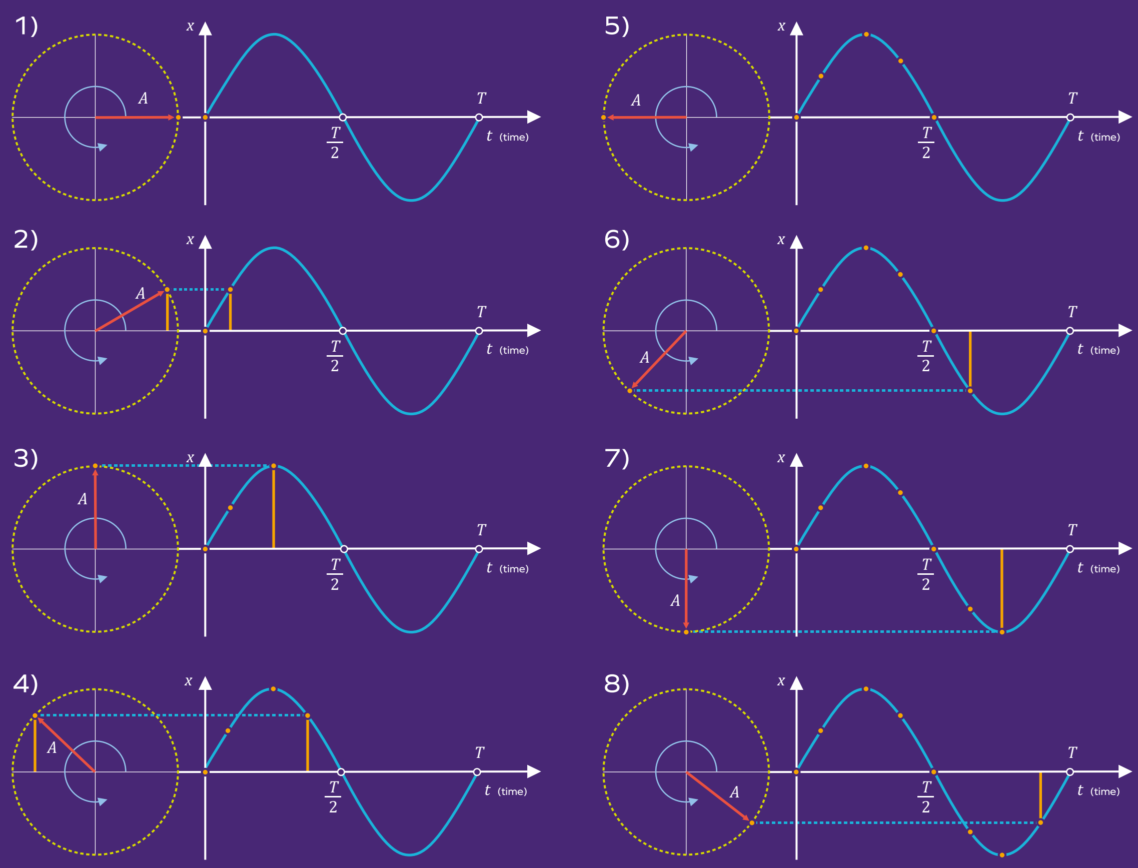


Fig. 3. Selected stages of circular motion and their corresponding moments in the cycle of medium (or field) oscillation.

3. Phase and phase shift

In both Fig. 2 and Fig. 3, the deviation from the equilibrium state at the moment $t = 0$ is zero. When we consider only one wave or one oscillatory motion, we can always set the moment $t = 0$ to the moment when the equilibrium state is just reached (at least temporarily). When there are more oscillations or we deal with a larger number of superimposed waves, then in general it is not possible to synchronise all of them so that the equilibrium state occurs everywhere at the same time. Therefore, it is worth introducing the concept of the oscillation phase.

By **phase**, we mean a physical quantity that specifies the stage of oscillation within a cycle for a vibrating object. For example, we can say that the oscillatory motion at two different points in space has the same phase if, at both locations, the wave has reached its equilibrium state or, at both, the wave crest. If such synchronisation does not occur, we say that there is a **phase shift** between the oscillations, understood as a phase difference (Fig. 4).

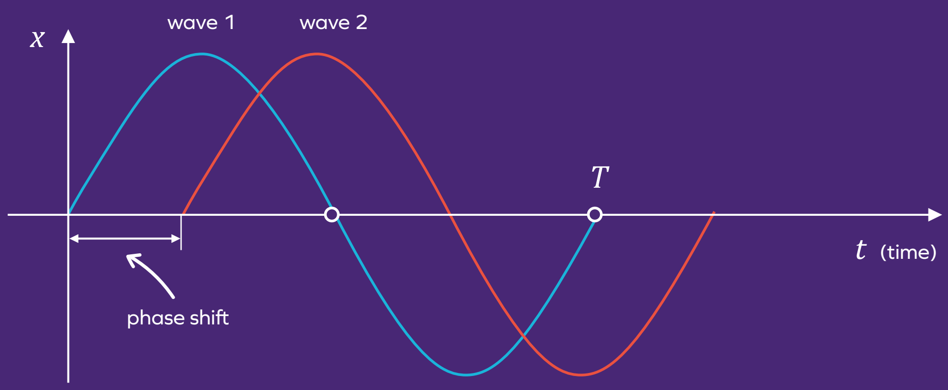


Fig. 4. Phase shift between the oscillations in two different points.

4. EM wave and antenna length selection

All the concepts we introduced earlier for the harmonic acoustic wave also apply to the harmonic electromagnetic (EM) waves that we can excite, for example, in a dipole antenna using an alternating current source (Fig. 5). Therefore, also in the case of EM waves, we can talk about wavelength, frequency, amplitude, speed and phase

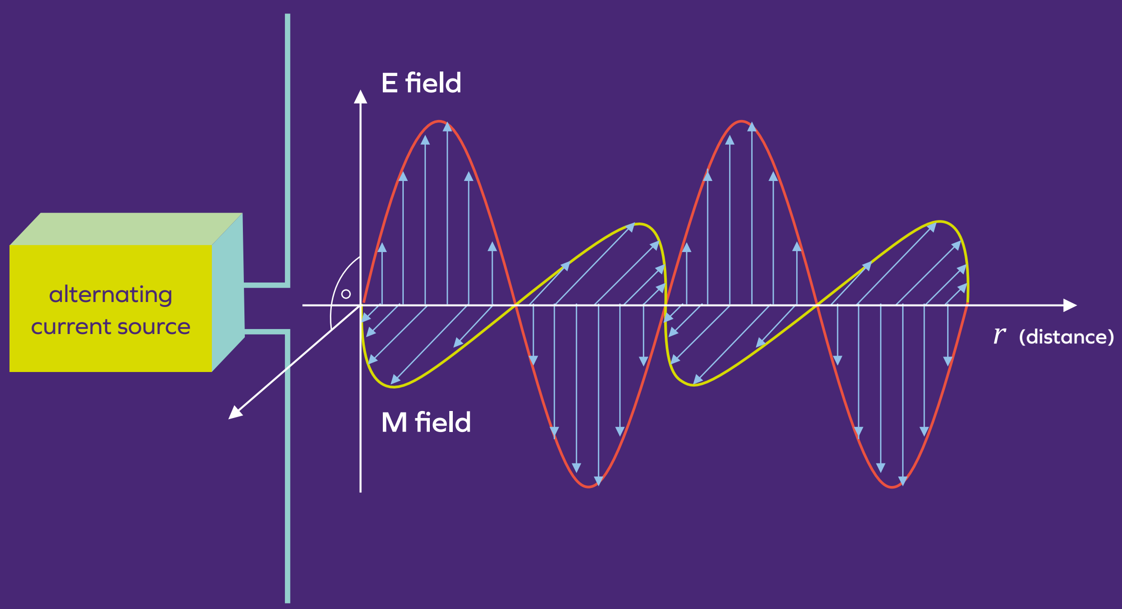


Fig. 5. Electromagnetic wave as vibrations of the EM field.

Since a time-varying electric field leads to the excitation of a changing magnetic field, both of these fields appear together in the propagating wave. Moreover, the vectors of these fields are perpendicular to each other in a vacuum and are perpendicular to the direction of wave propagation. We say that an EM wave is a **transverse** wave. An acoustic wave, on the other hand, is a **longitudinal** wave - in its case, the direction in which the air is compressed/rarefied is consistent with the direction of wave propagation.

It turns out that for the emission of an EM wave by a dipole antenna to be energy-efficient, the antenna length l should not be less than half the wavelength of the emitted wave:

$$l \geq \frac{\lambda}{2}$$

This does not mean that a shorter antenna will not emit a wave. However, it may require more energy from the power source to obtain the required EM wave energy (which translates into, among other things, the transmission range). The antenna length is often selected so that it is exactly half the wavelength. This creates specific conditions under which the electrical voltage at the ends of the antenna periodically changes into a crest and trough of the wave induced in the antenna wires, while the electric current assumes a maximum value at the centre of the antenna (so-called standing wave). An animation illustrating this phenomenon can be viewed in "External Materials" (1).

Warning: *for other types of antennas, the condition introduced may have a different form. For example, the length of a monopole antenna fed at one end should be greater than one quarter of the wavelength. These relationships will become important in Lesson 7, where we will introduce the concept of signal modulation.*

Experiment

Prepare a large container with water - it can be shallow, but it is important that its walls are widely spaced apart. In a home setting, this could be a bathtub with a small amount of water. The container should be illuminated from above so that the shadow of the waves on the water surface is clearly visible on the bottom.

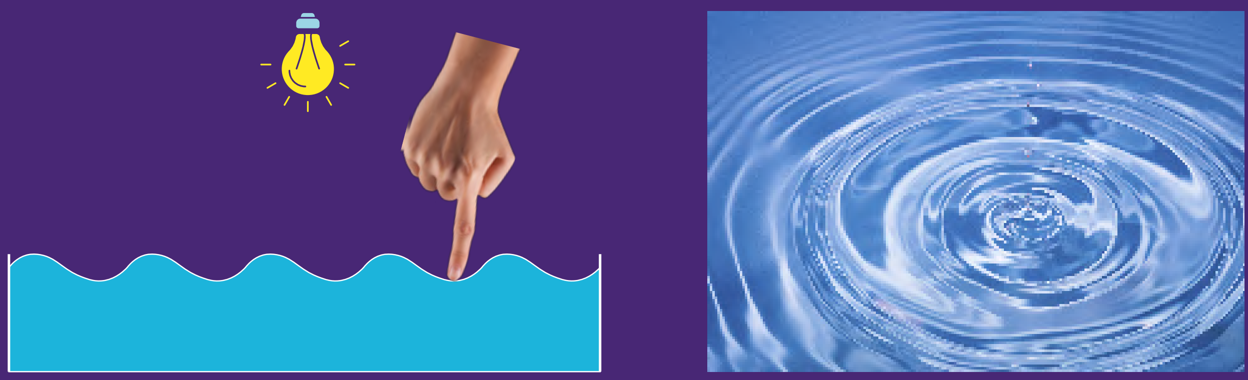


Fig. 6. **Experiment with the waves on the water surface.**

1. Touch the surface of the water with your finger. You should see the shadow of a circular surface wave on the bottom. Notice that the wave travels at a specific speed regardless of how deep you immerse your finger.
2. Start moving your finger rhythmically in the water (about 2 times per second). What physical quantity is related to the speed at which you move your finger? Observe the successive crests and troughs of the spreading wave. Can you estimate how big the gap is between successive crests? What physical quantity corresponds to this gap?
3. Now increase the speed of your finger movement (around a few times per second). What physical quantity has increased? Notice the distance between the wave crests. Has it changed from point 2? Has it increased or decreased?
4. Given the formula and assuming that the wave speed is constant, does the direction of changes of the physical quantities appearing in it agree with observations?

Discussion. *According to the formula* $v = \lambda f$ *an increase in frequency at constant wave speed must lead to a decrease in wavelength. Since frequency and wavelength are inversely proportional, doubling the frequency means halving the wavelength.*

Glossary

Wave motion – a propagating disturbance of a certain elastic medium or physical field.

Acoustic wave – a series of compressions and rarefactions that spreads in the air.

Wavelength – the distance in space between successive crests or troughs of a wave. Unit – metre [m].

Wave amplitude – the maximum magnitude of the disturbance caused by a wave measured from the equilibrium state. The unit depends on the physical nature of the wave.

Wave speed – the speed at which wave crests move in space. Unit – metre per second [m/s].

Vibration period – the duration of a full cycle of medium (or field) oscillation, i.e. the time between the passage of two successive wave crests. Unit – second [s].

Vibration frequency – the number of complete oscillation cycles per second. Unit – hertz [Hz].

Harmonic wave – a wave whose displacement follows the projection of a uniform circular motion.

Phase – a physical quantity defining at what point in the period a given object undergoing vibrations is. It is usually assumed to be a dimensionless quantity (as a part of the period).

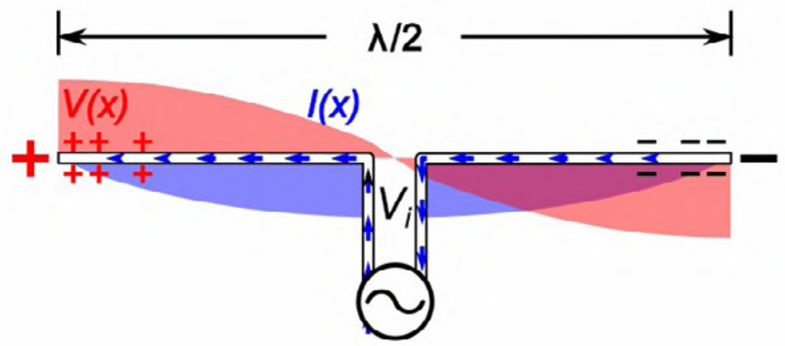
Phase shift – phase difference between two oscillatory movements.

Transverse wave – a wave in which the direction of disturbance of the medium (or field) is perpendicular to the direction of wave propagation.

Longitudinal wave – a wave in which the direction of the disturbance of the medium (or field) is consistent with the direction of wave propagation.

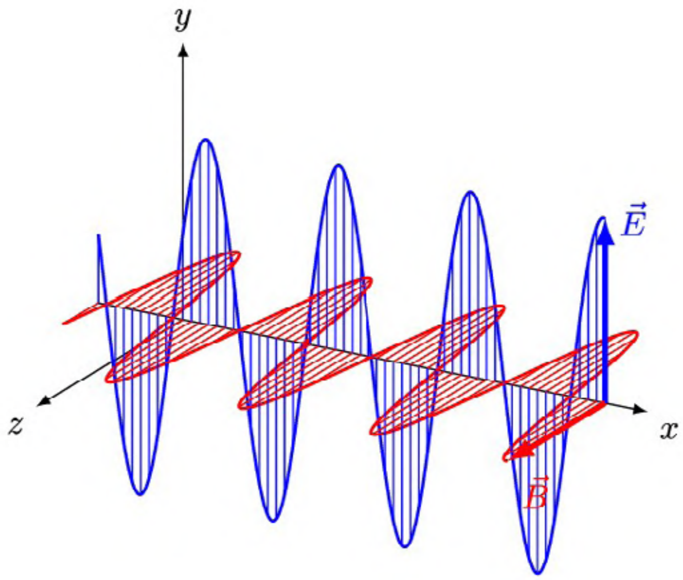
External materials

1. Animation of a standing wave in a dipole antenna.



Scan QR code

2. Animation of an electromagnetic wave and its two components – changes in the electric and magnetic fields.



Scan QR code

Homework

1. An acoustic wave is propagating in the air with a frequency corresponding to the note C_1 on a piano keyboard, i.e. $f = 261.6\,\mathrm{Hz}$. Assuming that the speed of sound is $v = 340\,\mathrm{m/s}$, determine the wavelength.

   Given:

   $f = 261.6\,\mathrm{Hz}$

   $v = 340\,\mathrm{m/s}$

   To find:

   $\lambda = ?$

2. Let us assume that we would like to transmit an EM wave with exactly the same frequency as the C_1 sound, i.e. $f = 261.6\,\mathrm{Hz}.$, through a dipole antenna. What can be the shortest antenna length that will allow for the energy-efficient emission of this wave? Assume the speed of the EM wave as $c = 3 \cdot 10^8\,\mathrm{m/s}$.

   Given:

   $f = 261.6\ \mathrm{Hz}$

   $c = 3 \cdot 10^8\,\mathrm{m/s}$

   To find:

   $l = ?$

Lesson 4

Signal spectrum

Objective

- Presentation of basic concepts related to spectral analysis of signals.

Learning outcomes

- The student knows the method for adding harmonic signals using vector diagrams.
- The student knows the basic objectives of spectral analysis.
- The student knows the basic ways of representing a signal: in the time domain and in the frequency domain.

1. Addition of harmonic signals

In Lesson 3, we introduced the concept of a harmonic wave as a specific disturbance that can be fully characterised by specifying amplitude, frequency (or period) and phase. Since the harmonic waves discussed so far were related to a disturbance of the medium or physical field (acoustic or electromagnetic wave), we can identify them with a **harmonic signal** (see Lesson 1). Such a signal is an image of the uniform motion of a point in a circle. It is an example of a **periodic signal**, i.e. one for which we can indicate a time interval after which the changes in the signal value are exactly the same as in the previous interval. We will treat the harmonic signal as an elementary periodic signal (Fig. 1).

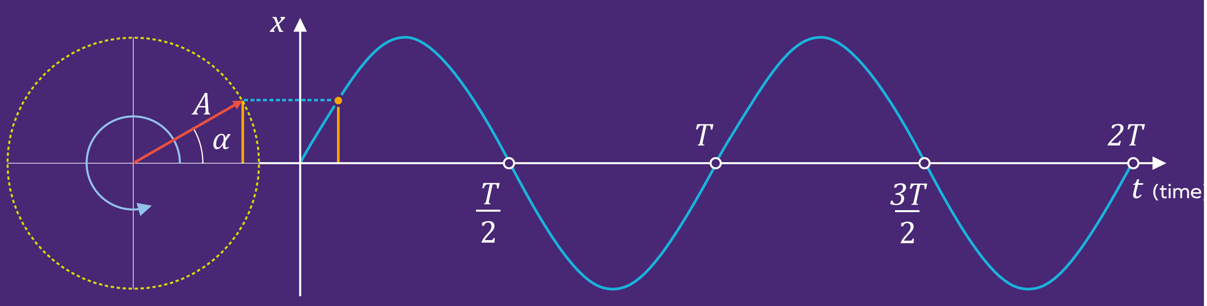


Fig. 1. Harmonic signal of amplitude A, period T and zero initial phase.

We can, of course, choose the parameters of the harmonic signal completely arbitrarily. Fig. 2 shows a signal with one-fifth the amplitude and one-fourth the period of the signal in Fig. 1. These signals are basically similar; they only differ in scale. However, everything changes if we allow certain mathematical operations on these signals, in particular: addition.

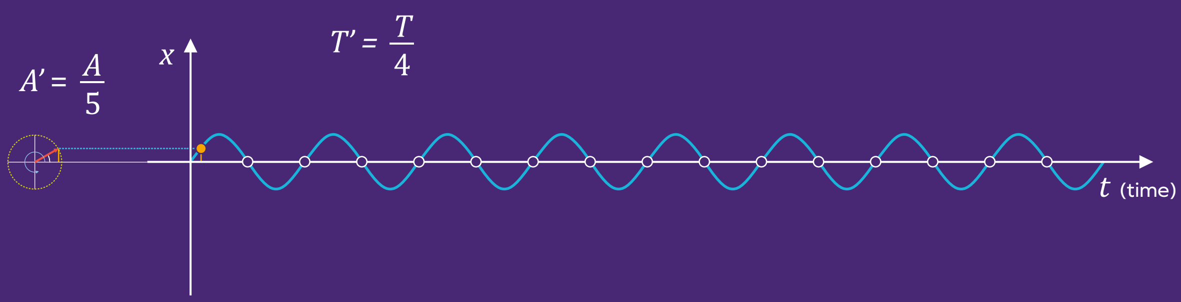


Fig. 2. Harmonic signal with five times smaller amplitude and four times shorter period.

We can do this addition very simply by taking the signal waveforms in time and adding the signal values at the appropriate time instants. For the purposes of later lessons, especially modulation analysis, let us see how adding harmonic signals is related to adding vectors.

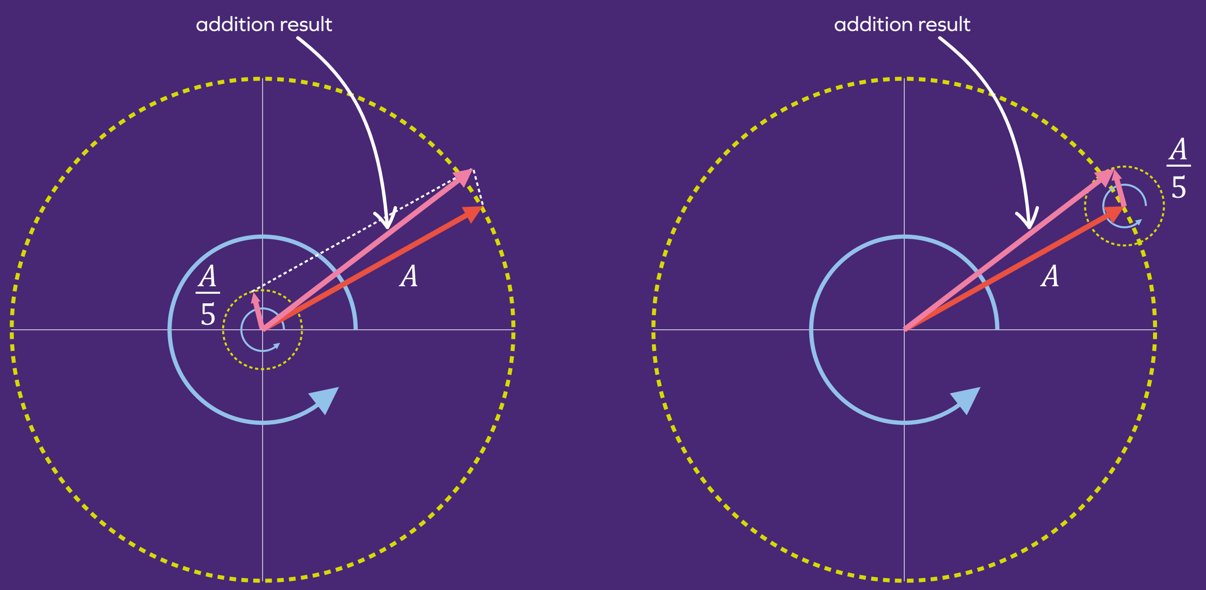


Fig. 3. Addition of two harmonic signals as the addition of rotating vectors. On the left, application of the parallelogram rule; on the right – the head--to-tail method.

A harmonic signal is represented not only by the motion of a point along a circle but also by a vector rotating around the centre of the circle and ending at the moving point (this is referred to as the so-called **position vector** of the point). The length of the vector equals the amplitude of the signal, and the time of one full rotation equals the period of the signal.

If we want to add two harmonic signals, we can plot both circles and vectors on one diagram. The position of the vectors will depend on the chosen moment of time – let us assume that it looks like in Fig. 3 on the left, with both vectors attached to one point. We add the vectors according to the parallelogram rule, that is, we construct a parallelogram whose adjacent sides are both vectors, and the result of the operation is a vector that coincides with the diagonal of the parallelogram. In this way, we obtain the resultant vector, the length of which depends on the mutual position of the vectors that we add (in this case, it is slightly greater than A). Similarly, as in Fig. 1 and Fig. 2, the value of the signal depends on the distance of the tip of this vector from the straight line passing through the horizontal diameter of both circles.

The result can be obtained more simply by using the so-called head-to-tail method of vector addition. In this case, we place the second circle so that its centre coincides with the tip of the first vector (Fig. 3 on the right). We find the resultant vector by connecting the tail of the first vector to the tip of the second, and it is obviously the same as when using the parallelogram rule.

By carrying out the above procedure for all time instants in the time interval of interest, we obtain the waveform of the resultant signal (Fig. 4).

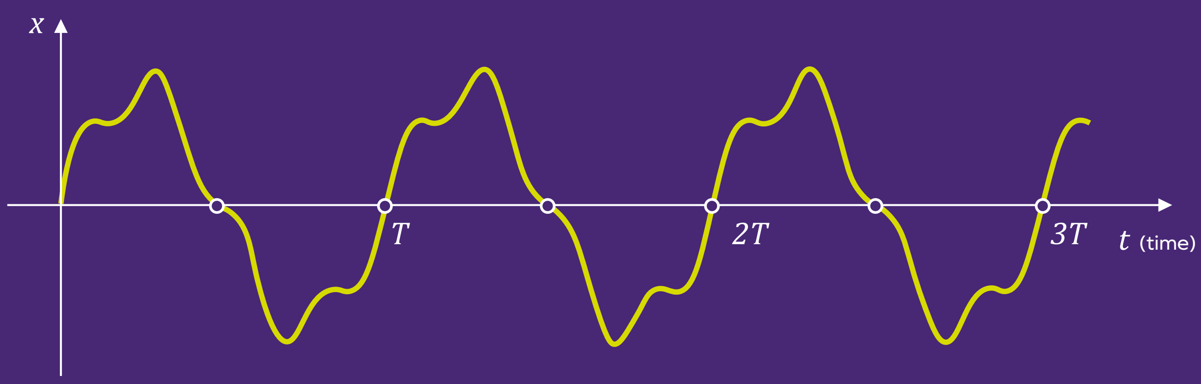


Fig. 4. Sum of two harmonic signals.

What can we say about this signal? It is still a periodic signal. As we can see, its period is equal to the period of the signal with the longer period (T). This is not a coincidence – it will always be the case when we add signals whose periods are multiples of each other (in our case, the signal with period T is four times the period of the shorter one). Moreover, the waveform of the signal is very similar to that of the signal with the larger amplitude. The signal with the smaller amplitude introduces slight yet clearly visible distortions into it.

There is no barrier to adding more signals. In Fig. 5 we can see an example of the sum of three signals with parameters (amplitude, period): a) A, T, b) $0.6A$, $0.5T$, c) $0.5A$, $0.25T$. Next to each of the graphs, an example of a vector diagram at a selected time moment is also attached. The sum of the individual harmonic signals presented in the last graph is a rather complex periodic signal with a period T.

Since the frequency of a signal is the reciprocal of the period, we can see that the frequency of the second signal is twice that of the first signal, and the frequency of the third signal is four times that of the first signal. We can therefore formulate the following theorem: *if we add successively the harmonic signals starting from a signal with period* $\boldsymbol{T}$*, we obtain as a result a signal also with period* $\boldsymbol{T}$*, provided that the frequencies of the successive signals are multiples of the frequency of the first signal.*

In a similar way, we can add any number of harmonic signals, even an infinite number, provided that the amplitudes of successive signals decrease sufficiently fast from one signal to the next.

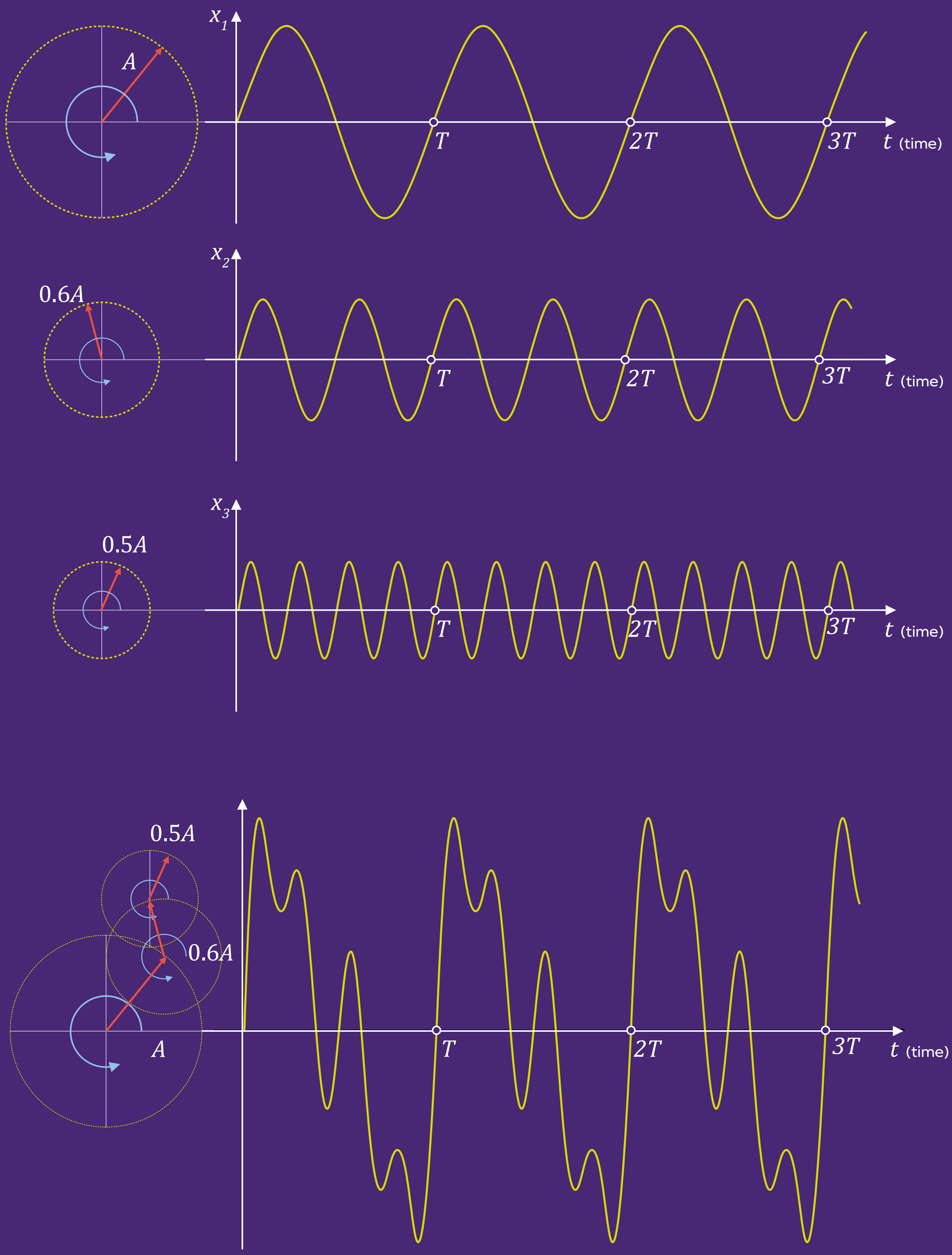


Fig. 5. Sum of three harmonic signals.

2. Spectral analysis

The complexity of the signal shown in Fig. 5 may inspire us to ask: maybe every periodic signal can be represented as a sum of harmonic signals? It turns out that yes!

Warning: *There are some exceptions, but in practice we don't have to worry about them.*

The procedure we will now consider is therefore somewhat the opposite of what we did above. Instead of adding harmonic signals together and generating increasingly complex patterns in time, we can take a given periodic signal and try to find such harmonic signals that, when combined (summed), will create the desired signal.

This procedure is called **spectral analysis** (other terms: frequency analysis, Fourier analysis). For a signal with a period T spectral analysis consists in determining the amplitude of harmonic signals (so-called components) with frequencies $f_1 = 1/T$, $f_2 = 2f_1$, $f_3 = 3f_1$, etc., which are components of this signal. The first of these components with frequency $f_1 = 1/T$ is called the **first harmonic**, the second – with a frequency twice as high – the **second harmonic**, etc. Some signals even contain an infinite number of non-zero harmonics (we will see an example shortly).

How to perform spectral analysis in practice, i.e. how to determine the amplitudes of harmonics for a given signal? There are some mathematical formulas that allow this to be done, but they are too complicated to be quoted here. However, we can easily show the effect of spectral analysis for the signals from Fig. 4 and Fig. 5, because we have previously selected the components of these signals ourselves.

So, for the signal in Fig. 4, the first harmonic f_1 has an amplitude that we have designated by A, and the second harmonic f_2 has an amplitude that is five times smaller ($A/5$). We do not need any further harmonics, so we can assign them an amplitude of 0.

For the signal from Fig. 5, the first harmonic f_1 has an amplitude of A, the second – f_2 – has an amplitude of $0.6A$, the third – f_3 – has an amplitude of zero (does not appear in the signal), and the fourth – f_4 – has an amplitude of $0.5A$. All subsequent harmonics have an amplitude of 0.

The information about harmonic amplitudes tells us exactly the same thing about a periodic signal as its waveform. We therefore have two ways to describe a signal:

1. Specifying the value of a signal in time for each time instant t. This is the so-called description in the **time domain**.
2. Specifying the amplitude of the harmonics of this signal. This is the so-called **frequency domain** description.

We are accustomed to the time-domain description, which involves plotting a graph with a horizontal axis corresponding to time and a vertical axis representing the signal's value. Similarly, for the frequency-domain description, we can create a comparable graph where the horizontal axis specifies which harmonic is referred to (or provides its frequency value), while the vertical axis indicates its amplitude. Conventionally, the amplitude of a harmonic is plotted on this graph as a bar (so-called spectral bar) instead of a point. Fig. 6 shows an example comparison of both descriptions for signals we already know.

The representation of a signal in the frequency domain is also called the **signal spectrum**.

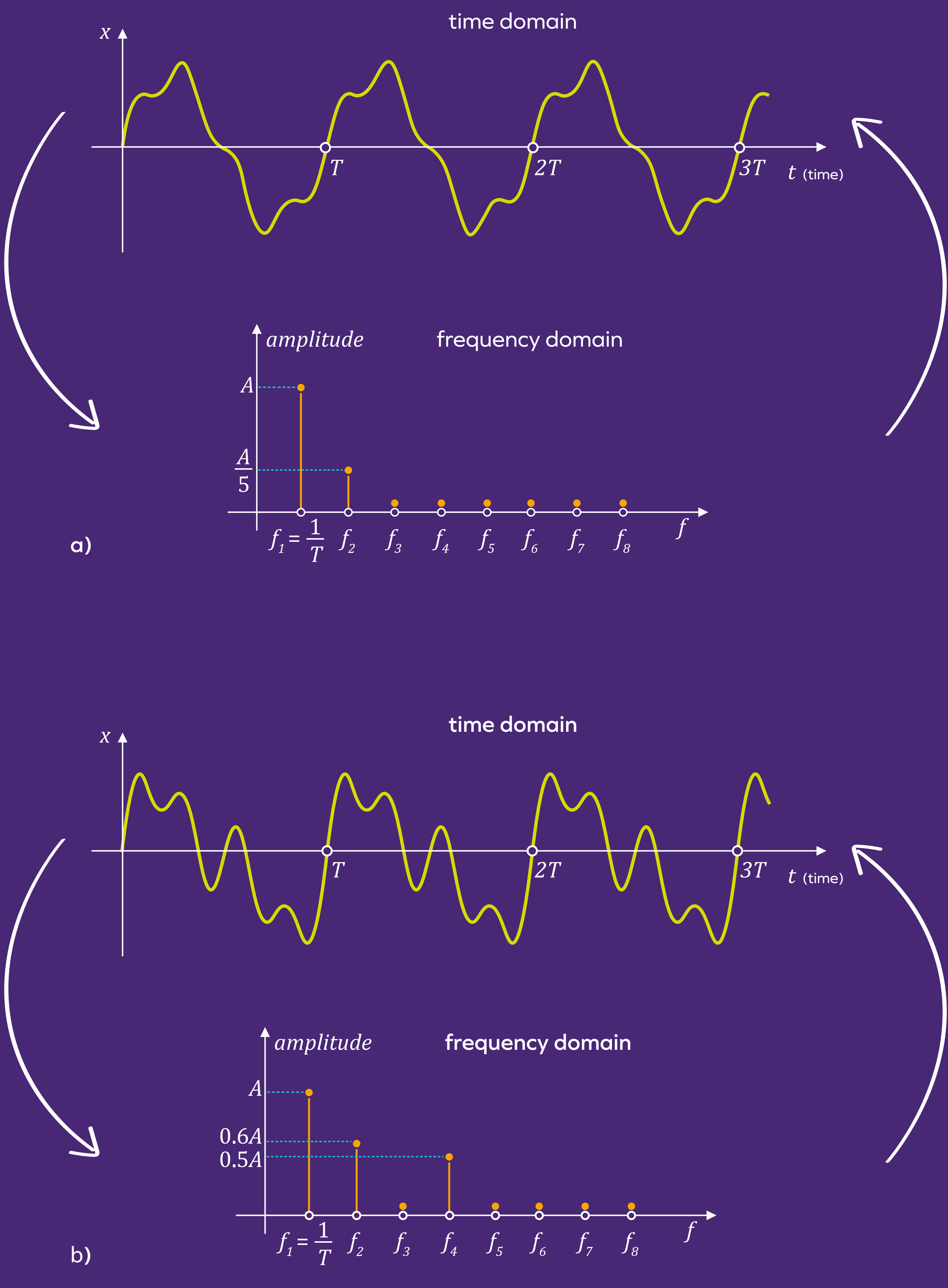


Fig. 6. Signal description in the time and frequency domains:
a) a signal with two harmonics, b) a signal with three harmonics.

3. Example - spectral analysis of a square wave signal

As an interesting example, let us now consider a signal that is traditionally called a **square wave**. During the period T this signal takes on only two values: A – in the first half of the period and -A – in the second half of the period (Fig. 7).

This signal differs significantly from those we have analysed so far. First of all, it is not a continuous signal - every half period we observe a sudden change in value. It does not resemble a regular and smooth harmonic signal in any way. We can treat it as an example of a digital signal with two levels (binary signal; see Lesson 1). However, it turns out that even a signal of this type can be subjected to spectral analysis and we can precisely determine its harmonics. Unfortunately, in this case there are infinitely many of them, so we have to be satisfied with approximations.

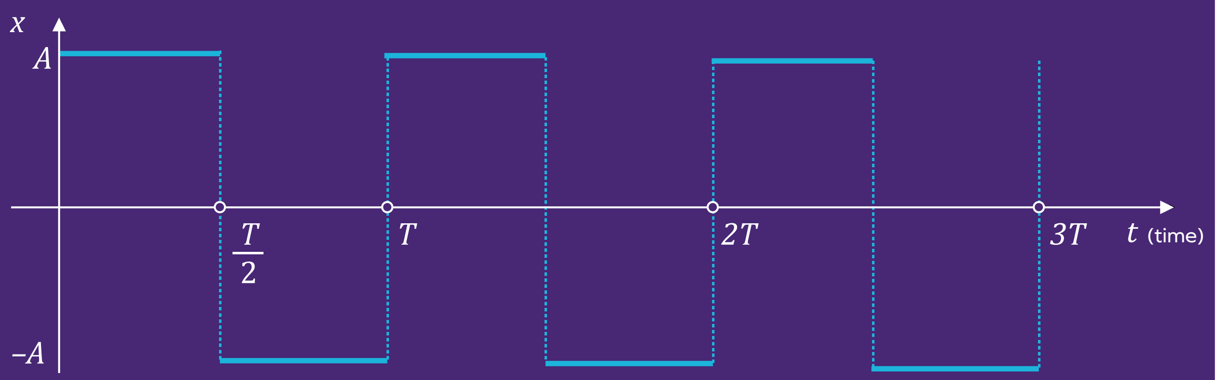


Fig. 7. Square wave signal of period T and amplitude A.

In Fig. 8 we can see the first several harmonics of a square wave signal. The first harmonic (red line) is a signal with a period T that quite well reflects the overall time variability and levels reached by the signal of interest. The second harmonic and the remaining components with even multiplicity (fourth, sixth, etc.) have an amplitude equal to zero. The third, fifth, seventh, and ninth harmonics are marked as signals with increasingly smaller amplitudes.

Learn more. *The amplitude a_n of the n-th odd harmonic can be calculated from the formula:*

$$a_n = \frac{4A}{n\pi}$$

The sum of the first five non-zero harmonics is highlighted in yellow. As can be seen, it is a fairly good approximation of the square wave despite visible ripples around the values it assumes. In advanced signal analysis, it is shown that after adding an infinite number of harmonics, the ripples disappear and we obtain exactly the waveform from Fig. 7.

The representation of the square wave signal in the frequency domain, i.e., its spectrum, is shown in Fig. 9. As we can see, all even spectral bars are equal to 0, while the odd ones gradually decrease to zero with increasing harmonic number.

There are many practical applications of spectral analysis, and we will not list all of them here. One of them is, for example, computer speech recognition. Based on the amplitude of individual harmonics in the voice signal, the computer can identify the pronounced phonemes (each has a characteristic spectrum) and then words and sentences.

From our point of view, spectral analysis will allow us to understand thoroughly the operation of modulation techniques (Lessons 7, 8, 9) and the division of frequency ranges between different users in mobile telecommunications (Lessons 5, 10).

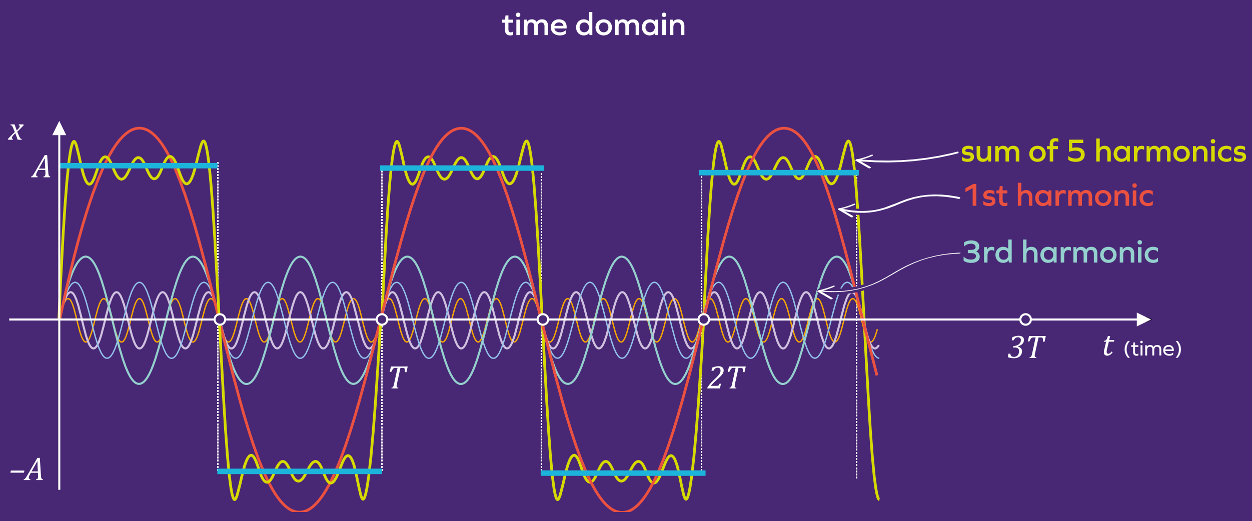


Fig. 8. Harmonics of the square wave in time domain.

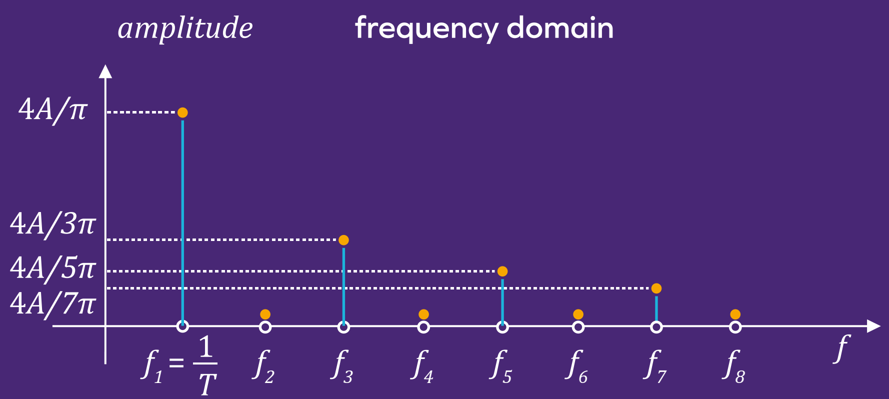


Fig. 9. Amplitudes of the first eight harmonics of the square wave (part of the spectrum).

Experiment

On the **compadre.org** website you can run the Soundanalyzer program to analyse the sound recorded by the computer's microphone (you must allow the application to install and access the microphone). The application window (Fig. 10) contains the signal spectrum in the upper part, and the time domain image in the lower part. The maximum frequency value is $10\,\mathrm{kHz}$.

Scan QR code

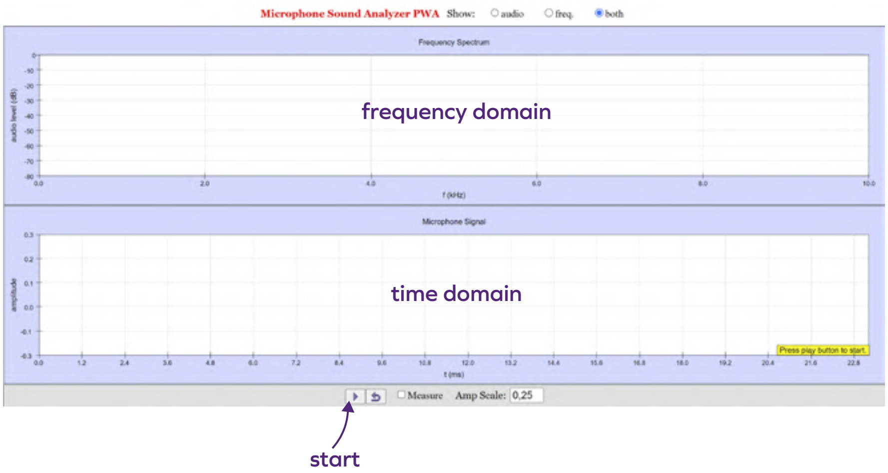


Fig. 10. Soundanalyzer application window.

Before we move on to the experiment, let us note that a normal voice message is not a periodic signal. However, we can divide it into very short fragments, which we will treat as segments of a periodic signal (samples), which we will then subject to standard spectral analysis. Since the signal can change significantly between successive samples, the image of the signal in the frequency domain (spectrum) can also change over time.

Use the app for the following observations:

1. Observe the time and frequency domains as you speak casually into the microphone. Which frequencies appear most frequently?
2. Compare the signal patterns for vowels and fricatives (especially "s"). Try to sustain the emission of these sounds for as long as you can. What differences in frequency distribution do you see?
3. Try producing one vowel (e.g. "u") while changing the pitch of the sound from lowest to highest. Does the observed frequency distribution reflect these changes?

4. If you have a keyboard musical instrument, use it to produce different sounds at different pitches. Some instruments allow you to select a pure tone (*sine wave*). Check if the frequency distribution matches your expectations.

Glossary

Spectral analysis – decomposition of a signal into its harmonic components.

Time domain – a method of describing a signal by specifying its value at each time instant.

Frequency domain – a method of describing a signal by specifying the amplitudes of its successive harmonics.

Harmonic – a harmonic signal constituting a component of a given signal. The frequencies of successive harmonics of a periodic signal are multiples of the fundamental frequency, which is equal to the reciprocal of the signal's period.

Harmonic signal – a signal whose value in the time domain can be represented as a projection of uniform circular motion or the uniform rotation of a position vector.

Periodic signal – a signal having a period, i.e. a time interval after which the signal value repeats.

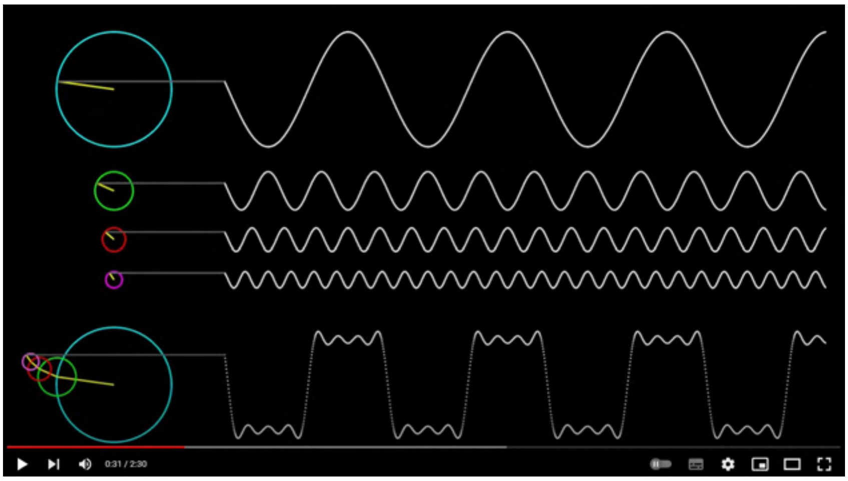
Square wave signal – a periodic signal that alternates between two values, with the transition occurring midway through the period. An example of a digital, binary signal.

Position vector – a vector fixed at a reference point and ending at the point whose motion is being tracked.

Signal spectrum – the distribution of amplitudes of the successive harmonics of a signal.

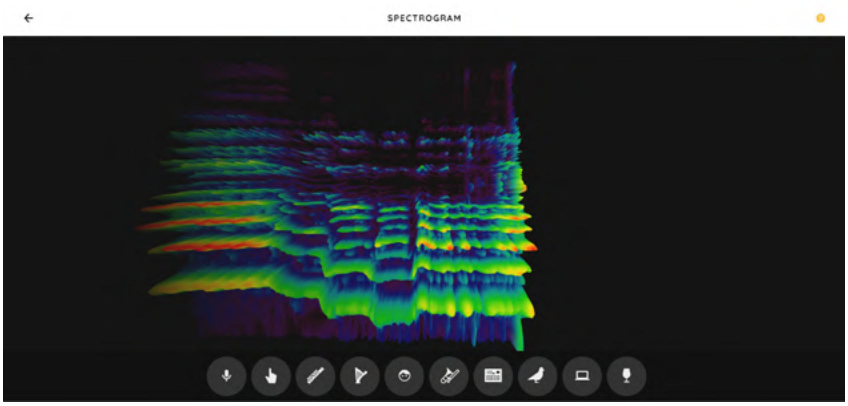
External materials

1. Animation illustrating the addition of the first four non-zero harmonics of a square wave signal (Fourier Series Square Wave).



Scan QR code

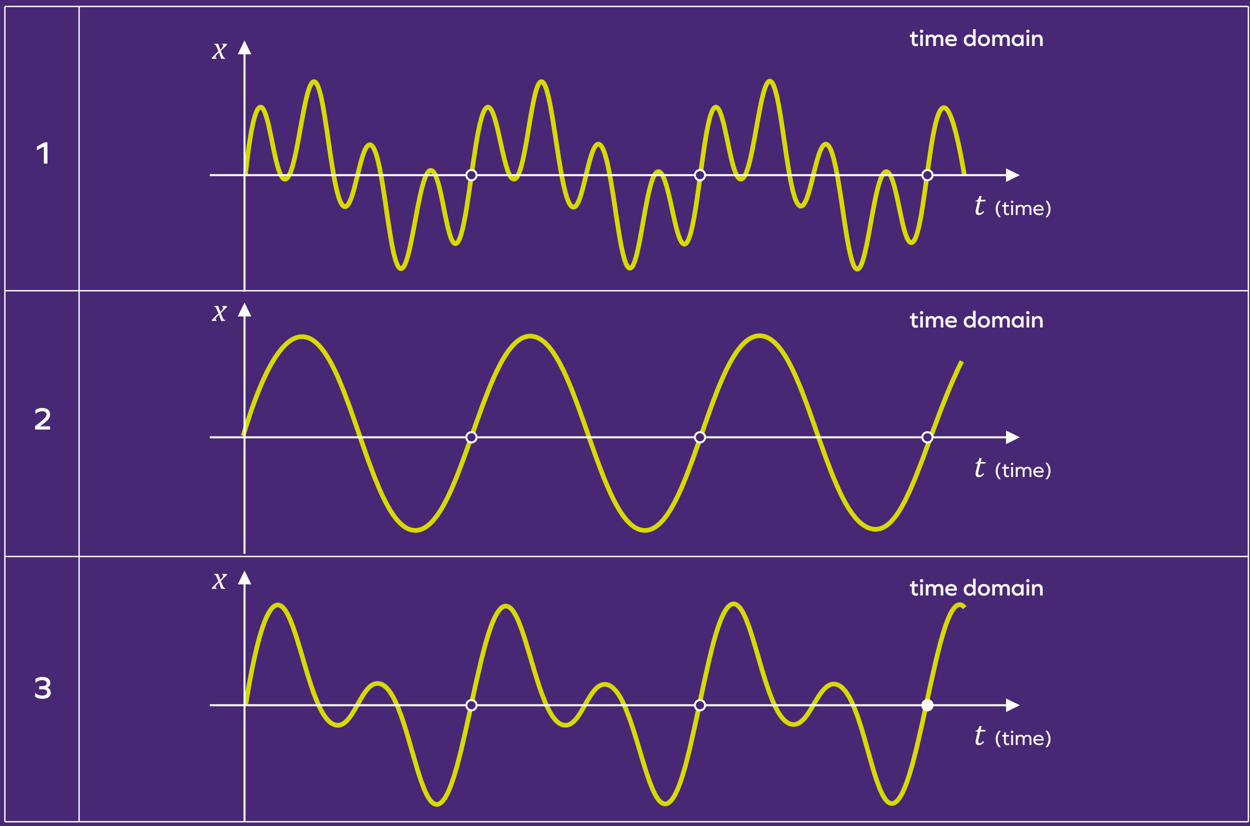
2. **A program for spectral analysis of sound from a microphone**. It generates a so-called **spectrogram**, which shows the frequency distribution of the sound recorded by the microphone in real time. The horizontal axis represents the time of sound sampling (the image dynamically shifts in the horizontal direction), while the vertical axis represents frequency. The colour corresponds to the amplitude of a given harmonic (red – high amplitude, blue – low). Since no scale is provided on the axes nor a quantitative assignment of colours to amplitudes, the spectrogram is solely illustrative. Let us also emphasise that the signal representation is shown here in the frequency domain but for continuously changing time samples.

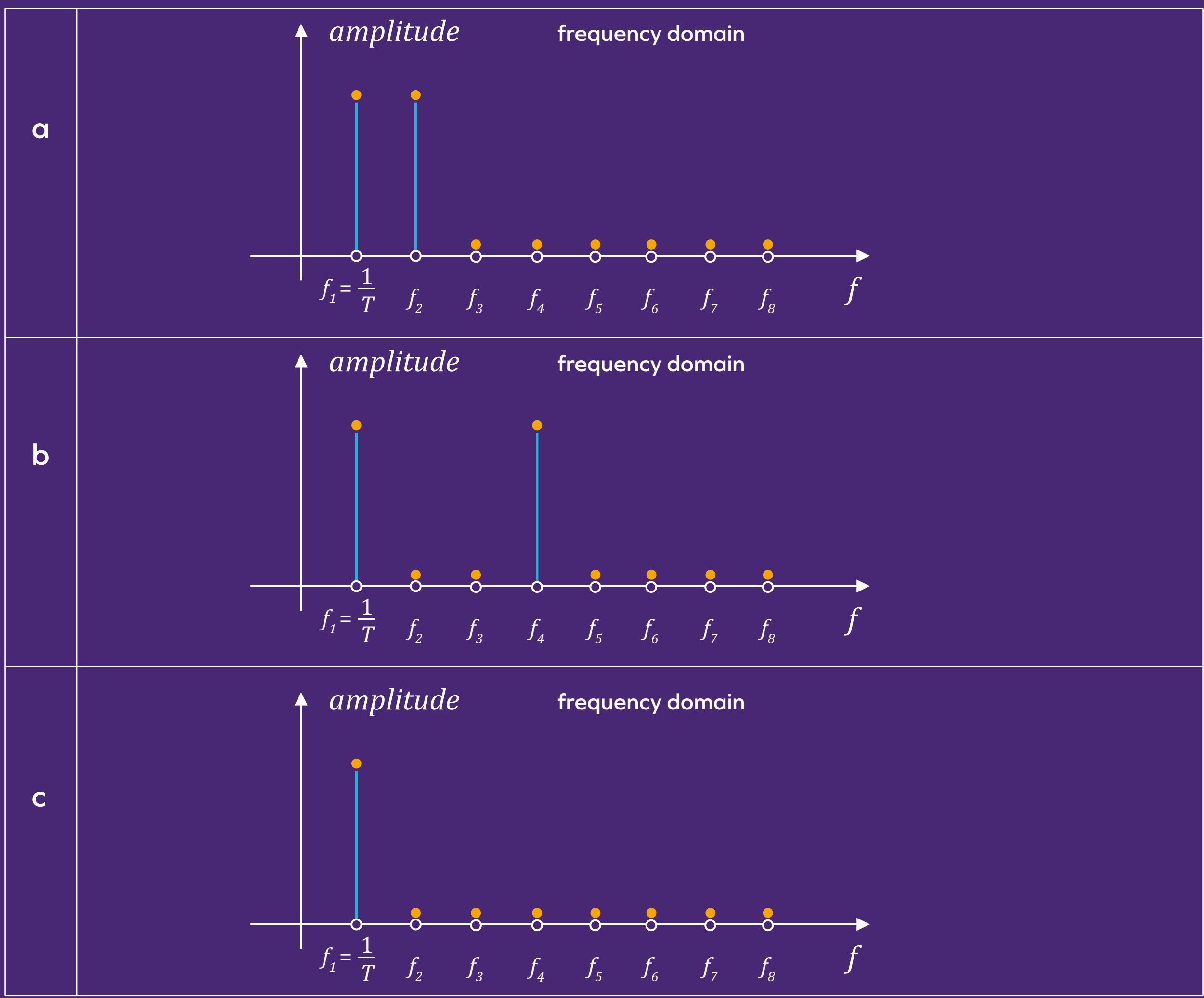


Scan QR code

Homework

Try to map the time domain representation of signals to the corresponding frequency domain (spectral) representation.


1
2
3
x
time domain
t (time)


a
b
c
amplitude
frequency domain
$f_1 = \frac{1}{T}$
f_2
f_3
f_4
f_5
f_6
f_7
f_8
f

Lesson 5

Signal filtration

Objective

- Presentation of basic concepts related to signal filtration.

Learning outcomes

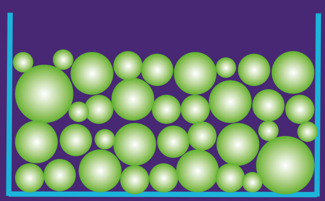
- The student knows the basic types of filtration: low-pass, high-pass and band-pass.
- The student is able to describe the effects of filtration in the frequency domain.
- The student is able to indicate the most important benefits of signal filtration in telecommunications.

1. Filtration – an illustrative model

Most of us have encountered the concept of **filtration** (or filtering) in our everyday lives. In many homes, we can find a water filter, the purpose of which is to purify tap water from possible contaminants. Filtration always consists of certain components being retained in the filtering device, i.e. the **filter**, while the rest are passed through. The most illustrative example of a filter is a sieve - a surface with holes of a specific diameter, which allows objects with a diameter smaller than the diameter of the hole to pass through, while retaining the rest. We also encounter the concept of filtration in signal analysis, but before we get to that, let us first look at filtration and what it is useful for, using the illustrative example of a sieve.

Let us assume that we have a container with a very large number of balls of different diameters. The balls in the container are randomly mixed, so it would be difficult for us to easily find specific ball sizes by rummaging through the container and selecting balls at random. However, we know how many balls of each diameter there are and we can present this information on a graph as in Fig. 1 (this graph not coincidentally resembles the signal spectrum). We can see that the most balls are $4\,\mathrm{cm}$ in diameter (the highest bar), the rest are less numerous, and this difference increases with the deviation from the $4\,\mathrm{cm}$ value.

If we now wanted to select only balls with diameters less than or equal to $3\,\mathrm{cm}$, from the container, we would not have to select the balls individually and make tedious measurements. We will use a sieve with holes whose diameter is equal to $3\,\mathrm{cm}$ (Fig. 2). After pouring the contents of container A onto the sieve, the balls of interest to us will go into container B placed under the sieve. Those that remain on the sieve can be discarded.



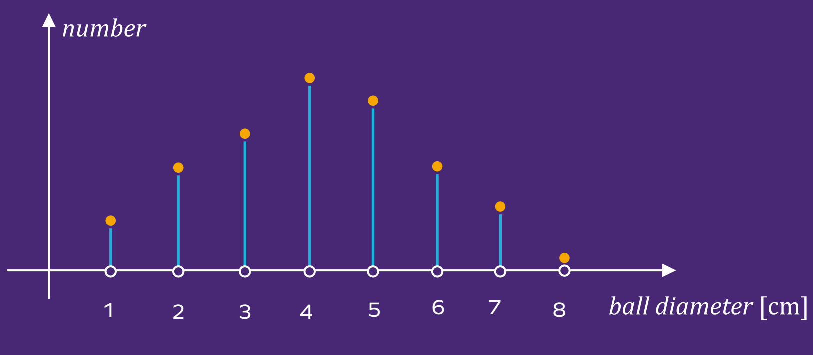


Fig. 1. Collection of balls of different diameters. The plot on the right side shows the number of balls for each diameter.

What will be the distribution of ball diameters in tank B fter this operation? Of course, there will only be balls with diameters of 1, 2 and 3 cm.

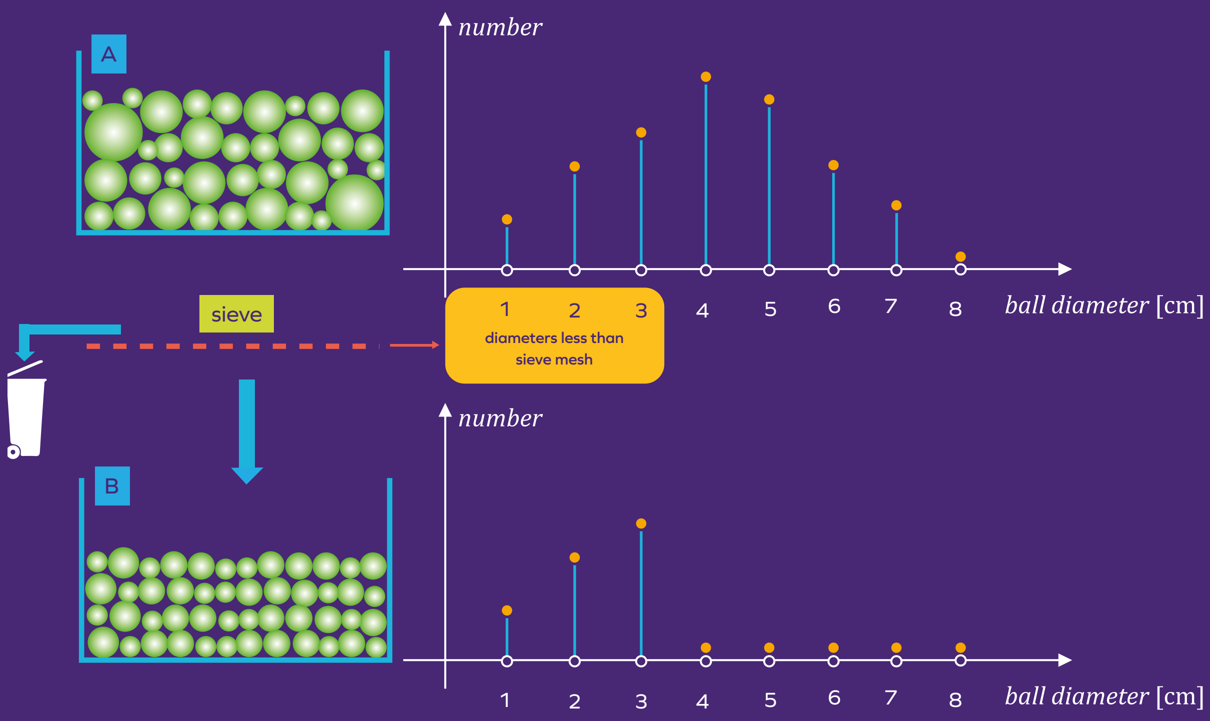


Fig. 2. **Using a sieve to select balls with diameters of $1-3$ cm.**

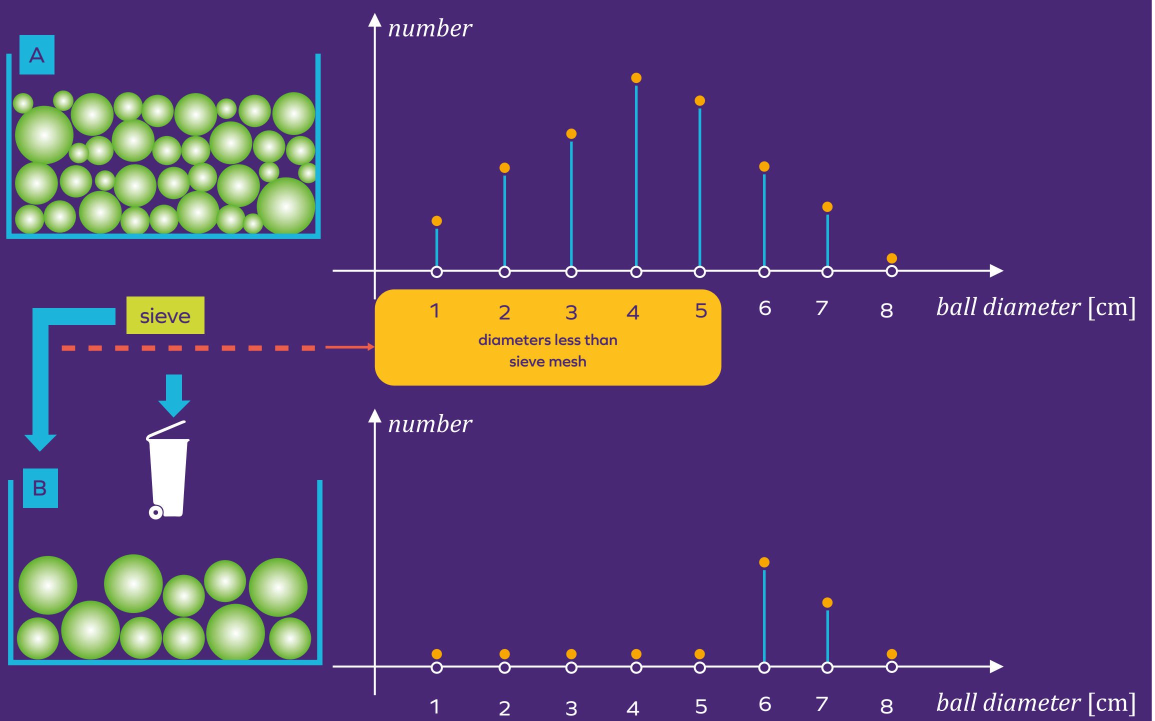


Fig. 3. **Using a sieve to select balls with diameters of $6-8$ cm.**

What if we wanted to select only balls with diameters greater than or equal to $6\,\mathrm{cm}$ from the container? Then we should use a sieve with meshes of $5\,\mathrm{cm}$ in diameter. This time, the sieve will let through balls with diameters equal to or less than $5\,\mathrm{cm}$, which we discard, while what remains on the sieve is collected in container B. The distribution of ball diameters in container B after filtration is shown in Fig. 3.

Exercise. Propose a way to use sieves – you can freely choose their number and size of holes – to select balls with diameters of $4-5\,\mathrm{cm}$.

2. Signal filtration concept

The filtering operation is so general that it can also be used in the context of signal processing.

Let us recall one of the signals analysed in Lesson 4 – its image in the time and frequency domain (spectrum) is shown in Fig. 4. The basic function of signal filtration is to select specific components from the spectrum of this signal and suppress the rest.

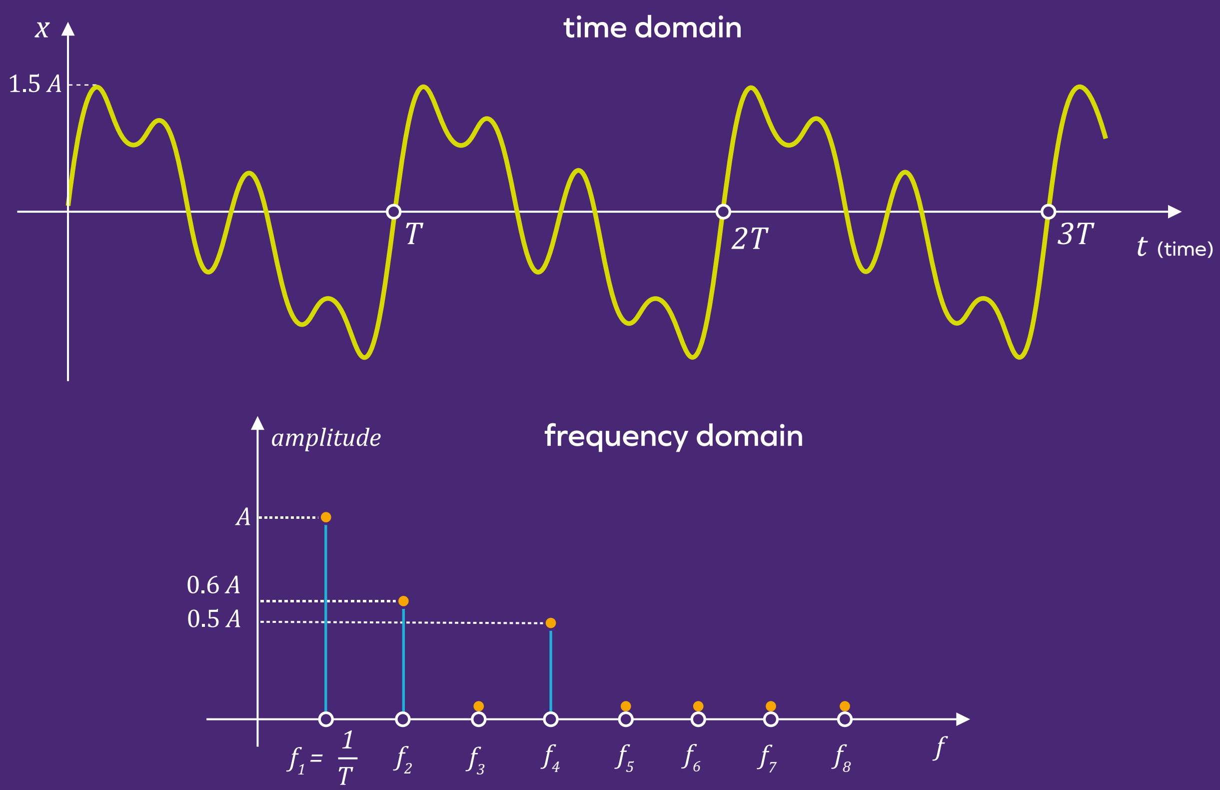


Fig. 4. Example signal represented in the time and frequency domains.

Filters used in signal processing can be divided into three basic groups:

- **Low-pass** – they transmit harmonic components of a signal *below* a certain, selected cut-off frequency.
- **High-pass** – they transmit harmonic components of a signal *above* a certain, selected cut-off frequency.
- **Band-pass** (or mid-pass) – they transmit harmonic components of a signal between certain, selected cut-off frequencies. In particular, a band-pass filter can separate one specific harmonic of a signal.

The effect of different filters on the signal in Fig. 4 is demonstrated in Fig. 5, Fig. 6 and Fig. 7 (the notation T as the period of the original signal is retained).

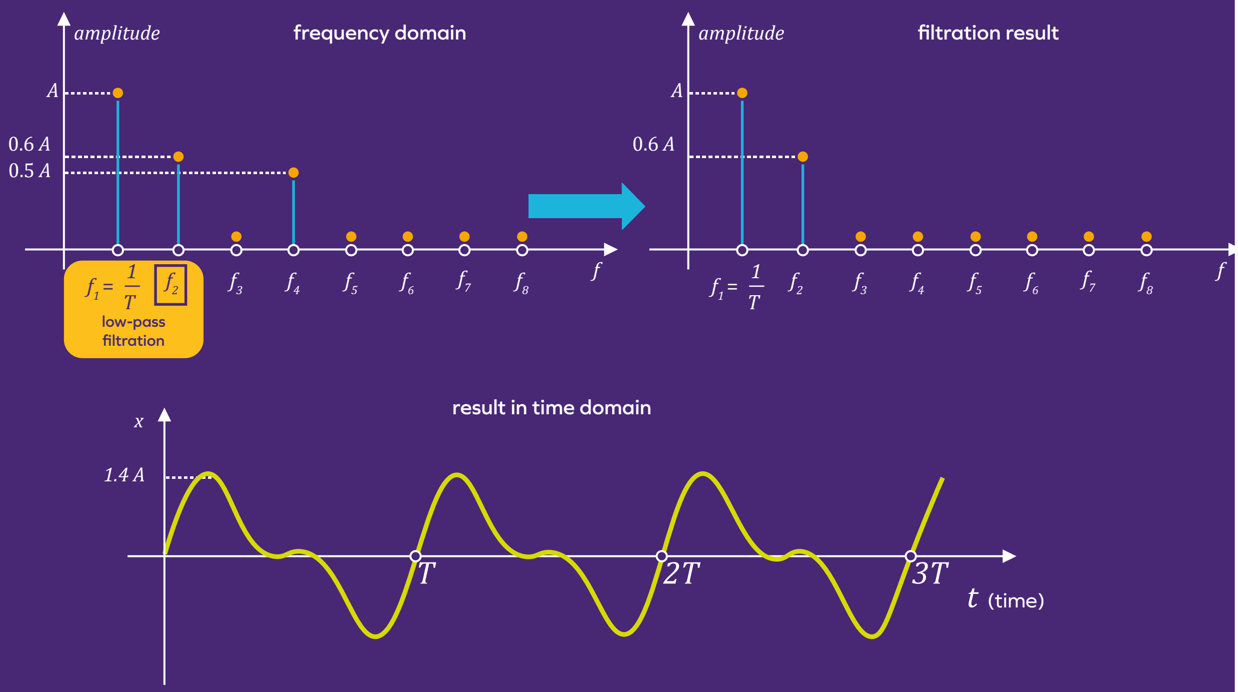


Fig. 5. Example of the low-pass filtration. Cut-off frequency – f_2.

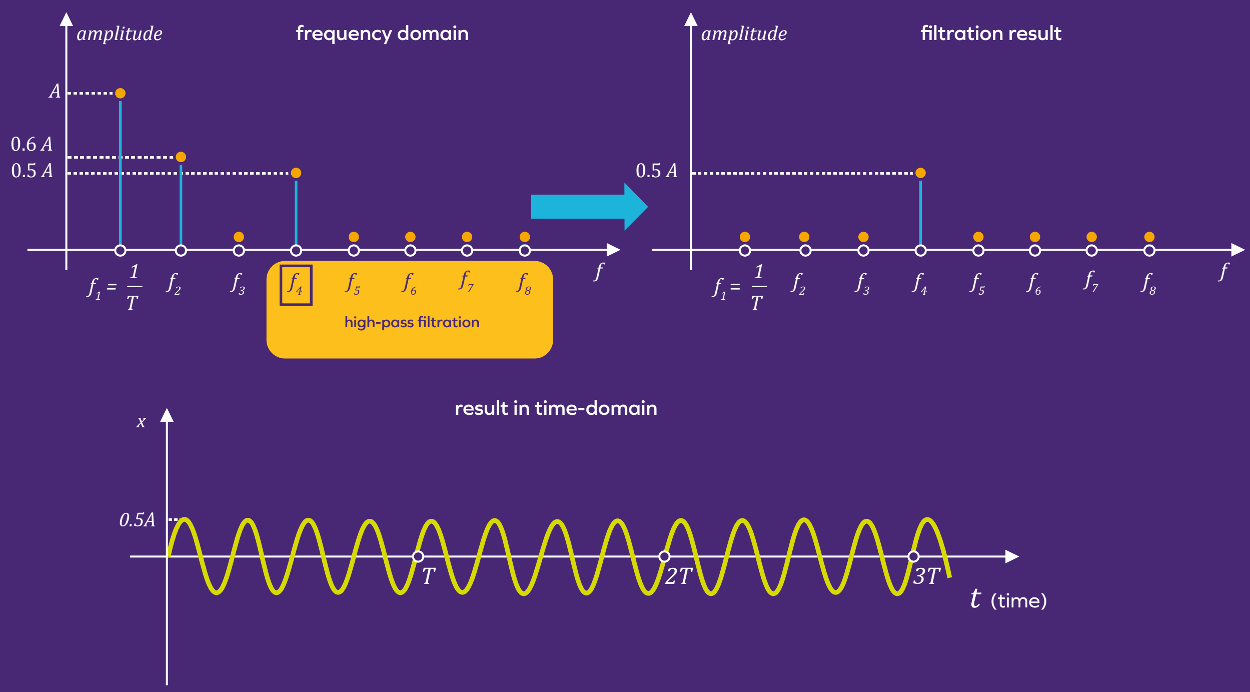


Fig. 6. **Example of high-pass filtration. Cut-off frequency – f_4.**

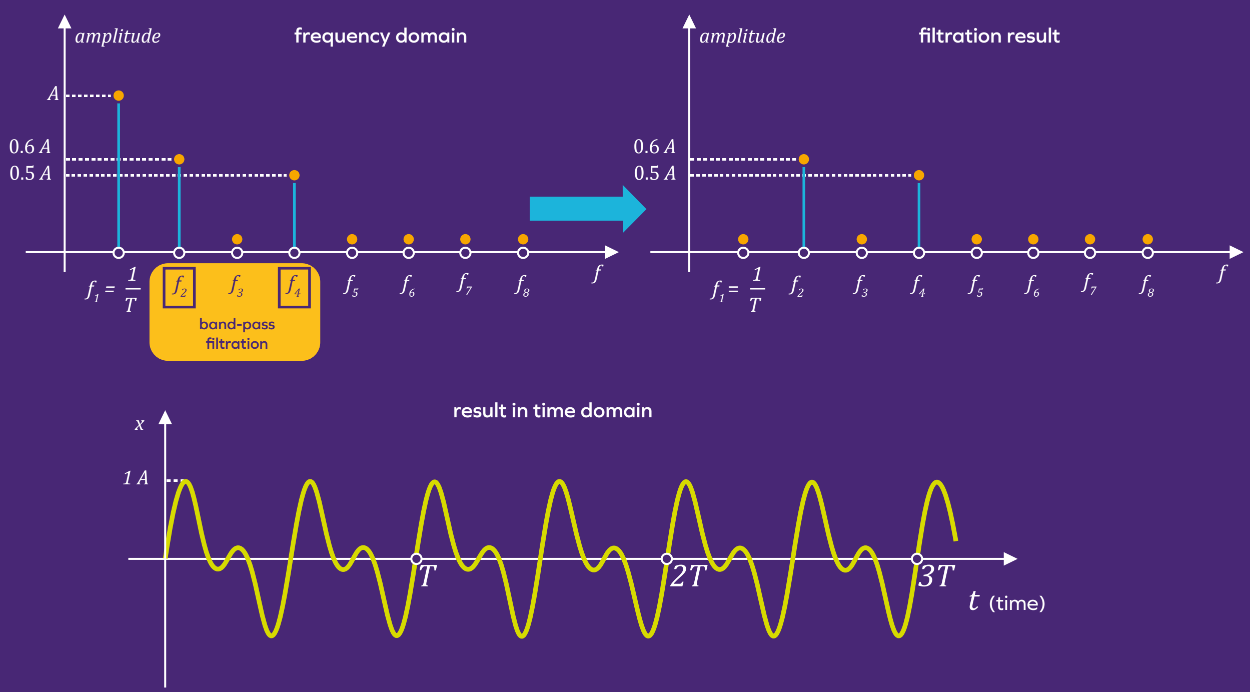


Fig. 7. **Przykład działania filtracji pasmowej. Częstotliwości graniczne – f_2 i f_4.**

Let us briefly discuss the effects of filters. A low-pass filter with a cut-off frequency of f_2 passes harmonics f_1 and f_2 unchanged (we assume that the cut-off frequency is passed by each type of filter), while all other harmonics are attenuated to zero (this fate befalls harmonic f_4). The image of the signal in the time domain can be seen at the bottom of Fig. 5. Note that the filter, by attenuating the fast-changing harmonic, has caused the signal to be "smoothed". This is a typical result of signal processing with a low-pass filter.

The high-pass filter has a somewhat opposite effect. Slowly changing components of the signal are attenuated. In the case of the filter in Fig. 6, only the f_4 harmonic is preserved. This filter therefore causes the separation of a single harmonic of the signal, but only because all the subsequent ones simply do not appear in this signal.

A band-pass filter with cut-off frequencies f_2 and f_4 is shown in Fig. 7. It suppresses the first harmonic of the signal, but since harmonics above f_4 have zero amplitude, the filter has no effect in this region.

3. How signal filters work

Most signal filters used in practice are implemented as a suitably constructed electronic system. In the case of electronic filters, it is of course necessary to first process the signal into electrical form. Elements such as coils or capacitors change their properties in an electrical circuit depending on the frequency of voltage changes and thanks to this we can process the electrical signal in the way we want, e.g. suppress harmonics above a certain limit value.

Many filter systems, especially band-pass filters, are based on the phenomenon of resonance. Let us assume that a certain physical system can vibrate – for example, a pendulum. Such systems usually have one frequency at which they vibrate if pushed out of equilibrium and left to themselves – this is the so-called **natural frequency**. For example, a swing (or any pendulum) once pushed will start moving in a periodic motion with a frequency that depends on its length.

If we want to stimulate a swing to oscillate with a large amplitude, we should act on it with a periodic force, which we call **excitation** (Fig. 8). Of what frequency? If the frequency is too high or too low, at certain swing deflections the excitation may be directed opposite to its movement, and thus may have a braking effect, not an accelerating one. The optimal excitation frequency is equal to the natural frequency, which should not come as a surprise. In such a case, we can speak of the phenomenon of **resonance** – the excitation accelerates the vibrating system exactly at the right moment, and the vibrations reach their maximum amplitude over time.

A filter built as a vibrating system with a specific natural frequency will show maximum response when the signal frequency ensures resonance. This allows us to isolate exactly those harmonics that interest us.

Fig. 8. **At what frequency should you push a swing for the best effect?**

An example of an acoustic wave filter is an ordinary wall. We have probably heard a loud conversation taking place in the next room more than once. It is not only quieter, which is associated with a general reduction in the signal amplitude, but it sounds completely different than if we were listening to it directly. What is the reason for this? An acoustic wave, containing a lot of harmonics, stimulates it to vibrate when it reaches the wall. The wall material usually has quite limited possibilities for vibrations at high frequencies. The amplitudes of high harmonics are therefore largely dampened after passing through the wall. For this reason, we perceive the sound of a given person's voice completely differently.

Question. What type of filter is the wall in this case?

It is worth adding that filters, especially signal filters, are hardly ever perfect. Harmonics that should be completely suppressed, although significantly weakened, can still get through the filter. This applies to filters of all types – low-pass, band-pass and high-pass.

4. Signal filtration applications

As we mentioned earlier, there are many applications for signal filters. Here we will focus on two – isolating a signal of a specific frequency (using a band-pass filter) and eliminating noise (using a low-pass filter).

In Lesson 2, we pointed out the electromagnetic field (EM) as a carrier of information. Since the EM field fills the entire space, we can ask how multiple users can use it at once. If one of them transmits a signal emitted into the space around the antenna, will there be no interference when another user wants to use his antenna and emit his signal? This does not have to be the case if the users agree on a so-called **frequency division**.

Let us assume for simplicity that we need to perform the division for only two users – A and B, whom we will also call senders. They can agree that the signal of sender A contains only harmonics from f_3 to f_5, and the signal of sender B – only harmonics from f_8 to f_{10} (inclusive). If they transmit simultaneously, the spectrum of the signal received by a certain receiving antenna may look like Fig. 9 (top). For convenience, we can present this type of spectrum in a simplified way as in Fig. 9 at the bottom – we then emphasise to whom a specific frequency range – i.e. band – belongs, without specifying in detail the amplitude values of individual harmonics, which can also change over time depending on the information content of the transmitted signal.

Can the receiver somehow select the signal from individual senders? Yes – all it takes is a band-pass filter with appropriate cut-off frequencies (see Fig. 10).

As we mentioned in Lesson 1, one of the disadvantages of an analogue signal is its high sensitivity to interference (noise). However, if we know the frequency range of the signal we are interested in and we know that the interference is in the upper part of the spectrum, we can eliminate it by using a low-pass filter. Unfortunately, in practice, noise also includes the frequencies of the useful signal and it cannot be completely removed.

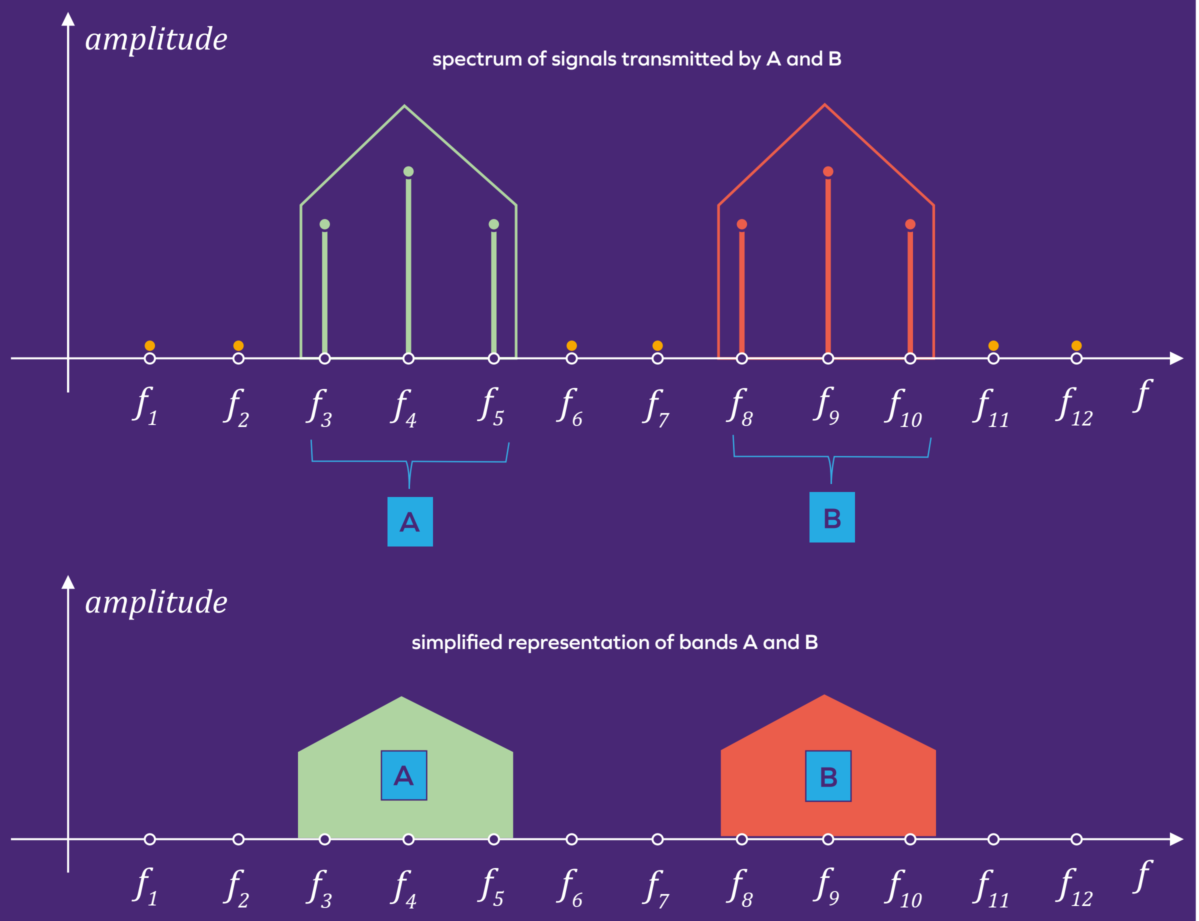


Fig. 9. Spectra of signals from both transmitters – A and B – under a fixed frequency division.

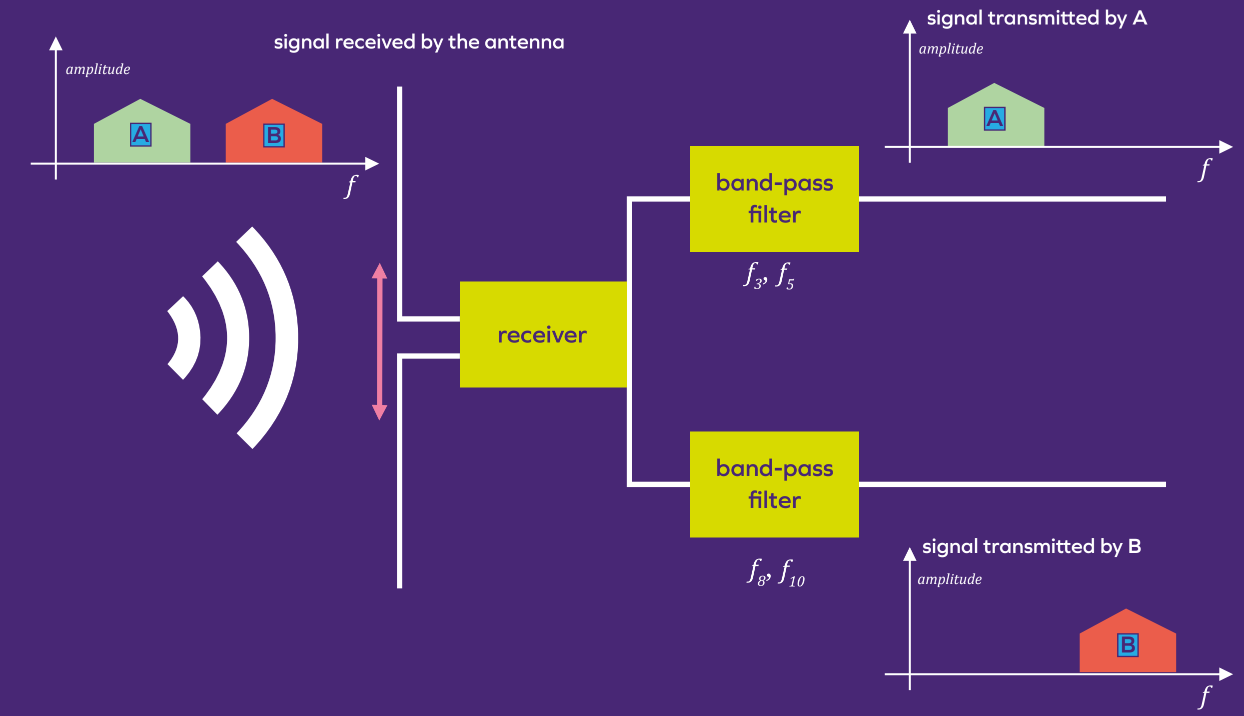


Fig. 10. **Using band-pass filters to select signals from different transmitters.**

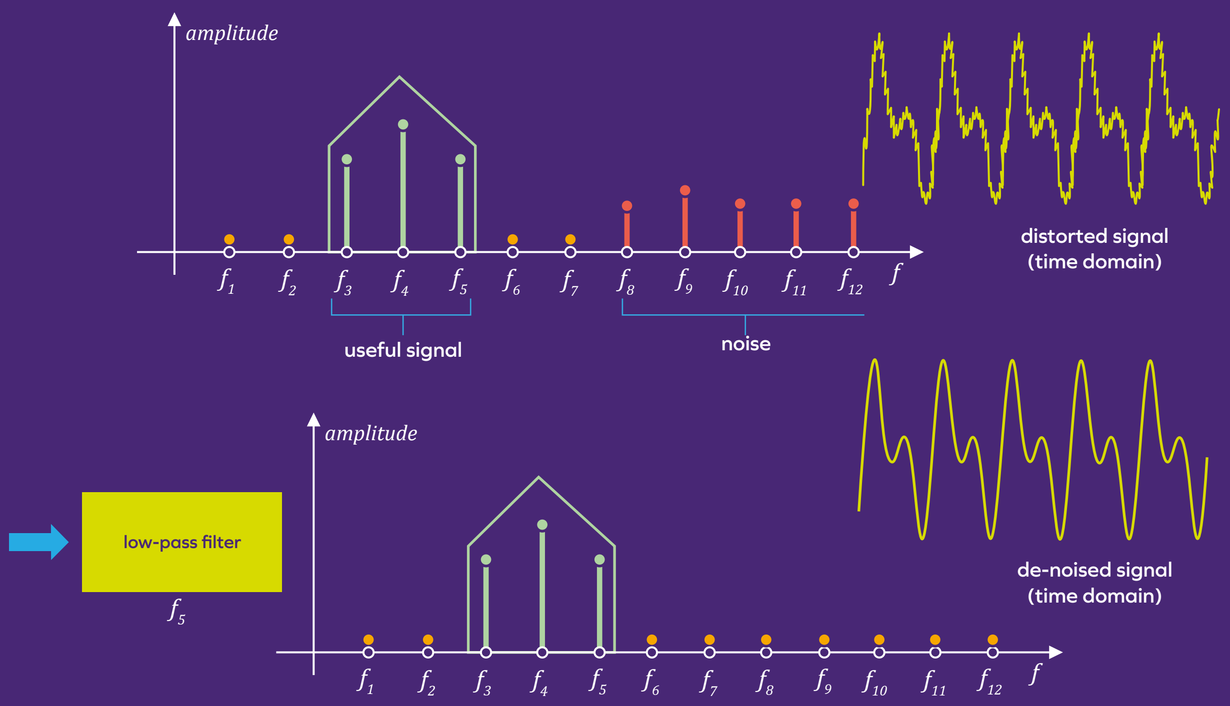


Fig. 11. **De-noising a signal.**

Experiment

Hang a thick string about $1.5\,\mathrm{m}$ long between two chairs as in Fig. 12. Near one of the chairs, hang three pendulums of different lengths (in the range of 20 to $60\,\mathrm{cm}$) – on the string – these can be nuts suspended on thin threads or lines. The distance between the pendulums should be about $20\,\mathrm{cm}$. Hang the same set of pendulums near the second chair. Make sure that the corresponding pendulums in both sets are the same length and that the string between the chairs is not too tight.

Fig. 12. Experiment with pendulums. Mechanical resonance.

1. Set the longest pendulum in motion in a direction perpendicular to the string. Observe what happens to the pendulums in the second set.
2. Stop all pendulums. Repeat the operation from point 1, but this time excite the intermediate pendulum.
3. Stop all pendulums. Repeat the operation from point 1, but this time excite the shortest pendulum.
4. Stop all the pendulums. Choose two of the pendulums from one set and make them swing simultaneously. Do the pendulums from the other set respond as you would expect?

Discussion. The movement of the pendulum in one of the sets stimulates the string on which the other set is suspended to vibrate. The string oscillations act as an excitation that drives the movement of all the pendulums. Excitation is most effective in resonance conditions, i.e. when its frequency is equal to the frequency of natural vibrations. This means that the pendulum of the same length as the pendulum that is forcing the movement is most stimulated. The other pendulums also enter into small vibrations, but their amplitude is small and is periodically damped due to the inconsistency of the excitation frequency with the natural frequency of a given pendulum. Here we are dealing with the so-called **mechanical resonance**, in contrast to the electrical resonance occurring in vibrating electrical circuits.

Glossary

Natural vibrations – movement performed by an oscillating system after a single stimulation. It usually occurs with a frequency characteristic of a given system (in the case of a pendulum, it depends on its length).

Filtration – an operation involving the selection of objects with desired features. Examples of filtration include sieving or selecting specific harmonics of a signal.

Filter – a device that performs filtration. It can be a mechanical or electrical system.

Frequency division – the assignment of a specific frequency band to different users.

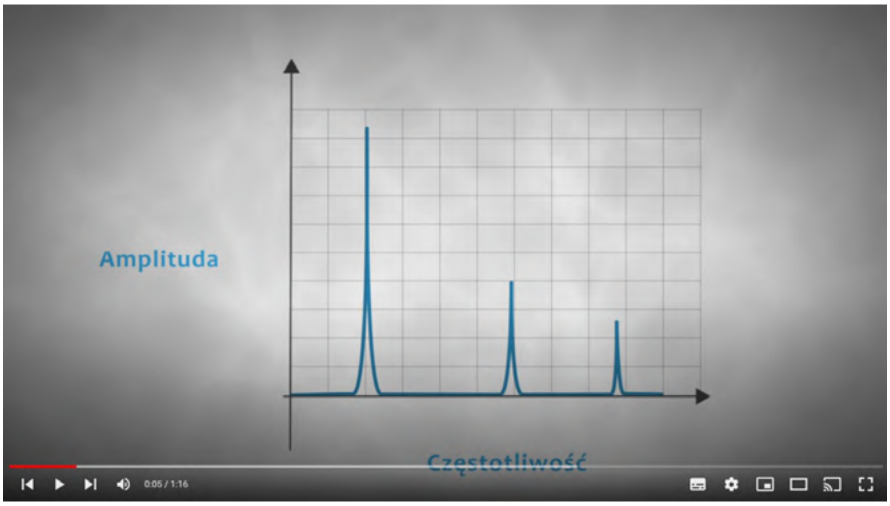
Band – a specific frequency range. A continuous fragment of the signal spectrum.

Resonance – a phenomenon that occurs when the excitation frequency is equal to the natural frequency of the system. This results in maximum excitation of the vibrating system.

Excitation – a force that excites a vibrating system. If it is a periodic force, its frequency is independent of the natural frequency of the system. It produces a maximum effect under resonance conditions.

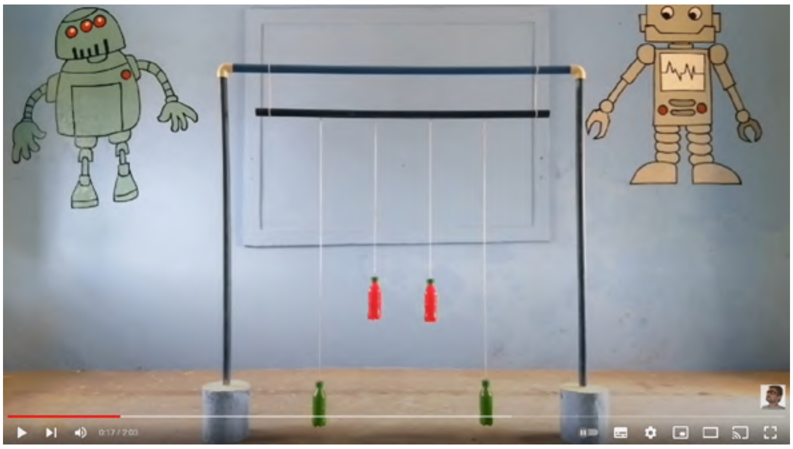
External materials

1. Signal filtration.



Scan QR code

2. Pendulum experiment illustrating mechanical resonance (*Resonant Pendulum | Science Experiments | science projects*).

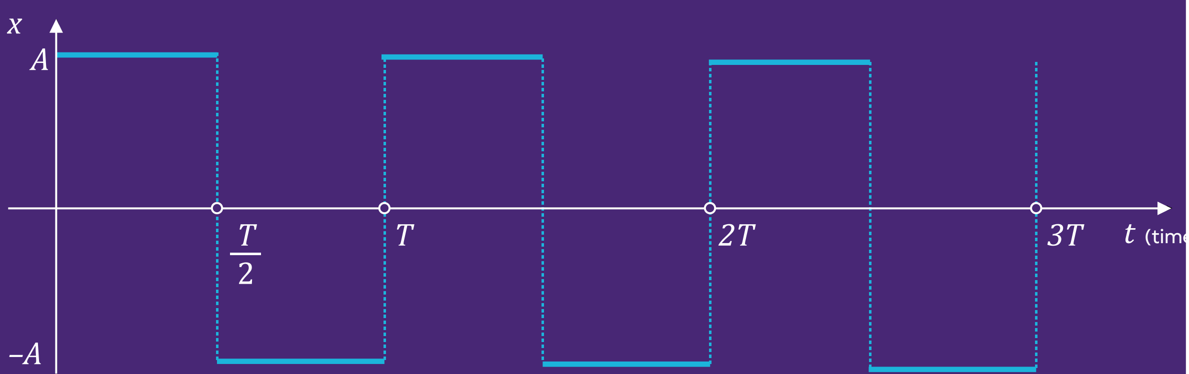


Scan QR code

Homework

1. You have two filters at your disposal: a high-pass with a cut-off frequency of f_5 and a low-pass with a cut-off frequency of f_8. Can you use them to build a band-pass filter with a range of f_5 and f_8?

2. Is it possible to obtain a harmonic signal with the same period from a square wave signal (see Lesson 4) with period T?



Lesson 6

Electromagnetic waves and obstacles

Objective

- Presentation of basic information about the phenomena accompanying the propagation of electromagnetic waves in an environment with material obstacles.

Learning outcomes

- The student is able to describe the behaviour of a wave impinging on a material obstacle.
- The student knows the special role of conductive obstacles in the transmission of electromagnetic waves.
- The student is able to describe the phenomenon of diffraction, its dependence on wavelength and its role in the transmission of information via electromagnetic waves.

1. Behaviour of EM waves at the interface between media

In Lessons 2 and 3 we mentioned the emission of electromagnetic (EM) waves as disturbances of the EM field propagating in a vacuum. However, when we want to use EM waves to transmit information, especially in urban conditions, the question arises: what happens when an EM wave encounters an obstacle in the form of another medium or material object?

When a wave hits an obstacle, some of the energy may be reflected (Fig. 1), which means that a wave is emitted from the obstacle at the point of incidence at the same angle as the incident wave, but on the opposite side of a straight line – the so-called **normal** – perpendicular to the surface of the obstacle at the point of incidence (marked in the figure by the vertical dotted line).

The part of the wave energy that has not been reflected, penetrates the obstacle, undergoes refraction depending on the properties of the obstacle material (the direction of the wave changes in relation to the incident wave). Part of the wave energy can be partially **absorbed** by the material medium, which causes a gradual decrease in the wave amplitude, after which the wave undergoes another refraction when leaving the obstacle on the other side (here too, partial reflection can occur, but for simplicity we will ignore this possibility).

According to the law of conservation of energy, the sum of the energy of the reflected wave, the energy absorbed by the obstacle, and the energy of the transmitted wave must equal the energy of the incident wave. How exactly the energy of the incident wave is divided between these different forms depends very much on the material of the obstacle and the wavelength of the incident wave.

2. Examples of obstacles to EM waves

A very special kind of obstacle is a metal object. Metal, as a perfect conductor, can provide such a distribution of electric charges on its surface as to cancel the electric field in its interior (if the field is not exactly cancelled, it causes further movement of charges until this effect is achieved).

This means that an EM wave cannot exist inside a metal object – a wave incident on its surface is almost completely reflected. What is more, for this phenomenon to occur, the surface does not have to be solid. An EM wave will be reflected even from a metal grid, the mesh of which is smaller than the wavelength of the incident wave.

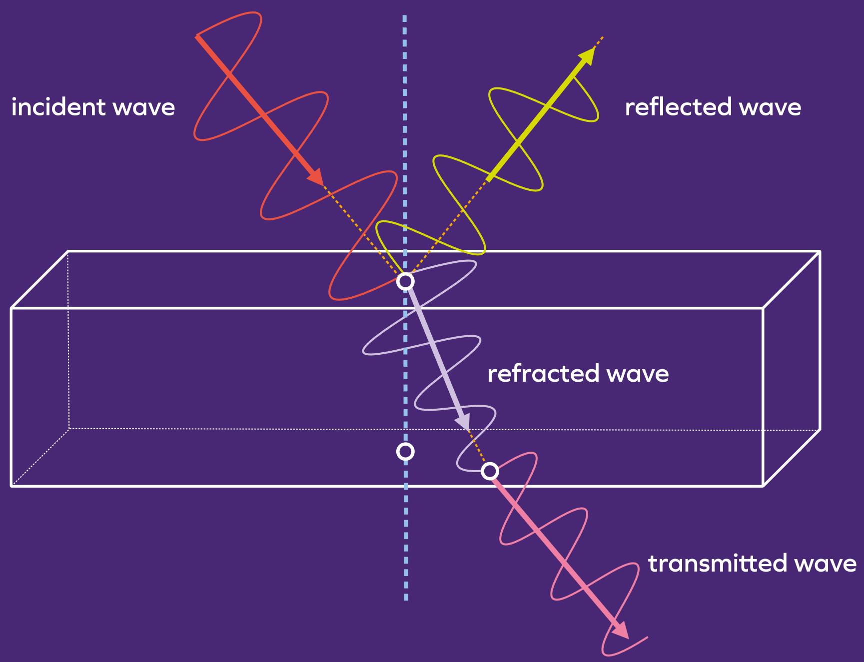


Fig. 1. General diagram illustrating wave behavior when passing through a material obstacle.

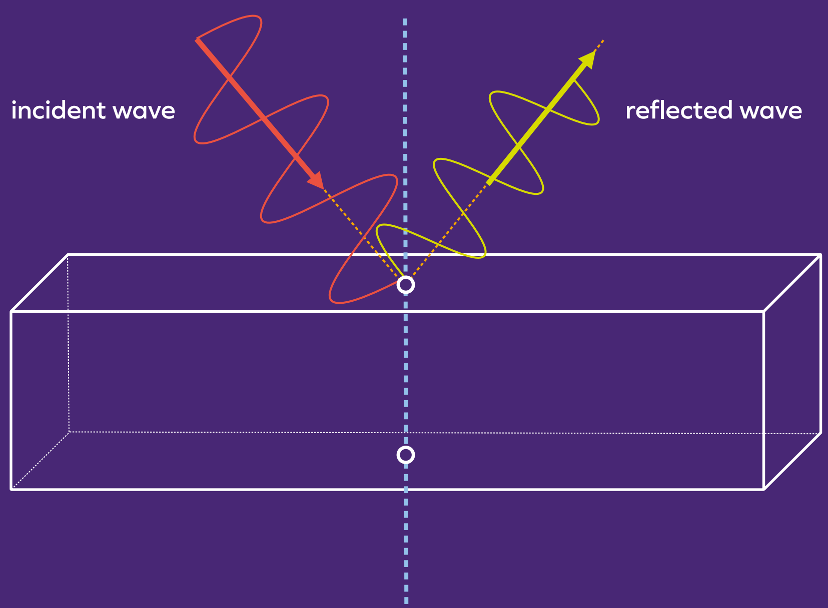


Fig. 2. Reflection of an EM wave from a metal surface.

For non-metallic (or generally non-conductive) obstacles, the situation is more complex. Very long EM waves, used, for example, in classic radio stations, penetrate a large part of the walls with negligible reflection and refraction (Fig. 3). If some of the wave energy is absorbed in a material medium, this energy is used to stimulate the atoms or molecules of this medium to move. This manifests itself as heat release and leads to its heating.

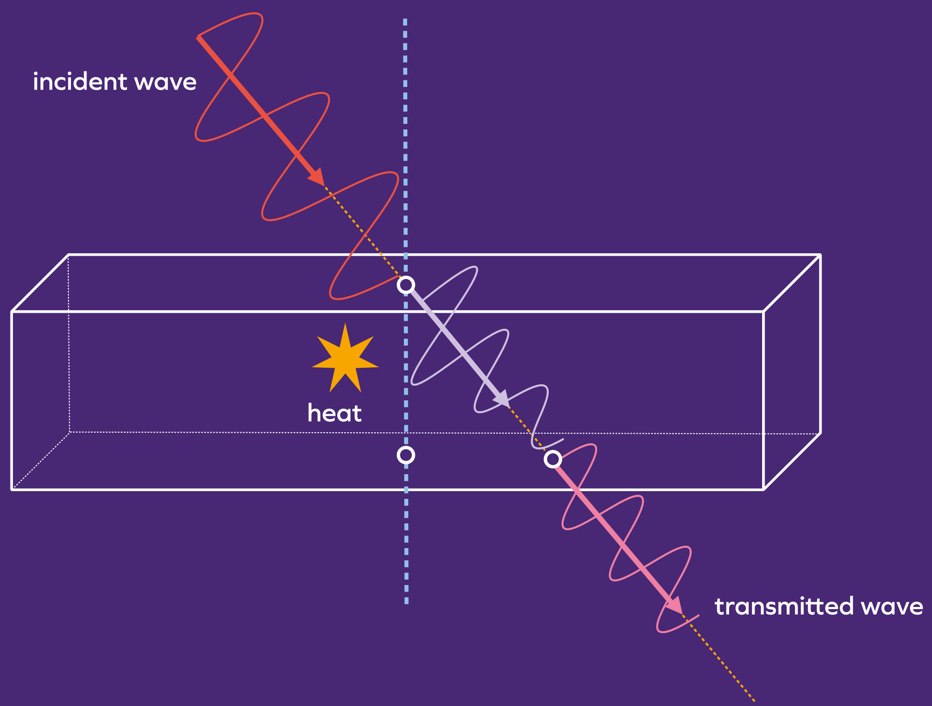


Fig. 3. Penetration of EM wave through non-conductive material obstacle with negligible reflection.

How vastly different the behaviour of EM waves in material media can be is illustrated by the example of light and ultraviolet radiation. Visible light is an electromagnetic wave, the wavelengths of which are in the range of 400 to 800 nanometres ($1\,\mathrm{nm} = 10^{-9}\,\mathrm{m}$). From everyday experience, we know that light penetrates ordinary glass perfectly (there is also partial reflection of about 4%). The same glass, in turn, absorbs to a very large extent ultraviolet UVB radiation, which is also an EM wave with a wavelength in the range of 280 to $315\,\mathrm{nm}$ (outside the visible range). For this very reason, it is not possible to get a tan through window panes.

Strong absorption of EM wave energy in a certain range is related to the existence of resonant frequencies (see Lesson 5) for vibrational and rotational movements of atoms and molecules that make up a given obstacle. The EM wave acts as an excitation, and in resonance conditions the absorbed energy is not radiated further as an EM wave (as is the case, for example, with wave reflection) but is used to stimulate the movement of the molecules of a given body, i.e. to heat it.

The aforementioned radio waves with lengths from $10\,\mathrm{m}$ to $10\,\mathrm{km}$ penetrate without significant losses even through thick concrete walls, which in turn can be almost impenetrable for microwave rays used, for example, in mobile telephony. Of course, walls made of so--called reinforced concrete, i.e. reinforced with steel rods or mesh, are hard obstacles for almost all EM wave ranges used in telecommunications. This is a major challenge for ensuring mobile communication in urban environments, between buildings and inside them.

3. Wave diffraction

To describe another important phenomenon related to wave behaviour near obstacles, it is worth specifying the concept of a **plane wave**. We deal with such a wave when **wave fronts** (or wave surfaces), i.e. lines connecting points of the same phase, are parallel to each other (Fig. 4).

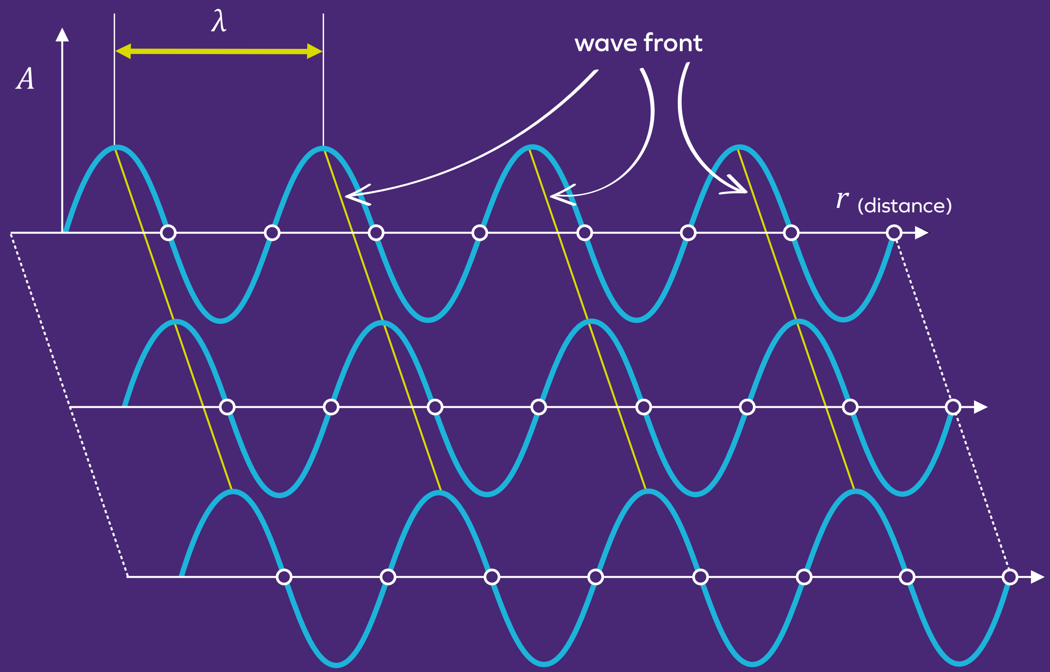


Fig. 4. Plane wave. Wavefront as a surface encompassing points of the same phase.

Any wave propagating in a sufficiently small segment can be considered a plane wave. Waves propagating in a circle on the surface of water are not plane waves, but in a small angular segment the wave fronts of such a wave can be considered approximately parallel. The advantage of the concept of a plane wave is that it allows for the unambiguous determination of the direction of wave propagation – perpendicular to the wave fronts.

One of the characteristic features of waves is that when they hit the edge of an obstacle that is impermeable to them, they are bent. Let us look at the waves, which in the lower part of Fig. 5 are plane waves and move vertically upwards. The wave fronts are marked in blue. When they reach the edge of the obstacle, the wave fronts are bent towards the obstacle, and thus the direction of the wave's travel changes. This effect, called **diffraction**, is all the more visible the longer the wavelength. The angle of deflection α udepends on the ratio of the wavelength to the size of the obstacle.

The diffraction phenomenon of a plane wave incident on a slit of width D is shown in Fig. 6 (the images were generated using the application linked in External Materials [1]).

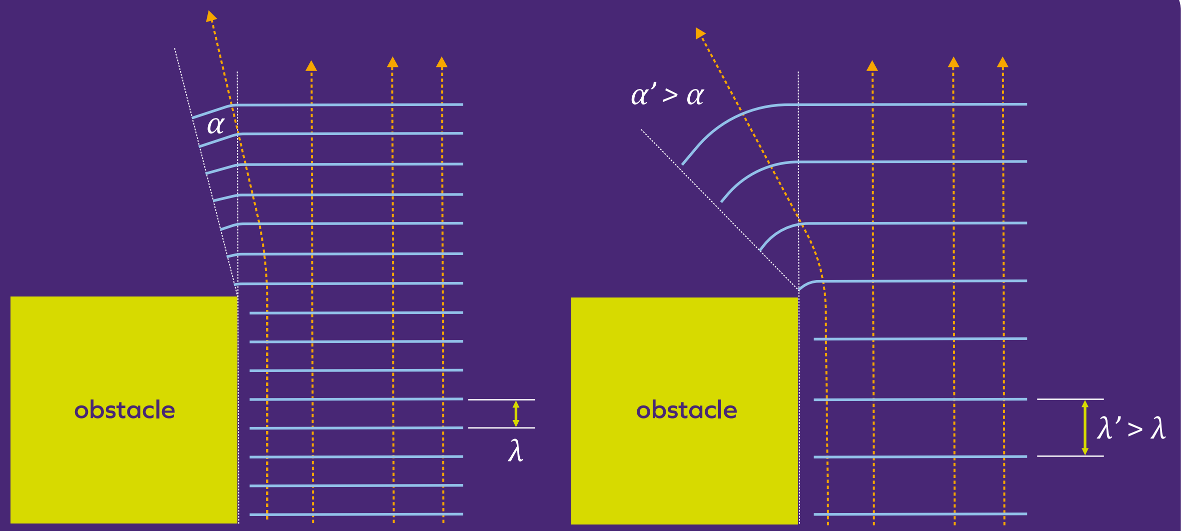


Fig. 5. Diffraction of a plane wave on an impenetrable obstacle.

As can be seen, when the wavelength is much smaller than the slit width, most of the wave energy is concentrated in the small angle region. The case of a wavelength larger than the slit width presents a completely different picture – the wave seems to spread behind the obstacle almost uniformly as a circular wave.

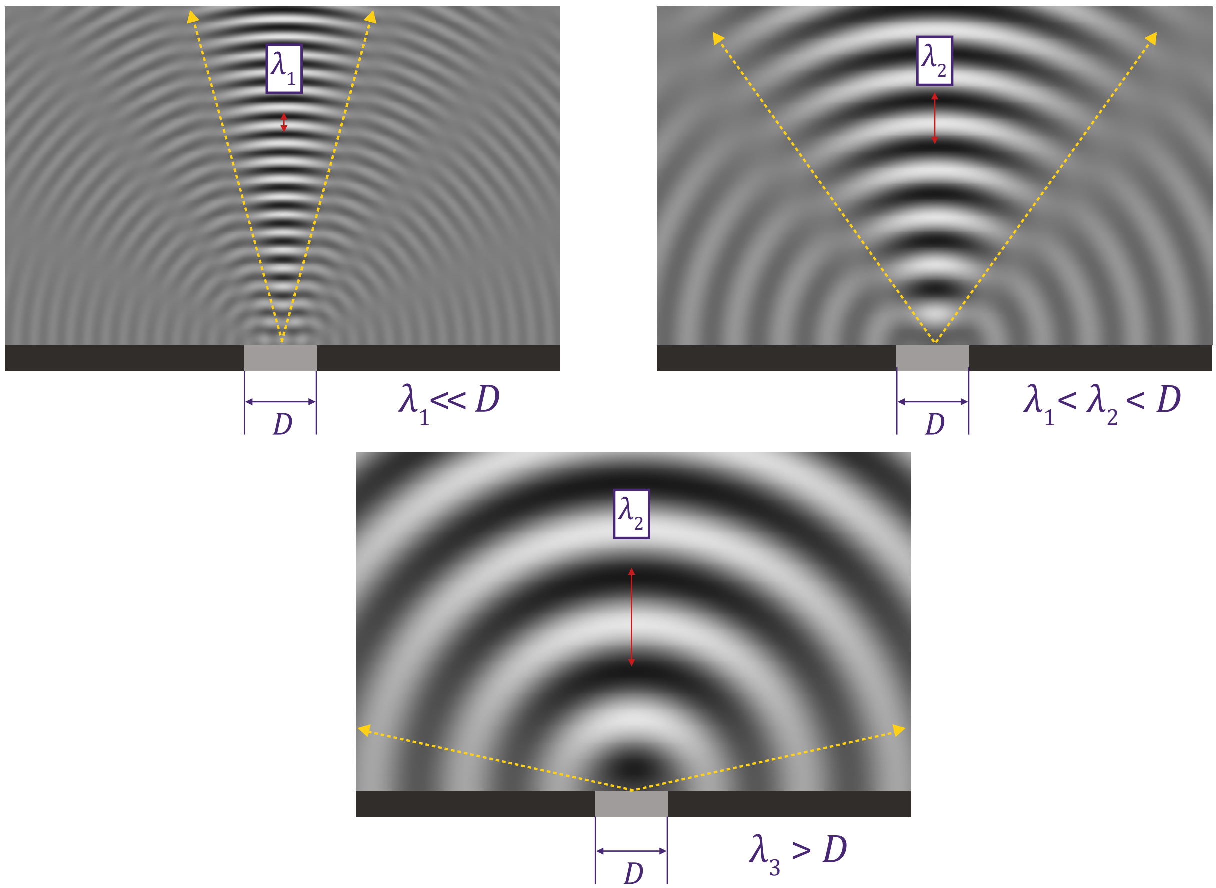


Fig. 6. Diffraction of a plane wave on a slit.

The phenomenon of diffraction of acoustic waves is very important in voice communication, because the wavelengths emitted by the organs of speech are comparable to the typical sizes of obstacles in our environment. Thanks to it, we can freely talk to someone standing just around the corner of the building, even if we do not see them.

The very short wavelength of visible light means that the bending of light on obstacles is negligible (but can be detected by the naked eye – see Experiment). This is also of great importance to us, because we can assume that light rays move in a straight line, which makes it much easier for us to accurately map our surroundings using our eyesight. Bats and other animals that base their behaviour on echolocation emit inaudible ultrasound, or acoustic waves with a very short wavelength, precisely to reduce diffraction effects. For a similar reason, ultrasound is used in ultrasonography (USG).

The bending of EM waves used in classic radio (wavelengths of the order of kilometres) allows them to reach hard-to-reach places outside the visibility area of the transmitting antenna (see External Materials [3]). As the wavelength of the transmitted signal shortens, and therefore its frequency increases (which we observe especially in mobile telephony), the bending becomes smaller and smaller, and it becomes increasingly important to reduce the number of obstacles in the path of signal transmission. We will deal with such issues in more detail in Lesson 10.

Experiment

Although the wavelength of light is very small ($400 - 800\,\mathrm{nm}$) and the observation of strong diffraction effects usually requires special tools (so-called diffraction gratings), there is a very simple experiment that not only allows the direct observation of light diffraction, but also confirmation of its dependence on wavelength.

To carry out the experiment, you will only need a small, bright object (preferably white) on a clearly darker background (the darker the better) – Fig. 7a.

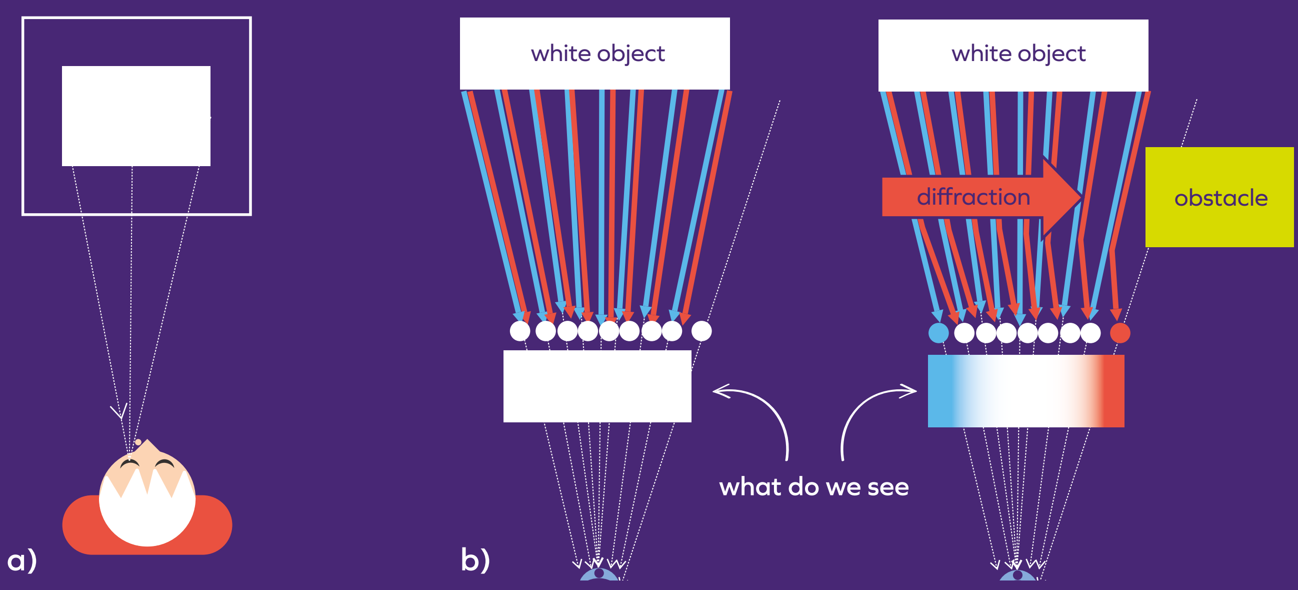


Fig. 7. Diffraction of light.

78

Close one eye (let us assume the right one, but that doesn't matter) and focus on the object by placing your finger on the left side of your nose so that it blocks your field of vision to the edge of the object (you can also turn your head to the left instead of blocking your view with your finger so that the edge of your nose acts as a block). Remember to keep your eyes focused on the bright object. Is there anything strange happening?

The edge of an object near an obstacle takes on a red tint, while the opposite edge takes on a blue tint. The effect is more visible the narrower the bright object.

Try to explain why this is happening.

Discussion

The observed effect can be easily explained by the diffraction of light and the fact that longer wavelengths are more susceptible to it. White light is basically composed of all the colours of the rainbow but let us assume for simplicity that just two colours - blue and red - create white when mixed. They represent a very small and a very large wavelength, respectively. Let us assume another simplification, namely that the diffraction effect of the red wave is significant, while for the blue one – negligible.

Now let us look at Fig. 7b. In the absence of any obstacles, light rays reach the eye in a mixed form, giving the object a white colour. What happens when an opaque obstacle appears next to the rays? The waves are bent in its direction, and the longer the wavelength, the more so. In the drawing, we present this effect in a simplified way as all the red rays shifting to the right, while the blue ones practically do not change. The central part of the object remains white as a result - the red rays continue to mix with the blue ones (although coming from a different part of the object). The situation is different at the edges. At the right edge, the bent red rays do not encounter their blue counterparts, and we see an excess of red in this place. At the left edge, the red rays no longer mix with the extreme blue ones - hence the blue shade appears in this place.

Glossary

Absorption – the phenomenon of wave energy being taken in by a material medium or object.

Diffraction – the phenomenon of wave bending at the edge of an impermeable obstacle. The amount of bending depends on the ratio of the wavelength to the size of the obstacle (e.g. the width of the gap through which the wave passes).

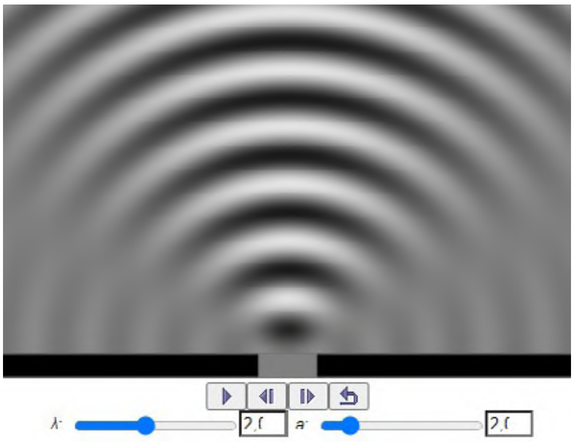
Wave front – a line (or surface if the wave propagates in space) connecting points of the same phase.

Plane wave – a wave whose wave fronts are lines (or surfaces, if the wave propagates in space) parallel to each other.

Normal – a line perpendicular to the surface of the body on which the wave is incident.

External materials

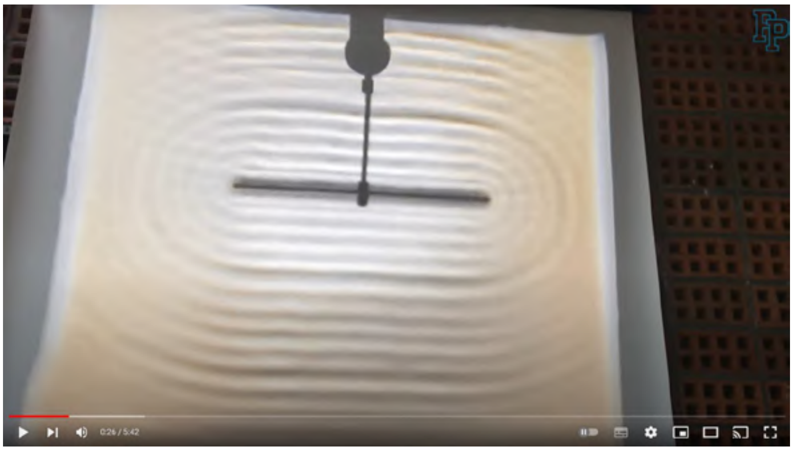
1. Online application enabling observation of diffraction of a wavelength λ on a slit of width α. Both parameters can be adjusted independently (*Diffraction Simulation ComPADRE*).



Scan QR code

2. Video: Diffraction phenomenon (experimental demonstration of the diffraction phenomenon in waves on a water surface).

Scan QR code



3. Animation showing the diffraction of radio waves and the influence of the wavelength on the scale of wave deflection by terrain obstacles (*Radio wave diffraction*).

Scan QR code

Homework

1. Bats use echolocation to detect and capture prey. What is the minimum sound frequency a bat must use to locate a $2\,\text{cm}$ moth? Assume the speed of sound is $340\,\text{m/s}$. Hint: you can assume that echolocation will work correctly when the object is comparable to the wavelength (or larger).

   Given:

   Speed of sound: $v = 340\,\text{m/s}$

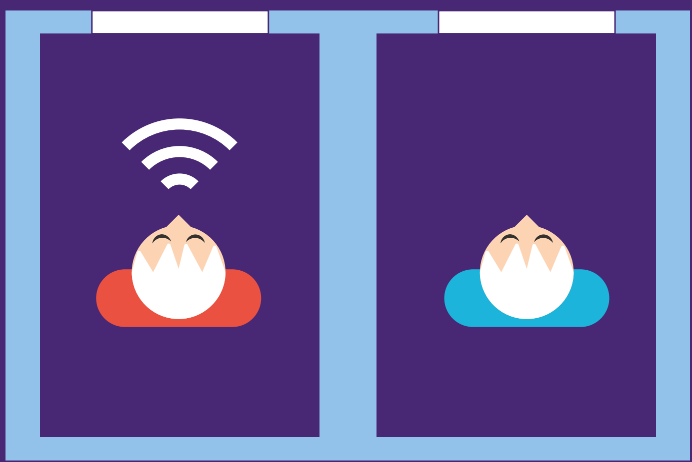
   Object size: $d = 2\text{ cm} = 0.02\text{ m}$

   Wavelength: $\lambda \approx d = 0.02\text{ m}$

   To find:

   Sound frequency: $f = ?$

2. Is it possible for individuals in two rooms separated by a soundproof wall to converse through open windows? Draw a schematic diagram of the propagation of the acoustic wave from the sender in the left room. Take into account that the wavelength of the sound waves in the voice message is large compared to the size of the obstacles.



Lesson 7

Amplitude modulation – AM

Objective

- To present the basic concept of amplitude modulation (AM) of a carrier wave.

Learning outcomes

- The student knows the role of modulation in transmitting information using wave phenomena.
- The student knows the basic types of amplitude modulation.
- The student can explain the structure of the spectrum of an amplitude-modulated wave with a harmonic signal.
- The student can explain the shifting of the modulating signal's spectrum near the carrier frequency and the presence of sidebands.

1. Harmonic wave does not carry information

In Lesson 3, we introduced the harmonic wave as an image of the uniform motion of a point in a circle. To describe its variation in time, it is enough to give its amplitude A, frequency f (or period T), and phase at time $t=0$. Can these parameters carry information? Let us specify that in the strict sense, a harmonic wave extends indefinitely in time into both the future and the past (Fig. 1). The wave parameters are therefore unchanging and there is no way to use them to transmit information.

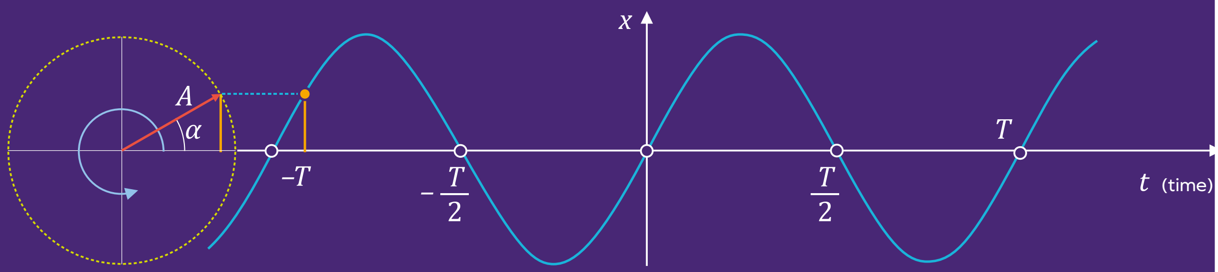


Fig. 1. **Harmonic signal – a periodic signal without a beginning or an end in time.**

If this seems counterintuitive – because, after all, one could encode a measured temperature value in the wave's amplitude – it is primarily because we rarely think of a signal as something extending infinitely into the past. Transmitting information via one of the wave's parameters would be possible if the wave were "activated" at time $t=0$ (Fig. 2a).

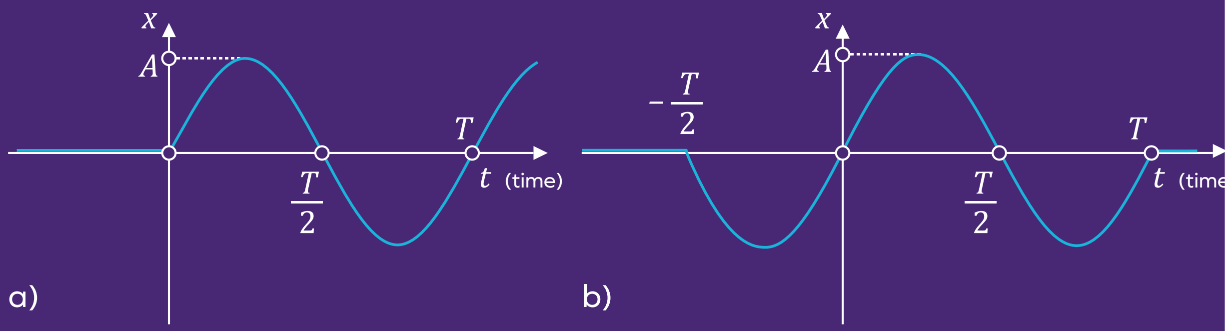


Fig. 2. **Non-harmonic signals composed of segments of a harmonic signal.**

We are then dealing with a situation in which the signal is not received up to a certain point, and then vibrations occur with a specific amplitude, the value of which can be freely selected according to the information content of the signal.

The signals shown in Fig. 2 are not harmonic waves (strictly speaking), although we can treat them as sections of a harmonic wave. The first one has zero amplitude up to time $t=0$, then A, while the second signal has amplitude equal to A only in the time interval from $t=-T/2$ to $t=T$. Outside this interval its amplitude is zero.

2. Carrier wave modulation

Changing the parameters of a harmonic wave over time is called **modulation**. The modulated wave ceases to be a harmonic wave, but we can still represent it as an image of the motion of a point in a circle, at least in certain time intervals. A wave whose parameters are changed over time is called a **carrier wave**, and its frequency is called the **carrier wave frequency** and is denoted by f_c.

In this lesson, we will focus on **amplitude modulation**, in which the parameter of a wave that carries information is its amplitude. In short, this type of modulation is often called **AM**.

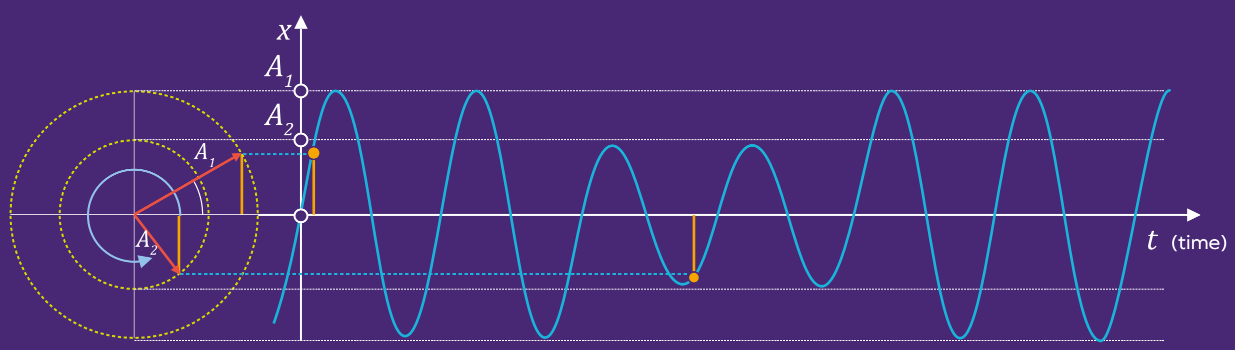


Fig. 3. Digital AM modulation of a carrier wave.

Let us examine at the simplest case of AM modulation, in which the signal amplitude can take only two possible values A_1 and A_2 (see Fig. 3). Such a signal can be represented as the result of the rotational motion of two vectors of different lengths but with the same rotation period. At the moment of amplitude change, the point jumps from one circle to another.

When the number of signal amplitude levels is finite, this modulation is called **digital** (we also speak of **amplitude keying**).

The signals shown in Figure 2 are examples of digitally modulated signals. They take only two values – A i 0.

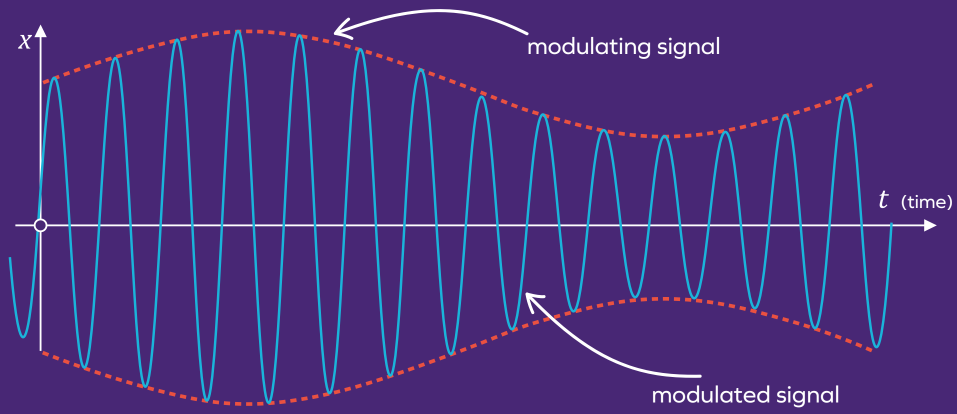


Fig. 4. Analogue AM modulation of a carrier wave.

A more general method of modulation is the continuous change of amplitude depending on the value of the analogue modulating signal (Fig. 4). The concept of an analogue signal was discussed in Lesson 1.

Question. If you are transmitting a Morse code message by turning a flashlight on and off, are you using some type of modulation?

3. Harmonic wave as a modulating signal

What does the spectrum, or frequency-domain representation, of an amplitude-modulated signal (see Lesson 4) look like? Suppose we have a continuously modulated carrier wave, as shown in Figure 4. Assume that the carrier wave has a period of T, so the carrier frequency is $f_c = 1/T$.

At first glance, the question seems very easy. Since the signal has a fixed frequency and only its amplitude changes, the spectrum should consist of a single line at frequency f_c, whose height varying with time would reflect the changes in the signal amplitude (Fig. 5).

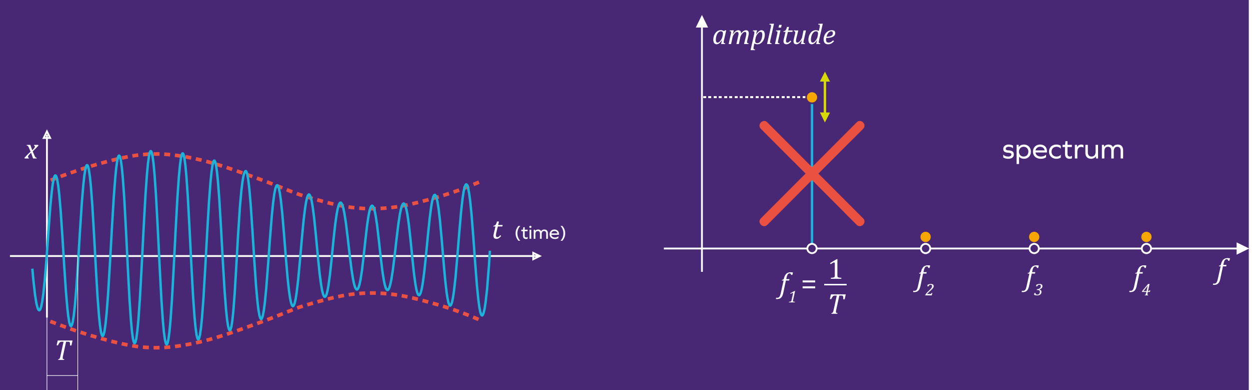


Fig. 5. A common misconception about the spectrum of an amplitude-modulated signal – a single bar with varying amplitude.

However, this is a false idea. The individual bars of the signal spectrum describe the amplitudes of the harmonic signals that make up the given signal. A harmonic wave has a fixed, identical amplitude at each time instant (recall Fig. 1). One bar is not able to reflect the temporal variability of the signal shown in Fig. 5, which has different amplitudes at different time instants.

Let us therefore analyse the spectrum of the amplitude modulated signal in more depth. For simplicity, let us first assume that the modulating signal is a harmonic wave with frequency f_m (independent of the carrier frequency f_c) – see Fig. 6.

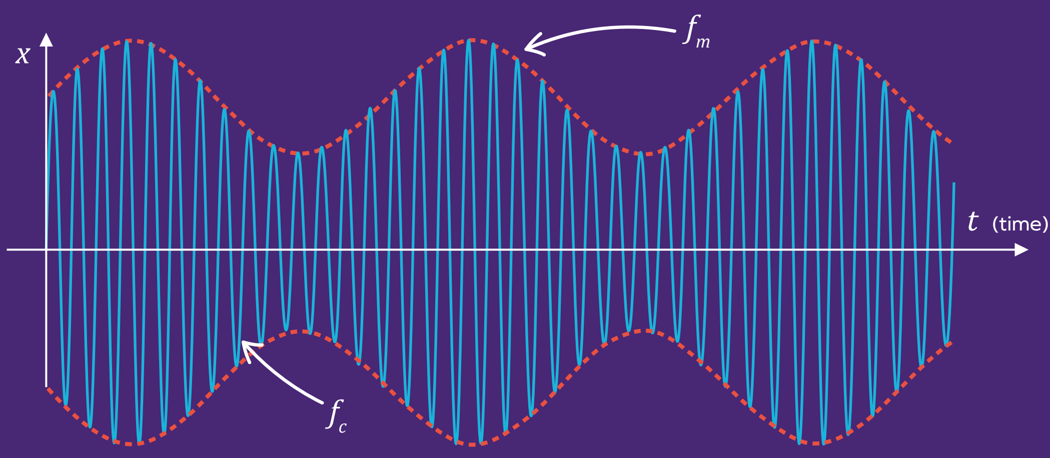


Fig. 6. Carrier wave modulated by a harmonic wave of frequency f_m.

Let us look at the circular diagram in Fig. 7a. We will adopt the direction of the vector representing the carrier wave with amplitude A_c as the reference direction and mark it with an orange arrow. This direction, along with the carrier wave vector, rotates at the carrier frequency f_c. Can we introduce one additional vector, representing the modulating signal, which would lead to a change in the length of the carrier wave vector to achieve AM modulation?

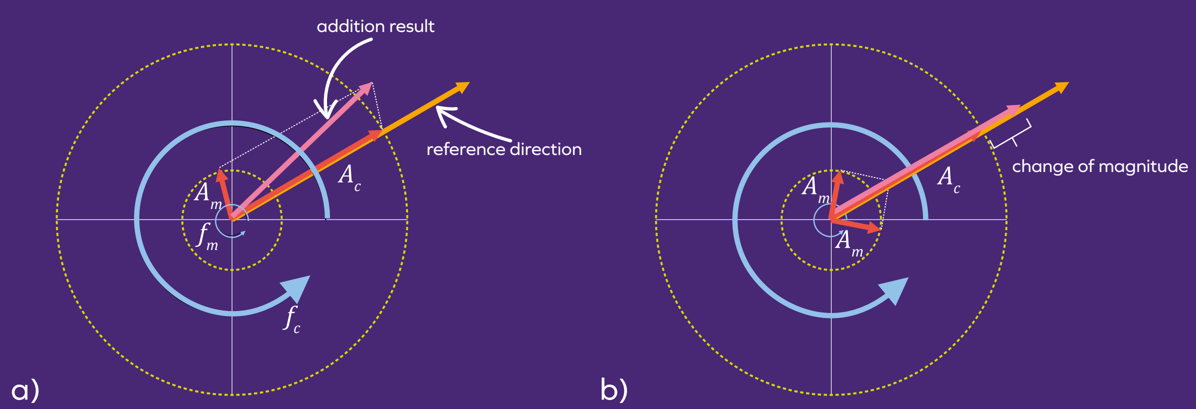


Fig. 7. Modification of the carrier signal's vector magnitude: a) one vector causes a change in direction, b) two vectors ensure the correct effect.

86

It is easy to see that we cannot. If this additional vector with amplitude A_m rotates with a different frequency than f_c, we will always find a moment when it does not coincide with the reference direction and as a result the sum of both vectors will change the direction, not only the amplitude (let us remember the parallelogram rule when adding vectors!).

Can we get the desired result with two additional vectors? Yes – let us look at Fig. 7b. If we introduce a second additional vector, which will be a mirror image of the first vector with respect to the reference direction, the direction of their sum will always coincide with the reference direction and the sum of all the vectors from the diagram in Fig. 7b will be a vector with the correct direction and changing amplitude.

At what frequency should the vectors corresponding to the modulating signal rotate? Let us simplify the analysis and move on to a reference frame in which the carrier signal vector is at rest – this is a frame rotating at frequency f_c (we use the principle of relativity of motion). For convenience, let us assume that the reference direction in this frame coincides with the vertical direction – see Fig. 8.

To ensure the amplitude variation with the desired frequency f_m, the first additional vector should rotate with this frequency, and the second vector – as its mirror image – also with the frequency f_m, but in the opposite direction (which we will denote for convenience as $-f_m$). Fig. 8 shows several possible configurations of the modulating signal vectors – their sum actually always coincides with the reference direction and ensures a periodic change of only the amplitude of the carrier signal with the frequency f_m.

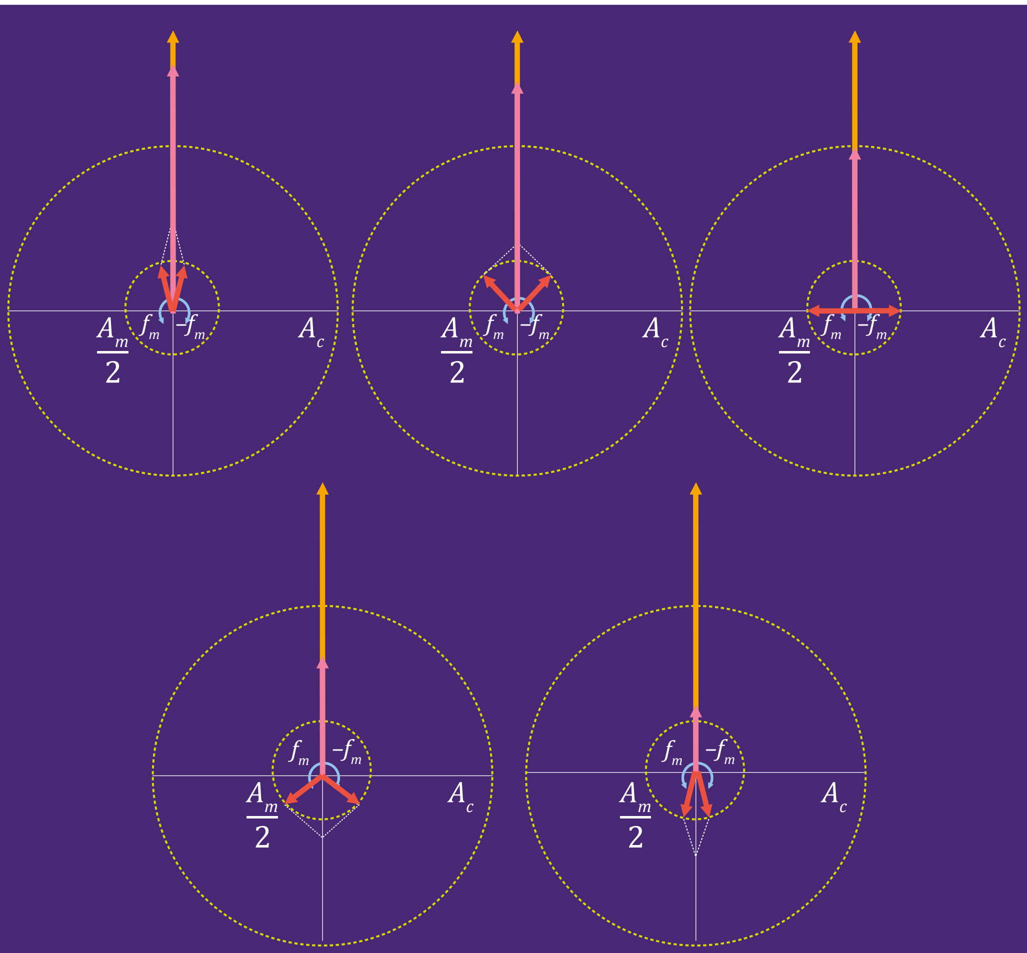


Fig. 8. Representation of AM modulation in the reference frame of the carrier signal vector with two vectors representing the modulating signal.

Note that when assigning a pair of vectors to a modulating signal, we must assign each of the vectors of this pair a length equal to half the amplitude of the signal $(A_m/2)$. Otherwise, when their directions and senses coincide, their sum will have a length of $2A_m$. We will take this into account from now on in the following diagrams.

Returning to the normal reference frame, we should give all vectors from Fig. 8 an additional rotational motion with frequency f_c. As a result, the reference direction together with the carrier signal vector will again start to rotate with frequency f_c, while the modulating signal vectors with frequencies $f_c + f_m$ and $f_c - f_m$ (see Fig. 9).

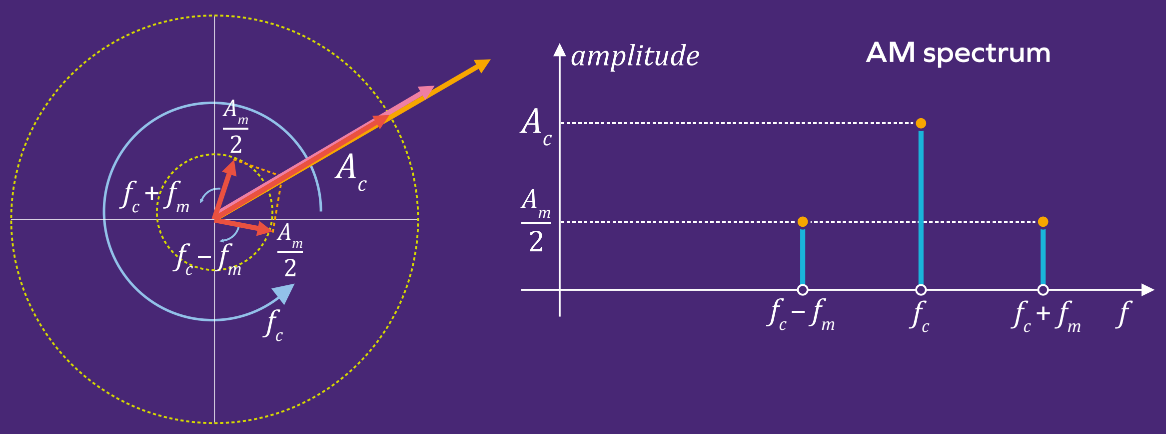


Fig. 9. Correct spectrum of a signal amplitude-modulated by a harmonic wave.

The correct spectrum therefore contains three lines – one with the carrier frequency f_c and two symmetrically located lines corresponding to the modulating signal with frequencies that are the sum and the difference of the carrier frequency and the frequency of the modulating signal and amplitudes equal to half the amplitude of the modulating signal.

4. Modulating signal with a given spectrum

One might think that the case of modulation by a harmonic wave discussed in the previous section is too simplified and not very practical. However, let us note that adding another harmonic to the modulating signal does not change anything in the whole reasoning, except that the circular diagram in Fig. 9 should be enriched with another pair of vectors, and the spectrum with two additional lines arranged symmetrically with respect to the carrier frequency.

In principle, we can assume that the modulating signal consists of a broad set of harmonics covering a certain frequency range, i.e., a so-called frequency band. This signal can represent, for example, a sample of a voice message, which, as we know from Lesson 4, presents a rather complicated image in the frequency domain.

Let us look at Fig. 10. In the left part of the spectrum, i.e. for low frequencies, the signal carrying information is visually illustrated as a series of bars covering a certain band. For simplicity, if we are not interested in individual bars, but only approximate form of the signal spectrum, we can represent it as the light blue shaded area. What happens when we modulate this signal with a harmonic carrier wave of frequency f_c?

Since each harmonic component of the signal with frequency f will be transferred to a point on the frequency axis with value $f_c + f$, the entire spectrum shifts just beyond the carrier frequency (note also the halving of the amplitude of each component). Each bar transferred in this way has its mirror image on the other side of the carrier frequency, so the modulated signal consists of two, symmetrically arranged bands called the **lower sideband** and the **upper sideband**. respectively. The bandwidth of the modulated signal is therefore twice as wide as the modulating signal.

Let us note, however, that both sidebands, as mutual mirror images, carry exactly the same information.

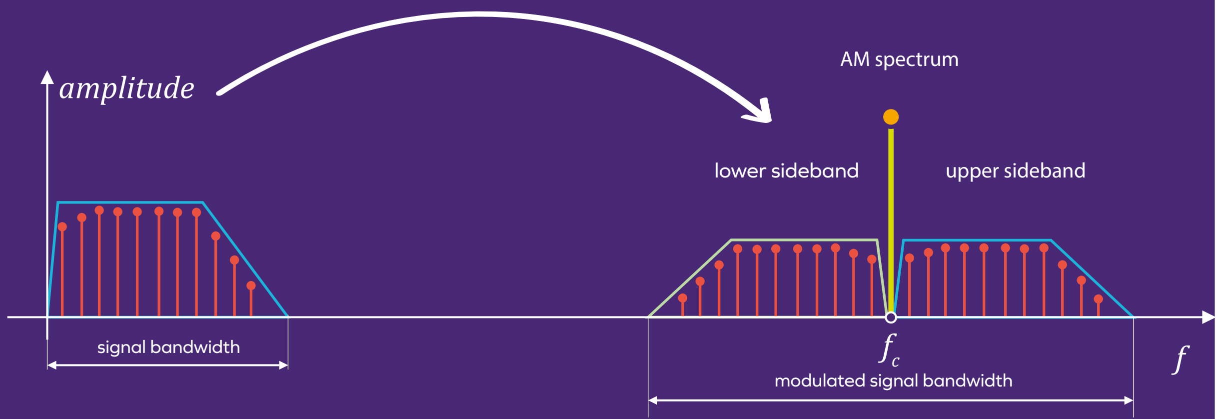


Fig. 10. Signal of a given spectrum after AM modulation.

Experiment

Prepare a large container with water—it can be shallow, but it is important that its walls are widely spaced apart. In a home setting, this could be a bathtub with a small amount of water. The container should be illuminated from above so that the shadow of the waves on the water surface is clearly visible on the bottom.

Place a Styrofoam ball or a ping-pong ball on the water surface, away from the walls (Fig. 11). At some distance from the ball, touch the water surface. At the bottom, you should see the shadow of a circular surface wave that reaches the ball and lifts it slightly.

1. Start touching the water surface rhythmically, trying to submerge your finger to a similar, small depth each time. Observe the wave that spreads. What happens to the ball?

2. Continue to rhythmically touch the water surface but occasionally change the depth of finger immersion. For example, for a few seconds, barely touch the surface, and then for the next few seconds, maintain a slightly greater depth of immersion, all while keeping the same hand movement tempo. Then return to the initial state with slight finger immersion. Are we dealing with modulation? If so, what type? How does the ball react to the changing wave?
3. Separate the area where you touch the water surface from the area where the ball is located using an opaque barrier placed above the water. Your friend on the other side of the container should observe the ball, and the barrier should prevent them from seeing your hand. Is your friend able to determine the depth to which you immerse your finger to generate waves solely by observing the ball's movements? Can information be transmitted in this way?

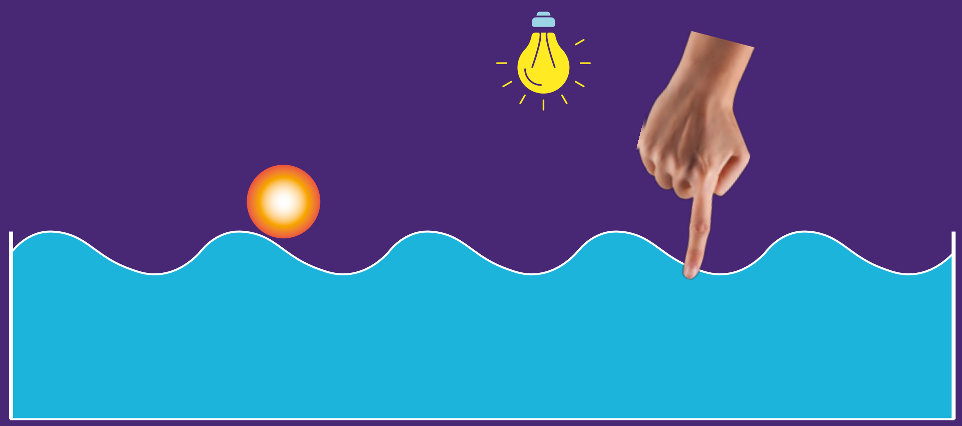


Fig. 11. Experiment with AM modulation of waves on the water surface.

Glossary

Modulation – the variation of wave parameters over time. It enables the transmission of information.

Carrier wave – a harmonic wave whose parameters we treat as reference parameters. It does not transmit information itself but can be modulated.

Amplitude modulation – modulation involving changes to the amplitude of the carrier wave over time.

AM – abbreviation for amplitude modulation.

Digital Amplitude Modulation – amplitude modulation, in which the number of possible amplitude values is finite.

Amplitude keying – another term for digital amplitude modulation.

Analogue amplitude modulation – modulation involving continuous change of the amplitude of the carrier wave over time by an analogue modulating signal.

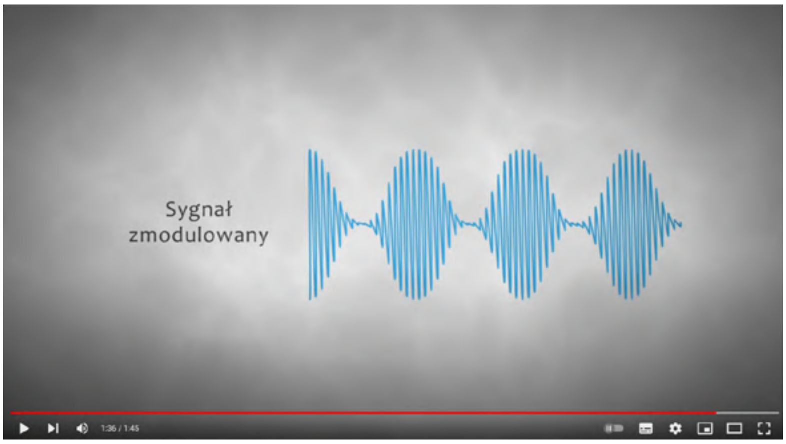
Lower sideband – the part of the modulated signal's spectrum below the carrier frequency.

Upper sideband – the part of the modulated signal's spectrum above the carrier frequency.

Bandwidth – the difference between the highest and lowest frequencies present in a given signal.

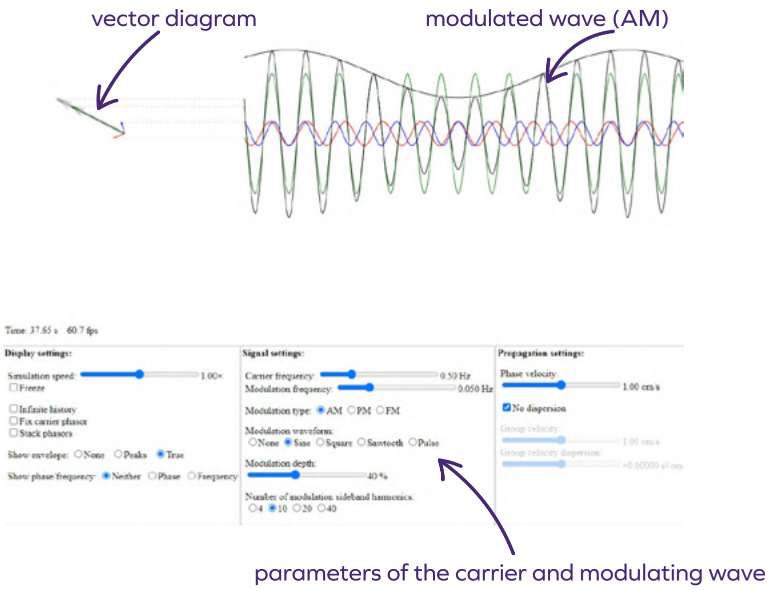
External materials

1. Electromagnetic wave modulation



Scan QR code

2. Online animation showing real-time generation of a modulated signal based on a rotating vector model.



Scan QR code

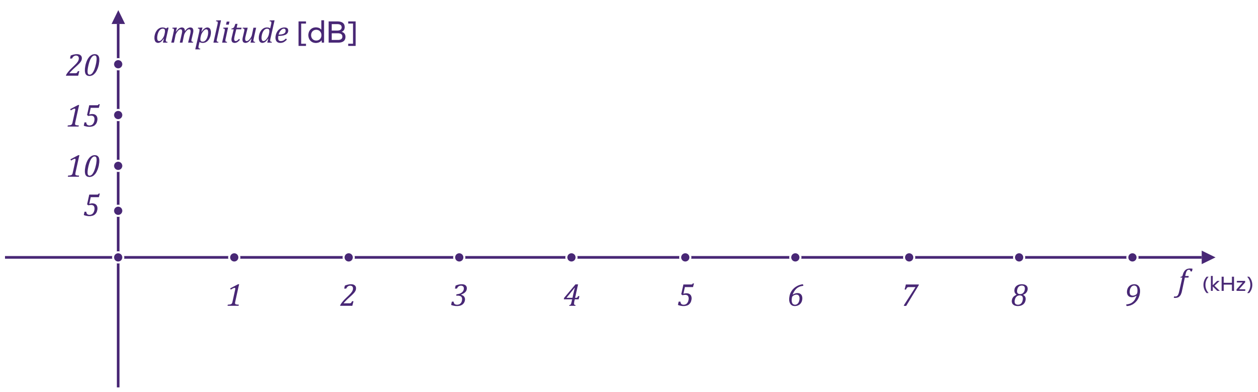
It is possible to independently set the carrier frequency, modulation frequency, modulation signal amplitude (via a parameter describing the so-called *modulation depth*). The simulation can be slowed down or sped up using the slider in the left window (*simulation speed*).

The generated waveforms represent, respectively:

- black lines – modulated signal and modulating signal (envelope);
- green line – carrier wave;
- red and blue lines – positions of the endpoints of the vector pair corresponding to the modulating signal (recall Fig. 9).

Homework

1. Draw an example modulation that would allow you to transmit the result of rolling a die three times – 3, 1, 2. How many amplitude levels should you use?
2. A signal containing a sound sample, consisting of two harmonic components with amplitudes equal to $10\,\mathrm{dB}$ and frequencies equal to $f_1 = 1\,\mathrm{kHz}$ and $f_2 = 2\,\mathrm{kHz}$, is used for AM modulation of a sound wave with amplitude $20\,\mathrm{dB}$ and frequency $f_c = 6\,\mathrm{kHz}$. Plot the spectrum of the sample and the spectrum of the AM modulated signal on one graph. Label the upper and lower sidebands.



Lesson 8

Phase and frequency modulation

Objective

- Presentation of the basic concept of phase modulation (PM) and frequency modulation (FM) of a carrier wave.

Learning outcomes

- The student knows the basic types of phase and frequency modulation.
- The student is able to explain the structure of the spectrum of a wave phase-modulated by a harmonic signal.
- The student is able to calculate the bandwidth of a frequency-modulated signal.

1. Phase and frequency modulation concept

In Lesson 7, we introduced amplitude modulation as a method for transmitting an information-carrying signal by varying the amplitude of the carrier wave. Is this the only way to modify a harmonic wave? Of course not. The other parameters of a harmonic wave, namely frequency and phase, can also change based on the information content of the signal.

Let us recall the circular diagram representing harmonic motion as an image of uniform motion of a point around a circle or rotation of a vector (Fig. 1). Amplitude modulation (AM) can be expressed on such a diagram by changing the length of the vector. Since the phase can be identified with the angle of rotation α, the change in phase is related to the modification of this angle with respect to the position resulting from uniform motion. We will call such a change **phase modulation**. The abbreviated term – **PM** – is also used.

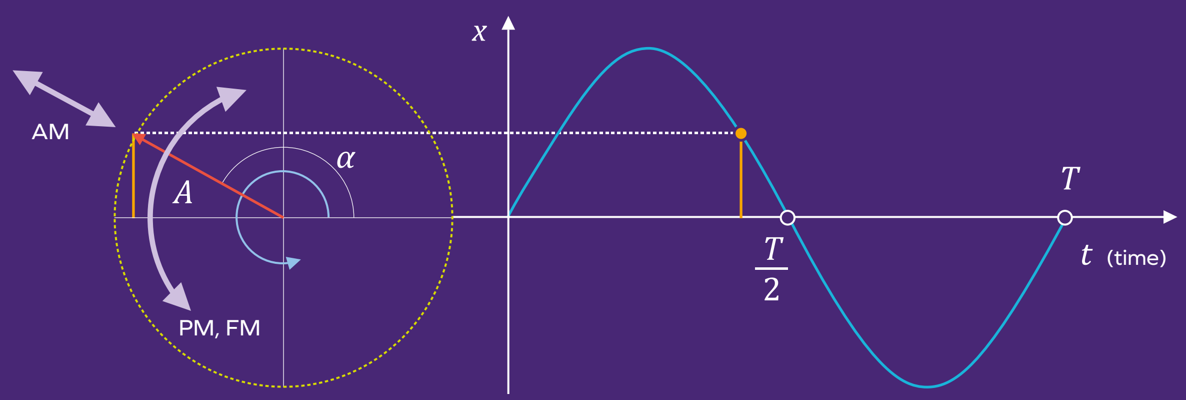


Fig. 1. Modulation as a variation of any of the harmonic wave parameters.

The frequency, equal to the reciprocal of the period T of rotation, can be understood as the rate of change of phase. A change in frequency will thus correspond to a change in the rotation speed of the vector in the circular diagram relative to the rotation speed assigned to the carrier wave. We will refer to this frequency modification as **frequency modulation** and also denote it as – **FM**.

In both types of modulation – PM and FM – the amplitude of the modulated signal is kept at the same level.

2. Analogue and digital modulation

All the modulation techniques introduced so far are shown in Fig. 2 for the case where the modulating signal has a finite number of levels (in this example specifically – 2). We then refer to it as **digital modulation** or **keying**. We have already discussed a special case of digital modulation for AM modulation in Lesson 7. Now we can see how the same signal controls the parameters of the carrier wave to obtain the individual types of modulation – AM, PM and FM.

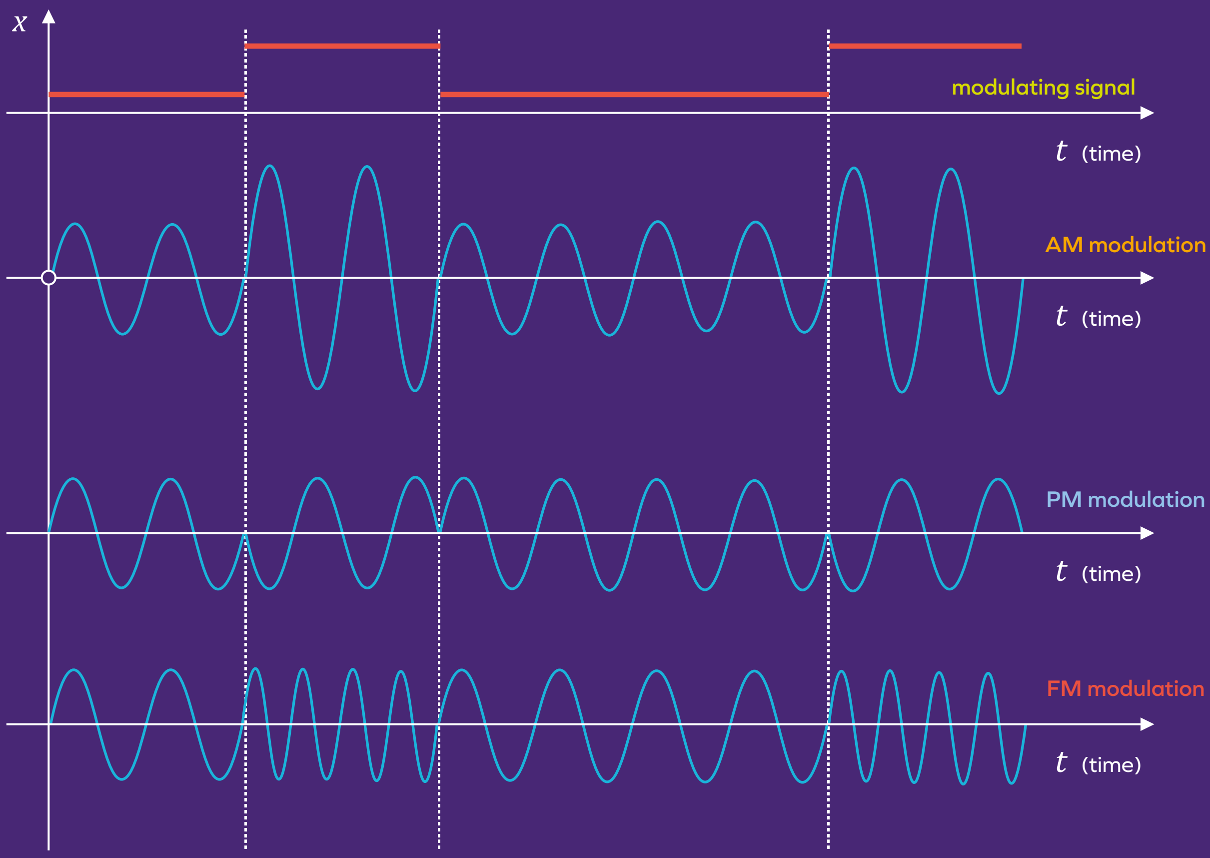


Fig. 2. Various techniques of digital modulation (keying).

As we can see, the amplitude modulated signal has two different amplitude levels. The phase modulated signal changes its phase by a certain fixed angle depending on the signal level. Here it is assumed that for the lower level of the modulating signal the phase is retracted by $180°$, while for the upper level it is advanced by $180°$ – this is equivalent to changing the phase to the opposite, i.e. the decreasing signal starts to increase, and the increasing signal starts to decrease (note the sudden change in the signal behaviour at the moments of changing the modulating signal level). In other words, the direction of the vector in the circular diagram changes to the opposite.

Frequency keying is characterised by assigning different signal frequencies to different levels of the modulating signal. Note that phase modulation does not alter the form of the carrier signal (it retains the same amplitude and frequency) but only shifts it in time at the moments when the level of the modulating signal changes. The effect of frequency modulation, however, is visible at any moment in time, even during long intervals where the modulating signal remains constant.

Note that in PM and FM modulations the signal amplitude does not change.

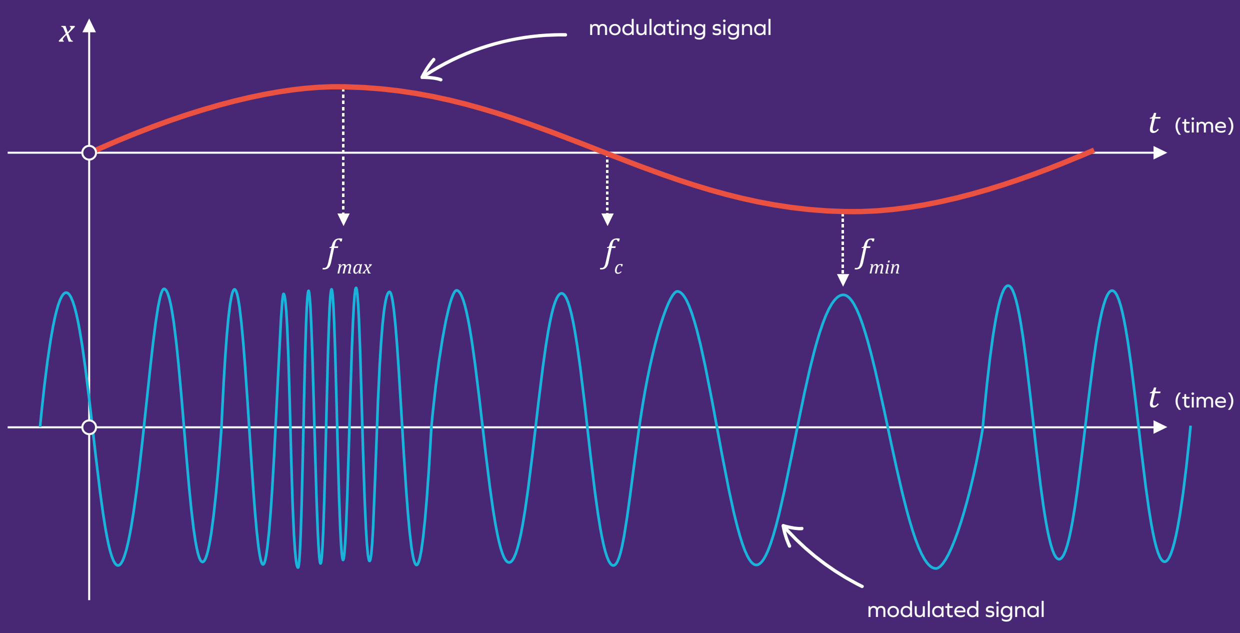


Fig. 3. Example of analogue FM modulation.

Analogue modulation, or modification of the carrier wave by a continuously changing signal, is shown in Fig. 3. The maximum value of the modulating signal is assigned a certain maximum frequency f_{max}, the minimum – the minimum frequency f_{min}, and at zero (neutral) signal value the frequency of the modulated signal is equal to the carrier frequencyj f_c. Between these values the frequency of the modulated signal changes smoothly.

The difference between the maximum frequency value of the modulated signal and the carrier frequency is called **frequency deviation** and is denoted by Δf:

$$\Delta f = f_{max} - f_c$$

Frequency deviation tells us how much the frequencies appearing in the modulated signal deviate from the carrier frequency. They are in the range of $f_c - \Delta f$ do $f_c + \Delta f$.

A particularly important reference value (we will see this when analysing the spectrum of the modulated signal) is the frequency of the modulating signal f_m. If the frequency deviation is smaller than the frequency of the modulating signal, i.e. $\Delta f < f_m$, then we are dealing with **narrowband FM modulation**. Otherwise – with **wideband FM modulation**.

The choice of frequency deviation is largely free, independent of the signal's carrier frequency, and dictated primarily by considerations of the transmission's sensitivity to interference. We will discuss this in more detail in Lesson 9.

3. PM modulation on vector diagrams

Can PM and FM modulation with a harmonic signal be visualised in vector diagrams as we did in Lesson 7 for AM modulation? Unfortunately, such an analysis of PM and FM is much more complicated. It can be carried out in this way only for PM modulation and then in the special case when the phase modifications are relatively small.

Let us see why this is the case. Recall Fig. 7b in Lesson 7. There, we introduced a pair of additional vectors representing the modulating signal, which were used to modify only the length of the vector corresponding to the modulated signal, thereby achieving amplitude modulation. If we wanted to do something similar for phase modulation, the length of the resultant vector should remain constant, while its direction should change. This means that the sum of the introduced vector pair should be perpendicular to the reference direction – see Fig. 4 (on the left; note that for greater clarity, we attached the pair of vectors to the end of the carrier signal vector, not at the centre of the circle).

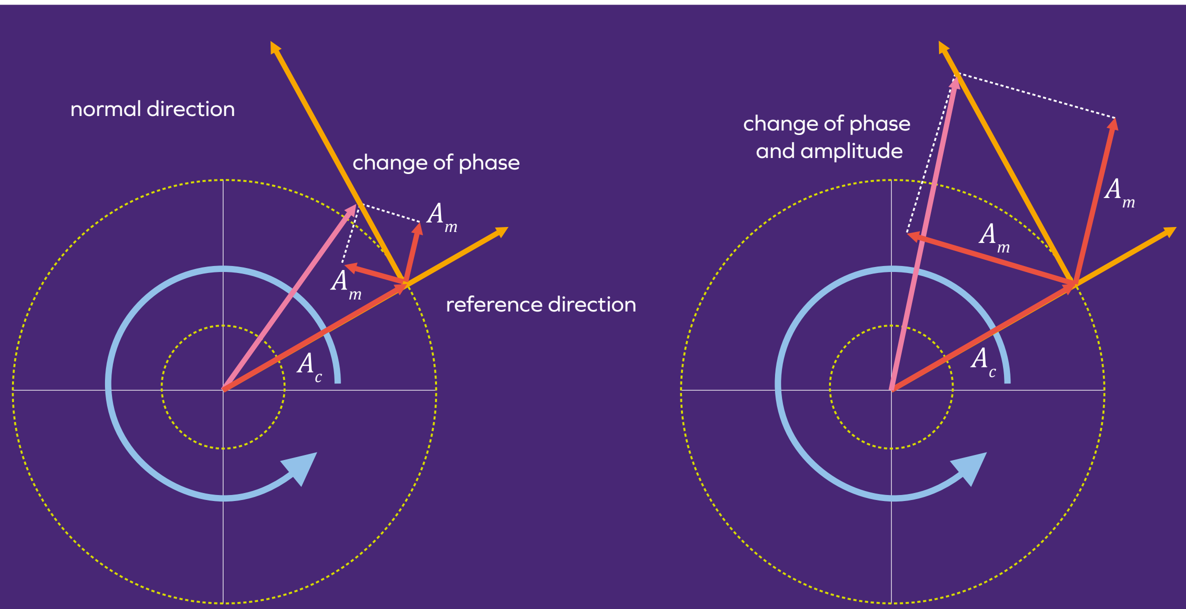


Fig. 4. Vector diagrams representing phase modulation for small (on the left) and large modulation depth (on the right).

Do we achieve a change in the direction of the resultant vector without changing the amplitude? Not quite. It is easy to see that the end of the resultant vector cannot lie on the circle of the carrier signal – it goes beyond it, and all the more so, the greater the amplitude of the modulating signal, i.e. the greater the change in phase (in other words – the greater the so-called modulation depth). If the change in phase is small, we can assume that our goal has been approximately fulfilled and only a change in phase occurs. Otherwise, as Fig. 4 on the right shows, the change in amplitude is significant and the PM modulation model we introduced cannot be considered correct.

Let us therefore assume that the depth of phase modulation is small and let us look at several phase modulation stages in a graph analogous to Fig. 8 in Lesson 7. Just as in that case we obtained a cyclic change of amplitude with the desired frequency f_m, there we obtain a phase change with negligible change of amplitude (Fig. 5).

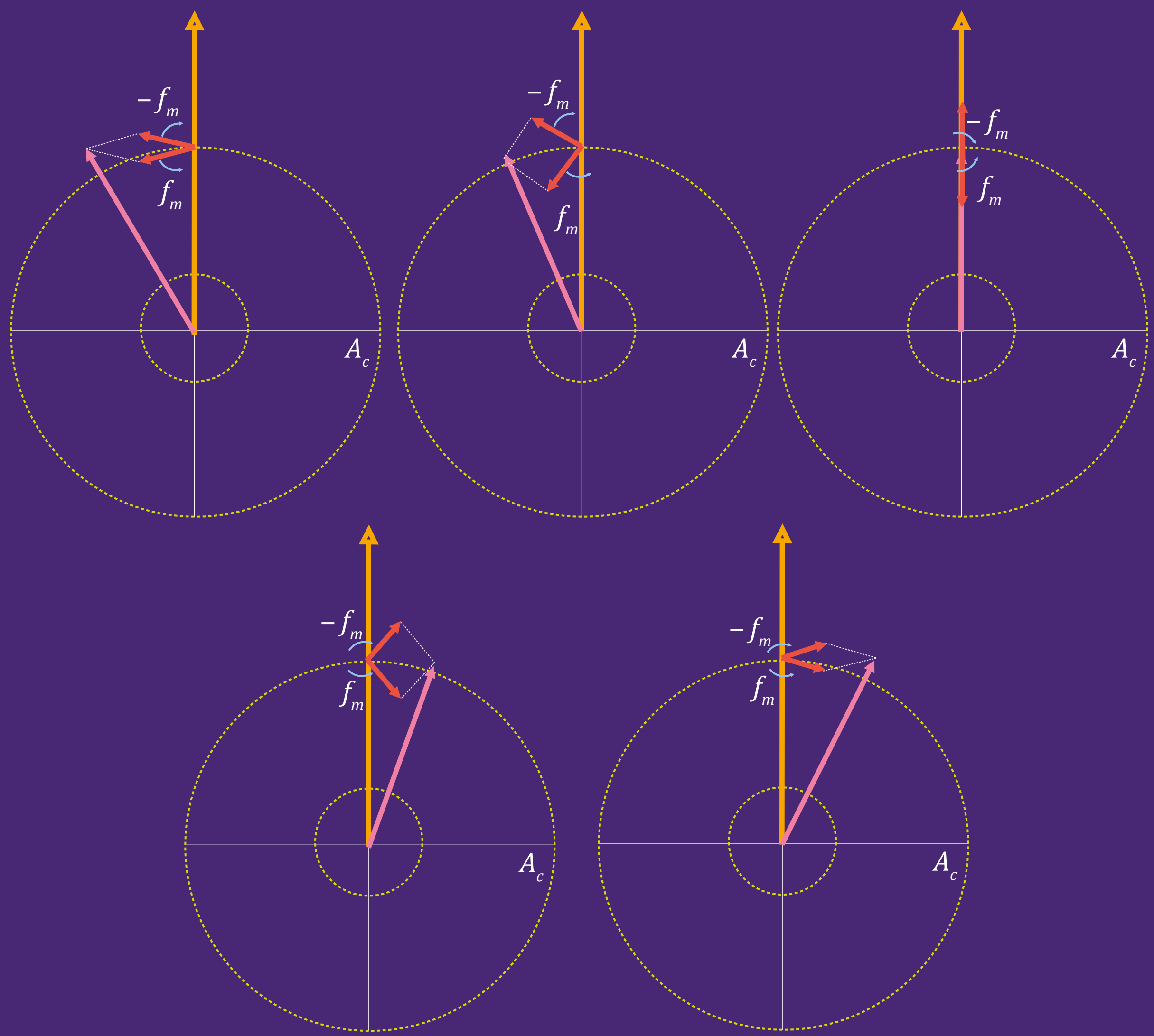


Fig. 5. Vector diagrams presenting a few selected stages of phase modulation.

98

What does a small change in amplitude in vector diagrams really mean? That one pair of vectors representing the modulating signal is not enough to obtain only a change in phase. In fact, you need an infinite number of these pairs, but the amplitudes assigned to them decrease very quickly. Therefore, the PM spectrum is not limited to only two additional, symmetrically placed lines, as in the case of AM modulation (see Fig. 9 in Lesson 7). However, with not too deep modulation, these spectra are almost identical.

4. FM signal spectrum

Let us assume that the modulated signal contains two frequencies f_1 and f_2 spaced Δf from the carrier frequency f_c. The form of the spectrum of such a signal depends on whether the modulation is narrowband or wideband (Fig. 6).

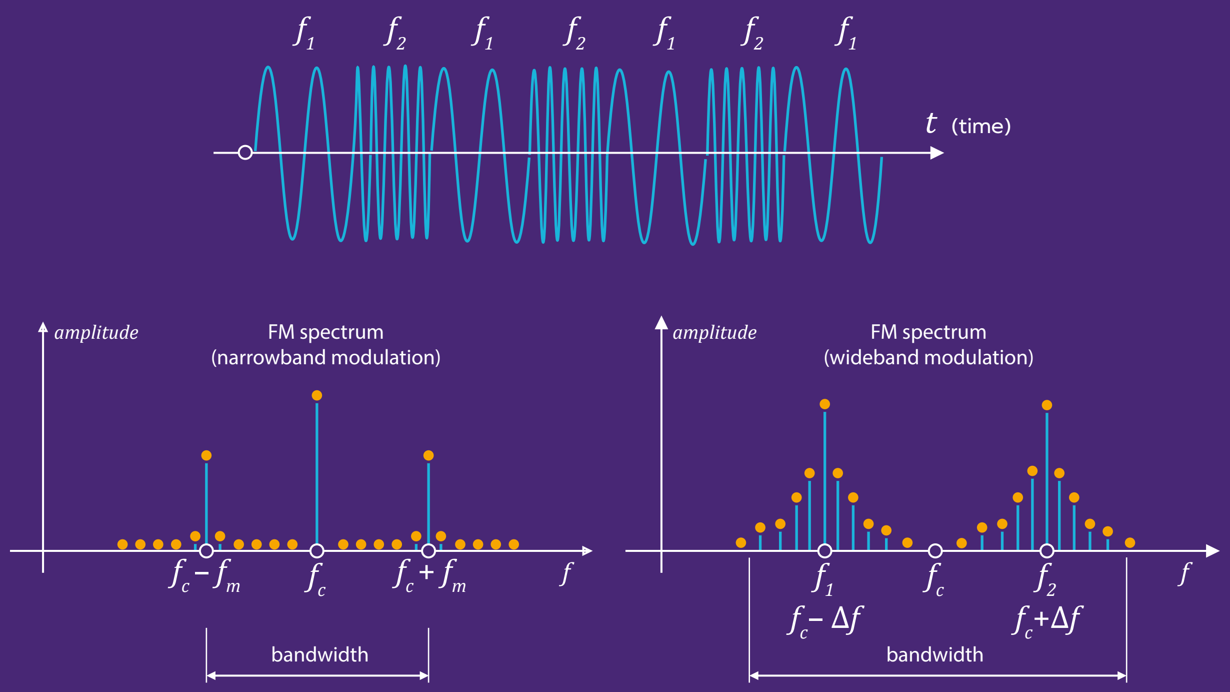


Fig. 6. Frequency modulated signal and its spectrum in the case of narrowband and wideband modulation.

In the first case, the frequency deviation is much smaller than the modulating signal frequency and the signal variability in time is mainly determined by the rate of change of the modulating signal. Therefore, in the spectrum we will see two strong, symmetrically located lines with frequencies $f_c - f_m$ and $f_c + f_m$, similarly to AM modulation (Fig. 6, left).

When the frequency deviation is large, i.e. the f_1 and f_2 frequencies clearly differ from the frequency of the modulating signal f_m, they are the ones that determine the content of the modulated signal. Many other bars appear because many harmonic signals of different frequencies are needed to obtain the waveform shown in Fig. 6 after their summation. Although the width of the FM spectrum is in principle infinite, the amplitude of the bars outside a certain range is so small that in practice they can be ignored.

Remark. If this is not intuitive, because at first glance the waveform of Fig. 6 contains only two frequencies, recall the spectrum of the square wave signal from Lesson 4. This signal is constant for half of the period, and also for the other half (although at a different value), and yet it contains an infinite number of harmonics with arbitrarily high frequencies. The harmonic signals that make up the spectrum are very simple waveforms – they have the same amplitude and frequency at every moment in time. Using them to generate a square wave or the waveform from Fig. 6, which differ significantly in character from the monotonic behaviour of harmonic waves, requires the use of a vast number of component signals.

In Lesson 7, we showed that AM modulation shifts the signal spectrum to the carrier frequency, and the signal bandwidth is doubled by creating two sidebands – lower and upper. We see a similar phenomenon in FM modulation (Fig. 7).

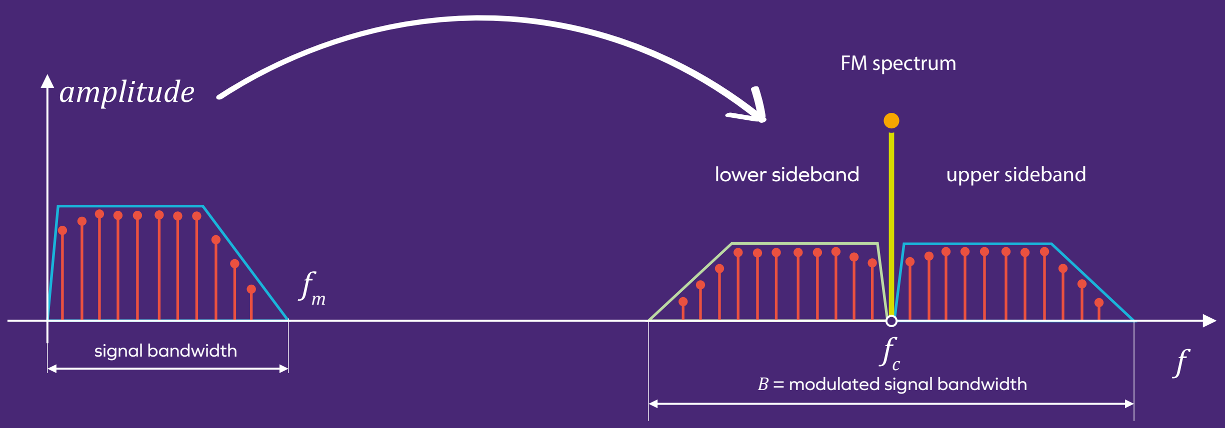


Fig. 7. Shift of the signal spectrum after FM modulation.

Despite the much more complicated situation with FM modulation, it turns out that it is quite easy to calculate the approximate bandwidth B of the FM modulated signal using the so-called Carson's rule:

$$B = 2\left(f_m + \Delta f \right)$$

where f_m is the maximum frequency in the signal carrying information, and Δf is the maximum frequency deviation. We can see that for small values deviation Δf we obtain approximately a doubling of the signal bandwidth, just like in AM modulation.

Experiment

We will conduct an experiment demonstrating FM modulation in a similar manner to that in Lesson 7.

Prepare a large container with water – it can be shallow, but it is important that its walls are widely spaced apart. In a home setting, this could be a bathtub with a small amount of water. The container should be illuminated from above so that the shadow of the waves on the water surface is clearly visible on the bottom.

Place a Styrofoam ball or a ping-pong ball on the water surface, away from the walls (Fig. 8). At some distance from the ball, touch the water surface. At the bottom, you should see the shadow of a circular surface wave that reaches the ball and lifts it slightly.

1. Start touching the water surface rhythmically, trying to submerge your finger to a similar, small depth each time. Observe the wave that spreads. What happens to the ball?
2. Continue to rhythmically touch the water surface, but occasionally change the speed of your hand movement, while maintaining the same depth of immersion of your finger. Is this modulation? What kind? How does the ball react to the changing wave?
3. Separate the place where you touch the water surface from the place where the ball is with an opaque partition placed above the water. Have your friend on the other side of the tank observe the ball, and the partition should prevent him from observing your hand. Can your friend tell the rate at which you are making waves just by observing the ball's movements? Can any information be transmitted in this way?

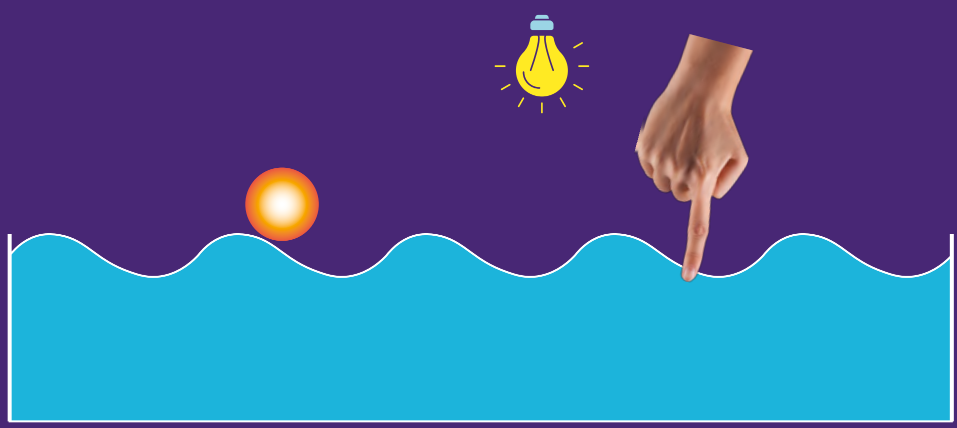


Fig. 8. Experiment with FM modulation of waves on the water surface.

Glossary

Frequency deviation – the difference between the maximum frequency value of the modulated signal and the carrier frequency.

FM – abbreviation for frequency modulation.

Frequency modulation – modulation involving changes to the frequency of the carrier wave over time.

Phase modulation – modulation involving changes to the phase of the carrier wave over time.

Broadband modulation – FM modulation in which the frequency deviation is greater than the frequency of the modulating signal.

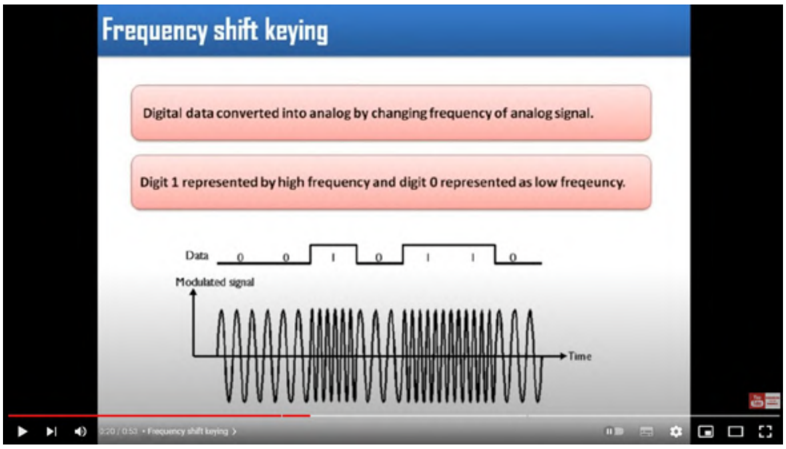
Narrowband modulation – FM modulation in which the frequency deviation is less than the frequency of the modulating signal.

PM – abbreviation for phase modulation.

Carson's Rule – a formula that allows you to calculate the bandwidth of an FM modulated signal based on the maximum frequency of the modulating signal and the maximum frequency deviation.

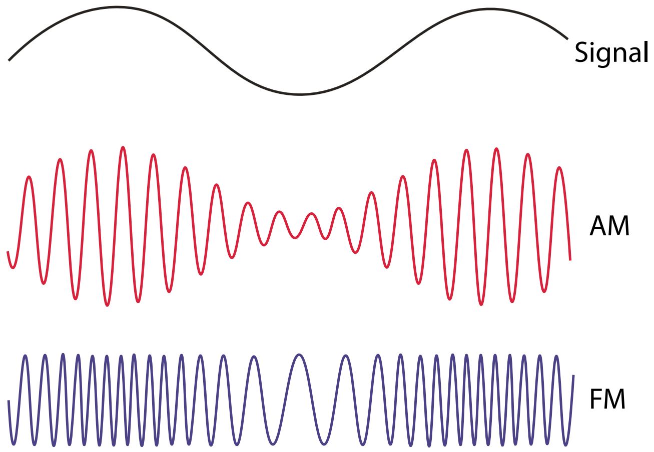
External materials

1. Animation illustrating digital modulation AM, FM and PM (*Animation of Digital modulation - Amplitude, Frequency and Phase shift keying*).



Scan QR code

2. Animated image of AM, FM modulation by harmonic signal.



Scan QR code

Homework

1. Draw an example of FM modulation that would allow the transmission of the result of rolling a die three times – 3, 1, 2. How many different frequencies should be used?

2. The signal containing a sample of a piece of music falls within the band with a maximum frequency of $f_m = 15\,\mathrm{kHz}$. This signal has been FM modulated onto a carrier wave of $100\,\mathrm{MHz}$ so that the maximum frequency deviation is $75\,\mathrm{kHz}$. What type of modulation is this – narrowband or wideband? What is the approximate bandwidth of the modulated signal?

   Given:

   $f_m = 15\,\mathrm{kHz}$

   $\Delta f = 75\ \mathrm{kHz}$

   $f_c = 100\ \mathrm{MHz}$

   To find:

   B = ?

Lesson 9

Applications of signal modulation

Objective

- Presentation of signal modulation applications and methods of shaping the modulating signal.

Learning outcomes

- The student knows the role of modulation techniques in frequency division among transmitters.
- The student is able to explain the difference in sensitivity to interference of signals modulated by AM and FM techniques.
- The student knows the relationship between signal quality and information transfer speed and signal bandwidth.
- The student knows the method of processing an analogue signal into a digital one and is able to use it in practice.

1. The role of signal modulation in frequency division

When we find ourselves in a room where many people are talking at once, it is difficult to communicate. With great effort, we can hear the person standing closest to us, but imagine hundreds of people in one room and each of them can be heard equally loudly. The problem is that everyone uses the same signal bandwidth (a few kHz wide) and the same medium – air. It would be similar if we wanted to directly transform the voice signal of the speakers in a certain area into an electromagnetic wave and send it into space in the hope that the recipient will find the signal intended for him in this chaos.

In Lesson 5, we showed that the solution to the problem could be the so-called frequency division, i.e. the transmitters can agree to transmit their signals in different frequency bands, and the receivers, using filtering, isolate the signal they are interested in. Now, using the knowledge acquired in Lessons 7 and 8, we can discuss a practical method for implementing this idea.

Let us assume that two transmitters – A and B – would like to transmit certain signals into space, whose bandwidths are limited by a certain maximum frequency, designated as $f_{\mathrm{max,\,A}}$ and $f_{\mathrm{max,\,B}}$, respectively. For example, if these are voice messages, it can be assumed that limiting their bandwidth from above by a frequency of $5\,\mathrm{kHz}$ should not significantly disturb their quality. Fig. 1 on the left shows schematically the spectra of these signals. If we plotted them on one graph, they would significantly overlap.

The senders agree to modulate their signals (e.g. with AM), using different carrier frequencies $f_{\mathrm{c,A}}$ oraz $f_{\mathrm{c,B}}$, respectively. According to what we learned in the previous lessons, after modulation the bandwidth of each signal will increase by creating sidebands. Therefore, special care must be taken to ensure that the separation between the carrier frequencies is sufficient to ensure the separation between the sidebands of the modulated signals of both transmitters.

When both modulated signals are now transmitted simultaneously, their spectra will not overlap. The receiver, using an appropriate band-pass filter, can separate a specific signal and, after **demodulation**, transfer it to the original audible frequency band.

There is another important benefit of modulation techniques. Recall that in Lesson 3 we discussed the minimum size of a dipole antenna for transmitting a signal at a given wavelength. When the signal is modulated from a high-frequency carrier wave, it moves on the spectral plot to a region of much higher frequencies. This shortens its wavelength and thus reduces the size of the antenna required. We will take a closer look at the actual gain of modulation in the Homework.

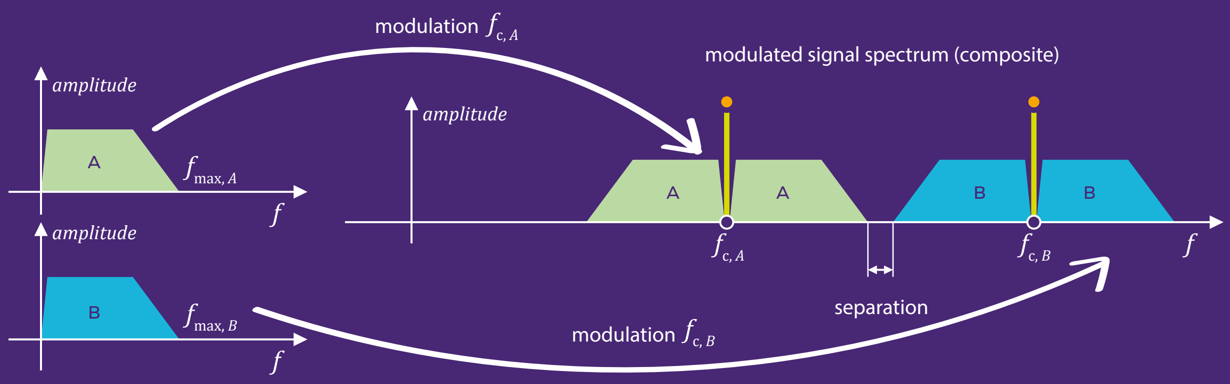


Fig. 1. Frequency division between two transmitters: A and B.

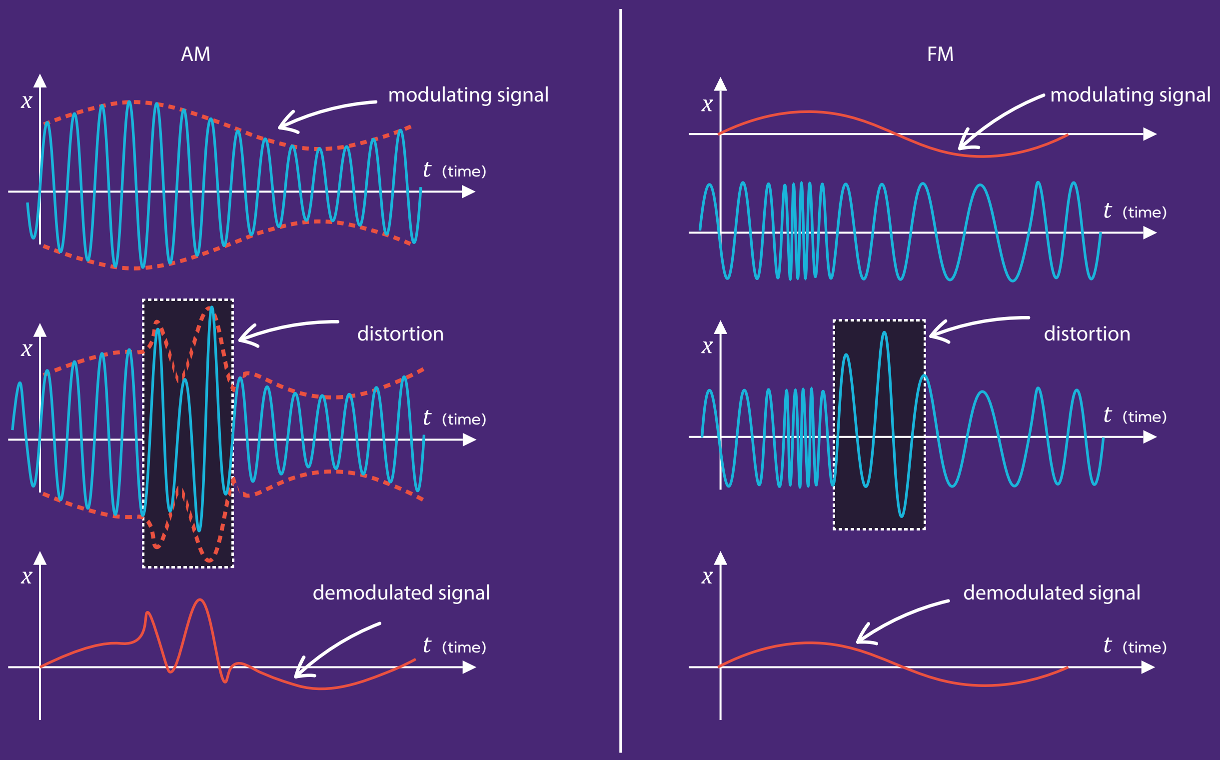


Fig. 2. Amplitude disturbances of the modulated signal: AM (on the left) and FM (on the right).

2. Modulated signal and interference

In Lesson 7 we introduced the AM modulation technique, which we discussed in some detail because it was easy to explain the principle of its operation in vector diagrams. The FM technique described in Lesson 8 was much more complicated. It also turned out that the bandwidth of an FM signal is effectively infinite and full of sidebands, even when modulated by a harmonic signal (especially for broadband modulation). We can therefore ask – why use FM modulation at all, and what are its advantages over AM modulation?

One of the greatest advantages of FM modulation is its low sensitivity to interference (noise). Most of the interference that we have to deal with when transmitting signals on electromagnetic waves (EM) are so-called amplitude interferences, i.e. those that mainly affect the amplitude of the signal. Such interferences include, for example, lightning discharges.

Let us look at how signals modulated using AM and FM techniques react to interference (Fig. 2). On the left, we see a carrier wave modulated with a slowly varying harmonic signal, which is clearly modified by amplitude interference over a certain period of time. This distorted wave reaches the receiver, which is unable to distinguish between amplitude variations related to information content and random interference that occurred somewhere between the transmitter and the receiver. The demodulated signal, i.e. recovered from the carrier wave and restored to normal form, still contains this distortion and may manifest itself as an unpleasant crackling sound in the loudspeaker.

And what about FM modulation? Let us look at the right side of Fig. 2. The frequency modulated signal also suffers from amplitude distortion. The difference is that, that this time the signal amplitude does not carry any significant information. Despite the clear interference, if the **frequency** of the modulated signal does not change (i.e. the intersection points with the time axis remain in the same places), the useful signal can be demodulated in the receiver in full compliance with the original signal.

Of course, this is an idealised case. In practice, some irremovable interference may occur, but the sensitivity of the FM modulated signal to it is much lower than in the AM technique. Moreover, the sensitivity to interference is lower, the greater the frequency deviation, which speaks in favour of using wideband modulation when we care about the quality of the transmitted signal (e.g. it is a music broadcast).

Fun fact. You may have heard the abbreviations AM and FM, which we introduced in the last lessons. In particular, the abbreviation FM can be found in the names of many commercial radio stations. It is typically used by stations broadcasting on radio waves using FM modulation – they are characterised by low interference and excellent quality of music broadcasts (note: the FM element is also popular in internet radio – in this case it has nothing to do with FM modulation; we are dealing here with a reference to the tradition of classical radio). In long-wave radio broadcasts using AM modulation, noise and crackling are more common, so they focus on news, reports, interviews, and other segments where transmission quality is less critical.

3. What does the signal bandwidth depend on?

Now let us talk a little more about the bandwidth of the original signal that we want to send to the recipient. It can be an analogue signal, e.g. a voice message or a piece of music recorded by a microphone. The human hearing range is from about $20\,\mathrm{Hz}$ to about $20\,\mathrm{kHz}$, which does not mean that sending the entire frequency band is always necessary. If we are sending only human speech and we want the message to be understandable, and not necessarily to fully convey the voice of the narrator, a bandwidth from about $300\,\mathrm{Hz}$ to about $4\,\mathrm{kHz}$ should be sufficient. Further narrowing of the frequency band leads to an increasing deterioration of the sound quality, until the message is completely incomprehensible.

The transmission of music, especially sophisticated music with a wide range of instruments, is much more demanding. Music that is primarily perceived as an aesthetic experience requires high quality and therefore a much wider bandwidth of up to $15\,\mathrm{kHz}$.

Similarly, for many other analogue signals, we can always limit the bandwidth by using low-pass filtering (Lesson 5), and the signal will be better represented (the higher the quality) the higher the filter cut-off frequency we use.

In modern telecommunications, it is particularly important to transmit digital signals, including binary ones (with two levels – conventionally 0 and 1). How to determine the bandwidth of a digital signal? Let us look at the most critical case, when the levels change alternately with a certain period T_1. This case is the most important, because it leads to the highest possible frequency in the signal. The digital signal then has the form 101010101… etc. We can therefore interpret it as a square-wave signal (see Lesson 4).

Let us recall that the spectrum of such a signal is characterised by the presence of only odd harmonics, the first of which has a frequency $f_1 = 1/T_1$.

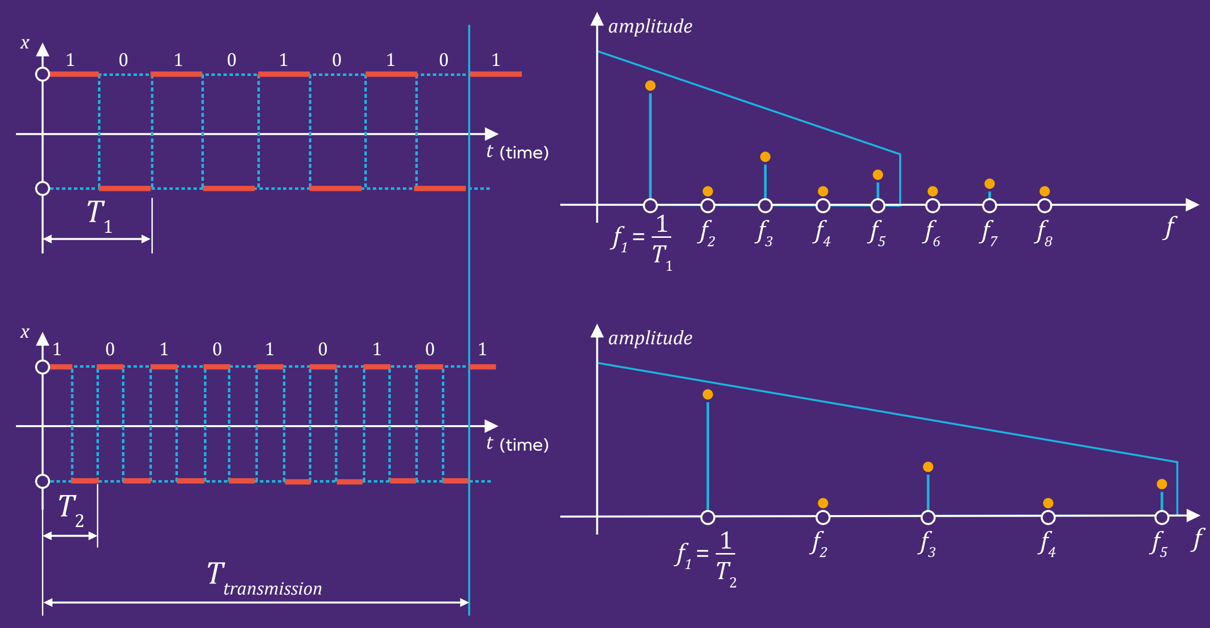


Fig. 3. Digital signal bandwidth.

To recreate the form of a digital signal, in practice we do not need all the harmonics. Let us assume that we limit ourselves to the first five. The digital signal band can then be presented as in Fig. 3 (upper part).

What does the period T_1? tell us? We can understand it as the time to transmit 2 bits of information (see Lesson 1), because each of the binary signal levels can be understood as 1 bit. At a given transmission time $T_{\text{transmission}}$, the smaller the value of T_1, the more bits we can transmit.

What happens when the demand for higher information transfer speeds increases? For example, users may expect the transfer time of photos to be reduced by at least two times or to be able to watch a movie on a streaming service in better quality (e.g. 4K). It is then necessary to reduce the T_1.

In the lower part of Fig. 3 we can see a digital signal with a shorter period T_2, which allows for sending twice as many bits in the same time. As we can see, due to the increase in harmonic frequencies, the signal spectrum is also significantly broadened.

Remember – the demands for higher quality analogue signals or higher data rates in digital signals always lead to an increase in the signal bandwidth. We will discuss the consequences of this phenomenon in more detail in Lesson 10.

4. Analogue to digital conversion

Finally, let us look at converting an analogue signal to a digital one – the so-called **analogue-to-digital conversion** (A/D for short).

In Lesson 1 we learned about the advantages of a digital signal, including its tolerance to minor interference. In modern telecommunications, dominated by digital techniques, it is important to convert any signal to digital form – most often binary (two levels). Is this even possible?

A/D conversion occurs in two stages:

- **sampling** – conversion of a signal continuous in time into a sequence of samples taken evenly every certain, fixed period of time (in other words – at a certain fixed frequency called the **sampling rate**);
- **quantisation** – a specific number of bits is assigned to the signal value in a given sample; at this stage, the signal is converted into a digital signal with a finite number of levels directly dependent on the number of bits.

Let us look at Fig. 4, where a continuous curve corresponds to the course of an analog signal over time. In the first step, we select a certain number of points on the curve. In this way, we perform sampling. The more points there are, the better we will be able to represent the shape of the curve, but also the more bit sequences will be assigned to the selected section of the signal (i.e. more information will have to be processed and then transmitted).

To perform quantisation, we need to determine how many bits we will assign to describe the signal value in the sample. Again, the more bits there are, the more precise we will describe the signal value. Unfortunately, this also has negative consequences, similar to the case of increasing the number of samples.

In the example in Fig. 4, we assumed that we would use 3 bits for quantisation. Knowing the range of the analogue signal (in our case from 0 to 7), we can now assign the values of these bits to individual samples.

In Lesson 1 we showed a clever method of guessing a number within a certain range by successively halving the variation intervals based on the information obtained.

Let us use this method to determine the bits in the first sample. We divide the entire range of variation in half. Is the signal value in the upper half (i.e. above 3.5)? As we can see, the signal value in the first sample is about 0.2, so the answer is "no" and we assign the value 0 to bit 1. Since the sample value is in the lower half of the variation, we divide this range in half this time. Is the sample value in the upper half of this new division (i.e. above 1.75)? No, so we assign the value 0 to bit 2 and again deal with the lower half, which we divide into two parts and similarly for bit 3 we obtain the value 0. The three bits of the first sample therefore have the values 000.

Using the diagram in Fig. 4, we can read the bit values almost immediately by drawing a horizontal line from a given sample to the left and reading the correct value based on the intersection of the line with the columns of individual bits. In this way, we establish three-bit sequences for all the samples in Fig. 4.

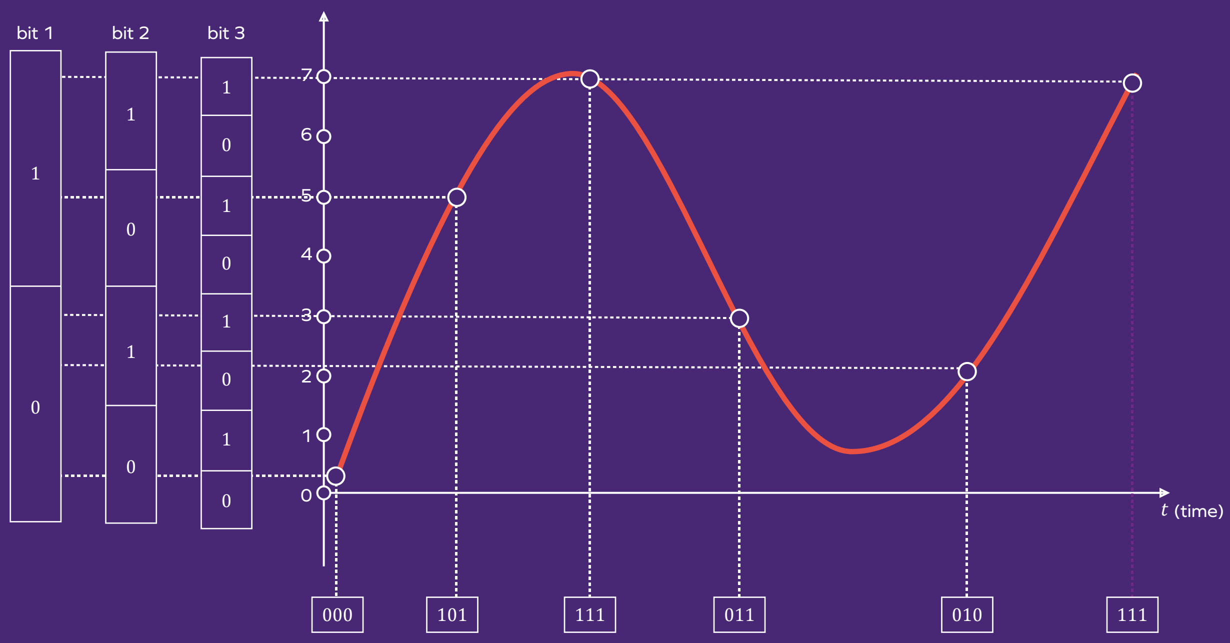


Fig. 4. Sampling and quantisation of analogue signal.

The digital signal that is the combination of all the three-bit sequences is shown in Fig. 5. Note that it does not resemble the curve in Fig. 4 at all.

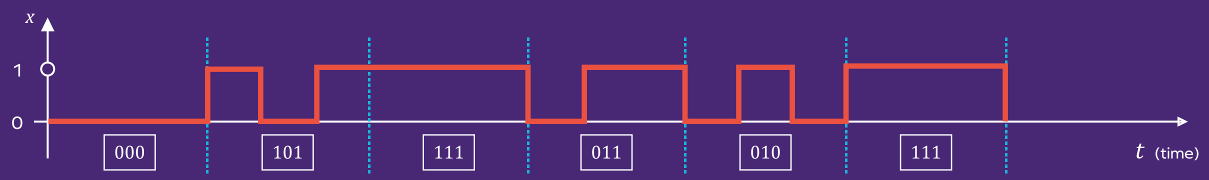


Fig. 5. Digital signal obtained after conversion of the signal from Fig. 4.

Can we convert in the opposite direction (digital to analogue, D/A)? Yes, it is enough to have information about the time interval between samples and, for the appropriate moments of time, assign the value of the analogue signal based on the bit sequence specified for a given moment (using the diagram in Fig. 4, but in the opposite direction than before). After reproducing the sequence of samples, we draw a curve as smooth as possible through them. The accuracy of the entire process depends on the sampling frequency and the number of bits per sample in the quantisation process.

Learn more. It turns out that if an analogue signal has a bandwidth limited from above by the frequency f_u, tthen it is enough to sample it at a frequency twice as high as f_u, to recreate it perfectly from a sequence of its samples. This is stated by the so-called sampling theorem. Unfortunately, quantisation always introduces some inaccuracy, which is why when the requirements for the quality of A/D conversion are high, a large number of bits per sample are used (even 24 bits in high-quality audio systems).

Experiment

Let us see how A/D and D/A conversion works in practice. To conduct the experiment, we will need three students. The first person – the "transmitter" – after printing the diagram from Fig. 6 draws on it any curve that could represent the analogue signal (it is necessary to ensure that for a given moment of time the signal has only one value). The curve should not be too "twisted" – it is best if its smoothness resembles the curve from Fig. 4.

Fig. 6. Diagram for the experiment with A/D and D/A conversion.

The second person – the "A/D converter" – receives a diagram with a curve from the "transmitter" but does not show it to the third person. Then he tries to assess how many samples are necessary to properly represent the signal. Is it enough to take samples every third square, or every two? Or maybe it will be necessary to sample the signal every one square? After making a decision, for each sample, the "A/D converter" converts its value to a sequence of three bits, proceeding as we showed in Stage 4 of this lesson. Then he writes down on a separate sheet of paper the sequences of bits for the subsequent samples, separated by commas (e.g. 110, 010, 000, etc.). After finishing his work, he gives the sheet of paper with the sequences (but not the diagram!) to the third person. He must also provide information about how many squares apart the samples were taken.

The third party – the "receiver" – will now try to recreate the original signal, acting as a D/A converter. With their own printed copy of the diagram from Fig. 6 and information about the intervals between samples, they try to recreate the value of the analogue signal in each sample based on the sequence of bits and the time of sampling. The result should be a series of points on the graph, which then need to be connected with a smooth curve. Only after this task is completed can the "transmitter" and "receiver" compare the original signal with the recreated signal.

Questions for discussion:

- Are the curves similar? How different are they?
- Could the differences be the result of too sparse sampling? Is it worth repeating the experiment with denser samples? Of course, if the experiment is repeated, the "receiver" should be another person who did not see the original signal.
- Can the differences be the result of too few bits? What can be the maximum difference in sampling moments between the original signal and the one recreated by the "receiver"?
- Can you guess how to modify the diagram in Fig. 4 to add "bit 4" for the quantisation stage? How much will the amount of work increase in A/D and D/A conversion?

Glossary

A/D – abbreviation for analogue-to-digital conversion.

D/A – abbreviation for digital-to-analogue conversion.

Sampling rate – the frequency at which samples of an analogue signal are taken; the inverse of the time interval between samples.

Demodulation – reconstruction of the original information-carrying (modulating) signal from a modulated carrier wave.

Analogue to Digital Conversion (A/D) – conversion of an analogue signal into a digital signal (bit sequence).

Digital to Analogue conversion (D/A) – conversion of a digital signal (bit sequence) into an analogue signal.

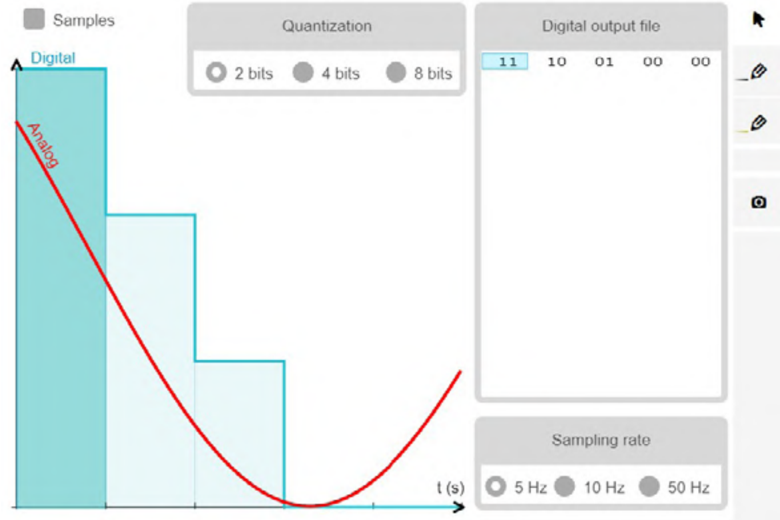
Quantisation – conversion of the signal value in a sample into a sequence of bits; after quantisation the number of signal levels becomes finite.

Sampling – taking a series of samples of an analogue signal, usually at regular time intervals.

Sampling Theorem – an analogue signal can be fully reconstructed from a sequence of its samples if the sampling frequency is twice the maximum frequency in the bandwidth of this signal.

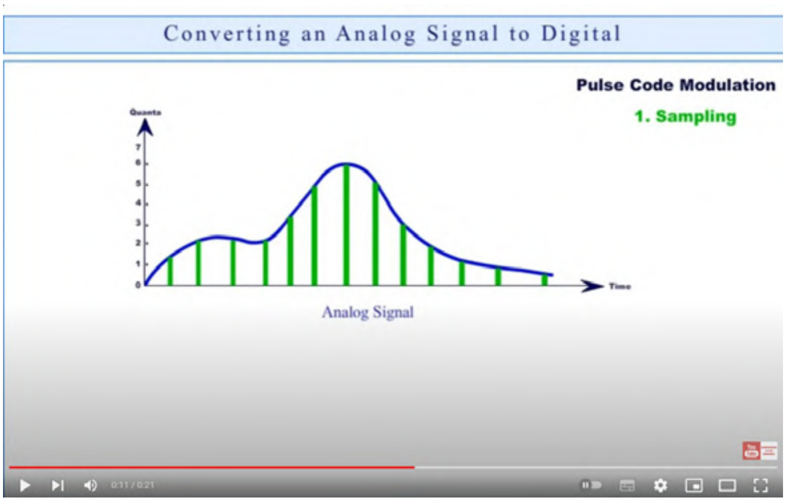
External materials

1. Online application demonstrating A/D conversion. For a given analogue signal (red curve) you can freely change the number of quantisation bits (Quantisation: 2, 4, 8 bits) and the sampling rate (Sampling rate: 5, 10, 50 Hz). The bit sequence after conversion is shown in the "Digital output file" window.



Scan QR code

2. Animation of A/D conversion (Animation of Analogue to Digital Conversion).



Scan QR code

3. What is the difference between AM and FM radio?

Scan QR code

4. FM modulation in classical radio broadcasting and comparison with AM modulation.

Scan QR code

Homework

1. Suppose we want to transmit an EM wave with exactly the same frequency as the C_1 sound on a piano keyboard, i.e. $f = 261.6\,\text{Hz}$. through a dipole antenna. What is the minimum length of the antenna that would allow energy-efficient emission of this wave? And if we amplitude-modulate (AM) a carrier wave of frequency $f_c = 2\ \text{GHz} = 2 \cdot 10^9\ \text{Hz}$? Take the EM wave speed as $c = 3 \cdot 10^8\,\text{m/s}$. Hint: see Lesson 3.

   Given:

   $f = 261.6\ \text{Hz}$

   $c = 3 \cdot 10^8\ \text{m/s}$

   $f_c = 2\ \text{GHz} = 2 \cdot 10^9\ \text{Hz}$

   To find:

   $l = ?$ (without modulation and with modulation)

2. As can be seen in Fig. 4, with three bits per signal sample, we obtain eight levels of the digital signal. What would it be like with four bits? Can you guess the general formula for the number of levels after quantisation when there are N bits.

Lesson 10

Cellular network

Objective

- Presentation of the basic concept of a cellular network and its functioning.

Learning outcomes

- The student knows the principles of selecting a carrier frequency for a signal with a given bandwidth and target antenna size.
- The student understands the need to raise the carrier frequency when increasing the signal bandwidth under frequency division constraints.
- The student can explain the benefits of cellular division of the area in mobile networks.
- The student knows the principles of establishing connections in cellular networks.

1. Carrier frequency selection

In previous lessons we have shown how modulation techniques can be used to divide frequencies between different transmitters so that they can use the electromagnetic (EM) field simultaneously to transmit their signals. Now let us consider what determines the choice of the carrier frequency value.

In Lesson 3, we gave a formula for estimating the minimum size l of an antenna for transmitting a signal in the form of an EM wave. This size depends very simply on the length of the emitted wave – we obtain it by multiplying the wavelength λ by a multiplier depending on the antenna type (½ for a dipole antenna, ¼ for a monopole antenna, etc.). Let us assume that we are dealing with a dipole antenna:

$$l \geq \frac{\lambda}{2} = \frac{c}{2f}$$

Here we have used the formula relating wavelength, frequency and the speed of light c. If we assume a specific, acceptable for us antenna size, we can ask about the condition that the frequencies of the emitted, modulated signal must meet. After a simple transformation of the above formula:

$$f \geq \frac{c}{2l}$$

Therefore, for a given antenna size, the frequencies in the modulated signal band should be greater than the minimum value min $f_{\min}$:

$$f_{\min} = \frac{c}{2l}$$

Let us recall the form of a modulated signal (see Lessons 7, 8, 9). Its bandwidth consists of two sidebands located around the carrier frequency, and the width of the sidebands depends on the bandwidth of the modulating signal. In the case of AM modulation, the width of the sideband is exactly equal to the modulating signal bandwidth, while in FM modulation it can be larger, especially in wideband modulation. If the entire bandwidth of the modulated signal is to be above the minimum frequency, then in the extreme case the left end of the lower sideband can coincide with it (Fig. 1, top). Therefore, the carrier frequency must be at least equal to the minimum frequency plus the width of the sideband.

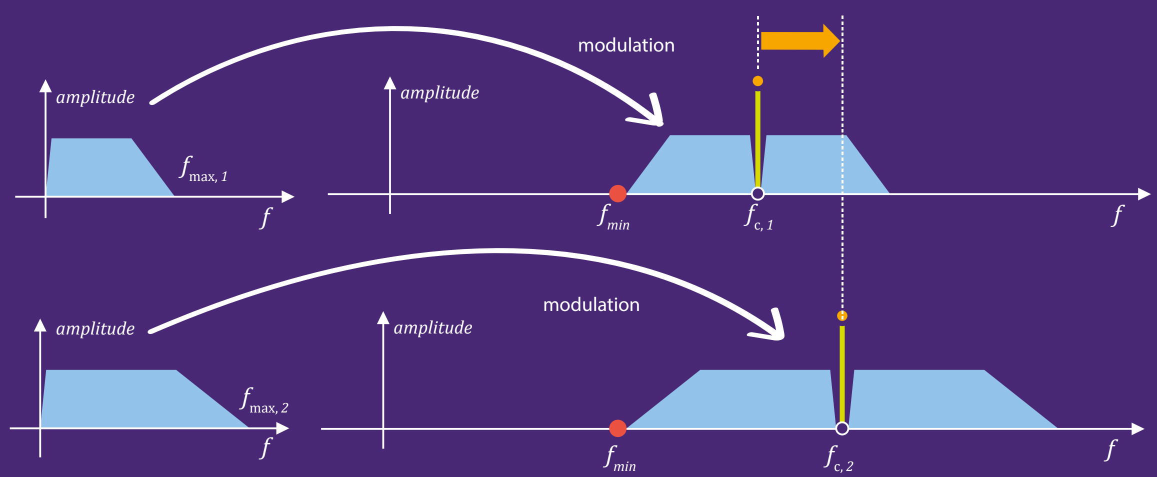


Fig. 1. Dependence of the carrier frequency on the signal bandwidth and the minimum frequency in the modulated signal's spectrum.

116

What happens when the bandwidth of the information signal is widened, for example by increasing the quality of the transmitted audio signal or the speed of the transmitted digital data (see Lesson 9)? This will of course lead to widening of the sidebands, so the carrier frequency will also have to increase (Fig. 1, bottom).

The limitation imposed on the modulated signal bandwidth resulting from the minimum frequency is just one of the possible limitations that we encounter in practice when developing mobile telecommunications systems. Let us assume that a certain sender has been given a frequency range that allows the transmission of two different signals. In Fig. 2, we have marked them as A and B. Frequencies below this range and a certain frequency band above it have been assigned to other transmitters, so we can consider them as occupied frequency bands (from the point of view of the transmitter we are talking about). The bandwidth of both modulated signals is such that the sidebands almost touch the occupied frequency bands, which means that the transmitter has used the full space available to him.

When there is a need to expand the bandwidth of the information signal, and thus the sidebands, or to increase the number of transmitted signals, the only thing the transmitter can do is to move to a completely different part on the frequency axis, usually to much higher carrier frequencies not used by other transmitters so far. This is exactly the phenomenon we observe in subsequent generations of mobile telephony.

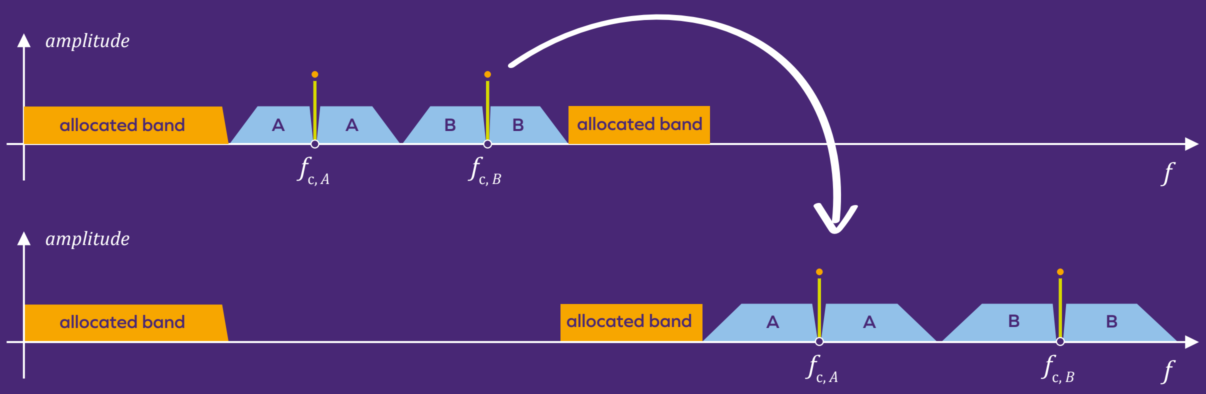


Fig. 2. Necessity of shifting a given transmitter's carrier frequencies when broadening the signal bandwidth under conditions of limited spectrum availability.

2. Allocation of transmission channels

The EM field as a medium for transmitting information is available to the same extent to any transmitter who is able to generate an EM wave. Since high-frequency EM waves lose their value as practical information carriers beyond a certain point (e.g. due to strong absorption or safety reasons, which we will discuss in more detail in Lesson 11), the space on the frequency axis is limited. Due to the vast number of transmitters, it is therefore necessary to legally regulate access to individual ranges. Here, we have a situation similar to road traffic in conditions of high traffic on the road network. Without the regulations introduced by the Highway Code, it would be difficult to imagine driving a car in the city. In Poland, the Office of Electronic Communications (UKE) is responsible for managing the EM wave frequency bands and allocating them to individual transmitters.

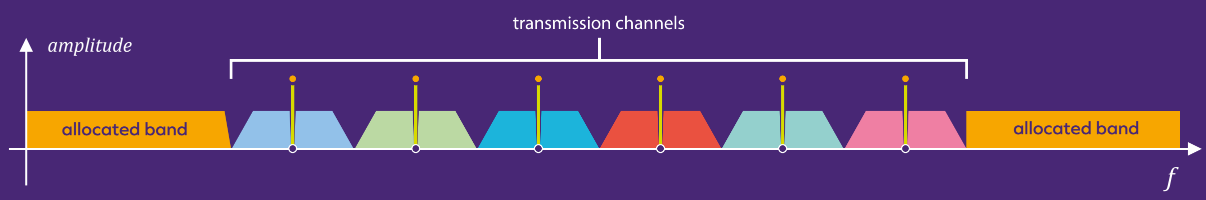


Fig. 3. Transmission channels assigned to a given transmitter.

Certain frequency bands are allocated to specific state services (e.g., the police), radio stations, television, aeronautical and maritime radio navigation. Other bands are allocated to radio amateurs and various generations of mobile telephony operators.

Although you can find freely available frequency bands, the division of the EM wave frequency axis is highly organised. A given transmitter, e.g. a mobile network operator, receives a certain band at their disposal, in which they have the right to transmit. Taking into account the bandwidth of the modulated signal, we can say that a given transmitter has a certain number of so-called **transmission channels** (Fig. 3). IThe larger this number, the greater the number of simultaneous telephone calls a given operator can handle. However, this is always a limited number, which is of exceptional importance for the way a mobile network is built, as we will see further in this lesson. In particular, it necessitates the introduction of so-called **cells**.

Learn more. For the record, it is worth mentioning that modern mobile telephony uses much more advanced modulation techniques than the AM and FM modulations mentioned in previous lessons. Thanks to them, it is possible to use the modulated signal frequency band more efficiently, i.e. to obtain a higher speed of digital information transmission than in simple amplitude modulation. Some of these techniques combine amplitude modulation with phase modulation, thanks to which it is possible to "stuff" more bits of information in different parameters of the modulated harmonic wave.

3. Cellular division of the area

In mobile telephony networks, individual phones do not communicate directly with each other, but always through a so-called **base station**, i.e. a set of antennas mounted on a single mast. Base stations of the **GSM**, system, but also of later generations of mobile telephony, are often called **BTS**s (abbreviation of Base Transceiver Station).

Could a mobile telephone system covering a very large area (e.g. an entire country) consist of a single base station and telephones connecting to it via EM waves? Such a system would be impractical for at least three reasons.

The first reason is that the distance between the base station and the telephone could potentially be very large. As we remember from Lesson 2, the intensity of the EM wave decreases inversely proportional to the square of the distance from the source. Even if we place the base station in the centre of the area, covering its edges with a signal with sufficient energy for uninterrupted reception would require a very powerful transmitter (Fig. 4). This could result in exceeding safety standards near the mast (see Lesson 11) and it would be difficult to ensure a small number of terrain obstacles for the wave between the mast and the receiver (see Lesson 6). Moreover, a high-power transmitter would also be required in the telephone itself, which could violate safety standards, especially since it is a device used in very close proximity to the body. It would also be associated with faster battery consumption.

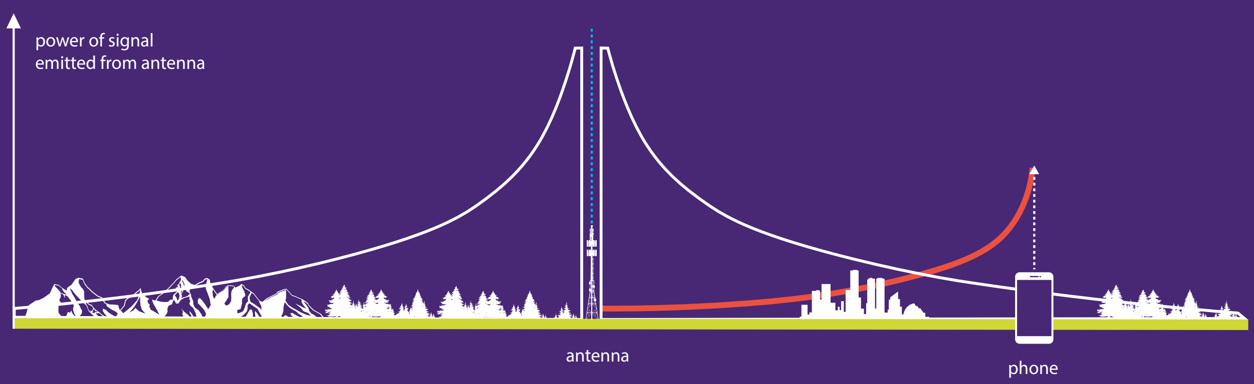


Fig. 4. Mobile network with a single base station.

The second reason is related to the limited number of transmission channels. If users across the entire country had to share a relatively small number of channels available to a given operator, situations would frequently arise where a connection attempt would be made while all channels were occupied by other users. This would result in a dropped call and a "network busy" message.

The third reason concerns the problem of increasing the network coverage area. If the network operator wanted to include users from the vicinity of the coverage area, this would involve even greater demands on the base station transmitter power and increased competition for limited transmission channel resources.

Fortunately, all these problems can be solved by dividing the area into so-called **cells**, i.e. smaller areas, which are assigned separate base stations (Fig. 5). The reduced area serviced by a given base station allows for reducing the power of signal transmitters in both base stations and telephones. Moreover, the frequency bands used in one cell can be used in other sufficiently distant cells. This is possible thanks to the smaller transmitter power – the signal reaching distant cells is weak enough not to interfere with signals emitted within them in the same frequency band This significantly reduces competition for transmission channels and also allows for easy expansion of the network with additional cells.

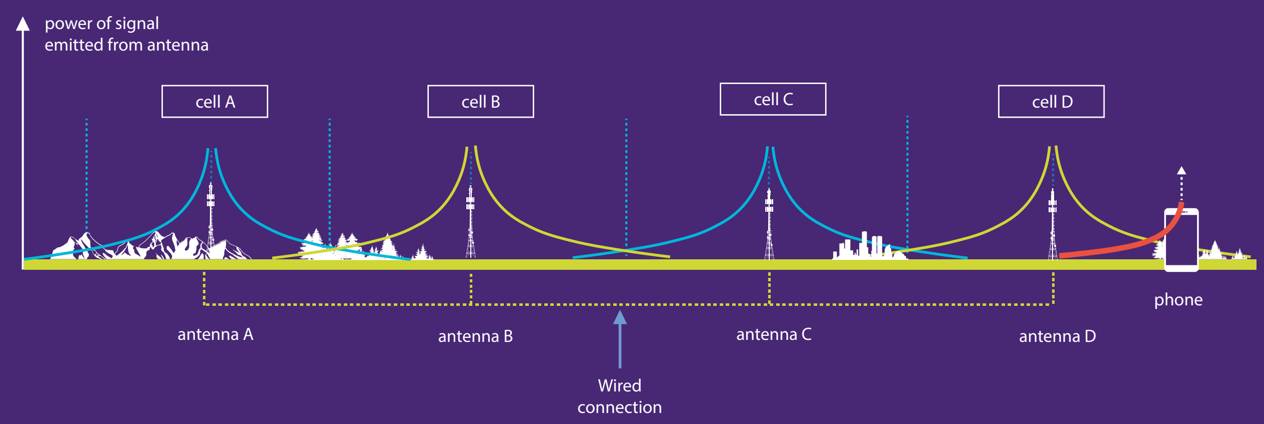


Fig. 5. Concept of a cellular network.

Since the carrier frequencies used in mobile telephony are very high (in 5G networks they can reach up to $26\,\mathrm{GHz} = 2.6 \cdot 10^{10}\,\mathrm{Hz}$), the diffraction effects of EM waves are small (see Lesson 6) and it is advisable to minimise the number of obstacles between the transmitter and the receiver of the signal as much as possible. By dividing the area into cells, it becomes much easier to select the location of the base station and the height of its antennas to ensure optimal visibility (and thus direct connection with phones) across as much of the covered area as possible. This is particularly important in urban environments.

How do base stations connect to each other? Fortunately, unlike phones, they are not mobile objects, so it is sufficient to connect them with a wired network (note: sometimes dedicated, highly directional radio links are used, but for simplicity, we will focus on wired connections).

The actual division of the area is often done using hexagonal cells, as shown in Fig. 6. Hexagons have the advantageous property of being able to tile a plane without gaps, and unlike equilateral triangles or squares, their shape closely approximates a circle.

4. Mobile phone calls

Let us take a closer look at Fig. 6. The signal emitted by a base station in a given cell always leaks into the neighbouring cells to some extent. This is partly a disadvantage – after all, we introduced cell division, among other things, to be able to use the same transmission channels in different parts of the area. Signal leakage into adjacent cells means that they cannot use the transmission channels that are active in their immediate vicinity.

On the other hand, maintaining the signal at the cell boundaries allows for a smooth "handover" of a telephone call when a network user crosses these boundaries, and this can certainly happen, because the mobile network was designed with mobile users in mind. So when we move from one cell to another during a telephone conversation, we maintain a connection with the previous base station for some time and only when the signal from the station serving the new cell becomes strong enough is the connection handed over to it.

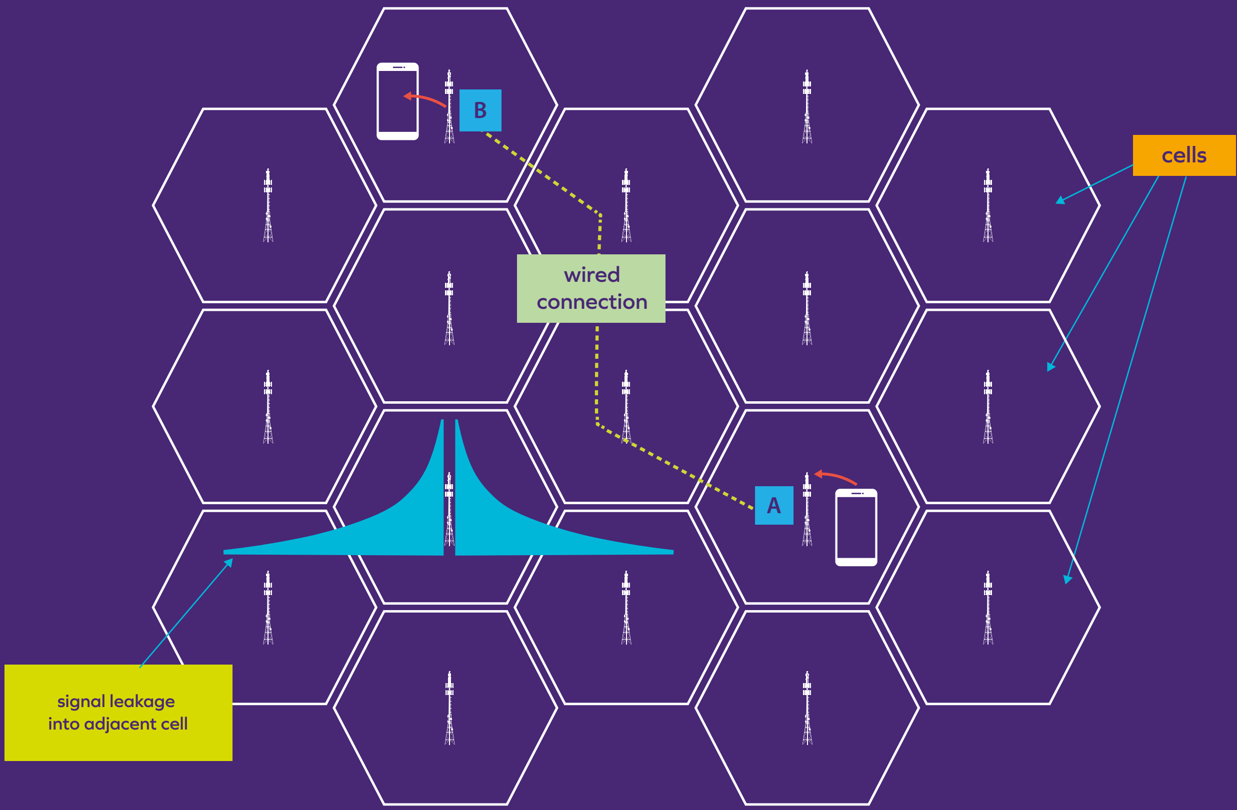


Fig. 6. Cellular division of a given area.

How does the connection process itself work? A phone located in cell A sends an appropriate signal with a connection request to the base station of its cell. The base station reads the necessary information (including who is calling and whom they are calling) and relays it via wired connection to the network management centre. The centre locates the cell where the call's recipient is located and establishes a wired connection between the base stations of cells A and B. Then, base station B initiates a wireless connection with the target phone.

And how does the management centre locate individual phones? Well, each phone periodically establishes a brief connection with nearby base stations to report its location, ensuring that the cell where the phone is located can be immediately identified when needed.

Experiment

We will conduct an experiment that will allow us to assess the effectiveness of distributing EM wave energy in a given area in two ways: in one, we will use a strong source placed in one location, and in the other, two weaker sources placed evenly. In this case, the EM wave will be light, and its source will be light bulbs.

Let us darken the room as much as possible and place two identical bulbs as close to each other as possible and at a distance L from the opaque screen A (Fig. 7). Note the degree of illumination of the screen with the current arrangement of sources.

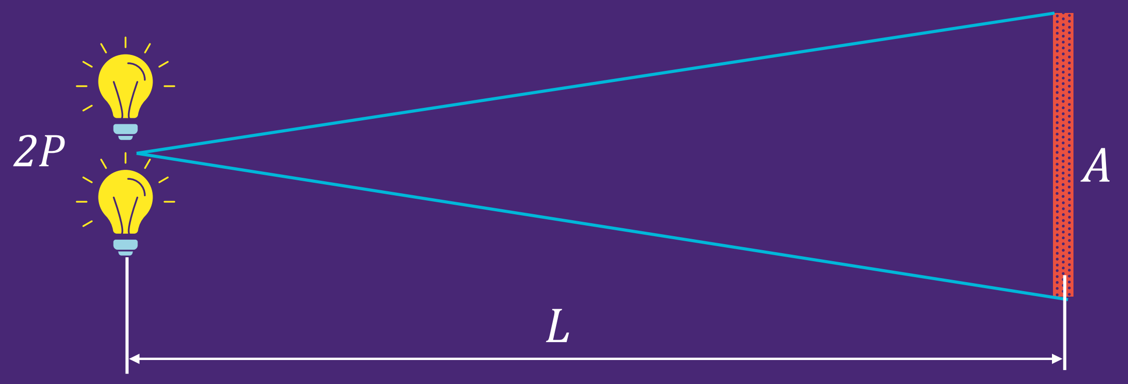


Fig. 7. **Lighting using bulbs placed in one location.**

Next, let us introduce the second screen – B – and place it with one of the bulbs as shown in Figure 8. With this arrangement, both screens should be illuminated to the same extent. Can you tell which of the arrangements illuminates the screens more?

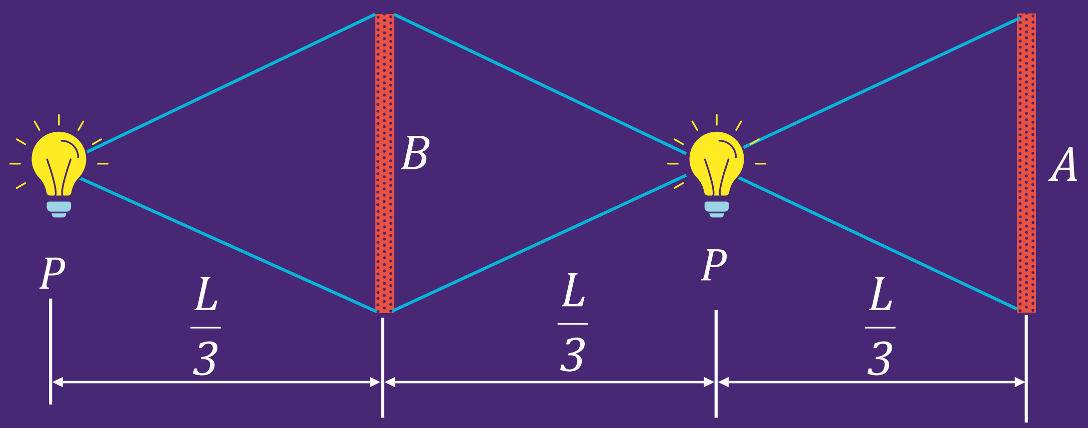


Fig. 8. **Lighting using evenly spaced bulbs.**

Discussion. Let us recall the experiment from Lesson 2. There we used the formula for the intensity of a light wave at a certain distance r, when the source has power P. Let us use the same formula to estimate the illuminance of screen A in the first setting:

$$I_1 = a\frac{2P}{L^2}$$

We have substituted $2P$, as the source power here, because we are dealing with two identical light bulbs located at a distance L from the screen (a is a certain multiplier here, the value of which will turn out to be irrelevant).

And what is the illumination intensity of screen A in the second setting (for screen B it will be the same on both sides):

$$I_2 = a\frac{P}{(L/3)^2} = a\frac{9P}{L^2}$$

Therefore, the ratio of both intensities is equal to:

$$\frac{I_2}{I_1} = \frac{a\frac{9P}{L^2}}{a\frac{2P}{L^2}} = 4.5$$

We can see, therefore, that the illuminance in the second case is over four times greater! So, when we want good lighting over a larger area, it is more advantageous to distribute light sources evenly rather than concentrating them in one place.

The same applies to mobile network base stations. From the perspective of covering a specific area, it is better to deploy many evenly distributed antennas with lower-power transmitters than a single antenna with a transmitter whose power equals the sum of the powers of the individual antennas.

Glossary

Transmission channel – a frequency band allocated to a given transmitter around a specific carrier frequency.

Cell – a designated area within a mobile network operator's coverage zone, served by a base station and assigned a unique set of transmission channels.

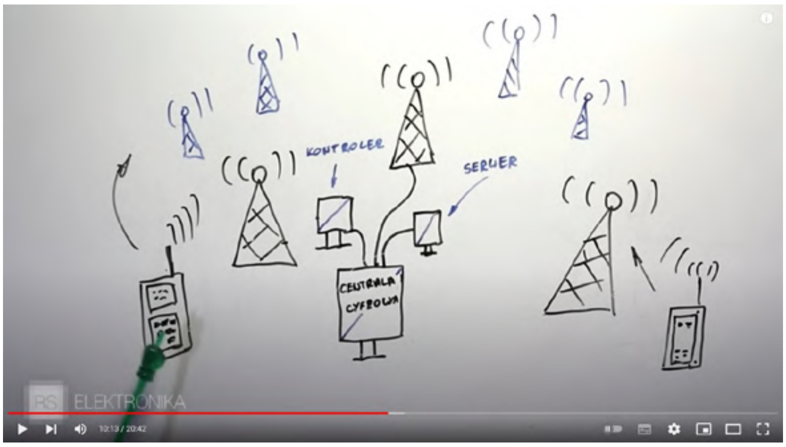
Base Station (or BTS) – a set of antennas connected to transceiver equipment that enables wireless communication between phones and the mobile network. Base stations are interconnected via wired infrastructure or dedicated radio links.

BTS (base transceiver station) – the standard abbreviation for a base station in mobile networks.

GSM (Global System for Mobile Communications) – one of the most widely adopted standards for mobile telephony.

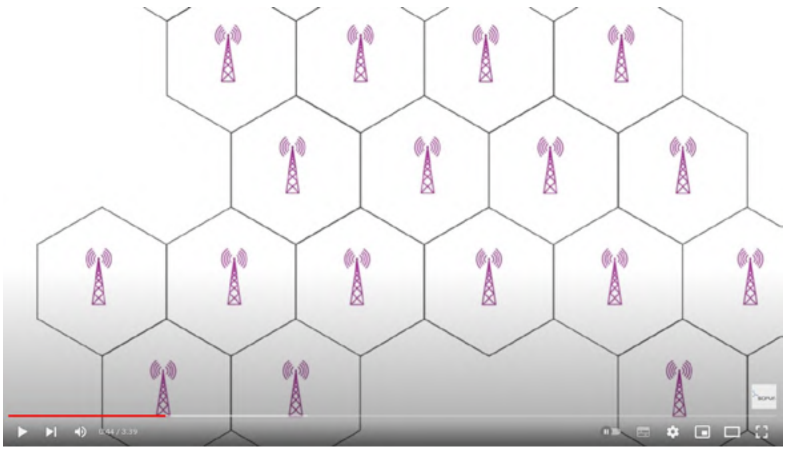
External materials

1. How mobile networks work.



Scan QR code

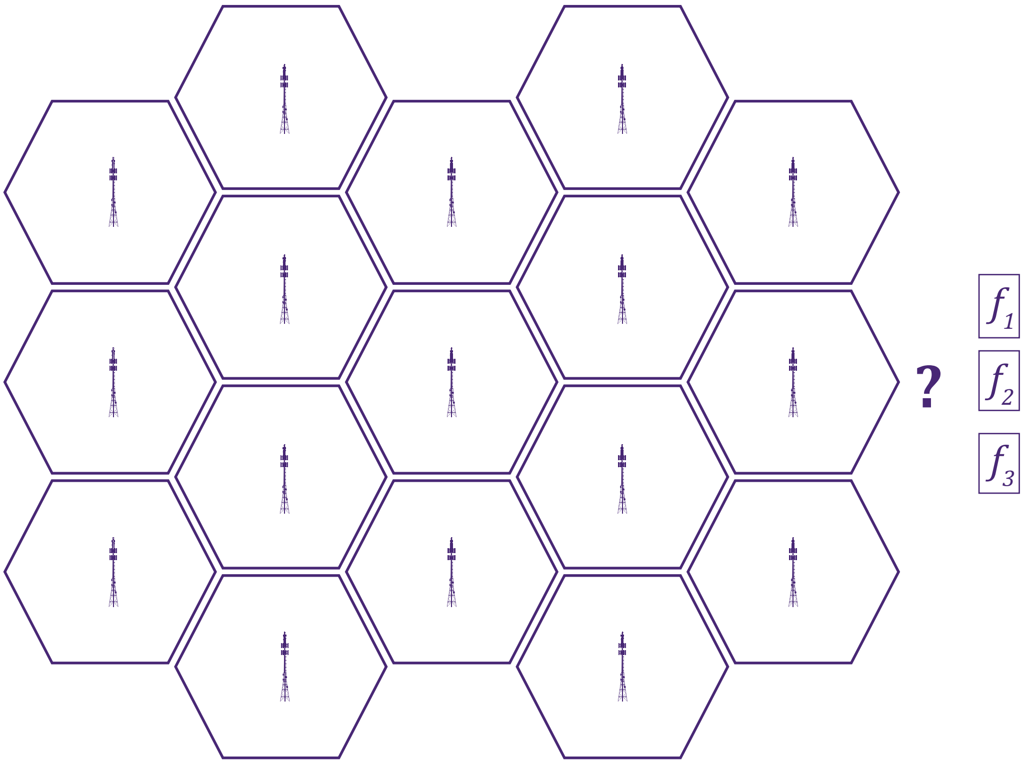
2. What is roaming and how does it work? – an educational film not only about roaming itself, but also about the basics of how a mobile network works.



Scan QR code

Homework

1. For the cellular network in the figure below, assign transmission channels. Adjacent cells cannot be allocated the same channel due to signal leakage across cell boundaries. Each channel will be conventionally referred to by its assigned carrier frequency (from the set: f_1, f_2, f_3, etc.). Can this task be achieved using only two channels (f_1 i f_2)? Justify your answer. What about three channels (f_1, f_2 i f_3)? If yes, enter the appropriate symbol into each cell.
2. Calculate the minimum carrier frequency for a $50\,\mathrm{MHz}$ bandwidth signal and a $1\,\mathrm{m}$. dipole antenna. AM modulation will be used to transmit the signal. Assume the EM wave speed as $c = 3 \cdot 10^8\,\mathrm{m/s}$.



Given:

$B = 50\ \text{MHz} = 50 \cdot 10^6\ \text{Hz} = 5 \cdot 10^7\ \text{Hz}$ – bandwidth of the modulating signal.

$l = 1\ \text{m}$ – antenna length.

To find:

$f_c = ?$

Lesson 11

Electromagnetic field and health

Objective

- Presentation of issues related to the influence of electromagnetic fields on living organisms, including humans.

Learning outcomes

- The student knows the frequency ranges of various types of electromagnetic radiation.
- The student knows the effects of ionising and non-ionising radiation on living organisms.
- The student knows the typical frequencies and signal powers used in modern mobile telephony.

1. EM waves - frequency ranges

In this final lesson, we will address the impact of electromagnetic waves (EM) on living organisms, including humans. The growing use of EM waves in modern mobile telecommunications sometimes raises concerns, especially when the term **radiation** is used, as it is often associated with significant health risks.

Let us get the facts straight. The EM field fills the entire universe, and the disturbances spreading in it have a huge number of natural and artificial sources, i.e. those caused by human activity. We call these disturbances EM waves or EM radiation. As we remember from previous lessons, we can assign a frequency (or wavelength) and amplitude to the spreading wave, which can be directly related to the energy (or power) carried by the wave. If we ask about the effect of an EM wave on living organisms, we must be aware that it depends to a large extent on these two parameters – frequency f and energy.

Let us first examine the former. Figure 1 shows the frequency ranges for the basic types of EM radiation (it is worth noting that the boundaries between them are conventional and that the scale on the axis is exponential - moving from one point to the next one means a tenfold increase in frequency). The waves with the lowest frequencies, below $10^8\,\mathrm{Hz}$, are identified with radio waves. The next range – so-called microwaves - includes typical frequencies found in mobile telephony (approx. $10^9 - 10^{10}\,\mathrm{Hz}$) or in microwave ovens. The infrared range largely overlaps with so-called thermal radiation (we will talk about this in more detail in the last part of this lesson). EM waves from the range of $4 \cdot 10^{14} - 8 \cdot 10^{14}\,\mathrm{Hz}$ are what we know as visible light- from red (the lowest frequency) to violet (the highest frequency) through all the colours of the rainbow. The range above visible light is called ultraviolet (ultraviolet radiation from the sun is responsible for our tan, among other things), while EM waves above a frequency of $10^{18}\,\mathrm{Hz}$ are called X-rays or Röntgen radiation (used in medicine to x-ray tissue).

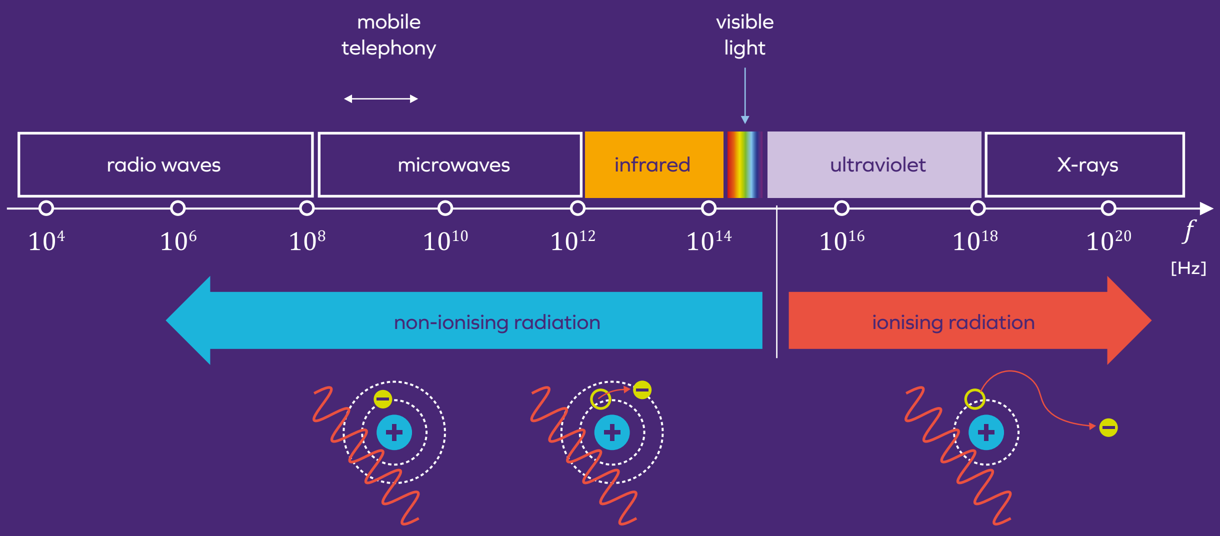


Fig. 1. Ranges of different types of electromagnetic radiation.

Radiation below 10^{15} Hz is defined as **non-ionising**, and above – as **ionising**. This division is of great importance from the point of view of the impact on living organisms.

2. Ionising and non-ionising radiation

To better understand the interaction of EM waves with atoms and molecules, which constitute matter - including living organisms - it is useful to know that, in addition to the wave description we are already familiar with, EM radiation also requires the assumption that it consists of particles called **photons** (see Fig. 2). Photons, like EM waves, travel at the speed of light c, have no mass, and are characterised only by energy E which depends on the frequency f of the EM wave according to a very simple formula:

$$E = hf$$

where the coefficient h is the so-called Planck's constant, equal to $h = 6,6 \cdot 10^{-34}\ \mathrm{J \cdot s}$. It is the energy of the photon that determines what happens when an EM wave passes through an atom. A change in the amplitude of the wave leads to a change in the number of photons emitted by the source but does not change the energy of an individual particle.

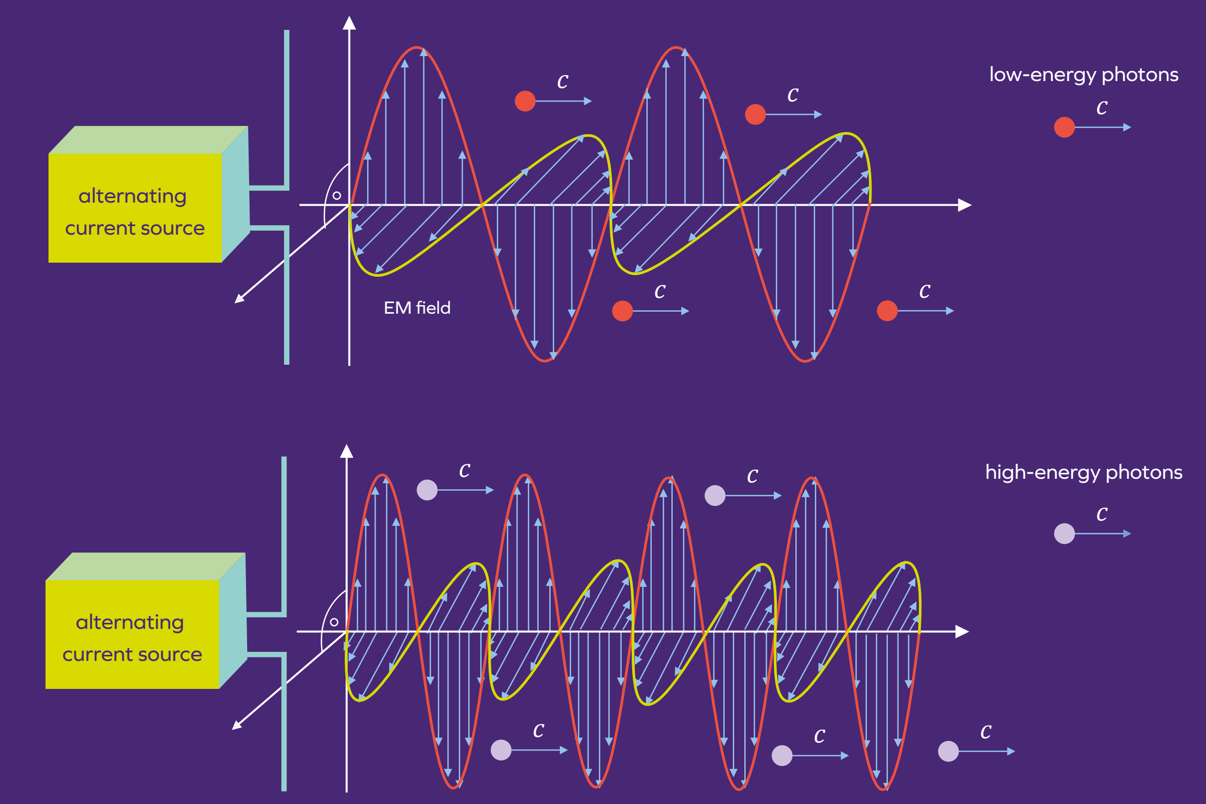


Fig. 2. EM radiation as a wave phenomenon and a collection of photons.

In Lesson 6, we discussed various phenomena accompanying the passage of EM waves through matter. Let us now examine this in more detail from the perspective of a collection of photons (see Fig. 3). A photon passing near an atom may, through interaction with its electrons, cause one of them to become excited to a higher energy level - but this can only occur if the photon's energy matches the gap between these levels. In such a case, a phenomenon similar to resonance occurs (see Lesson 5): the photon is absorbed, and the atom transitions to an excited state (Fig. 3b). If resonance does not occur, the photon penetrates through the atom (Fig. 3a) – meaning the EM wave passes through the matter without any disruption.

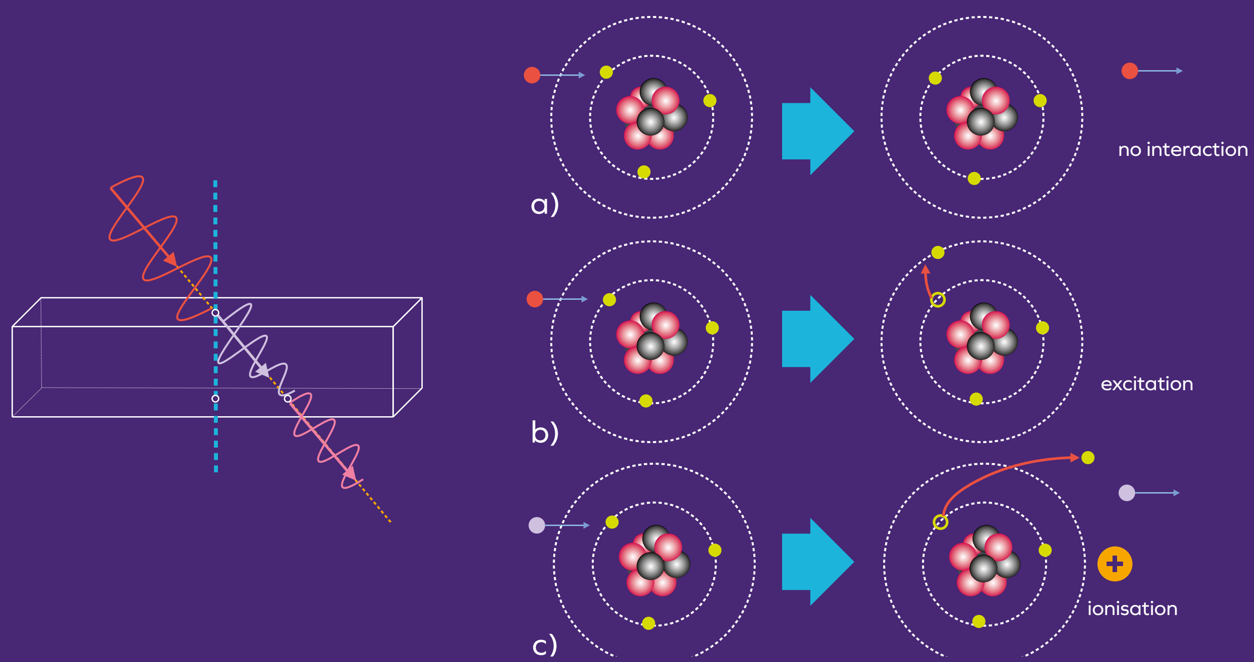
High-energy photons, i.e. those whose energy is sufficient to detach an electron from an atom (i.e. greater than the **ionisation energy**), knock the negatively charged electron out of the atom, transforming it into a positive **ion**. This phenomenon is called **ionisation** and is highly harmful to living organisms – it can damage DNA and disrupt the functioning of cells (including the development of cancer cells). Since the possibility of ionisation depends only on whether the photon energy is greater than the ionisation energy of atoms and molecules, and the photon energy depends only on the frequency of the EM wave, the division into ionising and non-ionising radiation can be made only based on the frequency value, as we did in Fig. 1. Does this mean that the radiation power does not matter? As we have already mentioned, the increase in the wave amplitude is associated with a change in the number of photons emitted per unit of time. Therefore, if we imagine high-energy photons as dangerous projectiles, then the change in the radiation power leads to an increase in the rate of fire of the gun. This means that more body cells can be damaged in a given period of time, with even a single bullet being able to cause significant damage (this is important with ionising radiation).

3. Thermal effect

We mentioned that one of the possible effects of electron-photon interaction is the excitation of an atom to a higher energy level (Fig. 3b). Such an excited atom (or molecule) can release energy in two ways:

- radiate energy in the form of a photon – the electron then returns to its original level and the EM wave is reflected (Fig. 4a);
- collide with another atom and transfer its excess energy as kinetic energ – this results in the absorption of the EM wave, and the atoms or molecules of the object increase their motion or vibration intensity (Fig. 4b).

From the point of view of the impact on a living organism, the second option is more interesting. Increasing the intensity of the movement or vibrations of the atoms and molecules of a given body is nothing more than an increase in temperature. In this case, the body is heated, which is why this phenomenon is called a **thermal effect**. For the thermal effect to be substantial, resonant photon absorption must occur, leading the excited atom to transition to a state of higher kinetic energy, and the temperature rise must be amplified by flooding the object with a "sea" of photons. The scale of this effect therefore depends critically on the radiation power and the type of matter exposed to the EM waves.



Rys. 3. Different types of interaction of photons with electrons in atoms.

A microwave oven is a perfect example of the thermal effect. The frequency of the waves emitted, close to $2.45\,\mathrm{GHz} = 2.45 \cdot 10^{9}\,\mathrm{Hz}$ (i.e. in the range used in mobile phones), is close enough to the resonant frequency of absorption by water molecules to excite them to vibrate strongly, and thus heat the food in which they are located. The radiation power of a typical oven is up to $1000\,\mathrm{W}$.

While ionisation can only occur for radiation in the range above $10^{15}\,\mathrm{Hz}$, i.e. at frequencies 100,000 times higher than those used in mobile telephony, the thermal effect must certainly be taken into account and included in safety standards, compliance with which depends on the power of the emitted signals.

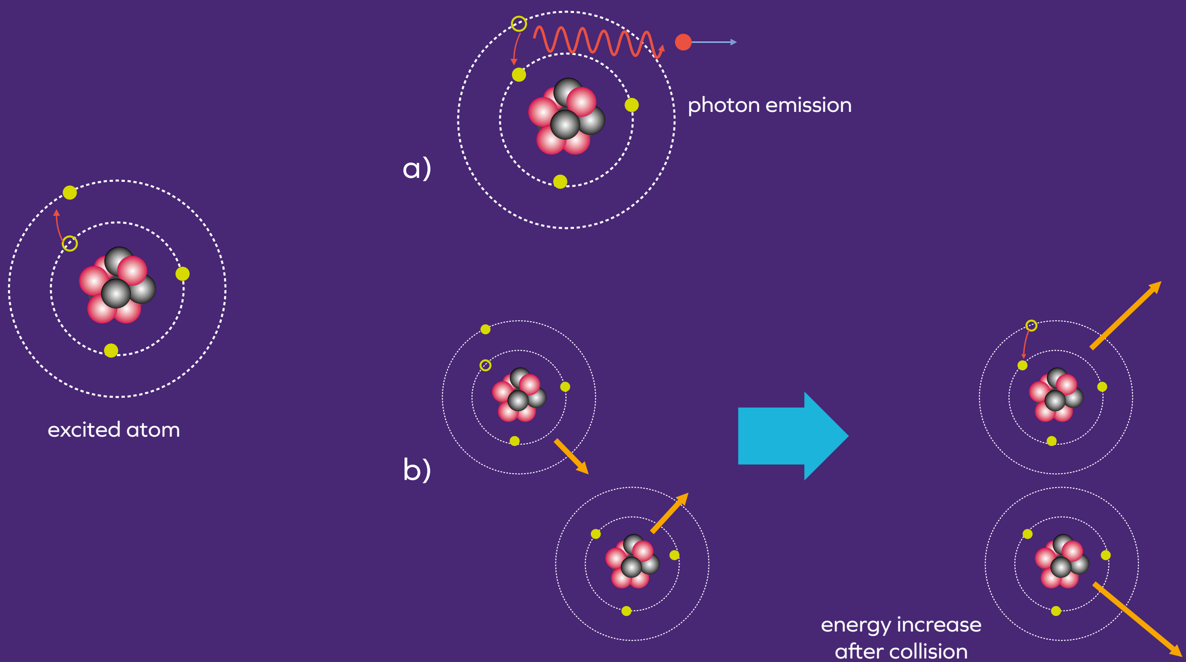


Fig. 4. Different ways in which energy can be released by an excited atom.

130

4. SAR coefficient

Having learned from the microwave oven example, we should be very cautious about using microwaves in mobile telephony. Overheating living tissue can be really dangerous, especially since it consists mostly of water, which is particularly susceptible to absorbing EM radiation in this frequency range.

For this reason, appropriate legal standards have been introduced, regulating – depending on the frequency – the permissible levels of exposure in publicly accessible areas. For mobile telephony frequency ranges, the maximum allowed levels are an EM radiation intensity of – $\mathbf{10\ W/m^2}$– and an electric field strength of – $\mathbf{61\ V/m}$. Since these quantities alone do not directly express the degree of energy absorption by living tissue, an additional coefficient called the Specific Absorption Rate **SAR** has been introduced. Measured in $\mathrm{W/kg}$, SAR estimates how much radiated EM power is absorbed per kilogram of body mass.

The safe SAR value accepted in Europe is $\mathbf{2\,W/kg}$ and ensures that the tissue temperature is not increased by more than $1\,^{\circ}\mathrm{C}$.

Although base stations usually raise the most controversy and concerns about their impact on the health of nearby residents, it should be noted that their transmitter power typically does not exceed $40\,\mathrm{W}$, and the intensity of the EM waves they generate decreases quite rapidly with distance. When compared to the transmitter power in a phone, which does not exceed $2\,\mathrm{W}$, and considering that the phone is usually held close to the head during a call, it becomes clear that the EM waves emitted by our own phones are of decisive importance (see Fig. 5). It is also worth noting that the phone's transmitter power is several hundred times lower than that of a microwave oven.

Each phone has its own SAR value, usually listed separately for the head and the whole body. Phones that do not meet the safety standard (i.e. violating the SAR $< 2\,\mathrm{W/kg}$) condition) are not allowed for sale. Typical SAR values of popular phones are much lower than $2\,\mathrm{W/kg}$.

It is worth noting that the SAR values provided by the manufacturer are maximum values, i.e. measured in conditions in which the phone transmits the signal with the highest possible power. Typically, the power of the transmitted signal is much lower. It depends on factors such as distance from the base station, the type of obstacles to the signal between the phone and the station, and even the way the phone is held.

From time to time, there are claims about harmful effects of EM radiation that are of a nature other than ionisation or thermal effects. However, it should be emphasised that none of these reports have been confirmed by solid scientific research.

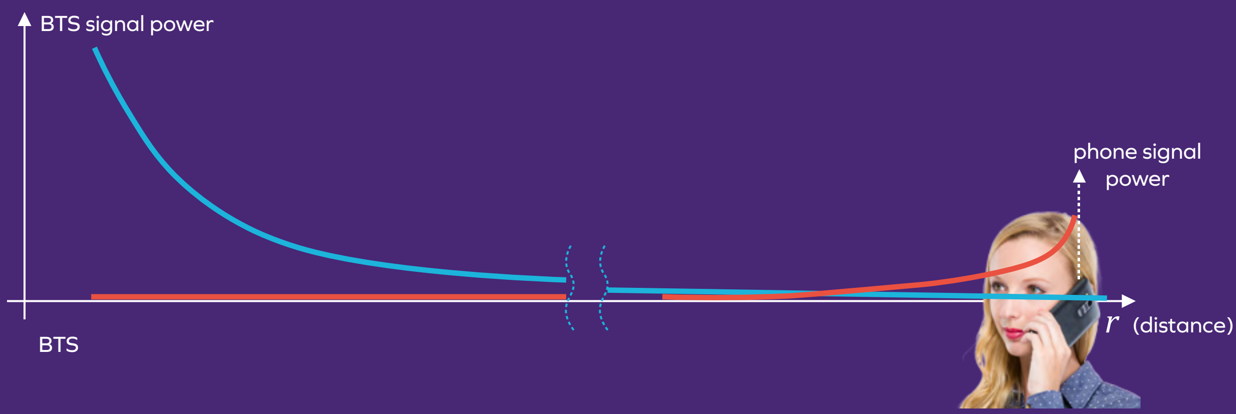


Fig. 5. Power of the signal transmitted by a BTS and by a phone as a function of distance from the antenna.

5. Thermal radiation

Finally, to emphasise that generating EM radiation is neither mysterious nor limited to human civilisational achievements (such as mobile telephony), let us note that every one of us is a transmitter of EM waves!

Any object with a temperature above absolute zero ($0\,\mathrm{K}$) is a source of so-called **thermal radiation**, which theoretically contains components at every possible frequency (see Fig. 6). According to the law of thermal radiation, its power is negligible for very low and very high frequencies, but there is a single maximum. This maximum becomes higher and shifts toward higher frequencies as the temperature of the object increases.

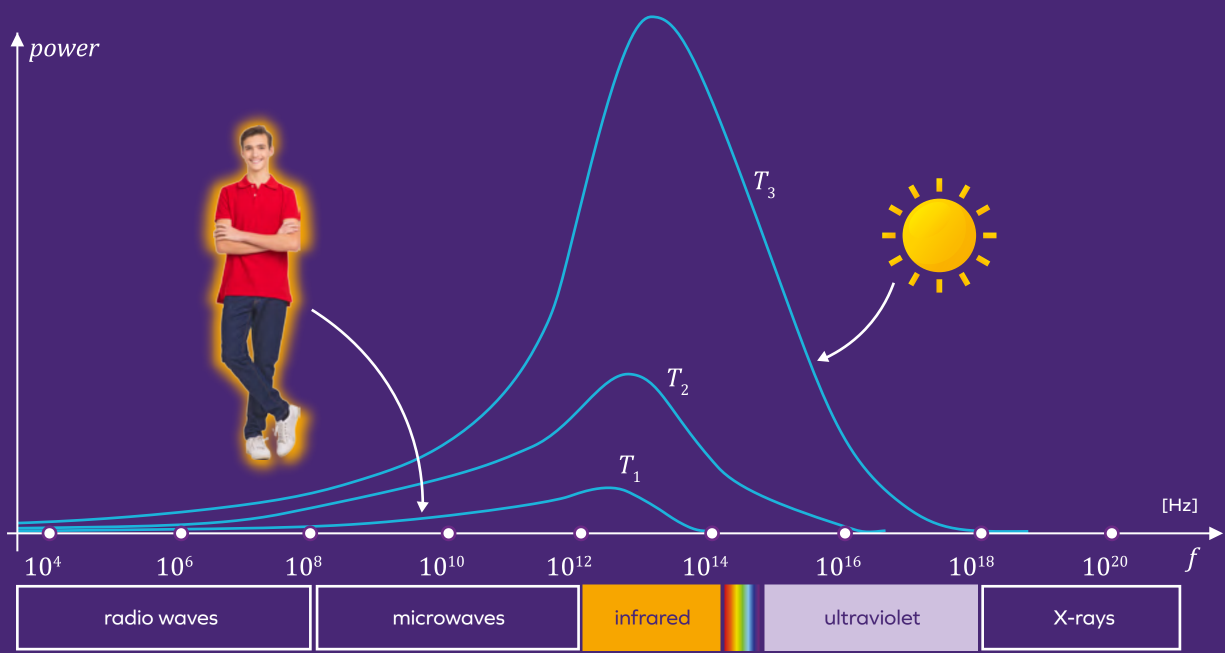


Fig. 6. Thermal radiation power for different frequencies depending on the temperature of the object.

For humans (temperature around $37\,°C$) this maximum falls in the range of infrared radiation, which is invisible to us. However, it can be perfectly observed by thermal imaging cameras.

Bodies heated to a sufficiently high temperature begin to glow (e.g., metal heated to what we call "red heat"). This occurs because the peak of the radiation curve moves into the range of visible radiation.

The sun, with a surface temperature of close to $5500\,°C$, emits not only infrared and visible radiation – a large part of EM waves falls in the ultraviolet range, which is a nice addition for every sunbathing enthusiast.

Experiment

We will conduct an experiment that will allow us to simulate, in a way, the phenomena of ionisation and the thermal effect. Let us prepare an open box with rather low walls and place several small balls (e.g., tennis balls) inside – see Fig. 7. The balls confined within the walls will model electrons, which cannot leave the box unless they are supplied with sufficiently high energy. Let us take one ball in hand, which will represent a photon.

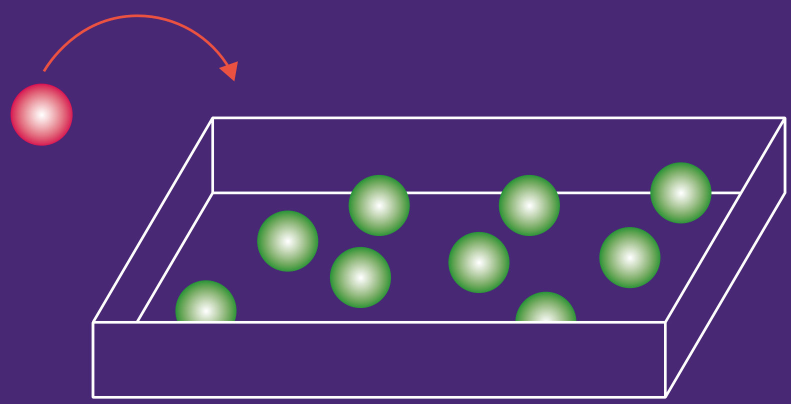


Fig. 7. Simulation of ionisation and thermal effect.

1. Let us throw a "photon" into the box without giving it too much speed. Notice that some of the balls, as a result of the collision with the "photon", will start moving and collide with other balls (Fig. 8a). In this way, we obtain a certain model of the thermal effect. The energy transferred by the ball was not enough to knock the "electrons" out of the box, but it increased their kinetic energy.
2. Let us throw the "photon" again, but this time with as much force as possible, trying to hit one of the balls in the box. With a bit of luck, you should be able to knock one of the "electrons" out of the box walls (Fig. 8b). This means that the energy of the "photon" was sufficient to achieve the "ionisation" effect.

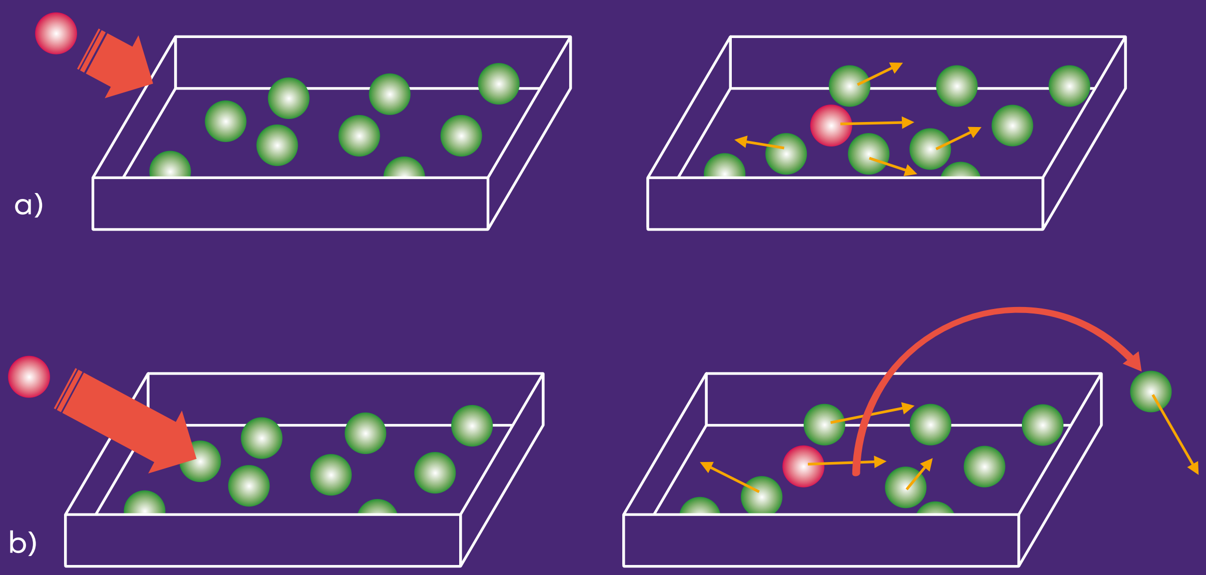


Fig. 8. Different behaviour of balls depending on the different initial energy of the 'photon' ball.

For discussion:

- What do the walls of the box represent in our model?
- Does increasing the wall height affect "ionisation"?

Glossary

Thermal effect – heating of matter due to its interaction with electromagnetic radiation.

Ionisation energy – the minimum energy that must be supplied to an atom or molecule to remove an electron and ionise it.

Photon – a particle of electromagnetic radiation. It has zero mass and carries energy proportional to the frequency of the radiation.

Ion – a positively or negatively charged atom. An ion is characterised by an excess or deficiency of electrons relative to an electrically neutral atom.

Ionisation – the conversion of a neutral atom or molecule into an ion by removing an electron.

Ionising radiation – a type of electromagnetic wave with sufficient energy to ionise atoms and molecules.

EM radiation – another term for electromagnetic waves

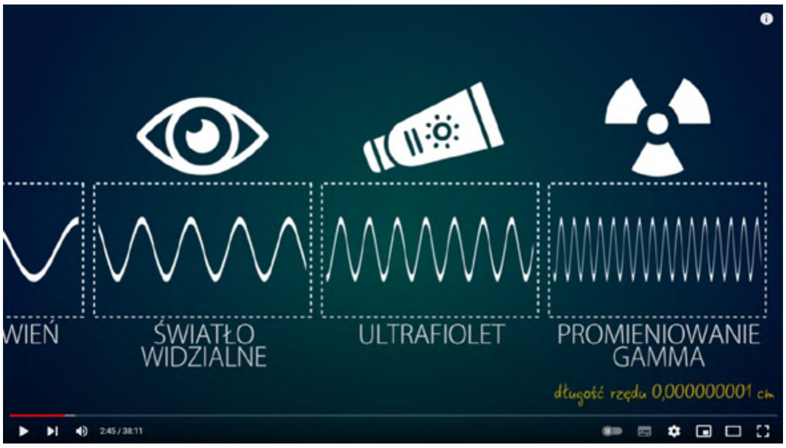
Non-ionising radiation – a type of electromagnetic wave with energy too low to ionise atoms and molecules.

Thermal (heat) radiation – emission of electromagnetic waves by a body at a temperature greater than absolute zero ($0\ \mathrm{K}$).

SAR – a coefficient defining the power of electromagnetic radiation absorbed by living tissue per kilogram of body weight. Unit – $\mathrm{W/kg}$.

External materials

1. 5G and mobile phones – are they a threat to us?



Scan QR code

2. 5G and health – is it harmful?

Scan QR code

Homework

1. Check your phone's SAR value. On some phones, this information can be accessed using the USSD code: #07#. If the code doesn't work, search the internet for your phone model along with the keyword "SAR".

2. Calculate the frequency of EM radiation to ionise the oxygen molecule. For this to happen, the energy of the photon must exceed the ionisation energy. What type of EM radiation is this? Assume the ionisation energy $E_i = 2.2 \cdot 10^{-18}$ J and Planck's constant $h = 6.6 \cdot 10^{-34} J \cdot s$.

Given:

$E_i = 2.2 \cdot 10^{-18} J$ – ionisation energy

$h = 6.6 \cdot 10^{-34} J \cdot s$ – Planck's constant

To find:

$f = ?$

Homework Solutions

Lesson 1. Problem 1.

1. Electrical signal (brain) – the sender decides on the content of the message.

2. Electrical signal (nervous system) – the sender's muscles are stimulated by the impulses causing tapping on the wall.

3. Vibration signal (wall) – the material from which the wall is made vibrates under the influence of impacts.

4. Acoustic signal (sound wave in air) – wall vibrations are transferred to air vibrations on the other side of the wall.

5. Vibration signal (eardrum in the ear) – the signal reaches the other person's ear.

6. Electrical signal (nervous system) – message is sent to the other person's brain.

7. Electrical signal (brain) – change of coding (tapping needs to be converted to flashing a flashlight).

8. Electrical signal (nervous system) – the other person's muscles are stimulated to press the flashlight button.

9. Electrical signal (flashlight electrical system) – a pressed button applies electrical voltage to the bulb.

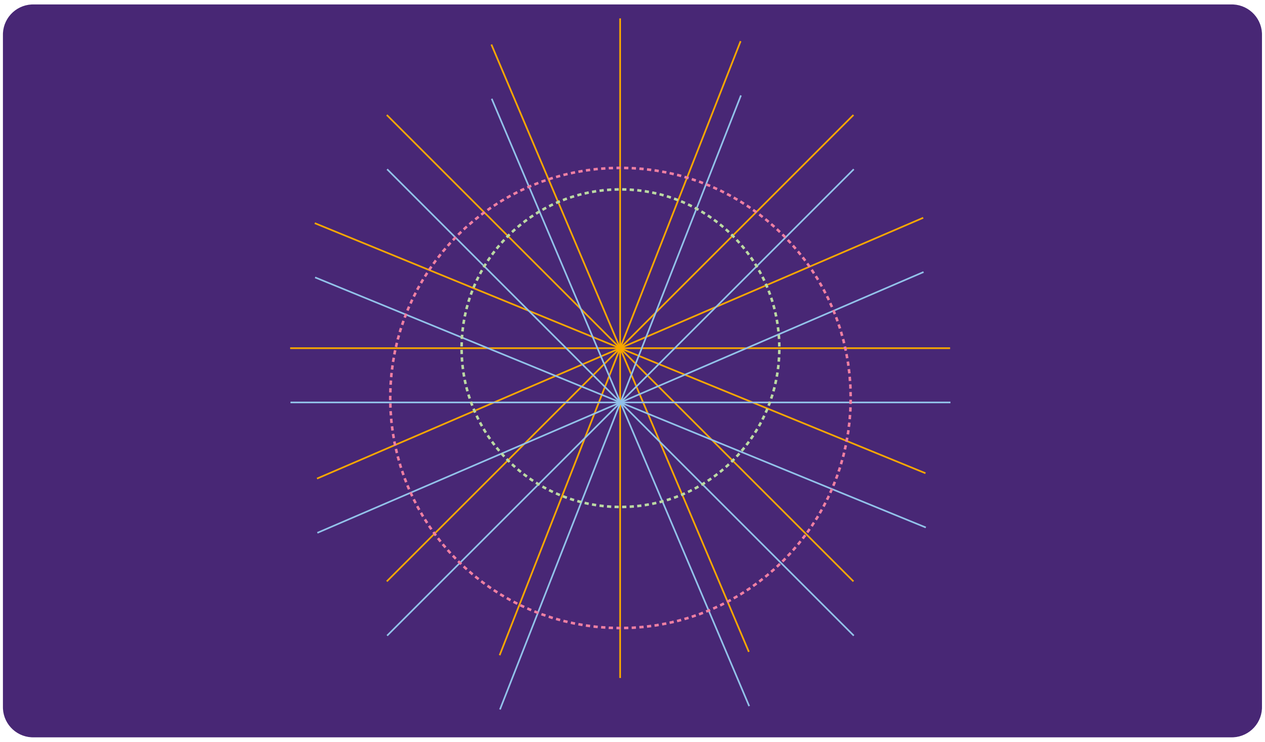
10. Light signal (bulb) – a bulb emits a signal in the form of light.

Lesson 2. Problem 1.

r	1	2	5	10	100	1000
$1/r$	1	0.5	0.2	0.1	0.01	0.001
$1/r^2$	1	0.25	0.04	0.01	0.0001	0.000001

We can see that with the increase of r the factor $1/r^2$ decreases significantly faster than $1/r$ and the more so the larger r is. At a large distance from the charge, the value of the field intensity calculated from Coulomb's law (variation of $1/r^2$) becomes insignificant compared to the value of the intensity in the EM wave region (variation of $1/r$).

Lesson 2. Problem 2.



Lesson 3. Problem 1.

Using the basic formula:

$$v = \lambda f$$

we transform it into the form:

$$\lambda = \frac{v}{f}$$

We substitute the given numerical values:

$$\lambda = \frac{v}{f} = \frac{340\ \mathrm{m/s}}{261.6\ \mathrm{Hz}} = 1.3 \frac{\mathrm{m/s}}{1/\mathrm{s}} = 1.3\ \mathrm{m}$$

Lesson 3. Problem 2.

We use the formula for antenna size:

$$l = \frac{\lambda}{2}$$

We assume equality here because we want the shortest possible antenna length.

From the formula for wavelength:

$$\lambda = \frac{v}{f} = \frac{c}{f}$$

we have

$$l = \frac{\lambda}{2} = \frac{c}{2f} = \frac{3 \cdot 10^8 \, \text{m/s}}{2 \cdot 261.6 \, \text{Hz}} \approx 5.73 \cdot 10^5 \, \text{m} \approx 573 \, \text{km!}$$

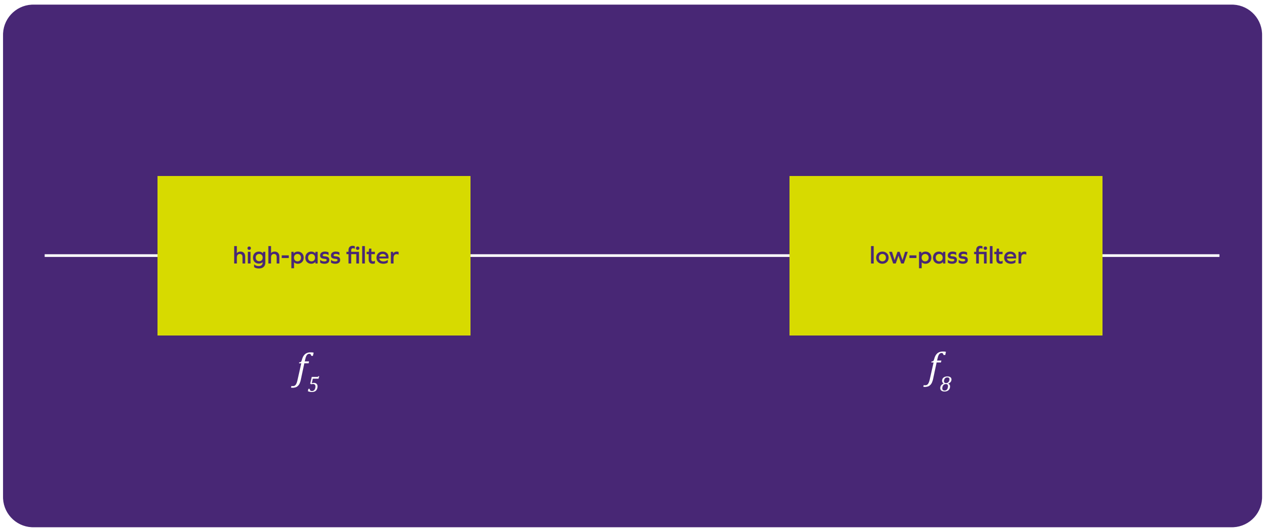
Efficient transmission of EM waves at such a low frequency seems practically impossible!

Lesson 4. Problem 1.

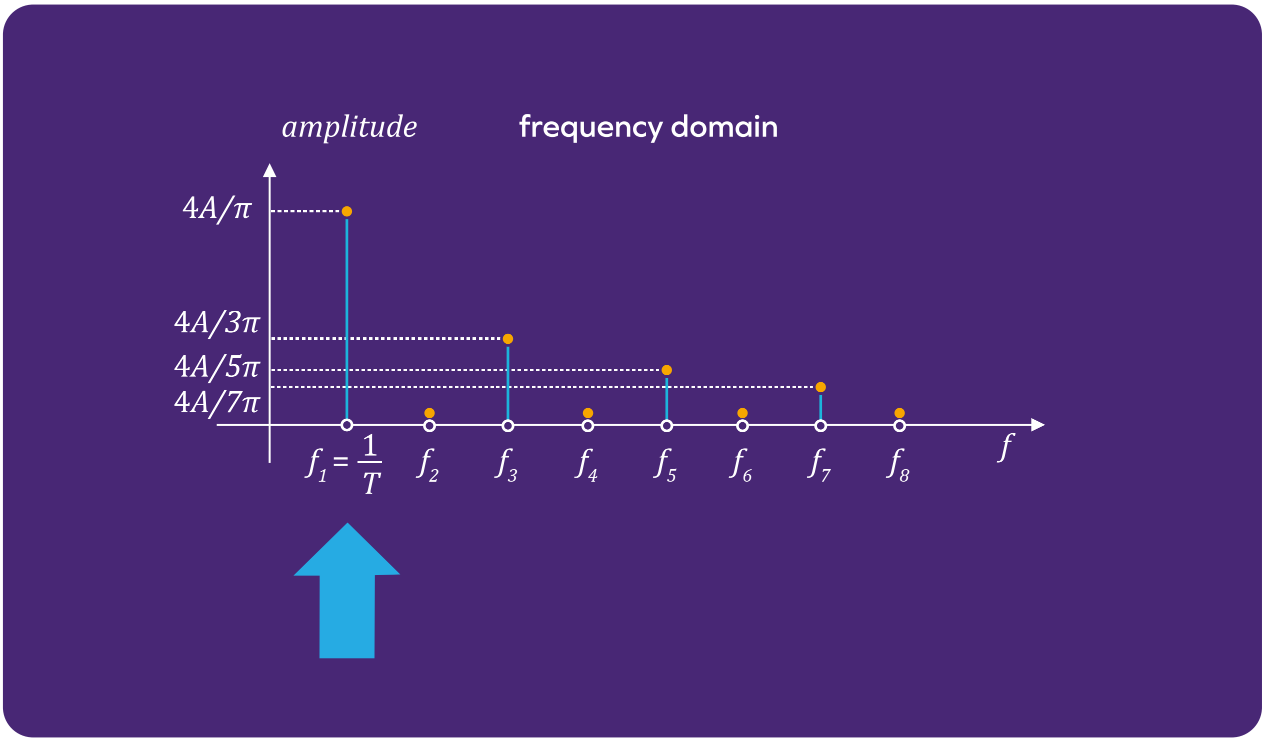
1 b, 2 c, 3 a.

Lesson 5. Problem 1.

Yes, it is possible. We can first process the signal with a high-pass filter with a cut-off frequency of f_5; then only frequencies above f_5 (inclusive) will remain in it. Then we pass the processed signal through a low-pass filter with a cut-off frequency of f_8. Then the harmonics below f_8. will be removed. So exactly those harmonics in the range from f_5 to f_8.



Lesson 5. Problem 2.



Yes. It is enough to use a low-pass filter with a cut-off frequency of $f = 1/T$, or a band-pass filter covering this frequency, but not reaching $f_2 = 2f_1$ (i.e. the second harmonic).

As we showed in Lesson 4, the spectrum of a square-wave signal (see below) contains a harmonic f_1 with frequency $f_1 = 1/T$, which has exactly the desired period.

Lesson 6. Problem 1.

Using the formula relating wavelength, frequency, and wave speed:

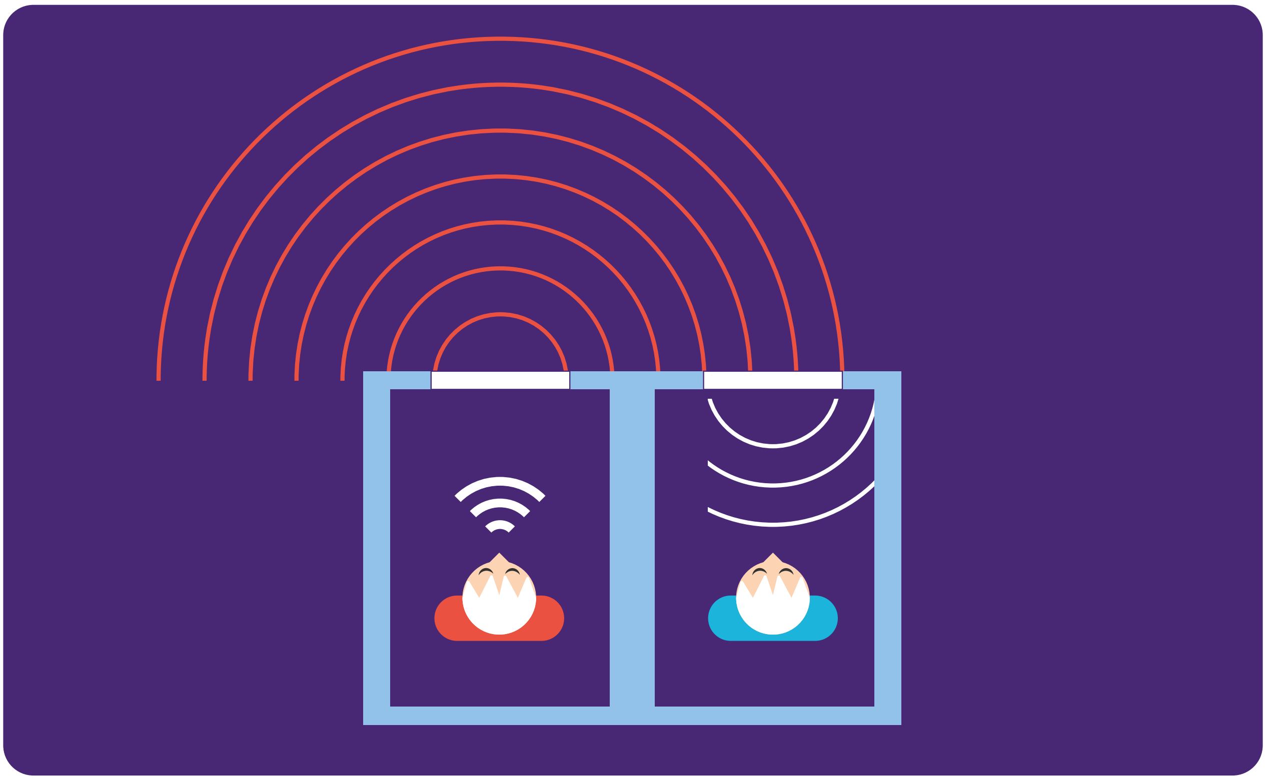
$$f = \frac{v}{\lambda} = \frac{340 \text{ m/s}}{0.02 \text{ m}} = 17000 \text{ Hz} = 17 \text{ kHz}$$

This is a frequency at the limit of human hearing. In practice, to reduce the impact of diffraction, bats emit ultrasound with a frequency of up to $200\,\text{kHz}$.

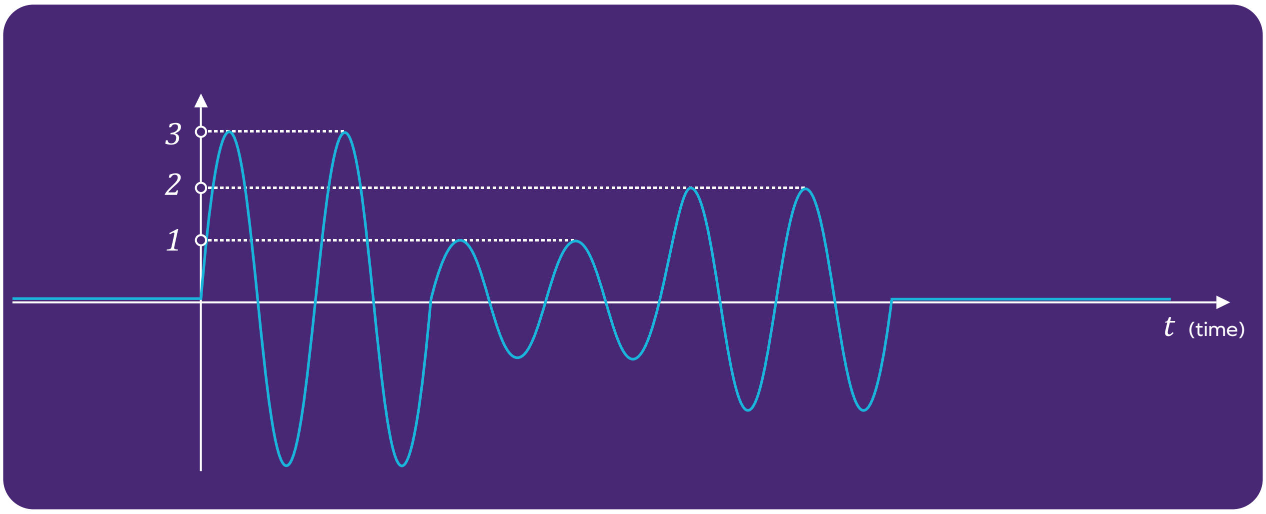
Lesson 6. Problem 2.

The acoustic wave diffracts at the edges of the window and propagates along the wall, and then diffracts again at the edge of the second window, as in Fig. 5.

As a result, it reaches the recipient in the adjacent room (of course, with a much smaller amplitude).

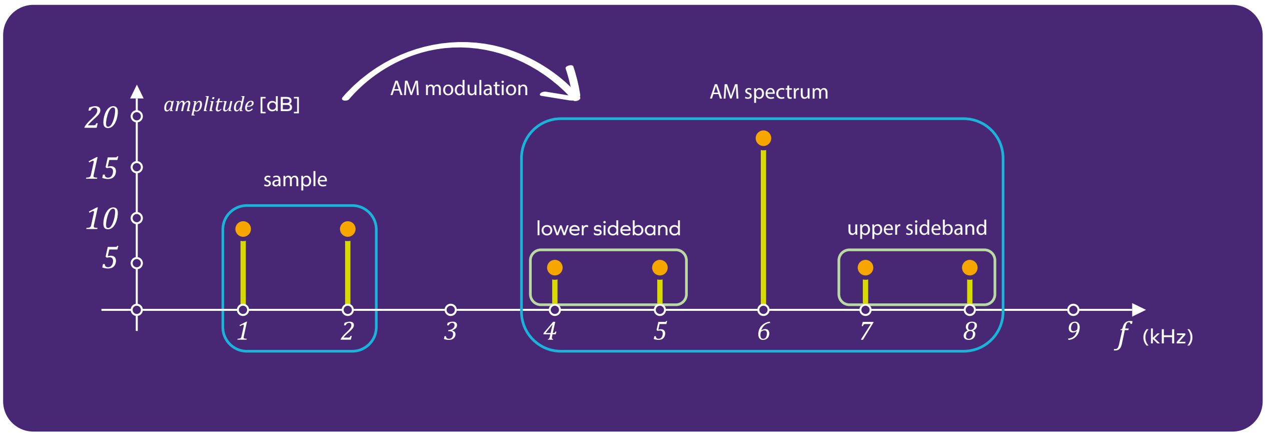


Lesson 7. Problem 1.



Each result of a die throw corresponds to one amplitude level. Formally, however, we should also consider level 0, at which no information is transmitted. Recognising a non-zero signal value allows us to state that the transmission of useful information has begun.

Lesson 7. Problem 2.

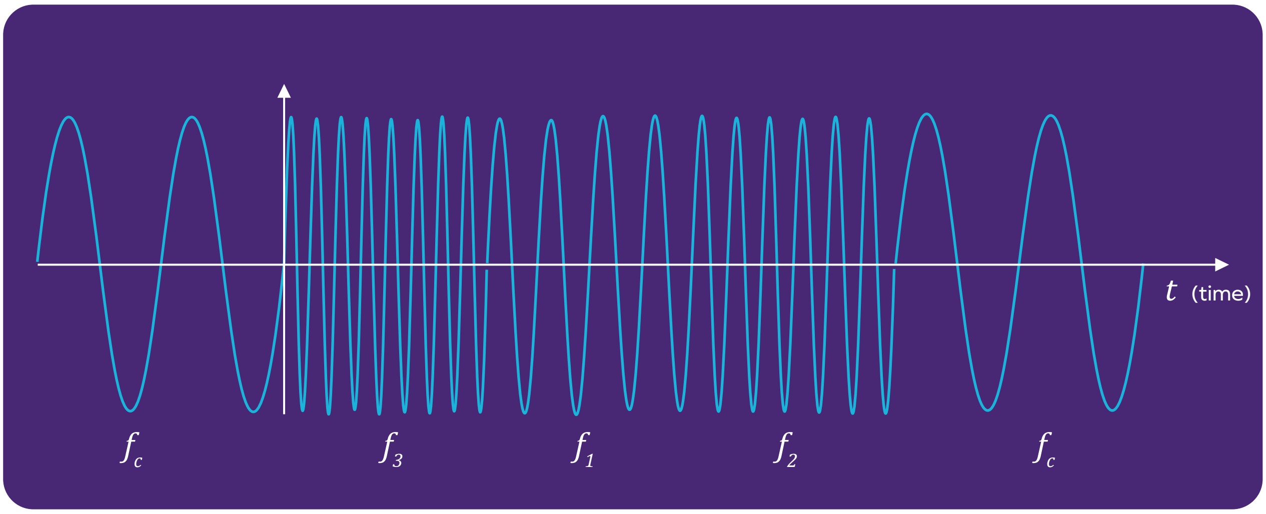


Both signal components appear in the spectrum of the modulated signal as a pair of bars arranged symmetrically with respect to the carrier frequency with the amplitude reduced by half. Their frequencies are calculated as:

$$\begin{cases} 6\text{ kHz} + 1\text{ kHz} = 7\text{ kHz} \\ 6\text{ kHz} - 1\text{ kHz} = 5\text{ kHz} \end{cases}$$

$$\begin{cases} 6\text{ kHz} + 2\text{ kHz} = 8\text{ kHz} \\ 6\text{ kHz} - 2\text{ kHz} = 4\text{ kHz} \end{cases}$$

Lesson 8. Problem 1.



In addition to the carrier frequency, we enter three different frequency values f_1, f_2 and f_3, which we assign to the respective dice result.

Lesson 8. Problem 2.

We check the modulation type by comparing the frequency deviation with the frequency of the modulating signal (here – the signal with the sample of the music piece). Because in our case:

$$\Delta f > f_m$$

and furthermore, because the frequency deviation significantly exceeds the frequency of the modulating signal, we are certainly dealing here with wideband modulation.

Using Carson's rule, we calculate the bandwidth B of the modulated signal:

$$B = 2(f_m + \Delta f) = 2(15 + 75) = 180 \text{ [kHz]}$$

The value of the carrier frequency is irrelevant to the questions asked.

Lesson 9. Problem 1.

When we solved this problem in Lesson 3 (without modulation) we came to the following antenna length estimates:

$$l = \frac{\lambda}{2} = \frac{c}{2f} = \frac{3 \cdot 10^8 \text{ m/s}}{2 \cdot 261.6 \text{ Hz}} \approx 5.73 \cdot 10^5 \text{ m} \approx 573 \text{ km!}$$

What does AM modulation change here? As we recall from Lesson 7, the spectrum of the signal obtained by modulating the carrier wave with a harmonic signal at frequency f contains two symmetrically arranged lines at frequencies $f_c - f$ and $f_c + f$. Since, according to the problem conditions, the carrier frequency is overwhelmingly greater than the frequency f of the modulating signal, we can assume with a very good accuracy that the entire spectrum of the signal is practically concentrated at frequency f_c.

So after modulation:

$$l = \frac{\lambda}{2} = \frac{3 \cdot 10^8 \text{ m/s}}{2 \cdot 2 \cdot 10^9 \text{ Hz}} = 7.5 \cdot 10^{-2} \text{ m} = 7.5 \text{ cm}$$

I think everyone will agree that an antenna of this length is slightly easier to construct.

Lesson 9. Problem 2.

Adding another bit will double the number of levels – each level in the "bit 3" column will be halved and the fourth bit will indicate which of the halves should be selected. So we will have 16 levels.

Since we have 2 levels for one bit, 4 for two, 8 for three, and 16 for four, and each subsequent bit will double the number of levels, it is easy to guess the general pattern:

$$\textit{num. of levels} = 2^N$$

Note that, for example, with N = 24 bits:

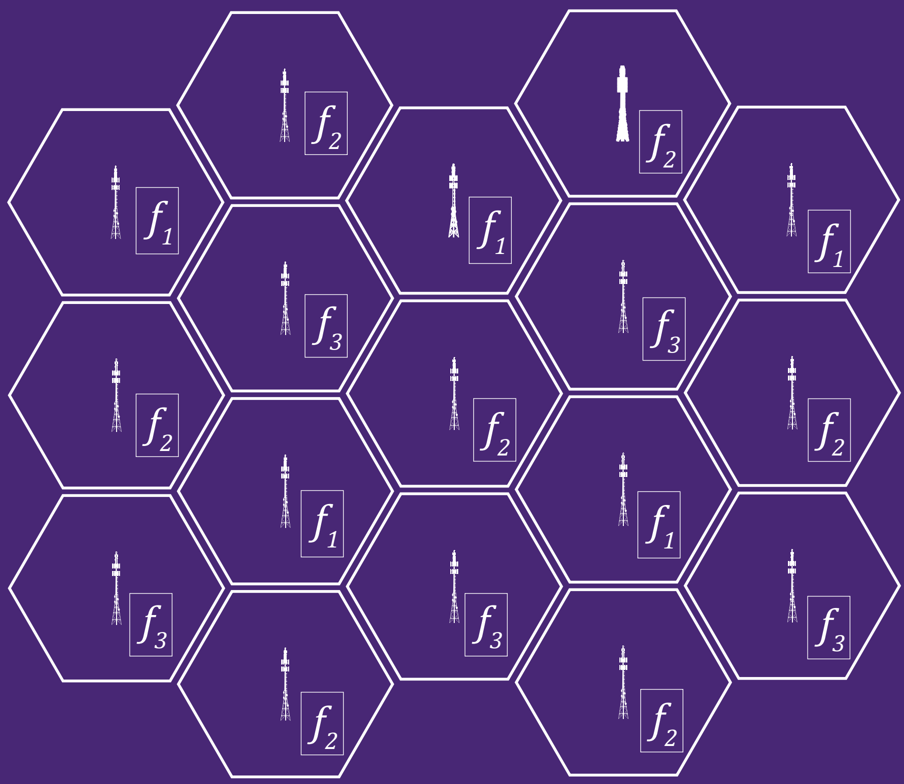
$$num.\ of\ levels = 2^{24} = 16777216$$

Samples quantised this densely may in practice be indistinguishable from the original.

Lesson 10. Problem 1.

The task cannot be accomplished using only two channels. If we assign channel f_1 to a given cell, then each of the neighbouring cells would have to be assigned channel f_2. Two adjacent neighbouring cells would then have the same channel assigned, which violates the conditions of the problem.

However, three channels are enough. The figure below shows one possible division.



Lesson 10. Problem 2.

We use the formula for the minimum frequency in the modulated signal frequency band:

$$f_{\min} = \frac{c}{2l} = \frac{3 \cdot 10^8\ \text{m/s}}{2 \cdot 1\ \text{m}} = 1.5 \cdot 10^8\ \text{Hz}$$

The carrier frequency is equal to the minimum frequency plus the width of the lower side-band of the modulated signal. In the case of AM modulation, the width of both sidebands is exactly equal to the bandwidth of the modulating signal. Therefore:

$$f_c = f_{\min} + B = 1.5 \cdot 10^8\ \text{Hz} + 5 \cdot 10^7\ \text{Hz} = 1.5 \cdot 10^8\ \text{Hz} + 0.5 \cdot 10^8\ \text{Hz} = 2 \cdot 10^8\ \text{Hz}$$

$$f_c = 200\ \text{MHz}$$

Therefore, a carrier frequency of at least $200\,\text{MHz}$ should be used.

Lesson 11. Problem 1.

Ionisation will occur when the energy of a photon with frequency f is at least equal to the ionisation energy. Let us transform the formula for photon energy:

$$E = hf \quad \rightarrow \quad f = \frac{E}{h}$$

Therefore:

$$f = \frac{E_i}{h} = \frac{2.2 \cdot 10^{-18}\,\text{J}}{6.6 \cdot 10^{-34}\,\text{J} \cdot \text{s}} = 3.3 \cdot 10^{15}\,\text{Hz}$$

This is the range of ultraviolet radiation.

Notes

Prefix	Symbol	Multiplier
exa	E	10^{18} = 1 000 000 000 000 000 000
peta	P	10^{15} = 1 000 000 000 000 000
tera	T	10^{12} = 1 000 000 000 000
giga	G	10^{9} = 1 000 000 000
mega	M	10^{6} = 1 000 000
kilo	k	10^{3} = 1 000
hecto	h	10^{2} = 100
deca	da	10^{1} = 10
deci	d	10^{-1} = 0,1
centi	c	10^{-2} = 0,01
mili	m	10^{-3} = 0,001
micro	μ	10^{-6} = 0,000 001
nano	n	10^{-9} = 0,000 000 001
pico	p	10^{-12} = 0,000 000 000 001
femto	f	10^{-15} = 0,000 000 000 000 001
atto	a	10^{-18} = 0,000 000 000 000 000 001

ISBN: 978-8-39161-468-6
9 788391 614686