Mitko Gorgiev

NEW THEORIES

1.On electromagnetism
2.On light and colors

© This book is a registered work. Printing or translation into other languages for commercial purposes without the author's permission is forbidden.

ON ELECTROMAGNETISM

The Cosmos, nature and life, which in fact are a unity, are built on principles. Hierarchically seen, the principles stand above the natural laws. The latter are consequences of the former; or, we could say that natural laws are projections of the principles. In various natural laws we can find the same principles. If man manages to apprehend and learn the latter, he will establish a solid foundation to serve as a reliable compass for his contemplation and understanding of nature, but also of himself as an inseparable part of it.

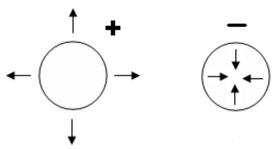
Contemporary science, despite its allegedly great achievements, has failed to perceive any of the important principles in the nature; hence, its chaos, incoherence, untruthfulness. Regardless of the praise and glory assigned to it by certain mass media, if one honestly asks oneself, one can hardly deny the impression that science, instead of making nature closer and more understandable, makes it more distant and incomprehensible.

This paper will begin with a presentation of a very important principle to immediately address what we have stated above.

We all breathe. Animals breathe too, and plants also breathe in some way. What is breathing? An initial observation could tell us that breathing is a constant expansion and reduction - a pulsation: when we inhale, our chests expand; when we exhale, they reduce in size.

The first two arithmetic operations a child learns in mathematics are addition and subtraction. If we add two to five, in reality it may mean that something that fills five volume units now expands by two and fills seven units. In turn, subtracting two out of seven means that seven contracts by two units and then fills five volume units. Therefore, we can label the expansion with the sign '+' and the contraction with the sign '–'. In this way, the act of inhaling we evaluate with plus, the act of exhaling with minus. In mathematics we can play through various computational tasks that often have nothing in common with our real world; but, if we want to stand on the ground of physical reality, we have to say that for each plus, a minus have to simultaneously arise somewhere. When we inhale, it means a plus in our chest at the expense of the surrounding atmosphere, which suffers a minus. This can be seen more clearly when we are inflating a balloon. The balloon is expanding – it means a plus arises within; but, at the same time, our chest is reducing - there is a minus in it. Let us take another example. A vacuum cleaner performs suction (–), but at the same time there is a vent on its plastic covering through which the air goes out (+). With a hair dryer we have the reverse.

Both the vacuum cleaner and the hair dryer are actually propellers (fans). The observer who stands in front of a fan will say that it blows, i.e. it exerts pressure (plus action), while the observer standing behind the fan will say that it suctions, i.e. it exerts depressure (minus action). In general, we could say that the plus denotes an action outwards, the minus denotes an action inwards (figure below).



Let us now consider a fan with only two blades. If the blades are completely flat, then, when the fan is turning, they will only cut the air like knives and there will be neither blowing nor suctioning. For this fan to function, it is necessary to twist the blades to a certain degree in the following way:

When this electrically driven fan, whose blades are twisted to the left, begins to turn to the right, then standing in front of it we will feel pressure, i.e. that it blows us (+); while, when it turns to the left, we will feel depressure, i.e. that it suctions us (–). If we twist the blades in the contrary direction (to the right), then at the turning of the fan to the right, we feel depressure (–), while at its turning to the left we feel pressure (+), or the reverse of the previous case[1].

Regarding the twist of the blades, the reader should think of wringing out a wet towel. If the right hand turns to the right, then we say the towel is twisted to the right; if it turns to the left, the towel is twisted to the left. The same applies to the fan blades.

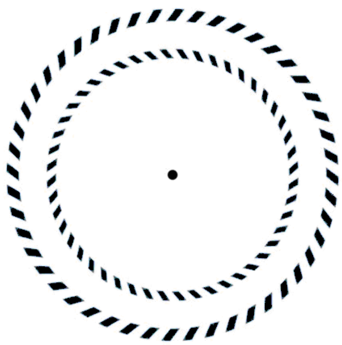
If we are to predict whether a fan will blow or suction if it starts turning to the right, first we need to look at the twist of the blades. If the blades are twisted to the left, i.e. like this '/' (fan blade viewed from above), then, when this blade turns to the right, it attacks the air first with its 'upper' part. Higher air pressure forms in front of this part than in front of the 'lower' part, so the air moves towards us, i.e. we are blown. What is important here to us for that what follows is to pay attention to the fact, that the blades of a fan (if it is a multi-bladed) which at a given moment are up blow us more on our left side, and those which at the same moment are down blow us more on our right side; that is, the flux is whirled rather than linear (flux ≈ flow).

We see that for an observer, whose position remains unchanged (i.e. standing in front of the fan the whole time), the following four cases may occur:

		The fan turns to the	
		left	right
The blades are twisted to the	left	–	+
	right	+	–

[1] The texts in frames should be footnotes, but since they often are pretty long, the author decided to put them in frames.

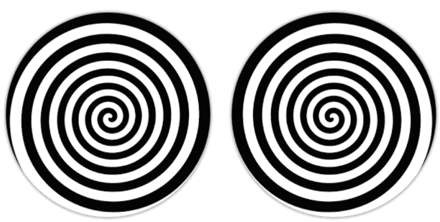
Look at the broken circles on the figure below. As you move the book closer and then farther away from you, looking constantly at the central point of the circles; or, if you place the book on a table, then lower and raise your head still looking at the central point (in this case the effect is stronger) - it seems as if the circles are turning in one direction upon coming closer, but in the contrary direction upon moving away. The experiment can also be carried out with only one circle. The dashes of the outer circle are 'twisted' to the left, those of the inner circle to the right. Upon moving the head closer, the outer circle turns to the left, the inner one to the right. When moving away, the opposite occurs. What does the movement away mean? It means nothing other than that the circle is 'blowing' at us, just as with the approaching of the head the circle draws us in. We see that we have here exactly the same conditions as in the previous case of the fans, so that the table above is valid here too.



The outer circle corresponds to the fan described in the text box above. When we move our head away, i.e., when it 'blows' at us, then it turns to the right.

The turning effect comes about only if the circles' dashes are somewhat slanted.

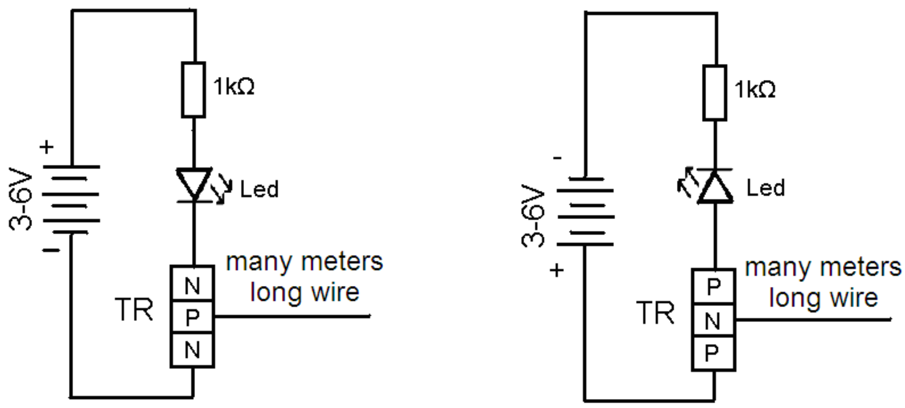
Let's consider these two spirals:



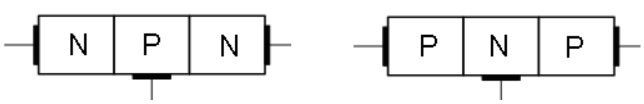
They differ only therein, that the first turns to the left (observing the black line from the center outwards), while the second to the right. If we make holes in the centers of the spirals, place them on spinning tops and turn the first one to the right looking unremittingly at it as it rotates, then we have the feeling as if it exerts pressure on us, i.e., as if it were pushing us (+), whereas when turning it to the left, we have the feeling as if it exerts depressure on us, i.e., as if it were pulling us in (–). With the second spiral occurs the same but in reverse order. We see that for an observer arise the same four cases from the previous table. (The spiral line turns to the left, the spinner turns to the right - then the spiral pushes us (+). This case corresponds to the fan's case described in the frame above. The same occurs with a screw or a household meat grinder. The thread of the screw turns to the left, and as the screw is turned to the right, it penetrates (+) into the wood.)

Let's go back to the fan again. Instead of an internal drive setting it in motion, it can also be turned by an outside force, as is the case with windmills. To see what happens here, we will make a simulation with a small fan (like those in computers), a hairdryer and a vacuum cleaner. If we bring an operating hair dryer close to the fan, it starts turning in one direction, and upon bringing an operating vacuum cleaner, it rotates in the contrary direction. The reverse happens if the fan blades are twisted in the contrary direction. There are four cases also here, two pluses and two minuses.

Everyone knows that something called ‘plus’ and ‘minus’ exists in both electricity and magnetism. We have all seen that ordinary 1.5V batteries have the mark (+) at the nipple and the mark (–) at the flat end. In magnetism, the two poles are called north and south, but they can be rightly called plus and minus. Which pole is here plus and which minus, we will see later. Do these plus and minus poles of electricity and magnetism show properties reminiscent of those we have just seen? To test this, we will carry out some experiments. Therefore we need two simple, almost identical electrical circuits, each with a battery, a resistor, an LED lamp and a transistor (figure below).



The circuits are independent of each other and do not differ absolutely in anything other than in polarity. What that means will be clear in a moment. The lamp serves as an indicator. When it lights up, it means that current is flowing through the circuit. The resistor (300Ω - 1kΩ) is solely in the service of the lamp, to prevent a stronger current causing damage. What remains is to briefly explain the element called transistor. Unlike the majority of elements in electrical technology that have two ends, i.e. two leads, this element has three ends, because internally it consists of three segments (figure below).

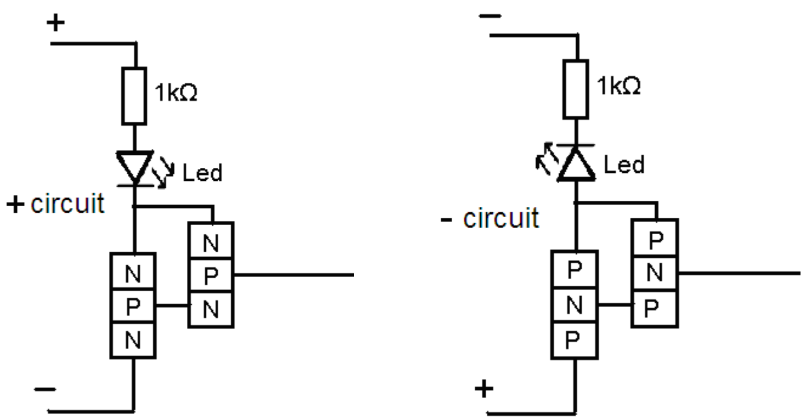


About the transistor we will now figuratively say only what we need for the experiments, but will explain more later. What matters most to us at the moment is its middle segment, which we will temporarily call a heart of the transistor. In the drawing we can see that the left transistor has a plus-heart (we call it + transistor), while the right one has a minus-heart (– transistor). We also see that the heart is a kind of bridge between the other two segments. In order to make the (+)transistor work, its heart should be actuated by (+)electricity. Thereby the bridge is established. If the heart is acted upon by (–)electricity, then it behaves indifferently. The reverse applies for the (–)transistor.

The lead from the heart we lengthen with a metal wire that is several or even many meters long, thus its end will be far from the circuit itself. Therefore we will be absolutely sure that the influence we are going to exert on the end of the wire affects only it and not any other element in the circuit. The end of the wire is loose, that is, not connected to anything.

However, in order to check what we have just said, that the (+)heart reacts only to plus-electricity, whereas the (–)heart to minus-electricity, we may hold the loose end of the wire with one hand and with the other hand first touch the (+)pole of the battery and then the (–)pole. We will see that the lamp lights up only in one of the two cases: in the (+)circuit [the circuit with the (+)transistor we will call (+)circuit] the lamp lights up only when we touch the (+)pole of its battery; and in the (–)circuit it lights up only when we touch the (–)pole of its battery. It is not advisable to connect the end of the wire directly to the (+)pole of the battery in the first case and to the (–)pole in the second for reason explained later in this paper.

What will be described now as an experiment can be done with these circuits' set-ups; however, for their greater sensitivity, in each of them we will add one more transistor [two (+)transistors in the first and two (–)transistors in the second circuit (figure below)]. It doesn't change anything except that we will save on effort needed to do the experiment, i.e., with less effort we will achieve a greater effect. If we still work with only one transistor per circuit the effect will be weaker, but it can be somewhat intensified if we attach the loose end of the wire to a wide metal plate - let's say a pot lid - and if we reduce the resistor's value to 100-200Ω.



Once the two circuits are ready, we take a vinyl gramophone record, a thin-walled glass, and a piece of woolen and silk fabric. We rub the vinyl plate with the woollen cloth and bring it close to the loose end of the wire of the (–)circuit. We will see that the LED will light up. It will also light up if we bring it close to the wire's loose end of the (+)circuit. But if we play a little bit, we will notice that there is a fundamental difference between the two cases: the LED in the (–)circuit lights up when we move the vinyl plate towards the wire, and the LED in the (+)circuit lights up when we move the plate away from the wire. Now, if we take the glass, rub it with the silk (or woollen) cloth, we will notice that the reverse happens: the LED in the (–)circuit lights up when move the glass away from the wire, and the LED in the (+)circuit lights up when we move it towards the wire. If we don't move the electrified objects, absolutely nothing happens. As mentioned before, this is quite feasible with only one transistor per circuit, yet the movements of the vinyl plate and the glass have to be much more energetic. But even in this experiment with two transistors per circuit we can notice that the faster we move the electrified objects, the stronger the lamps light up.

The cloths after the rubbing produce the reverse effect from the rubbed objects. Still, their effect dies out much faster than that of the vinyl plate and the glass.

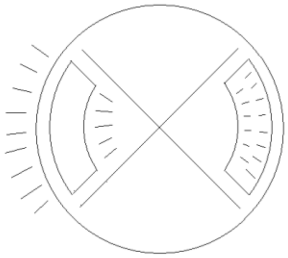
From this it becomes clear that vinyl and glass act completely opposite: vinyl stimulates the minus transistor by moving towards, and glass by moving away from the wire end; vinyl stimulates the plus transistor by moving away from the wire end, glass by moving towards it. We see that there are four cases here as well:

	towards	away
Glass	+	–
Vinyl	–	+

Let's carry out another experiment with these two circuits. We take a long isolated wire, wind it around a cylindrical object and then remove it, thereby obtaining a spiral-shaped wire. We connect one end of it to the two loose ends of the wires leading to the transistors of the (+) and (–)circuits (here, as before, we can do the experiment with only one circuit at a time). The other end of the spiral wire remains loose. Now we take a strong cylindrical neodymium magnet and quickly insert it, keep it inside, then quickly pull it out of the spiral. We notice that one lamp lights up upon inserting the magnet, while the other lights up upon pulling it out. As long as the magnet remains in the spiral, nothing happens. Then if we turn the magnet, insert it and pull it out with its opposite end ahead, the lamps light up in reverse order. They light up more strongly if the magnet is inserted and pulled out faster, if the spiral has more windings, if the magnet is larger and stronger, and if its diameter is not much smaller than that of the spiral. For this experiment to be carried out successfully as described here, we need a very strong magnet, many windings and very quick insertion and removal from the spiral. If these conditions are not fully met, then we don't leave one end of the spiral loose; instead, we connect it to the (–)pole of the battery in the (+)circuit, and to the (+)pole of the battery in the (–)circuit; thereby the experiment is carried out much more easily. We see that there are four cases also here.

That the positive electricity has the nature of expansion (blowing, pressure, explosion) and the negative electricity the nature of contraction (suction, depressure, implosion) can also be seen with naked eye. There is namely a whole group of so-called electrostatic generators, also called influence machines, similar but somewhat different from each other: Voss-, Toepler-, Holtz-, Bonetti-, and the most popular and widespread, the Wimshurst-machine. Since this machine is available to us, we will briefly describe it. The basic elements of this generator are two very close (about 5 mm), vertically placed glass or plastic circular plates (discs), metallic sectors of aluminum foil glued on the discs and two metal rods placed in the shape of the letter X, but one in front of the front disc (\), and the other behind the rear disc (/). Although the rods are on different sides, we will say that their X-shaped placement divides the discs into quarters, which we will call quadrants. We term the left and the right one horizontal quadrants, the upper and lower one vertical quadrants. The metal rods, which have the shape of the square bracket "]", end with metal brushes that gently scratch the plates (including the metallic sectors) when the discs rotate. They rotate in contrary directions; this is achieved so that during the manual rotation of the crank, the movement is conveyed by two belts, one of which (for the front disc) is in the form of the letter O, and the other (for the rear disc) is twisted in the shape of number 8. Electricity is generated solely by these elements. Therefore we consider the other parts of the machine as inessential. They are necessary only if we want to produce sparks from the already generated electricity; so, to prevent them from bothering us, we can even remove them. We will consider them later.

If we begin to rotate the discs by turning the crank to the right in a dark room (the most noticeable results can be seen at night in a room with a little exterior street light entering it), and if we do this for at least 10-15 seconds to let the eyes get used to the feeble light, we will notice that the horizontal quadrants emit a light flicker, whereas the vertical are completely dark. On turning the crank to the left the flicker relocates to the vertical quadrants, whereas the horizontal ones now remain dark. Looking even more attentively at the scene, we will notice *an essential qualitative difference* between what happens in the left and the right quadrant (i.e. the upper and the lower one when the crank is turned to the left). The flicker in one horizontal quadrant is directed from the metal sectors outwards, in the other one inwards. In other words, in the left quadrant *the metal sectors are dark* and the flickering light glows around them, but in the right quadrant *the metal sectors are illuminated* and around them it is dark (image below).



The sectors in the image are drawn as a whole, and not individually, because the light phenomenon appears as a whole; more precisely, as two wholes, one left and one right, and not individually in the sectors. We consider this as an ultimate proof of the essential difference between the plus and the minus of the electricity. We say that a proof is ultimate or final when we directly perceive the truth with our senses.

Without turning the generator, we move the wire of the (+)circuit with its loose end towards, and then away from one horizontal quadrant; then we do the same with the other quadrant. We can do also the reverse: move the generator with its left or right quadrant towards and away from the wire (as we did with the vinyl plate and the glass), which is basically the same. With the left quadrant, where the flicker was directed outwards, the lamp lights up only when the wire moves towards it; with the right quadrant the lamp lights up only when the wire moves away from it. If we do the same with the wire of the (–)circuit, then the reverse happens. We see that the (+)quadrant behaves like the glass, while the (–)quadrant like the vinyl plate.

Observing the described phenomena in the dark, we find that we don't actually need any detector to determine on which side is the plus-, and on which side the minus-electricity.

Whether the plus will appear in the left, and the minus in the right quadrant, or the reverse will happen, is left to chance. The plus and the minus may occasionally change sides.

From the history of electromagnetism it is known that Benjamin Franklin (1705-1790) is the man who was the first to introduce the terms "positive" and "negative", i.e. "plus/minus" in the field of electricity in the middle of the 18th century. Previously, the different types of electricity had been called "vitreous" (meaning "glass") and "resinous" (meaning "amber"), since the glass and the amber were the most often rubbed objects to produce the opposite electricities. At the time when Franklin gave his contribution, people had actually spoken of two types of electric fluids; however, Franklin argued that there is only one electric fluid, and the excess and the shortage of it in the objects he called "plus" and "minus". He said that bodies in normal condition have medium amounts of this fluid and are therefore neutral. When two objects are rubbed against each other, one allegedly transfers a part of its fluid to the other and thus the first becomes minus-, and the second object plus-electrified.

It remains a mystery how this type of thinking resulted in the glass electricity being called "plus", and the amber electricity "minus", although it has been recorded that Franklin is the man who assigned the plus to the glass, and the minus to the amber electricity. Still, this cannot be confirmed. In fact, on the basis of this kind of thinking (i.e., in the sense of "excess" and "shortage") it is impossible to reach a solution, which electricity is plus, and which minus.

Back then, as well as now, it is still considered to be arbitrary, a matter of convention; therefore it is said that there are no obstacles in naming the electricities the other way round. There exist even such opinions that this de facto should have been the reverse, because the convention is that the electric current through the wire flows from the plus to the minus pole (conventional current), while the electrons, which "appeared" almost one and a half century later and are allegedly the carriers of the electric current, were negatively charged and consequently moving in the contrary direction (electron current), so that with the reversed designation the irreconcilable contradiction, which has since set in motion an “eternal” discussion, would have been avoided. From what we have presented so far, but also from what we are going to expound further, it becomes clear that the polarity of electricities is well chosen and there is no need to change it.

If we look at an image of a magnet with its lines of force in any textbook, we will notice that the directional arrows point outwards at its north (N→) and inwards at its south pole (S←). This should mean that the north pole is the positive, the south pole is the negative. And here, too, it is said in science that it is arbitrary. But since this in no way can be arbitrary, it remains to determine which magnetic pole is actually plus and which minus.

First, let's clarify what is magnetic north pole and what is magnetic south pole. Since we need a compass for that, let us briefly explain what kind of instrument that is. The Earth is a giant magnet with two poles, North and South. They do not quite coincide, but are pretty near to the Earth’s geographic poles. Each magnet, separated from the Earth and free to move, strives to align itself with the giant magnet. To illustrate this, we take a bar magnet and place it on a flat piece of styrofoam. Then we let the styrofoam with the magnet float in a water tank. We will see that however we place it on the water surface, the styrofoam always turns so that the magnet has a strictly fixed direction. If we check the direction, we will find that it is north-south. But not only that. If we mark the styrofoam at one end of the magnet with a red dot, and at the other end with a blue dot, we will see that, in addition to the strict direction, the orientation is also strictly determined: the red and blue ends always place themselves in the same position - one color dot always points north, the other south. Our magnet can only move in a horizontal plane. If it can move in a three-dimensional space, we would see that it is positioning itself in a north-south direction, always tilting at a certain angle to the earth's surface, lowered northwards and raised in the south (this is referred to as the angle of inclination). This we can prove again in our water tank. We take a ball of styrofoam, insert a non-magnetized sewing needle through the center and place the styrofoam ball in water; if the needle does not tend toward to one side, this means that its center of gravity is exactly in the center of the ball. Next, without removing the needle from the ball, we magnetize the needle by touching it with a magnet. When we place the ball back in the water, we notice that the needle except that it turns to where it is in north-south direction, it also dips to the north (that is, it is pointing in our direction if we are facing north). This angle is approximately 45-50° in our latitude. It shows that the needle wants to unite with the magnetic north pole, because it is closer. The further north we go, the greater the angle. It is 90° at the magnetic north pole (the magnetic needle is erected vertically), but at the equator the angle is 0°. We see that the pole of the compass facing north is actually its south pole.

In order to determine which magnetic pole is plus, which minus, the author tried to detect some difference in the jagged shape which tiny iron filings create when they adhere to the north and south pole of the magnet. There seemed to be a difference therein, that at the one pole the spikes looked as if they were single-spiked, and at the other pole they appeared double-triple spiked, similar to the anterior and posterior part of the arrow shape. But it was so unclear and uncertain that one could not rely on it at all. The undoubted result came when the author once played with a ring magnet from a loudspeaker and accidentally came up with the thought of filling the middle of the ring with the iron powder. The poles of the ring magnet are its two flat surfaces. Once its middle was filled with the iron powder and then it was tapped to allow the powder to freely take its shape, the difference between the one and the other side became clearly visible. At the north pole a form of suction was evident, and at the south pole a form of blowing. Hence, the plus pole with an action outwards is the magnetic south pole of the Earth, and the minus pole with an action inwards is the magnetic north pole of the Earth.

The convention in force today is that the pole of the compass pointing north is called the north pole. Hence, the Earth's magnetic pole close to the Geographic North Pole is called the Magnetic South Pole of the Earth, and the one close to the Geographic South Pole is called the Magnetic North Pole. In this work, contrary to the convention, we name the pole of the compass facing north its south pole.

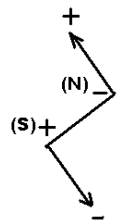
All the confusion actually disappears if the magnetic poles are simply called "plus" and "minus". The pole of the compass facing north is the plus magnetic pole. The compasses, whose needles have an arrow shape, give a very good picture of this because we term the front part of the arrow, which faces north, "plus", and the back part, "minus". (− >⟶ +)

> The front part of the arrow we consider as plus, the back as minus. The front part penetrates and exerts pressure, and the rear suctions, exerts depressure. This can be seen in the shape of the front and the back part of the arrow itself. It is the same with vehicles. Some cyclists risk their lives by driving directly behind large trucks to take advantage of the depressure in the slipstream and reach speeds up to 90-100 km/h on level roads. In videos that can be seen on Youtube, it seems like they are turning the pedals in a void, as if the truck pulls them, although they do not hold onto it.

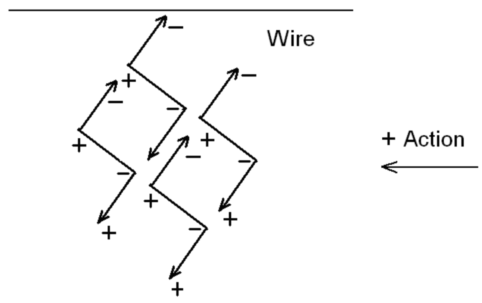
We will now introduce a theory which explains what happens in the wire leading to the heart of the transistor, as well as in every current-carrying wire. (At this moment, only on the basis of what has been so far presented, this may appear too early; in what follows, however, we will see other experiments and phenomena that contribute to the theory.) We call this theory "dynamistic" because it speaks of forces (δύναμις = force), in contrast to the current theory, which we call "materialistic" because it speaks of material particles, called electrons, supposedly moving through the metal wires. We call the theory dynamistic because in its basis lies *vibration of electromagnetic forces* (EM-forces). These forces are not of material nature. What was just said is well documented when we recall that the magnetic and the electric forces cannot be blocked by material bodies that are placed between the source of the force and the bodies they act on. For example, if we put a piece of iron near a magnet, the magnet will attract it even if we place a plastic, wooden or metal board between them. Likewise, radio waves penetrate walls without perforating them. This can be done only by something that is not of material nature. But even though they are immaterial, a material body is needed as their source. And in order to manifest themselves, they also need a suitable object to act upon; otherwise we would not be aware of their existence. Actually this is also the case with many other things in life. For example, the painter's abilities are immaterial, but a suitable physical body is necessary as their source. It can be only a human, not a monkey and not a wolf. Still, for these abilities to manifest themselves, they need a material body to act upon, and that is the artist's canvas.

Other terms necessary to understand the theory are "order" and "orientation". We can get a notion of these terms from several things: from magnetism, thread, wood, etc. When a magnet is brought in the vicinity of iron powder, the particles will adhere to the magnet with strictly *oriented order*. If we think of such a particle as a very small line segment, then it aligns itself not only in the same direction with the other particles, but also has a strict orientation of its plus and minus poles. We can imagine the particle as the smallest possible line segment and yet its properties will remain as described. In the thread we also have an *ordered* multiplicity of tiny little plant or animal fibers in the same spiral direction, except that there is no orientation here, that is, the fibers have no poles.

Now we introduce the electromagnetic force element, which is the basis of this theory. We put it this way:

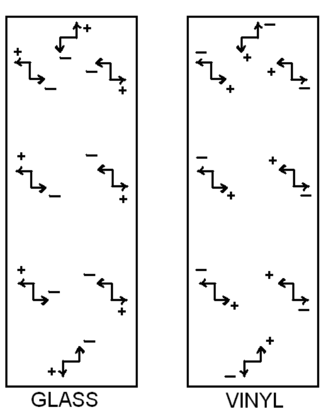


It has three segments. In the middle is the magnetic segment with its two poles, S(+) and N(–), and at its ends the electrical plus and minus segments, arranged at an angle of 90° to the magnetic segment. We have to imagine these elements in a huge multiplicity, evoked[2] by the movements of the aforementioned objects (vinyl, glass, magnet) and at the same time ordered according to a strict orientation of their electric and magnetic segments. (Figure below)



When there is no movement of the electrified objects towards or away from the wire, we cannot say that these elemental forces are still present in the wire only being chaotically distributed; rather, we should simply say that they are not there, or, to put it more correctly: they are latent. Here we can draw a parallel to the human. If we are offended, it can cause anger in us. Should we then say that the anger constantly exists in us but is, so to speak, only chaotically distributed throughout the body and therefore has no power, and at the moment of offense the chaos get ordered or concentrated and thus develops power? The author thinks that cannot be said. And as the anger of immaterial nature is, so is the hurtful word that has evoked it; for their manifestation, however, material bodies are necessary.

These forces appear not only in the wire, but also in the objects (vinyl and glass) that we rubbed with the woolen cloth. Their electrification can be represented as follows:



[2] For what we call here ‘evoke’ or ‘provoke’, in the current theory is used the Latin verb ‘inducere’, which means "bring in, lead in, introduce". From the explanations in this work, the reader will understand why we use the verbs ‘evoke’ or ‘provoke’. In Latin, they are ‘evocare’, ‘provocare’.

The plus segments of the elemental EM-forces in the glass are directed outwards from the object, the minus segments towards the interior of it, and therefore have no external effect. With the vinyl, it is the other way round.

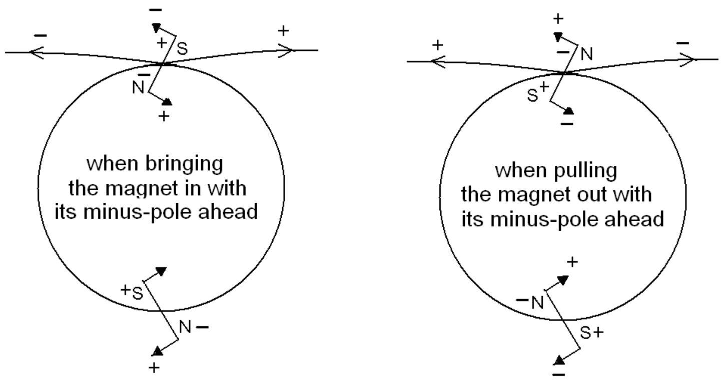
When we move the plus-electrified object towards the wire, its plus segments evoke the elemental EM-forces in the wire and at the same time arrange them in a *spirally whirled form*, doing this *<u>by acting on their plus segments.</u>* The ordered direction of the plus segments in the wire is the same as the direction of motion of the plus-electrified object. Like a gust of wind this effect propagates in a domino effect through the entire length of the wire. Hence, the plus segments of the EM-forces in the wire are oriented to the heart of the transistor, and if it is a plus heart, the lamp lights up. When we move the plus-electrified object away from the wire, it again evokes with its plus segments the EM-forces in the wire *<u>by acting on the same-named segments</u>*. Because this time the motion is in the opposite direction, the plus segments in the wire are oriented outwardly from its free end. This at the same time means that the minus-segments of the EM-forces are oriented towards the heart of the transistor. If this is a minus heart, then the lamp lights up. The aforesaid also applies to the processes with the movements of the minus-electrified object, only in this case the effects are reversed.

To explain what happens when we insert the magnet into or pull it out of the wire spiral, first we will present the following experiment. Through a thick copper or aluminum tube held vertically we drop a strong cylindrical magnet. We notice that the magnet in the tube falls much slower than out of it. We conclude that in the metal of the tube are evoked the EM-forces whose magnetic segments are so directed that they delay the fall of the magnet. This delay happens from two sides. While the magnet is falling in the tube, its lower end at every moment enters the remaining portion of the tube, and at the same time its upper end leaves the already traversed section. Both the one and the other effect must be slowing down the fall of the magnet; for, if the one slows it down and the other accelerates it, then these two effects would cancel each other out and the magnet would fall with the normal speed. Thus, when the magnet falls down with its minus pole ahead, in the part of the tube which is lower down it evokes the EM-forces whose magnetic segments are oriented with their minus poles upwards, therefore repelling the magnet (i.e. slowing it down); but, in the part of the tube that is higher up than the magnet (where its plus pole is), the EM-forces are evoked in the metal with their minus-poles of the magnetic segments oriented downwards, therefore attracting the magnet (i.e. slowing it down too).

That being said, we return now to the experiment with the spiral wire (which is a kind of a tube) and we can say that the insertion of the magnet into the spiral evokes the EM-forces in the wire *<u>by acting on their magnetic segments</u>*, which align themselves so that they try to prevent the entrance of the magnet; and that its pulling out evokes the EM-forces, which align themselves so that they try to prevent this, too. But the wire of our spiral is insulated with transparent lacquer, so the metal of the windings cannot touch directly; therefore, the magnetic and electric segments of the elemental forces in the spiral wire are not arranged so to form closed toroidal fluxes below the lower and above the upper end of the magnet (as we can describe the case of the copper tube), but they string together throughout the entire length of the spiral and continue onwards through the straight part of the wire. In other words, the magnetic spiral wind spreads through the entire path of the conductor. The insertion of the magnet with its minus-pole ahead will evoke the EM-forces with their minus-poles oriented outwards of the spiral, however, not at right angles with respect to the wire, but in the upper part pointing to our left, and in the lower part to our right side (thus, in the left part downwards and in the right part upwards)[3]. (Figure below)

---

[3] This direction of the EM-forces does not result from some properties of the wire metal, but from the inherent direction of the magnet's spin. We have here something similar to the push-and-spin mechanisms that we see in small toy carousels, in appropriately designed ashtrays, or spinning top toys.



This again means that the (+)E-segments will be directed to the right, and if the right end of the spiral (*the ends facing upwards*) goes into the plus and minus hearts of the plus and minus transistors, then the lamp in the plus-circuit lights up. Pulling the magnet out of the spiral will evoke the EM-forces with their (+)magnetic poles facing outwards, therefore the (–)E-segments will be directed to the right, so that the lamp in the minus-circuit lights up. If we now insert and pull out the magnet with its (+)pole ahead, the lamps light up in reverse order.

We can make a similar experiment with an analog or a digital am(pere)meter, but the result with an analog is more impressive. We connect the right end of the spiral to the red (+)input of the ammeter, the left end to the black (–)input. We place the range selector at the highest sensitivity position (mA or µA). The insertion of the magnet with its minus-pole ahead will cause a positive deflection of the pointer (i.e. to the right), while pulling the magnet out will cause a negative deflection (to the left). The same experiment made with a digital ammeter will show a minus sign in front of the digits upon pulling of the magnet out.

What we evoke in the wire with the oscillating movements of the vinyl plate, the glass and the magnet is nothing other than alternating current. But it can also be said that we produce what in digital electronics is called one (1) and zero (0). As we will see later, when the plus electricity is directed to the hearts of the transistors, it is a digital “1”, when the negative electricity is directed that way, it is a digital “0”. Or briefly: the (+)electricity is 1, the (–)electricity is 0.

Let us say something about the battery as a source of electric current. We can regard it as a container of dissolved agent (acid, base or salt) wherein two plates of different metals are partly immersed. Instead of one of the metal plates can also serve a graphite rod. As there is an exception to every rule, so it is in the electromagnetism. Coal is a special and unique case of a non-metal that is a conductor of electricity. Consequently, it is also a good electrode in a battery[4]. As the electricity was previously produced by rubbing a woolen cloth against vinyl or glass, so it also arises here through friction between the agent and the metal surface. The difference is that in the first case there was a mechanical, whereas here it is a chemical friction. But since two dry materials cannot rub chemically, the battery jar must contain water. In the first case everything had to be dry, in this case everything has to be well wet. Another difference is in the order of the EM-forces. In the first case, the entire vinyl record was electrified negatively. The positive segments of the elemental forces are directed inwards and therefore have no effect outwards. But that they also exist in the vinyl shows the fact that it can be electrified positively, however not by

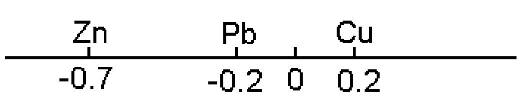
From where we draw the conclusion about the orientation of the magnetic segments in the upper part of the spiral to our left, and not right side, we will see later when we present other experiments.

[4] According to the etymological meaning of the words, what we describe here should not be called a battery, but a cell (an electrochemical cell). Battery is a multi-cell set. For example, the car battery is exactly that - a multi-cell set.

mechanical friction, but by something called influence. If we near a negatively electrified vinyl plate to a non-electrified plate, the latter becomes positively electrified. However, this positive electrification of the vinyl plate loses its effect very soon because this type of electrification is not its innate. The same applies to the glass, but with opposite polarity. We can see that certain materials have an inherent tendency to plus (glass, leather, nylon), but others to minus (amber, vinyl, silicone, teflon, PVC). With metals that are partly submerged in a dissolved agent, both polarities manifest simultaneously. The part of the metal plate outside the liquid is polarized in one sense, the immersed part in the opposite sense. The two metal plates of the battery can be imagined as two fans. The one that blows outside the liquid (positive electrode), that suctions inside it; the one that suctions outside the liquid (negative electrode), that blows inside it. When the electrodes are connected with a wire, a closed flux is created, which will circulate until one of the plates is consumed or until the agent has transformed into something else and thus has lost its aggressiveness. These two processes are interconnected and take place simultaneously. In the agent there is movement of matter. For comparison and better understanding, we use another natural phenomenon, the waves in the oceans. The water level rises and falls and we cannot say that the water moves toward the coast with the speed of the wave. However, an object floating on the water will move with each wave a little onwards until it eventually reaches the coast.

It is similar with the trembling of the EM-forces, that is to say, with each tremble the matter moves slightly through the solution. At the same time the negative electrode wears off, which is understandable, because its part inside the solution behaves as positive and *the movement is always from the positive to the negative*. If we open a "dead" carbon-zinc battery, we will see that the zinc jar (the negative pole) is completely dissolved, while the carbon rod (the positive pole) in the middle is quite good. The electroplating is also a result of movement of the matter on the electric waves. The positive electrode of carbon arc lamp connected to DC consumes itself, whereas the negative electrode remains pretty intact. Above we have said that certain non-metals have the tendency to plus-, others to minus electricity. For metals we cannot speak in the same sense of the tendency to plus or minus, because, as we see, the two polarities manifest here simultaneously. However, if we speak of the polarity of the metal plate outside the liquid, we can say that gold, silver, copper, platinum and carbon* among others have inherent plus tendency, while zinc, lead, aluminum, tungsten, iron etc. have minus tendency.

Let us now consider the following experiment. We need three metal plates, a copper, zinc and a lead plate and a small container with vinegar. First, we dip the copper and the zinc plate partly in the vinegar and connect them to a voltmeter: the copper plate to the (+)input, the zinc plate to the (–)input of the instrument. A voltage of about +0.93 V is measured. Then we do the same with the other two possible combinations. In the combination of copper and lead plate we measure about +0.43V when the copper plate is connected to the (+)input of the instrument, and in the combination of lead and zinc plate about +0.50V when the lead plate goes to the (+)input. In all cases the copper plate behaves like a (+)pole, the zinc plate like a (–)pole, only the lead plate in one case behaves like (–), in the other like (+)pole. On the horizontal axis of a graph we mark copper with +0.2, lead with –0.2 and zinc with –0.7 and see that the difference between copper and lead is 0.4, between copper and zinc 0.9 and between lead and zinc is 0.5, which is nearly consistent with the results of the voltage measurements. Therefore, we conclude that copper has the tendency to plus, lead the tendency to minus and zinc has a strong tendency to minus.



In the third case of the experiment (Pb-Zn), we saw that the lead plate behaved as a (+)pole. Here a reversal of the polarity of the lead plate takes place. The strong minus of the zinc plate reverses the weak minus of the lead plate converting it into plus. Let's compare that to the fans. Two fans with electric drives, one strong, the other weak, both suction, are moved towards each other. The stronger will then force the weaker to turn in the contrary direction, i.e. to behave as (+)blowing force in the interspace. If the strength of the stronger is 0.7 units and that of the weaker 0.2 units, then the strength of the stronger will decrease by 0.2 units in the interspace because part of its force it uses to overcome the power of the weaker. In the case of copper and zinc, there is another configuration because their forces add up (0.2 – (–0.7) = 0.2 + 0.7 = 0.9).

A reversal of polarity is also seen on magnets: a strong neodymium magnet reverses the polarity of a small and weak Al-Ni-Ko magnet, but in this case the reversal is permanent.

If we connect the poles of a 1.5 volt carbon-zinc battery with a good conducting wire (so-called short circuit), then the battery voltage will fall rapidly, but will remain at a certain value considerably less than the initial. If we break the circuit, the voltage soon returns to the initial value. Connecting the battery poles with a wire of high resistance, the voltage of the battery will not fall. We can compare this to a container filled with water. If an outlet is located in the lower part of its lateral wall, the force with which the water flows out (determined by the reach of its jet) depends only on the height of the water level, which also represents the pressure on the outflow point. The container is not large and, as the water flows out, the water level and thus the pressure decreases, which reduces the jet reach. Let's imagine that the water column is quite high and that, as we open the outflow, a steady flow of water into the container begins; this inflow is smaller than the outflow of water in the first moments. As the water level drops, so does the intensity of the outflow, thereby at a certain moment the inflow and the outflow equalize and from then on the level but also the jet remain unchanged. Let us imagine another situation. Instead of a water container we have a huge lake with sufficient inflow and an outlet in the lower part. The force of the water jet with unchanged size of the outlet will always be the same; it is so because the water drain has no, or only a negligible effect on the level of the lake.

Let's get back to the battery. The fact that we cannot get more than 1V from copper, zinc and an acidic solution is a good example of the limitations that exist in the real physical world, and that must always be kept in mind if one doesn’t want to lose touch with reality. Also, in our water example, we cannot construct a container that reaches sky-high even with the latest technology; the container will always have a limited height and at some point the water will overflow.

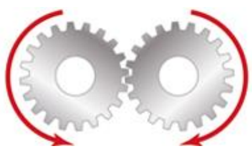
We compare the size of the outlet to the conductivity of the wire. The larger the outlet is, the greater the conductivity. The intensity of inflow into the container we compare with the friction intensity of the agent with the metal plates. If we make a battery with larger plates in a larger container, but also with a more aggressive agent (sulfuric acid, for example, is much more aggressive than vinegar), then we have more intensive friction (i.e. more intensive “inflow”) and the voltage to which the initial voltage will fall when connecting the battery poles with a good conducting wire will be higher compared to the case with the smaller plates and the less aggressive acid.[5] Hence, the current will be stronger too. With a huge battery,

---

[5] The described battery filled with vinegar is unusable for practical purposes because the vinegar is a very weak agent, i.e. the “inflow” in the battery is very low. No matter which bulb we connect to this battery, its voltage will immediately drop to zero and the bulb will not light up. Nevertheless, this battery shows the same voltage on the voltmeter as the one with the more aggressive agent, such as sulfuric acid. The measuring of the battery voltage we can compare with the measuring of the pressure at the lateral bottom point of an opaque container whose water level is unknown. At this point, we pierce a fine hole with a very thin needle and determine the water level in the tank according to the jet reach. Before that, through experiments with foreknown water levels (i.e. pressure values) we had set a scale of the jet reach at the same pinhole. With this scale we can also determine the pressure in the opaque tank, i.e. the unknown level when we make a hole with the same needle. Since the jet is very thin, the outflow won’t have a significant effect on the water level even of a very narrow tank. When we measure the voltage of a battery, we actually connect it to a great resistance (which we can think of as an extremely thin and long wire), measuring

the voltage will not drop at all (this is the case with the huge lake). To achieve a voltage greater than 1 volt, we need to connect two or more cells in series. So we can get higher voltages only in steps: 2, 3, 4 volts. The parallel connection of two cells (plus connected to plus, minus to minus) which deliver the same voltage is the same as if we would have increased the dimensions of the plates and the container of a single cell. Increasing the dimensions of a cell also affects its power output, inasmuch as with a larger cell a bulb will shine longer before it begins to dim. The unit of measurement for this value is called ampere-hour (Ah). If a battery has 10 amp-hours, it means that it can supply current of one amp for 10 hours. Of course, this is an ideal value, since the voltage and thus the current *gradually* drops, and so the luminosity. However, if the battery delivers half an amp for two hours, that's one ampere hour too.

When two gears mesh and move, then one turns in one direction and the other in the contrary direction.


A


B

Instead of the gears as in the image A, we imagine two helical gears with opposite teeth as in the image B. If both have their own drives, they complement each other, i.e. these two forces add up no matter whether they are of equal or of different intensity.

The connection of the electrical segments with the magnetic segment can be thought of in a similar way as the connection of three helical gears at angles of 90°. The two parallel gears have opposing teeth. One of these two attacks in the transverse gear overshot, the other on the other end undershot. This principle of acting of the forces can be encountered in many places, even when we turn a bottle screw closure: the thumb presses into one, the forefinger into the opposite direction of the closure.

When the electric current in the wire is evoked by the movement of the (+)electrified object, this excitation causes only the (+)E-segments to start running, which in turn sets the magnetic segments in motion and through them also the (–)E-segments. But a (+)E-segment also participates in the movement of the adjacent parallel element through its connection to its (–)E-segment (Figure below). Therefore, we can talk about series and parallel connections of the force elements. When we evoke the current through movement of the magnet, this sets in motion only the magnetic segments, which in turn move the plus and minus E-segments. In the part of the wire where the magnet has no direct effect, the action propagates mainly through the E-segments.

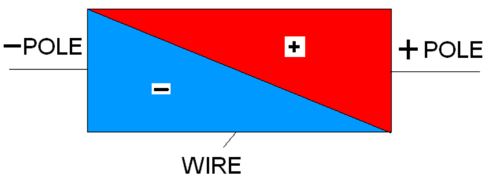
---

the strength of the current through it on the basis of the scale previously set. Although the bulb is also an extremely thin wire, its resistance is still low because of its shortness. Therefore, its connection to the vinegar battery is comparable to a larger hole in a narrow tank with water, in which there is also a very low inflow, resulting in the level falling immediately to zero.

Although the electromagnetic element is represented by straight lines, it is only a symbolic representation. Each line represents a flux, and the many elementary fluxes unify themselves in a single electromagnetic flux (principle of self-similarity).

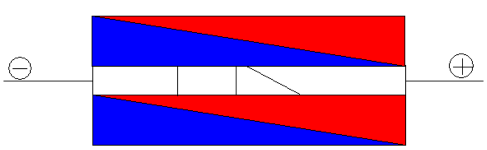
Probably it seems inconsistent that we draw the EM-force element so that the arrows of both E-segments point from their sources outwards on the one hand, while on the other hand we say that the one force has a suction effect. Hence, its arrow should have been drawn in the opposite direction. However, the direction of the arrows does not refer to whether the force acts from the source outwards or inwards, but rather to the effect of the action of both E-segments on the M-segment, that is, on its righting with respect to the wire line.

When the battery poles are connected with a wire, then current of equal magnitude flows through every cross-section of it. But not everywhere is the intensity of the plus- and minus-electricity equal. The plus is the strongest near the positive pole and, as we move away from it through the wire, its strength continuously decreases. The same applies to the minus, but starting from the other pole. Figuratively, we can represent it this way:



The quadrangle has the same width everywhere. This means that the current strength is the same throughout the length of the wire. The red field indicates the strength of the plus, the blue field the strength of the minus. Since the (+)E-forces have the contrary spin to the (–)E-forces, they complement each other in a similar sense as the helical gears mentioned above; therefore, the flux is the same through the entire length of the wire.

Two identical wires are connected in parallel to a battery. A current of equal strength flows through both of them. If we now connect them with a third wire at a right angle (figure below), no matter in which section of the wires, the instrument won't detect any current in the cross wire. However, if we fix this wire not at right angle, but slightly askew, we can measure a small current through it. The greater the inclination, the stronger the current. The current flows through the cross wire in our figure from bottom to top, because the plus/minus ratio down is greater than up. Let's assume that both rectangles are one centimeter wide. The lower end of the cross wire is connected at a position where the plus/minus ratio is 0.7 / 0.3 = 2.3, while the upper end of it is connected at a point where the ratio is 0, 5 / 0.5 = 1. Since 2.3>1, the current flows through the cross wire from bottom to top. The above applies in any case, no matter what kind the wires are. This arrangement in the electrical engineering is known as "Wheatstone bridge".



Therewith it should be born in mind that the indicated distribution of the intensities of the plus and the minus in the parallel wires will no longer persist after connecting the slanted wire.

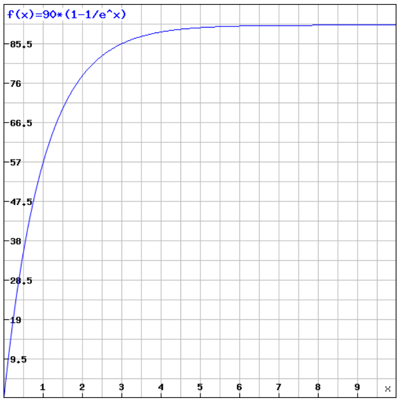
Now we take a compass and put it on a table. We ourselves are turned north, the compass is in front of us and it will be so in all that follows. On its housing we attach a copper wire parallel to the compass needle, i.e., in north-south direction. If we connect the ends of the wire to the poles of a standard carbon-zinc battery (3-6V) and the positive pole is nearer to us, we will see that the compass needle turns to the left (to the west) and stabilizes at an angle of 40-45° in respect to the original north-south direction. (It is desirable to use a compass whose needle is immersed in oil because it stabilizes very quickly in this way. This circuit should not be kept closed long because the battery drains rapidly). Whether we shift the wire to the left, to the right, or up – but still preserving the wire's parallelism to the needle – then nothing changes except the magnitude of the needle's deflection. If we place the wire under the compass, again parallel to the needle, we get the same deflection; but, now it is to the right (to the east). Exchanging the poles of the battery causes the same to happen, only the other way round. This shows that there is a whirling magnetic field around the wire. The established theory claims that the movement of electrons through the wire forms a ring-shaped magnetic field (like many rings strung on the wire) with a direction of exactly 90° to the wire line.

If we consider the previous experiments without carrying out more, we could think of four possibilities. First, the magnetic field is exactly in the direction the compass needle points (in our case at about 42°). Second, the magnetic field is between 42 and 90°. Third, the magnetic field is exactly at 90° (today's established postulate). Fourth, the magnetic field is oriented over 90° (this possibility comes down to the second). The whole problem of where the magnetic field is directed is due to the influence of the Earth's magnetic field. If we could "switch" it off for a moment, we would immediately see where the magnetic field around the wire is directed and there would be no problem. The first variant could be considered in case the magnetic field in the immediate vicinity of the wire is much stronger than the Earth's magnetic field, so that the force of the latter has only a negligible effect on the needle deflection. The other three variants come into consideration if the Earth's magnetic field is strong enough to influence the needle. In this case, the angle at which the needle deflects is a result of two magnetic fields: the magnetic field of the wire pulls to one direction, while the Earth's magnetic field pulls to the opposite direction; and the result, i.e. the angle at which the needle stabilizes, is somewhere between the directions of the both fields.

All of the variants mentioned, except for the variant asserted nowadays, lead back to a single conclusion: the magnetic wind through the wire is spiral-shaped. We affirm here the second variant: namely, that the twist of the magnetic spiral is not as large as the deflection of the needle shows, but larger; however, the Earth's magnetic field causes this twist to appear smaller in the deflection of the needle. To get a clearer idea of the shape of the magnetic wind in the wire, it is helpful to find a lace with colored parallel threads along it. If we twist this lace, we see that the parallel lines become spirals. The more we twist the lace, the greater the angle between the lines of the spirals and the line of the lace as a whole. There is something similar in the current-carrying wire. The stronger the current, the more twisted are the chains of the magnetic segments; on the other hand, the more aligned are the electrical segments in the direction of the wire line. At a stronger current, the magnetic spiral chains practically reach an angle of 90° relative to the wire, but never ideally. Likewise, the chains of the electrical segments practically reach the angle of 0°, but also never ideally.

The twisting of the magnetic forces in the wire occurs, like many other phenomena in inanimate nature, according to mathematical rules. The dependency that applies in this case can be expressed by the following formula:

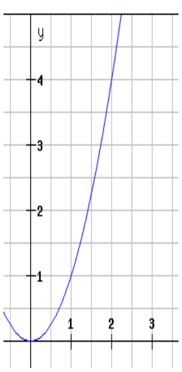
$y = 90° \, (1 - e^{-x})$ or $y = \pi/2 \, (1 - e^{-x})$



Before we explain what this formula means, let's make a comparison. Suppose we have an elastic rubber rod that a strong person can twist to the maximum it can withstand. At the beginning, he achieves a considerable effect with little effort. As he approaches the point of maximum twist, the effort required will increase steeply, and the visible effect will drop just as steeply. So now we have the opposite situation than at the beginning: great force, but little twisting. When it comes to the limit point, the two values tend to the extreme: extreme force, minimal twist. Something similar also happens in the wire. In the beginning, when the current starts to increase, a small increase causes a large twist of the magnetic forces: that is, a large increase in the angle of the spiral field with respect to the wire line. As the twist approaches its maximum (i.e. 90°), an extreme current increase causes only a minimal increase in the angle. We can imagine the curve in our graph as a hill being climbed. In the beginning it is very steep; as we move forward, it becomes flatter and flatter, but never ideally flat. At the beginning we cover a small distance in the horizontal, but manage a large distance in height. When we are almost at the top, we cover an enormous horizontal distance, but manage only a negligible height difference. The enormous distance that we cover in the upper "flat" part is actually the enormous increase of the electric current; in contrast to that, the extremely minimal gain in height is actually the approach of the magnetic spiral to a 90° angle. If in the above formula instead of x we put the current sign I and, instead of y, the angle α, it becomes:

$$\alpha = 90 * (1 - 1/e^I)$$

Let's briefly explain where this formula comes from; at the same time we will recall the math a little bit. As mentioned before, the physical quantities in nature depend on each other and the dependencies can be expressed through mathematical formulas. We have already mentioned one. Let's take a simpler and seemingly purely geometric formula: the area of the square as a function of the side length, $y = x^2$. (Figure below)



If we look at the graphical representation of this formula, we see that the curve at the beginning is little steep and later becomes steeper. The steepness of the curve at a given point is actually the tangent at that point. Since we are talking about mathematical functions of real physical quantities, we must also express the steepness in real numbers from $-\infty$ to $+\infty$ and not in degrees (0-90° or 0-360°), so the tangent in the total horizontal has a steepness of 0, while the tangent line which will be fully upright will have a steepness of $+\infty$; if the tangent is at an angle of 45° with the X-axis, then the steepness is 1. In mathematics this is called tangens, which is another word for tangent. If the steepness is 1 (=45°), we have covered the same distances in the horizontal and in the vertical, i.e. their ratio is 1. We see that the steepness is the mathematical relationship between the vertical and horizontal line segments if the curve at a given point continues to run as a straight line (tangent), i.e. the inner physical dependence that this curve represents suddenly ceases to apply. We get an idea of this when we turn a keychain on a string: the string suddenly breaks, and the hitherto effective physical laws cease to apply; thus, the keys fly in a straight line pursuing the tangent of the point when the forces stopped acting.

If we look at a tangent of a given point on the curve, we will see that the farther we move on the tangent from the point of contact, the farther we move away from the curve itself, and vice versa - the closer we approach the point of contact, the smaller is the difference between us and the curve. When we are very close to the point of contact, we can say that the difference is negligible, or that the tangent and the curve overlap on this very small line segment. Therefore, we can calculate the steepness at a given point if the values of the function at two near points - and that is the height difference on the vertical axis - are divided by the distance difference on the horizontal axis. If we denote the latter as Δx, then the height difference becomes $f(x + \Delta x) - f(x)$, and so the steepness (S) = $[f(x + \Delta x) - f(x)] / \Delta x$. Applying this to the function x^2, we obtain:

$$S=\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

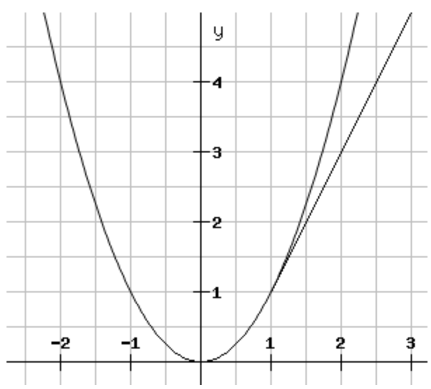
$$S=\frac{(x+\Delta x)^2-x^2}{\Delta x}$$

$$S=\frac{\cancel{x^2}+2x\Delta x+\Delta x^2-\cancel{x^2}}{\Delta x}$$

$$S=\frac{2x\Delta x+\Delta x^2}{\Delta x}$$

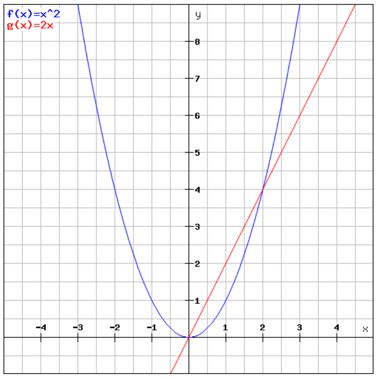
$$S=\frac{\cancel{\Delta x}(2x+\Delta x)}{\cancel{\Delta x}}$$

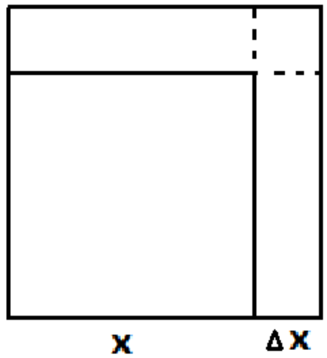
$$S=2x+\Delta x$$



As we have said, we assume Δx to be very small, practically zero, so only 2x remains for the steepness. We see that the steepness is a continuous function as well as the basic function, which one could have expected. This means that for any given point x we can now not only specify the value of the function, but also of the steepness at that point.

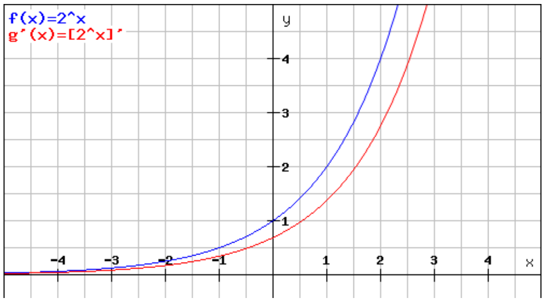
If we draw both functions together into a coordinate system, we see that the steepness is greater than the basic function from the beginning to the point x=2; they intersect in the point x=2, y=4, that is, they are equal, then the steepness is less than the basic function. What does this mean, and what does “2x” mean actually? Let's look at the figure below (right). There we see a smaller square, which has grown up a bit in the way that it is extended to the right and up. The increase can be divided into three sections, two rectangles and one smaller square. The rectangles have equal areas, x times Δx, and the small square is Δx^2. If we extend the initial square only slightly, then Δx is very small, so we can neglect Δx^2 in comparison to the areas of the two rectangles. Their total area is $2x\Delta x$. We see that this 2x means the expansion of the square in the direction of its two sides where the rectangles are attached. Even if we say that the square "pumps" itself on all four sides equally, there is absolutely no difference compared to that if we say the square expands on only two adjacent sides. If we look for the steepness of the function x^3 in the above-mentioned way, we get $3x^2$. The function x^3 is the cube content. The increase in cube content is due to the expansion on three (3) of its sides, which are squares (x^2).

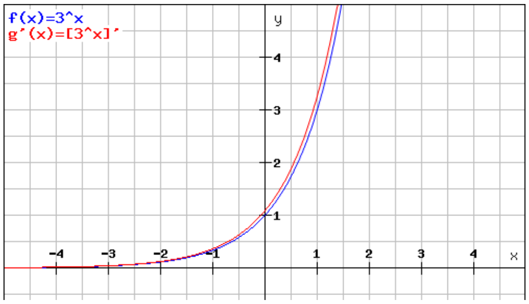




The fact that the steepness of the function x^2 is greater at the beginning and later smaller than the value of the function itself comes from the fact that for a very small square, even the small growth is larger than its initial area. For a large square, however, a small growth is much smaller than the initial area. In this context we can also make a comparison with life. The physical and mental changes of a child between the ages of one and two will be incomparably greater than those of an adult during the same period of one year. It's similar to spending money on different goods. An increase in price of one thousand euros when buying an apartment is easier to agree to than to the same increase when buying a computer.

Let's look at the two exponential functions 2^X and 3^X. If we draw their graphs along with their steepnesses, it will look like the figures below. The blue curve is the function, the red is its steepness. In the first function 2^X, we see that the steepness is always smaller than the function; but, in the case of the 3^X function, the steepness is always greater than the function. Somewhere between the numbers 2 and 3 there will be a number 2,…., which will form a function whose two curves, function and steepness, are congruent. This function can be found in the number 2.7182818 ... or 2.7182818^X . This is the so-called Euler number, denoted by the letter e.

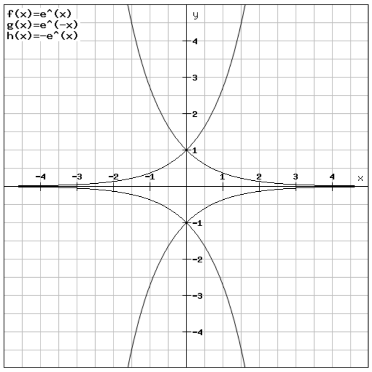




The function e^x is the only mathematical function that has a steepness identical to itself. Is that supposed to mean something? It means that the physical processes that unfold under this function have such inner constitution and dependence that the general increase is a true reflection of the increase in each elementary part of the physical process. As we have previously said about the square and the function x^2, we have seen that the steepness at the beginning is greater, but later smaller than the function, and that the overall increase can be reduced to the increase at two adjacent sides of the square. However, "e^{x}" is a different story: the wholeness is a true copy of each of its elements. In nature and in mathematics there is something called self-similarity. A beautiful example of this term in nature is Roman broccoli. Considering it, we see that the whole plant is a true reproduction of each of its parts. We observe the same in magnetism. However much we divide a magnet, its parts will again behave like the wholeness.

If we change the signs before “e” and its exponent “x”, i.e. if we use the four possible combinations (+ +), (+ −), (− +), (− −) we get four variants of the function e^x, namely e^x, e^{-x}, $-e^x$, $-e^{-x}$.

If we plot them in a single coordinate system, we get the following symmetric image:



Except for the $-e^x$ variant, we will see that all other variants can serve in the representation of the physical correlations of the electric current. We have used the variant (− −) or $-e^{-x}$ for the formula developed above. Since this function asymptotically approaches zero and we need approaching to 1, we will add 1 to this function and raise it by 1, giving “$-e^{-x}+1$” (or $1-e^{-x}$). The maximum value 1 of this function in our case is 90°, thus we multiply the entire expression by 90 and get $90 \cdot (1-e^{-x})$.

In the function e^x, the increase in growth is equal to the increase in increment; while in this function, $-e^{-x}$, the increase in growth equals the negative increment, or the decrease. The same speed with which the function grows, its increment decreases. In other words, the steepness of the function ($-e^{-x}$) is the function (e^{-x}).

For these reasons in the formula for the angle of the magnetic field with respect to the wire line we have used the exponential function with the Euler’s number. If we put in it the values 1,2,3,4,5,10 and 30 for the current strength, we get the following results:

I = 1, then	α = 56°
I = 2	α = 78°
I = 3	α = 85,5°
I = 4	α = 88,4°
I = 5	α = 89,4°
I = 10	α = 89,9959°
I = 30, then	α = 89,99999999991577°

We see that the angle very soon (0 to 5) reaches almost 90°. What are these values 1,2,3,4,5,10 and 30 for the current? They are not amperes, the unit of measurement for the current strength, as we will see soon.

Therefore let us say something about the measurement units. With the emergence of the modern empirical science, it has become necessary to abandon the old measures such as inch, foot, etc. and replace them with precise units. Considering the fact that everything that is measured in the last instance comes down to measuring space and time[6], there was a need first and foremost to have a precise measure of space, because the measure of time the humans possessed for a long time.

But nature is in perpetual motion and everlasting change and there is nothing permanent and unchanging on which man can rely in establishing the units of measure. Therefore, he must conceive them arbitrarily. In this way he has set the meter, namely by means of a metal rod which is kept in Paris. The length of the rod he has divided into 10, 100 and 1000 equal parts, which he calls decimeters, centimeters and millimeters. When the meter was set, mass came next. What helped here was water, easily available and easy to produce in pure form, since in the case of measures an important moment is that they should be reproducible everywhere and as easily as possible. A container in the form of a cube with the inner length of one decimeter is filled to the brim with pure water and placed on one side of a balance scale. On the other pan, pieces of pure metal were placed until the scale had ideally balanced. The container was afterwards completely emptied and put back on the scale. Then some of the metal pieces were removed until the scale had balanced again. People then said that the taken pieces of metal weigh as much as one dm^3 of water, a liter. They fused the pieces of metal in a single piece and called it kilogram.

We see that man is forced to set the measurements arbitrarily. When it was discovered at the end of the 18th century that two plates of different metals partly immersed in a container filled with diluted acid produce a force then called electromotive force (EMF) (now called voltage), and that different metals in different agents produce different intensities of this force, one had to find a unit of measure for that force. As a physical standard of measurement (etalon) the so-called Daniel cell was selected. The electromotive force this cell gives off was called a volt (1V). Due to its stability, the Weston cell was later chosen as a volt standard. Since these two cells differ by about 0.1V in voltage, the agreement of what should be considered as one volt was also changed.

But this is only the potential force. The current (i.e. the kinetic force) that this cell will give off depends on the material and on the dimensions of the wire by which the plates are connected. The reason for this dependence was called resistance. Thereby one had to set a unit of measurement for this quantity. The unit was established with the help of mercury because this metal was easy to produce in its purest form. In addition, it has a large specific resistance, thus with smaller amounts considerable resistance can be produced. A mercury-filled, circular glass tube with wires protruding from the ends was chosen as a standard for resistance (one ohm/1Ω). The tube was one meter long (later modified to a length of 1m and 6 cm after a settlement between the USA and Europe), and its area in cross-section is one millimeter squared. These two units of measurement, volt and ohm, served to define the strength of the current. When resistance of one ohm is connected to a source of one volt, then the current flowing in this circuit is one ampere (the voltage of the source must not fall in the course of this). We see that the unit of the current strength could have been a different one if a different kind of cell was chosen instead of the Weston cell; or if, instead of mercury, one had taken another metal, or the same but with other dimensions of the wire.

In order to match the above formula with the unit ampere, it is necessary to insert a constant before the "I". Its value must be significantly greater than one, because even at considerably lower currents than 1A, the angle is almost 90°.

---

[6] Measurements with all analog instruments are based on measuring space, whereas measurements with all digital instruments are based on measuring time. For example, at measuring temperature with a mercury thermometer we measure the space for which the mercury has expanded under the influence of heat. That we measure time with a digital thermometer (but also with a digital instrument for whatever physical quantity), we will see later.

Lower currents than 1A should not be considered as small currents, because 1A is a quite large unit. For example, when 30mA flow through the human body, it causes a severe electrical shock, which can be fatal to some individuals. This does not mean that if we interrupt a circuit in which 30mA flow and then touch each of the two ends with one hand, we would be killed. In this case, a significantly lower current than the 30mA would flow through our body because it is an additional, fairly great resistance in the circuit. As small currents one can consider those in the order of micro- and nanoamperes.

Let's clarify one more thing about magnetism. As we can easily notice, the poles of the same name repel and the different ones attract each other. If we bring two identical bar magnets together like the Roman numeral II with the plus poles up and the minus poles down, they will repel one another and we cannot bring them together. In order to bring them together in this arrangement, the poles must be reversed against each other. So we get "one" magnet with bipolar ends. The two magnets would never come together in this way if they could move freely, i.e. if they were not forced to come together like that. If allowed to move freely, they join together in a row so that the magnet becomes stronger. If we now take a bar magnet and slowly lower it towards the compass, so that the magnet and the needle are always parallel, then the needle will either not deflect or suddenly will turn by 180°. It makes another movement to connect in series, but it is imperceptible because the needle is fixed in the compass housing. So it is clear that the compass needle aligns itself reversed to the bar magnet. Based on this, we might think that the needle under the wire would settle reversed to the magnetic field of the wire. This is however not the case. The needle places itself in accordance with the magnetic field of the wire and thus strengthens the field. The following experiment shows this. If the minus of the battery is "up" and its plus is "down", the needle makes a deflection to the left, i.e. to the west. Now we break the circuit and coil the wire in few circles. When current flows through this coil, it behaves inside like a permanent magnet.[7]

The coil with its center now we bring near the lower end of the compass needle and then connect the coil to a battery so that the plus pole is on the right and the minus pole on the left. Let us recall how the first configuration was: the wire over the compass, plus "down", minus "up" - deflection to the left; now, in the second configuration, as if we turned the wire counter-clockwise by 90°, we had a part of it transformed into a coil and placed this coil "below" the compass needle. The upper segments of the turns of the coil have the same magnetic field under them as the straight wire in the first configuration, except that this time the magnetic field is rotated 90° counter-clockwise. If we now connect the ends of the coil to a battery, we will see that the needle turns 180°.

It follows that the upper segments of the turns have under themselves a magnetic field whose positive pole points in our direction (that is, to the south). It means that when the wire was straight and positioned in north-south direction, then the positive pole of its magnetic field under it was oriented to the left (that is, to the west).

In order to better understand the proof that the magnetic field in and around the wire is spiral-shaped, we will first mention something that is well known from everyday life. If two children sit on a roundabout seesaw with two seats, it will be much easier for us to turn them from outside than to do it from the very rotation axis. In the latter case, we may not be able to move them at all. We mention this because the experiment to be described now is to set the magnetic needle in motion by acting on its axis.

In order to show that the direction of the magnetic field of the current carrying wire differs from 90°, we will position the wire exactly 90° over the compass needle, i.e. in east-west direction. The needle and the

[7] When we make a coil from a straight wire, we actually cause an inversion. In the first case, we have a straight wire and a spiral magnetic field and, in the second, a spiral wire and a straight (*but twisted*) magnetic field.

wire must form an exact cross. But to demonstrate this, we need a new strong battery (let's say 9V), because we need – recalling the example with the roundabout seesaw – a great force to cause a movement of the needle from its very axis. There is something similar also here: the magnetic field will attack the needle at its axis point. The strong battery, in turn, short-circuited only with copper wire, will give off a strong current and the magnetic field of this current is practically at 90° with respect to the wire. The circumstances are very tricky, which is why we have to be very precise. The procedure consists of two important points. First, the wire must be precisely at right angle to the needle. Second, the wire should begin to bend on both sides some distance from the needle for its ends to meet in the battery. These curves must be far enough from the needle. If this is not the case, the magnetic field of the curves will affect the needle, which will make the experiment inaccurate. When this experimental setup is ready, we close the circuit. If the positive pole of the battery is left, then the needle twitches very weakly but still noticeably to the right (to the east); and, if the positive pole is right, then the needle moves to the left - at the beginning very slowly and with great difficulty, and later accelerating to settle at an angle greater than 90°. The Earth's magnetic field prevents it from turning further. What is indicative in this experiment is that the tiny twitch when the plus is left is always to the east; while, with plus right, the big deflection is always to the west. We consider this to be a sufficient proof that the magnetic field around the wire is not at right angle to the wire line. If it were at right angle, then the compass needle should not move in either case.

The simple experiment we are about to describe now is of essential importance to understanding of electric current, but is not mentioned in the science of electromagnetism anywhere. Only in the text of Hans Christian Oersted after the discovery of the magnetic effect of the current carrying wire, a brief remark is made, which can be understood in this direction (we will quote it further down).

Let us take two pieces of wire of equal thickness, but of different metals, which have an enormous difference in their specific resistances – say, one of copper, the other of kanthal[8]. If we connect the two wires in a series and position this "one" wire in north-south direction, place a compass under each piece, then connect the ends to a battery, we will see that the deflections of the two needles are different. The needle under the copper wire deflects more than that under the kanthal wire.

On the other hand, when we connect two pieces of kanthal wire of different cross sectional areas and then place a compass under each one, we notice again that each of the two needles makes a different deflection. Under the thinner wire piece we see a bigger deflection than under the thicker one. Moreover, one also feels a greater warming of the thinner piece.

The explanation of these experiments is as follows. If we have several wire pieces of different metals and of same thickness connected in series, then the deflection of the magnetic needle is larger next to the wires of electrically more elastic metals[9] (silver, copper, aluminum), because the more elastic material has a smaller resistance. In such material the electric segments tend more easily to the wire line, which on the other hand means that the magnetic spiral is pushed closer. As the elasticity of the metal becomes lower, so the electric segments exert a greater resistance to being inclined to the wire line, which in turn means that the magnetic spiral is further pulled apart and therefore the deflection of the magnetic needle is smaller. Now that the electric spiral is more compacted, it means that the electric flux has a longer path to go through this than through a wire segment of a more elastic metal connected in series. Given that the electric flux is the same throughout the circuit, it follows that in the less elastic metal the trembling of the E-forces must be faster in order to keep pace in the transfer of the flux with the more elastic metal. The

[8] Kanthal is a metal alloy which is electrically high resistive.

[9] As it is so natural to speak about elasticity of materials in a mechanical sense, so it is to speak about their elasticity in an electrical sense.

faster vibration of E-forces leads to more frequent friction between the plus and minus E-segments, resulting in a greater release of heat.

When several wire pieces of a same metal of varying thicknesses are connected in series, then in the thinnest piece of wire the electric wind will flow fastest (similar to how the air flows fastest in the narrowest pipe section in a series of differently wide pipes); therefore the greatest leaning of the electric segments to the wire line takes place here, which at the same time means the greatest righting of the magnetic segments in respect to the wire line (that is, the most compacted magnetic spiral), manifesting itself in the largest deflection of the magnetic needle. Since the vibration of the EM-forces is the fastest in the thinnest wire in order to keep up with the flux transfer in the thicker wires, the greatest heat is also generated here and, in extreme case, light too.

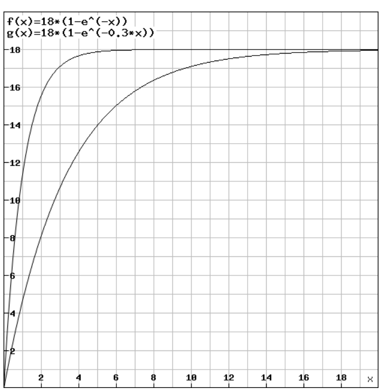
Considering that the electric flux through all the wires of a closed series circuit is the same, it follows that the magnetic flux through them, or the strength of the magnetic field around them, must be the same everywhere. As we know, physical forces can be represented by vectors. The direction of the vector represents the direction of the force; its length represents the strength of the force. In our case, if we should represent the magnetic field around the various wires of a circuit by vectors, they will be everywhere with equal length, but not everywhere with the same direction.

Since the cross sectional area of the wire and the specific resistance of the material act inversely proportional to the angle of the magnetic spiral, we will write these two under the fraction bar of the exponent, in which previously we had only the current; thus, we obtain the following formula for the angle α:

α = 90° (1-e^(-k*I/ρ*S))

The letter k stands for a constant that adjusts the formula with the currently accepted units of measurement, and which in our estimation should be significantly greater than 1. This formula calculates the angle of the magnetic spiral in each part of a series circuit.

If we plot the two following exponential functions α = 1- e^(- a·I) and α = 1- e^(-b·I), substituting for a = 1 and for b = 0.3, we get the following figure:



The number a=1 is an arbitrary chosen value of "k/ρ·S" for one of two wires connected in series, b =0.3 is such a value for the other wire. From the graph we see that with increasing current the angles of the magnetic spiral in both wires become practically equal. If the resistivity ρ and the cross sectional area S of the second wire increase, the lower curve will become even lower and even farther along the X-axis it will approach the maximum value, that is, at greater current it will approach the first curve. Accordingly, there will be practically (but not factually) no difference between the angles of the magnetic fields around the two wires at stronger currents. Such differences should be noticeable at lower currents. But at lower currents, however, not only the angle but also the strength of the magnetic field decreases, so that then it has no strength to overcome the inertia of the compass needle and its tie to the Earth's magnetic field to cause its deflection. However, when the strength of the magnetic field is great enough to cause a deflection of the needle, then the angle is already very close to 90°. Therefore, with two wires connected in series, which don't have a big difference in their resistivities, the difference in the angles of the magnetic needles cannot be noticed with the naked eye, but only with precise instruments. For those that have an immense difference in their resistivities, the difference in the angles is perceptible to the naked eye.

The different deflections of the magnetic needles over different pieces of wire of a closed electric circuit could be regarded as a manifestation of Bernoulli's principle in the case of electric flux.

As we have already mentioned, there are indications in Oersted's work that point in this direction, namely, in his writing of 21 July 1820 with the discovery of the phenomenon that a magnetic field builds up around a current-carrying wire. Oersted, however, mentions this only very fleetingly and since that time it appears nowhere else. He writes: "The uniting conductor may consist of several wires, or metallic ribbons, connected together. <u>The nature of the metal does not alter the effect, but merely the quantity.</u> Wires of platinum, gold, silver, brass, iron, ribbons of lead and tin, mass of mercury, were employed with equal success."

The underlining is from the author of this work. The effect of which Oersted speaks, evident from the exposition before, is the deflection of the compass needle. This means that for different wires connected in series everything is the same ("equal success"); only the quantity, that is, the angle of deflection changes.

The external manifestation of the magnetic wind in the wire is the magnetic field around the wire, the external manifestation of the electric wind in the wire is the heat and possibly the light around the wire. Through the external manifestation we can measure the intensity of what happens inside, that is, the current strength. Light does not develop in every electrical circuit and measurements through heat generation are difficult to carry out. What remains is magnetism. In practice, with analog instruments, the current is measured almost exclusively through the strength of the magnetic field.

Although we sometimes use the phrase "current flows", there is nevertheless nothing material flowing through the wire; rather, it blows through it an immaterial electric and magnetic swirling wind, **both from the positive to the negative pole of the battery; namely, the magnetic wind in clockwise- and electric wind in counter-clockwise direction.** Motion of matter on the electric waves exists only inside the battery. We speak of waves because every single vibration of the forces is a wave. The vibrations come therefrom, that the electrical resistance of the material relentlessly opposes the ordering of the EM-forces, but the power of the electric source restores the order anew. We can think of the electromagnetic wind as of something that comes in rushes, similar to the flux caused by a fan comes in rushes. Every single blade of the fan grasps and moves onwards a certain portion of air. Because of the high frequency of these rushes, it appears to us as a continuous flux.

For current to flow through a wire, no closed circuit is needed. We have seen this at the beginning of this work with the free end wires of the so-called plus and minus electric circuits. We can see the same when we use the phase tester (one-contact neon test light) to determine which wire is the phase. The phase tester is a series connection of a high value resistor (hundreds of kilo-ohms) and a small neon lamp. It is sufficient to touch the end of the phase tester with a finger when its other end is touched to the phase to make the lamp light up. Why is this necessary? We will use a comparison to explain it. If a turned on hair dryer or vacuum cleaner is brought close to a fan that does not have its own drive, then the fan is turning. But if we attach the fan to a wall, and if we bring the hair dryer or the vacuum cleaner close to it, it will not turn. The reason is that behind the fan there is no free space *filled with air* in which the flux can spread or from where it can suction. The same happens with the electric flux. Touching the end of the phase tester, we are becoming the "air space", actually the body with sufficient electrical conductivity whereto the flux can spread or wherefrom it can suction, and consequently the lamp lights up. In German it is called Masse - material conductive mass (English: ground).

Interesting is another experiment with the described plus and minus transistor circuits, which, because of its importance, will be analyzed also later when we will talk about the semiconductors. Instead of a long wire from the heart of the plus circuit, this time with a long wire we extend either of the two other leads of the plus transistor. From the heart of the transistor there is no wire at all. Apart from that, the experimental setup remains as in the basic experiment from the beginning of this paper. If we move an electrified glass toward or move it away from the wire's free end, which is far from the circuit itself, then nothing happens. But if we now connect to the lead of the heart a wire piece of 10-15cm or more (the other end of the piece is free) and again act with the electrified glass on the end of the first wire, then the lamp lights up. Now the situation is opposite to the one in the basic experiment. In that experiment the lamp lit up when we moved the glass toward the heart wire; here it lights up when we move it away from the non-heart wire. But this could not happen without the small piece of wire from the heart of the transistor. This short wire is actually mass (ground), material conductive “space" in which the flux - evoked in the long wire by the movement of the glass – can spread and thus more powerfully excite the heart from the other side, passing first through the N-part of the transistor.

The largest conductive body, i.e. mass, is the Earth. This mass is most commonly used for grounding electrical equipment, lightning rods, antennas, long-distance power lines, etc. But it is a good mass only if it is sufficiently moist. In long dry summer months it is not a very good mass. The Gobi Desert, the driest place on Earth, would be a bad mass too.

Although the electric flux does not need a closed circuit, still the flux in a closed circuit is much stronger because the blowing of the positive from one side meets the suction of the negative from the other, thus multiplying the effect many times. The order of the aforesaid can also be reversed.

If we connect two light bulbs, one 100W and the other 60W, parallel to 220V-240V, the first shines more brightly than the second. If we connect them in series, then the 60W bulb shines more brightly than the 100W bulb. In the second case both shine much weaker than in the first. Ordinary light bulbs consist of very thin tungsten wire; that is, they are resistors with certain resistance. In principle, any wire of such dimensions would shine, but would immediately melt and thus break the circuit. Tungsten has the highest melting point among metals (3400°C) and will not melt even at temperatures as high as those in the wire (over 2000°C).

We see opposite situations here. In the first case, there is equal voltage applied to both bulbs but different currents through them. In the second case, the same current flows through both bulbs, but the voltages at their ends are different.

The energy consumed in the first case is obviously greater with the 100W bulb and smaller with the 60W; in the second case, it is reversed. The total energy consumed in the first case is significantly greater than that in the second.

The 100W light bulb can be thought of as a shorter wire and the 60W as a longer wire, both of same thickness (we can think of the 100W bulb as a thicker and the 60W as a thinner wire, both of the same length; but, here we stick to the first conception).

From everyday life we know that, when transporting a weight from A to B, physical work is done. The further apart places A and B are, the greater the work done. But the work done also increases the heavier the weight is. Work is therefore a product of the weight of the object (the force) and the distance over which it has been transported. If we denote the work with the letter A, the weight with F and the distance with s, then A = F * s, or A = m * g * s.

Just as by the air flow through a pipe we can speak of an amount of air passed through it (expressed in m^3) due to the action of a propeller, so can we speak of a quantity of electricity passed through a wire thanks to the action of a battery. If we denote this quantity by the letter Q, then this Q when divided by the time (Q/t) will result in the magnitude of the electric current, just as m^3/s would give us the magnitude of the air flow through the pipe. When the electricity has flowed through the wire a certain length, then the work done, i.e. the energy consumed, is proportionally dependent on the length. However, since the electric flux usually encounters different resistances in the various wires that make up the circuit (as opposed to the transport of weight where the resistance, i.e. the gravity, is the same everywhere), the length of the path is of no significance, whereas the voltage difference between two given points of the circuit is crucial.

Therefore, the term for the consumed energy will be Q*U. Above we used the formula F*s for the energy consumed, where F is the product of the quantity (mass m) and the resistance (gravitation g) against which the quantity moves; the distance (s) was expressed separately. In the case of the current, the quantity Q is noted separately, whereas the resistance against which the flux moves and the path it travels are summarized in U, because the greater the voltage difference between two points in the circuit, the greater the resistance. On the other hand, since we cannot measure the quantity Q, but the current I, which is Q/t (hence Q=I*t), we write I*U*t instead of Q*U, and can thus determine the energy consumed in a certain resistor (or the work done for the transfer of the flux through it).

If two workers carry equal parcels over a certain distance AB, but the first moves faster than the other, then the former will carry more parcels in a certain time. He does more work in a unit of time. We say that his effect (power) is greater. We can also put it another way: they carry the parcels at the same speed, but the first carries heavier parcels: again, the former achieves a greater effect. This physical quantity, i.e. the effect is F * s / t.

As we said above, we can think of the light bulb of 100W as a shorter wire, that of 60W as a longer wire, both of equal thickness. If both are connected in parallel to 220-240V, then stronger current flows through the shorter wire, so that the power I * U is greater here (although the path is shorter, however, a significantly bigger "load" is transferred here). When both wires are connected in series, then the same current flows through both bulbs, so that the power U * I is of course greater in the longer wire (the "load" travels a longer path).

In the case of the parallel connection, the current that the source gives off is significantly greater than in the case of series connection (in the first case both lamps shone much brighter). Therefore, we characterize the parallel as plus connection (expansion, intensification of effect) and the series as minus connection (contraction, attenuation of effect).

These two types of connection we can imagine as a thicker, cylindrical piece of plasticine, which becomes thinner and longer as we roll it between our hands, and then we knead it back into the original shape. During the first process we make a series connection out of a parallel. When we knead it back, we make a parallel connection out of a series. If we imagine the two processes as incessant pulsing, then the first action is minus, the second plus. We equate the thickening with the parallel connection because it makes no difference from the point of view of the current source, whether we connect two identical wires in parallel or melt them together in a wire of the same length and greater thickness.

The resistance of a wire can be determined by the following formula: $R=\rho \cdot l/S$, where ρ is the resistivity of the material, l the length and S the area of its cross section. The lowest specific resistance has silver, the highest mercury. Because mercury was used to determine the measurement unit of resistance, its resistivity is also the reference point 1 with respect to which the resistivities of other materials are determined (it is actually slightly lower than 1 because of those added 6cm). The value ρ for materials which have a higher resistivity than mercury (that is, alloys) is greater than 1, for the materials with lower resistivity the value is less than 1. From the formula we can see that the longer the wire is, the greater the resistance; the thicker the wire, the smaller the resistance. We can also say that in the length l we have the series connection and in the cross section area S the parallel connection.

Instead of speaking about the resistance of the wire, which we have determined as minus quantity, we can also speak of the conductance of the wire, which will be a plus quantity. The conductance is G=1/R.

With several wires connected in series (minus connection), their resistances (minus quantities) add up; with several wires connected in parallel (plus connection), their conductances (plus quantities) add up.

$R=R_1+R_2+R_3+... R_n$ series connection (minus)

$G=G_1+G_2+G_3+...G_n$ parallel connection (plus)

CAPACITOR

A very important component in electrical engineering is the capacitor. We can represent it principally as two metal plates between which there is air, ceramics, glass, various plastics, paper or other materials commonly called insulator, dielectric or nonmetal. These materials are actually those where the EM-forces can be evoked by mechanical friction. But as we have already seen, in mechanical friction, the two materials are electrified opposite. The dielectric of the capacitor however has the two types of electricity at the same time on the opposite sides. We can test this by placing a vinyl plate between two metal plates and connecting them to the two outputs (collectors) of the Wimshurst generator. We turn the generator for some time and then remove the vinyl record, after which we bring it closer to the electricity detector and then away from it, first with one side and then with the other side. We notice that the two sides show opposite polarity.

In the capacitor occurs twisting and, simultaneously, tension of the EM-forces. The twisting means electric current through it in one direction. We can compare this in the mechanical sense with the twisting of a rubber rod. As we have described above, one hand turns in one direction, the other hand in the contrary direction. If we let go of this tense rod, it untwists, of course turning in the contrary direction as in the twisting. With the capacitor, it means electric current in the contrary direction.

We have to stress here a clear difference between what we have called twist of the EM-forces in the current-carrying wire and that in the capacitor. When we talked about twisting of the EM-forces in the wire, we just meant their spiraling order. At breaking the circuit, these forces disappear immediately. With the concept of the EM-forces in the capacitor, we mean something similar as with the rubber rod. After the capacitor is disconnected from the battery, the twist of the EM-forces can untwist, i.e. the accumulated energy can be regained as electromotive force. The difference to the rubber rod is that the rod unturns just after it has been released, while in the capacitor the twist remains even after disconnection from the electric source (which is particularly noticeable in the case of larger capacitors). By itself, the capacitor unturns slowly. To untwist the capacitor immediately, it is necessary to connect its ends with a more or less conductive wire. When connecting its ends to a good conductive wire, it will untwist immediately, that is, we will get stronger but short lasting current, while with a bad conductive wire the untwist will take longer and we will get weaker, but longer lasting current. The untwisting also manifests itself therein, that in this case the current is counter-directed.

If we connect a capacitor to a battery and also add an ammeter in the circuit, we will notice that stronger current flows at the beginning, then it weakens and dies out at the end. Nothing happens from that moment on (because this happens very quickly, we place a great resistor in series, slowing down the process). The capacitor is now maximally twisted. This doesn't mean that it cannot be twisted any further, but that it is as much twisted as the battery voltage. Just as a stronger person can turn a rubber rod more than a weaker one, so a stronger battery can twist the capacitor more than a weaker one. We can test that the capacitor is twisted to the source voltage by disconnecting it from the battery and then measuring the voltage at its ends.

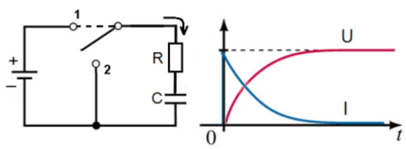
Here, too, we have exponential growth of the voltage (at the same time exponential decrease of the current strength) according to the same function as in the increase of the angle of the magnetic forces with respect to the line of the wire.

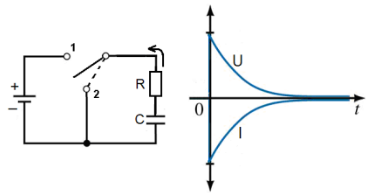
$$U = U_{batt}(1 - e^{-t/RC})$$

$$I = \frac{U_{batt}}{R} * e^{-t/RC}$$

$$I = -\frac{U_c}{R} * e^{-t/RC}$$

$$U = U_c * e^{-t/RC}$$





The difference is that there the dependence is on the current, here on the time. The reason for this is that the capacitor, as an elastator[10] that twists and untwists, is a time-dependent element.

The similarity with the twisting of the rubber rod is obvious: at the beginning the rod twists quickly - the initial quick twisting corresponds to the initial strong current; and, as the twist (i.e. the tension) increases, the speed of the twisting, that is, the current drops. The formula for the current through the capacitor is nothing else than the steepness of the function for the voltage.

As the resistance of a wire depends on its material and its dimensions, so does the capacitance of a capacitor on the same parameters. And just as we spoke about resistance and conductance of the metal wire, so we speak about capacitance and elastance in regard to the capacitor.

The larger the area of the dielectric plate which is sandwiched between the metal plates, the larger the capacitance and the smaller the elastance. The larger the thickness of the plate, the greater the elastance and the smaller the capacitance. Therefore, the formula for the capacitance is:

$C=\varepsilon*S/d$

S is the area of the dielectric *that touches the metal plates*, d the thickness of the dielectric and ε the permittivity of the dielectric (the opposite of elasticity).

In order to understand the nature of the capacitor from the aspect of its dimensions, we will try again the comparison with the rubber rod. Suppose we have two rods of same rubber and of equal cross-sections, but one is shorter, the other longer. The longer rod has certainly a greater elasticity at bending as well as at twisting. The rod is attached at its ends to two counter-rotating devices which twist it, just as we twist the rod with our hands. Each of the two devices turns 360°, which would be the same as if one end were fixed and the device at the other end turned 720°. If we clamp into the device the longer rod, then its twisting for two full circles will be easier than it will be with the shorter rod for the same twist. The effort

10 At the end of the 19th century the english electrical engineer Oliver Heaviside (1850-1925) introduced the term "Elastance" for the capacitor (comparing it with a spring). This term is inverse to the "capacitance", i.e. E=1/C.

in the second case is greater. But when we remove the rods from the device, the shorter will release more energy than the longer one.

Let us now think of two rods of same material and of equal length, but of different cross-sections. If we clamp the thinner rod into the device, it will be easier to twist it for the two full circles than the thicker one; but on release, the latter will give off more energy.

The last experimental setup contains two rods of equal dimensions but of different kinds of rubber. The different rubbers have different elasticity, so for the twisting of the rod with the less elasticity the device will use up more energy.

In the first variant, the shorter rod has the greater capacity (smaller elasticity); in the second variant, the thicker rod has the greater capacity. It is possible that we twist one of the rods not for two but for twenty full circles. It may happen that the rod tears and permanently loses its elasticity. With stepwise tryouts, we can determine to what degree of twist we can go before the rod tears.

What we have just described is well comparable with a capacitor. The voltage up to which the capacitor is twisted corresponds to the number of circles for which the rubber rod is twisted. The thickness of the dielectric plate between the metal plates corresponds to the length of the rod; its area corresponds to the cross section area of the rod. The maximum number of circles up to which we can twist the rod undamaged corresponds to the maximum voltage we are allowed to connect the capacitor before its dielectric is destroyed, an indication usually printed on the capacitor itself. In order to increase the maximum voltage that it can withstand, we have to increase the thickness of the dielectric plate. But this will reduce its capacitance. In order to keep its capacitance the same, we should at the same time increase the area of the dielectric plate. With these two operations all of its dimensions increase. Therefore, we sometimes see two similar, say electrolytic capacitors, with the same capacitances, but different dimensions. The bigger one can withstand a higher voltage.

The capacitance of the capacitor also depends on the nature of the dielectric, expressed in the formula by the variable ε. It is always greater than 1, because the reference point 1 is for the air, as the most elastic dielectric. This is followed by various types of plastic, glass, paper, etc.

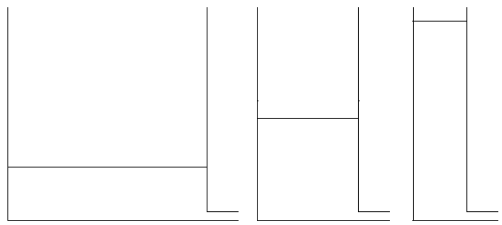
Connecting two or more capacitors in series is similar to sticking multiple dielectric sheets (i.e. plates) together; it is like increasing the thickness of the dielectric plate of a single capacitor. The total elastance increases and is a sum of the elastances of the individual capacitors.

Two or more parallel connected capacitors behave similarly as if the dielectric's area of a single capacitor is increased. The total capacitance increases and is a sum of the capacitances of the individual capacitors.

If we hook up to a battery three series-connected capacitors of same material and equal dielectric's thicknesses but different areas and then measure the voltage across each capacitor, we will see that the greatest voltage is at the ends of that which has the smallest area, the lowest voltage on the one with the largest area. In other words, the one with the smallest area is the most twisted.

Here we have a kind of accordance with what we have discussed about the twisting of the EM-forces in series connected wires of same metal but different cross sections. The greatest twisting is in the thinnest wire, the smallest in the thickest - but, with one difference: this twisting in the wires refers to the magnetic forces, while in the capacitors it refers to the electric forces. Here we can also see the opposition, the polarity between magnetism and electricity, between the metal and non-metal.

Even in the mechanical sense it is quite obvious; namely, when we bind two or more pieces of elastic band of different elasticity. Upon the stretching of this composite band, the most elastic portion will stretch the most and, when twisted, it will twist the most.



In view of the fact that the same current has flowed through the three capacitors, it follows that the same amount of electricity has flowed into their different twisting. The upper picture shows three vessels with equal amounts of water but with different pressures at their outlets. If these vessels are empty and we want to fill them through their outlets with the help of a pump from a well, then we will have to expend the most energy (work) on the third vessel, since here builds up the greatest pressure against which has to be pumped. Although the amounts of water in each vessel are equal, the greatest amount of energy is stored in the third vessel. Similarly, in the capacitor with the smallest capacitance, at the ends of which the greatest voltage is measured (it is most twisted), the most energy is stored.

If we connect two capacitors in parallel, then different currents flow through them during the twisting, but both twist up to the same voltage. Through the greater one flows stronger current (with a “greater capacitor” meaning the one with greater capacitance). It stores a greater amount of energy. Here again we see the opposition between the parallel and the series connection, similar to the above with the bulbs of 100 and 60 watts.

As we have seen, we can detect current flow through the circuit only in the first few moments as soon as we connect a capacitor to a battery, namely until the capacitor is twisted up. After that, there is no current. However, if we leave the capacitor connected to a 1.5V battery for several days, we will find that the voltage of the battery has dropped. The voltage will decrease faster with a greater and slower with a smaller capacitor. Based on the comparison with the rubber rod, this is easy to explain. When we twist the rubber rod with our hands and hold it so, it seems like a static state. But is it really so? Each elastic body tends to return to the state of zero tension, and so the rod tends to untwist. We prevent this tendency with our hands. If we have twisted the rod considerably, we will notice that our hands tremble under the strain. It is because the rod tends to untwist with its elastic force, which minimally it succeeds in doing, but we twist it again and again. Keeping this condition for a longer time will cause our force to subside and thus also the degree of twist to decrease. A kind of such trembling also takes place in the capacitor. The twisted electricity unturns a bit, but the power of the battery twists it again. And this is nothing other than a weak alternating current which the instrument cannot detect. After several days, however, the voltage of the battery is decreased, as it has happened with the power of our hands. A greater capacitor uses up the battery faster because it “pushes” back with greater force. In comparison with a smaller capacitor, the current in the greater capacitor is slightly stronger.

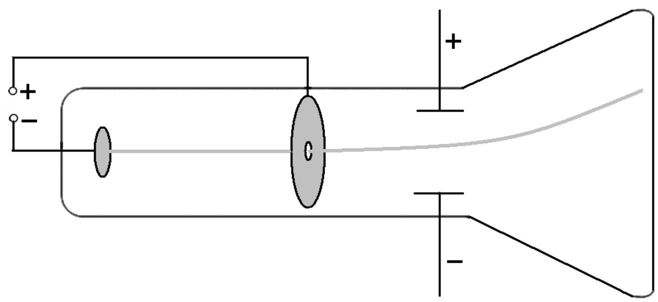
The ammeter cannot detect this weak alternating current, but the circuits with the two transistors described at the beginning of our paper are capable of doing this. If we install a small capacitor in the branch leading to the heart of the transistor (its other end is connected to the plus pole of the battery in the plus circuit), the LED will light up briefly and strongly at the moment when the capacitor is connected and then shine weakly continuously. The first strong flash is due to the twisting of the capacitor from null to the maximum it can reach in that branch. So first we have a brief strong DC to the heart of the transistor and later continuous weak AC. As we replace this capacitor with an ever greater, the small AC will grow, which can be seen in the ever increasing brightness of the LED lamp. It goes without saying that this current stimulates the heart of the transistor only with its positive half cycle, but since the frequency is high, no flickering of the lamp can be noticed.

Lightning, from the greatest in the strongest thunderstorm to the smallest in the taking off a pullover, is a result of the tearing apart of the excessively twisted (i.e. tense) electromagnetic “rope”. Just as the tearing of an overstretched rope of plant fibers or metal wires is accompanied by an ear-piercing, crackling noise and shows itself in an irregular zigzag-like shape, so too does this happen when the electromagnetic “rope” tears apart. The sudden jerk when tearing causes very brief but powerful friction between the plus and minus E-segments of the EM-forces, which produces enormous heat and light phenomena. Due to the great heat in the "channel" (i.e. in the lightning), it becomes for a very brief moment also a good conductor; that is, it comes to a short circuit between the two polarities.

When a capacitor is twisted beyond the prescribed voltage (breakdown voltage), it explodes. This tears the dielectric apart and nothing happens anymore. But if, as in the Earth's atmosphere, air is the dielectric and lightning strikes, this can be repeated countless times because the air is constantly renewed.

When we place two electrodes in a glass tube and connect them to high DC voltage, lightnings start to jump over. Then, as we pump the air out of the tube, the lightnings turn into a continuous light flux, and, at the end – when a considerable degree of vacuum is reached - it turns into invisible electricity flow. By pumping the air out, the matter which can be twisted diminishes, so that the electricity gradually turns from the state of twisting and tearing into the state of continuous flow.

Vacuum tubes are often called cathode ray tubes (CRT) because it is claimed that through their space moves negative electricity, namely electrons, which supposedly dissolve out of the negative electrode, called cathode, and overflow to the positive electrode, called anode. What is here adduced as a proof that this is a current traveling from the negative to the positive electrode is the deflection of the beam towards the positive electrode of two additional opposite electrodes in the so-called Braun’s tube (image below). This tube is what is used in CRT televisions, monitors and oscilloscopes.

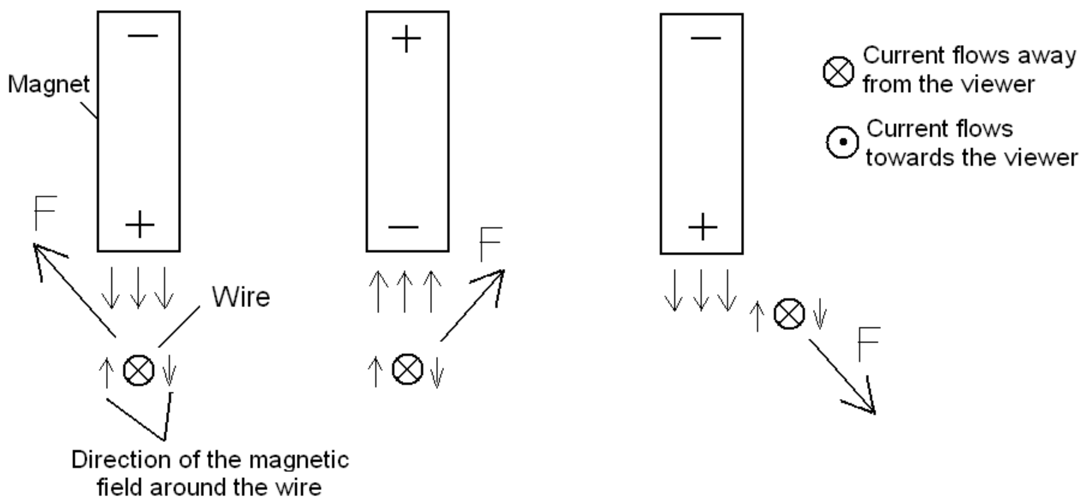


On the left side of our image is the negative electrode (cathode) and a little to the right the positive, the anode, which is in the form of a metal disk with a small hole in the middle. To the right of the anode are the additional electrodes (they are built into the tube) which, when connected to a high DC voltage source, deflect the beam from its straight line upwards to the positive electrode. The beam itself is actually invisible, but is made visible by adding a small amount of some inert gas (neon, argon, etc.) into the tube.

What the author of this paper regards as contradictory in the assertion of flowing negative electricity are two things. The first is of principle nature: one of the basic principles of nature is that movement is always from the positive to the negative and not contrariwise. The second is a matter of fact: let us examine the nature of electricity around the right part of the tube in the image above; let us examine it in front of the screen of a CRT television, monitor, or oscilloscope - we will always find that the detector shows intense positive electricity.

That it is impossible, negative electricity to flow towards the screen and on its other side to appear positive electricity, shows the following experiment: we electrify a vinyl plate by rubbing it (as we know it is negatively electrified) and place it behind a big glass window. Then we test the nature of electricity on the other side of the window. The detector shows presence of negative electricity just as it would have indicated without the glass. Glass does not change the nature of electricity on the other side.

Before we present our explanation of this phenomenon, let us consider a few more experiments. We put a stiff copper wire on a table. Parts of its length don't touch the table. Above a wire section that does not touch the table we hold a strong cylindrical magnet with its positive pole down, so that the wire lies exactly under the middle of the magnet. Then we connect a new battery to the ends of the wire so that the positive pole is closer to us and the negative pole further away from us. At the moment of connection we will notice that the wire makes a strong deflection to the left and up. As soon as we turn the magnet over and repeat the same, the wire will make a strong deflection to the right and up. If we hold the magnet again with the positive pole down, now not directly over the wire, but left over it, however still close to it, we will notice that the wire after connecting to the battery makes a jerky movement to the right and down. How is this explained? In the first variant, the permanent magnet "blows" down; the magnetic wind in and around the wire blows clockwise spirally (over the wire to the right, below it to the left) from the plus to the minus pole of the battery; it blows down on the right of the wire, up on the left of it; on the right of the wire both magnetic winds coincide (the effect intensifies), and on the left of the wire they collide (the effect weakens); the wire moves to where the effect only intensifies, namely to the maximum, and that is to the left and up. In the third variant, in which both winds only collide, the wire deflects to where the adverse effect is maximally attenuated or quite ceased, namely to the right and down.



Now, facing a CRT oscilloscope, we let its beam run slowly and uniformly from left to right (visible as a bright dot moving horizontally from left to right in the middle of the screen); then, exactly over the center of the screen, we place a magnet with its positive pole down. We will notice that the dot no longer moves horizontally but that it slopes downwards and passes through the center. When we turn the magnet upside down, the dot slopes upwards, passing through the center. If we compare this observation with what we have just said about the experiment with the copper wire and the magnet, we find the same thing happening in both cases. We conclude that the rotational direction of the magnetic wind generated by the beam in the oscilloscope coincides with that of the wire, as long as the positive pole of the battery is closer to us. So it's also the oscilloscope's plus side closer to us when we stand in front of it.

The (+)pole of the magnet points downwards, the beam of the oscilloscope approaches it from the left. On the right side of the beam its magnetic wind blows down, i.e. both winds match; so, the beam is shifted upwards. When it goes to the right side of the screen, also on the right side of the magnet, then their winds collide, so the beam is shifted downwards.

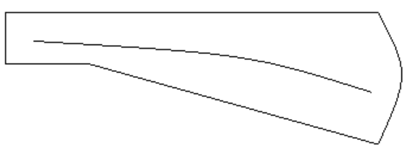
We explain this phenomenon as follows: the positive electricity radiating from the anode spreads to the right into the broader part of the Braun's tube in the figure above. Since the anode is a disc with a circular hole in the middle, this electricity, with the help of the suction minus cathode on the other side of the anode, forms a vortex which is directed to the opening of the anode and continues to the cathode. This electromagnetic tornado is actually the beam that is visible when a small amount of an inert gas is introduced into the tube.

So when we stand in front of an oscilloscope and the bright dot lies still in the center of the screen, then it flows in the tube around the bright dot invisible positive swirling electricity towards us and from the very dot begins a vortex in the opposite direction towards the hole of the anode and onwards to the cathode. The bright dot is actually the eye of this EM-tornado. (Even with toys that cause a vortex in a water-filled container by means of a small electric motor located at the bottom, it can be noticed that the movement of the water around the vortex is directed upwards, but in the vortex downwards).

The fact that the vortex is deflected to the positive of the two additional electrodes does not contradict our explanation, because we claim that this is not something that can be simply accommodated under the postulate “plus attracts minus”, but rather a positioning of a motion consistent with ambient influences whereby maximum effects are achieved (we could observe something similar in the previous experiment, where the wire was deflected to the left and up while the magnet with its plus pole was positioned over it). For the effect of the vortex to reach the maximum, it is deflected to the positive electrode when additional electrodes are inserted in the tube.

In the above-mentioned toys, the water vortex is fully upright when the electric motor is positioned right in the middle of the bottom of a cylindrical or slightly conical vessel. However, when the motor is displaced to one side of the vessel, the vortex is curved towards the opposite side. In this way, it strives to achieve the maximum effect, in this case to capture the largest possible amount of water and make it spin (YouTube video: “Discovery kids tornado lab extreme weather toys” by uploader “dFunKidsToys"). In our case the electromagnetic vortex makes a curve to the positive electrode, and so it seeks to capture and spin the largest possible amount of positive electricity.

It can also be assumed that a non-symmetrical conical glass tube would make the vortex curved even without the additional electrified electrodes (image below).



Another detail indicating that this is a kind of vortex is the shape that the bright dot takes when turning off the oscilloscope. It dissolves circularly. Something similar is also noticeable on the water surface of the mentioned toy after switching off the electric motor.

• • •

Earlier we showed that inserting a magnet into and removing it out of a coil causes production of alternating current. This electricity is produced regardless of whether the ends of the coil are closed or open. When closed with a high resistance wire, we will feel greater resistance to the insertion and removal of the magnet than to open ends, i.e. we have to make a greater effort to move the magnet. The more we diminish the resistance of the wire, the greater the resistance we feel - that is, the more force we have to apply to maintain the frequency of our movements.

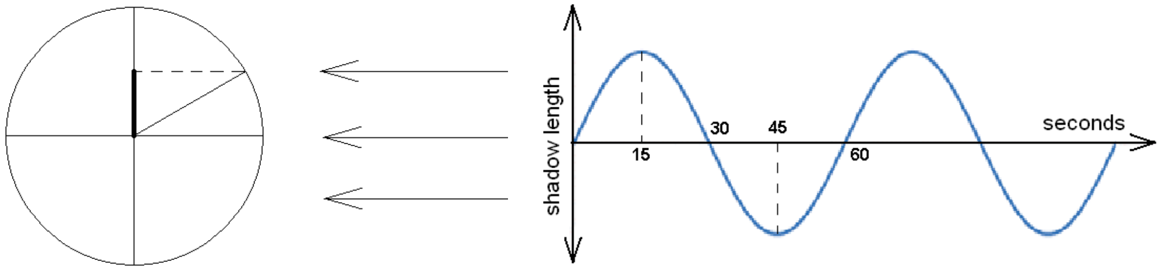
When the coil is closed with a wire of high resistance, the evoked and ordered EM-forces in the coil have the opportunity to close in a circle through the wire, i.e. the plus from one side to meet with the minus from the other side. In such a closed loop, the electromagnetic wind intensifies, which at the same time means that the magnetic segments get the opportunity to become more rightened with respect to the wire line. So they can exert a greater resistance on the magnet movement in and out. As the resistance of the wire decreases, the process described intensifies and the evoked current increases. Imagine that the magnet is moving in and out 50 times per second (50herz) and that we are maintaining the speed constant while at the same time someone else is reducing the resistance of the wire continuously. As the described operation proceeds, our effort to maintain the speed will continually increase. We can compare this with an upside down bicycle. First we turn the pedals without the chain, which corresponds to the movement of the magnet at open ends of the coil. Then we put the chain on the biggest sprocket on the rear wheel and turn the pedals. This corresponds to the connection of the coil ends to the high resistance wire. Then we change the transmission to ever smaller sprockets, which makes the turning of the pedals harder. This corresponds to the reduction of the wire resistance. But we can also imagine that while riding a bike more and more load is continuously added in a trailer behind it and we should still keep a constant speed.

Something similar happens also in the generators of the power stations. Electricity consumption varies at different times of the day: it is lower overnight and increases after the beginning of the day to reach its maximum in the afternoon. As the consumption increases, the so-called turbine wicket gates at a hydropower plant need to be opened wider to allow greater water flow. Every house, every factory, that is, every consumer, can be considered one of the thousands parallel resistors connected to the generator. The more resistors connected in parallel, the lower the electrical resistance in the circuit. Principally, in the synchronous generator happens nothing different from the processes in our small coil.

As we have to insert and pull out the magnet with increasing effort at the resistance reduction in the circuit in order to keep the same frequency and thus the same voltage, so also in the hydroelectric plant more water must be let to the turbine blades to maintain the speed of the generator, since the turbine with the generator sits on the same axle. This is automatically regulated when the changes are gradual. Sudden changes must not happen, as this would reduce or increase the frequency and the voltage (if a large consumer is suddenly switched on or off). The result would be a disaster in the system. Therefore, such consumers like big smelteries, whose furnaces draw thousands of kilowatts, must report both their startups and their shutdowns at the control centers. The sudden connection of a larger consumer in the network can be compared to the aforementioned bicycle trailer in which, while riding at a constant speed, a significant weight (say 100-200kg) is placed. The sudden switch off can likewise be compared to the instantaneous loss of the weight in the trailer.

When we insert the magnet into the coil and pull it out again, the intensity of the evoked current is variable. As the magnet approaches it, the current gradually increases, to reach its maximum at the moment of immersion in the coil. As the magnet continues to penetrate the coil, the speed drops and so does the current. When the magnet comes to a standstill, the current goes to zero. When pulling out the same happens: at the moment of “emergence” the current reaches its maximum. But this time it has the contrary direction. Ideally seen, the thus generated current is a sinusoidal alternating current.

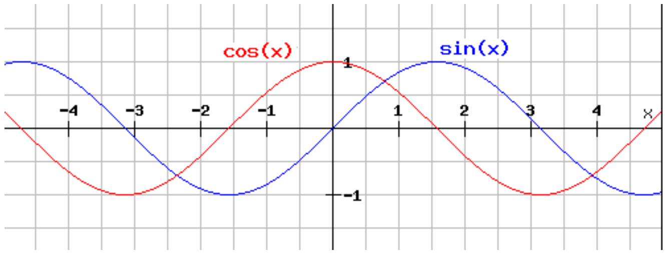
Let's briefly explain what a sinusoid is. In the picture below we see a circle with radius 1. No matter how large the radius is, its length can be always considered as 1. The radius of this circle begins to rotate uniformly like the second hand on a watch, but starting from 3 o'clock backwards to 12 o'clock. The horizontal line of the drawn cross denotes the X-axis, the vertical line the Y-axis. Let us now imagine that there is a light source to the right of the circle that emits parallel rays towards it. These rays will make the rotating radius cast a shadow on the Y-axis. At the beginning, the length of the shadow is 0, i.e. just a point. In the first five seconds the radius has moved from 3 to 2 o'clock, that is, by 30°, with the shadow rising from 0 to 0.5, so it has grown by 0.5. From the 5th to the 10th second, it shifts from 2 o'clock to 1 o'clock, and the shadow has risen to 0.87. From the 10th to the 15th second it has moved on from 1 o'clock to 12 o'clock, again by 30°, and the shadow has risen from 0.87 to the value 1. In each of the three intervals the radius moves 30°, but in the first 5 seconds the shadow grows by 0.5 and in the last 5 seconds only by 0.13. In other words, for the same shift of the radius, the shadow once grows more, another time less. If we had chosen instead of 3 intervals a larger number of them, say 9 intervals per 10°, the ratio between the growth of the shadow in the start interval and in the final interval would have been even more drastic.



The radius then moves uniformly towards 11, 10 and 9 o'clock. We imagine now that the light source has moved to the left side of the circle. The same happens again, but in reverse order: the shadow initially reduces its length very slowly and then faster and faster. When the radius arrives at 9 o'clock and moves on towards 8 o'clock, the shadow goes to the negative side, i.e. it gets minus values.

The length of the shadow we call sine. If we plot the seconds in a coordinate system on the X-axis, with the length of the shadow at a given second on the Y-axis, and then connect these points with a line, we get a sinusoid.

Instead of on the right and then on the left side, let us imagine the light source first over and then under the circle while considering the shadow the radius casts on the X-axis. Again, we get the same figure, except that the shadow now starts at 1 and drops to 0 and so on. The resulting curve is called cosinusoid. It has the same shape as the sinusoid, but is only shifted to the left. It is said that these two curves have a phase difference of 90°, or that the sinusoid lags by 90° (π / 2) behind the cosinusoid. These 90° phase differences transformed to time in our particular example is 15 seconds.



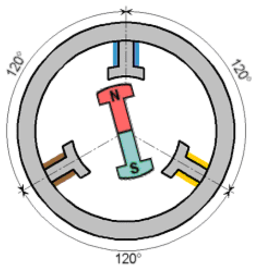
Let us now consider not the whole shadow as a line segment, but only its peak as a point. It moves up and slows down to come to a stop at the top. Then it moves back down, accelerating to reach maximum speed at the center of the circle. Then it continues further down into the negative part and slows down again. The point behaves similarly to a pendulum.

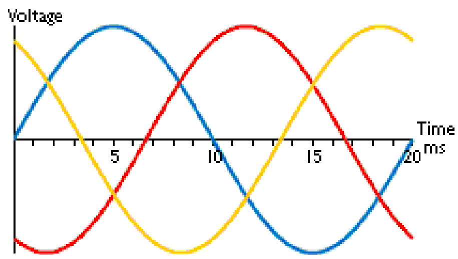
With the insertion of the magnet and also with its removal from the coil, we have generated electricity directly in its windings. But this can also be done indirectly without inserting the magnet into the coil, in such a way that we place in it a cylindrical piece of iron with a diameter almost like that of the coil. If we now just move the magnet towards and then move it away from the iron core (holding it with one hand to prevent it from coming out of the coil and sticking to the magnet), we will see that the same happens to the ammeter as happens by inserting and removing the magnet from the coil. The iron cylinder may be longer than the coil and protrude from it even several centimeters, without changing anything in the effect; we get again the same current by the same speed of magnet movement. This clearly shows that the effect is now not exerted directly on the wire but indirectly through the iron core.

Let us score an elongate notch in a piece of styrofoam plate into which we put a cylindrical piece of iron, then we place the plate on the surface of water in a plastic vessel and bring a strong magnet from outside to the wall of the vessel. Our floating piece of styrofoam with the iron will be attracted to the wall. However, if we pull the magnet back quickly and suddenly, our “boat” will make a rapid movement in the opposite direction. In both cases the iron moves contrary to the movement of the magnet. This shows that as the magnet approaches the iron, magnetic forces of attraction are evoked in the iron, and forces of repulsion upon the magnet's removal. We can also say that as the magnet approaches, the iron experiences twisting of its magnetic forces and, at the removal of the magnet, untwisting. Since the spin during the untwisting is reversed, counteraction happens, i.e. repulsion.

A cylindrical magnet has a powerful effect only at its poles. It is very small on its cylindrical surface. Since the iron in the coil can be well fixed and not move at all, we conclude from the previous experiments that it doesn't act on the coil wire through its ends but through its cylindrical surface. Therefore, we consider it legitimate to talk about twisting and untwisting of the EM-forces in the iron core.

Since a linear oscillating movement of the magnet is difficult to be carried out in practice, the electric current in generators is generated by its rotational movement. If we place a strong magnet to rotate on an axle in its center and position around it three coils with iron cores at angles of 120° (image below), we get so-called three-phase synchronous generator. The coils are called stator, the magnet is called rotor.





When a pole of the rotating magnet approaches a coil, it evokes a current in one direction; when it moves away from the coil, then current is generated in the contrary direction. For better understanding, it is helpful to compare this with the rotation of the radius in a circle as shown above. There we have explained that when the imagined point, which moves along the Y-axis during the rotation of the radius, approaches the topmost mark, then its speed drops, coming to a standstill at that mark. We can apply the same to the rotating magnet. Although it rotates at a constant speed, we only consider the speed of the pole on the imagined axis in front of the coil, which is exactly the same with that of the imagined point of the radius. At the moment when the magnet is directly opposite the coil, the current in the latter comes to zero, because the approaching speed of the magnet comes to zero. As the magnet begins to move away from the coil, current begins to flow in the opposite direction and increases as the speed increases, to reach maximum value when the magnet reaches a 90° angle with respect to the coil. After this moment, this pole of the magnet loses its influence on this coil and its role, but now from the maximum current dropping down, takes over the opposite pole, as it begins to approach the coil with maximal speed from the other side.

The same happens with the other two coils, so we get three independent current sources with sinusoidal characteristics. The three sinusoids are shifted by 120° from each other.

If we connect the outgoing wires of the coils to another three equal coils arranged in the same way and in the middle of which there is also a magnet, then this magnet begins to turn. Thereby the synchronous generator is connected to a synchronous motor. We see that there is no difference between the two arrangements. Again, the coils are called stator and the magnet rotor.

The graph shows three sinusoids for the three coils. Let's choose any point on the X-axis and at this point look at the current values for each coil. Their sum is always zero. This is quite the same as the following: in a circle we draw three radiuses like the Mercedes sign. Then we draw their shadows on the Y-axis. If we add their length, we get zero.

$\sin x + \sin (x+1/3*360°) + \sin (x+2/3*360°) = 0$

$\sin x + \sin(x+120°) + \sin(x+240°) = 0$

$\sin x + \sin(x+1/3 * 2\pi) + \sin(x+2/3 * 2\pi) = 0$

$\sin x + \sin(x+2\pi/3) + \sin(x+4\pi/3) = 0$

The magnet in the picture turns counterclockwise. We can conclude this from the graph of the sine curves. When the magnet is exactly opposite the blue coil, then its current is zero. The next encounter exactly opposite a coil will occur between the opposite pole of the magnet and the one of the other two coils. Since after a zero point of the blue curve it follows a zero point of the yellow curve, it means that this pole will move towards the yellow coil.

In order to cause reverse rotation of the magnet, it is sufficient to exchange the leads of any two coils. In this way, we will also change the arrangement of the sinusoids, i.e. after a zero point of the blue curve, now a zero point of the red curve will follow.

The fact that the sum determined above was zero means that the torque from the coils upon the magnet is the same at every moment. Let us take for instance the already mentioned moment when the magnet is exactly opposite the blue coil at the given arrangement of the sinusoids. This coil has no effect. The yellow coil attracts the lower pole of the magnet, the red coil repels that pole with the same intensity as the yellow one. We can recognize this by the equal but opposite sign values of the yellow and the red sinusoid at the moment when the blue sine curve is zero.

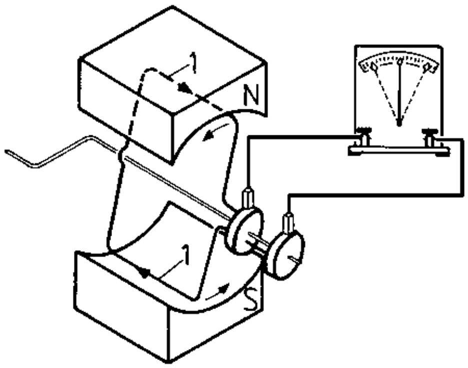
Consider another moment, when the magnet in our image is in horizontal position. The action of the blue coil is then maximal and attracts the right pole of the magnet, the red coil has opposite current and therefore attracts the left pole, the yellow coil has the same current and repels the right pole of the magnet. This results in smooth rotation without turbulence.

Here we have in principle a very similar picture as with the front and rear sprocket of a bicycle. The pedals are the water turbine, the front sprocket is the rotor of the generator, the rear sprocket is the rotor of the motor, the chain is the copper wire and the teeth are the magnetic fields. The difference is that the chain of the bicycle must be stretched mechanically, while the copper wire must be electrically “taut” and can be many kilometers long.

What we have said so far about the synchronous generator and the synchronous motor is not in line with that presented in textbooks on electromagnetism. There it is stated that in the position of the magnet exactly opposite a coil, its current is at maximum. Let's see if it would be possible to get a smooth rotation of the magnet. Therefore we take the already mentioned moment: the magnet exactly opposite the blue coil with its current at maximum. Until this moment the coil has attracted the given pole, then the pole goes to the left side of the coil; this would still have the current in the same direction, which means that it would still attract the pole and thus act against the direction of rotation. At the same moment (i.e. when the magnet is exactly opposite the blue coil) the red and yellow coils have equal currents in the same direction and both act on the opposite pole of the magnet. Thereby both will exercise an attractive force. It follows that the yellow coil attracts the lower pole of the magnet in the direction of rotation and the red coil attracts it against the direction of rotation. We see that in two places, both up and down, contradictory effects take place. When the upper pole of the magnet has passed the blue coil a little bit, then of the three coils only the effect of the yellow one on the magnet will be in the direction of rotation, making the whole assembly impossible.

The difficulties in understanding generators and motors have at their core the lack of consistency that results from what is called in physics “Faraday's law of induction”. This law states that the induced voltage in a wire loop equals the rate of change in the magnetic flux embraced with the loop, or $U=d\Phi/dt$. In the textbooks, to illustrate the law, is often given an example of a loop in the form of a rectangle which rotates in a homogeneous magnetic field.

To explain how the law is applied in this case, we will draw a comparison. If we hold a ring in front of our eyes as if we wanted to see through it, it has the shape of a circle. If we turn it 90°, we only see a line. In every other intermediate position of the ring, we see an ellipse. In the first position the ring has the maximum area in front of our eyes; in the second, the minimum (i.e. zero). In any other position, the area has an intermediate value. If the ring before our eyes begins to rotate about its axis starting from the second position (0) and has turned 180°, then the area we see in the course of rotation will be a sine curve of half a period.



Let's go back to our picture. When the rotating loop is in the horizontal position, the magnetic flux through it is maximal, and in the vertical position it is zero. When the flux is maximal, then the speed of its change (the steepness of the sine function, or $d\Phi/dt$) is zero, and when the flux is zero, the speed of its change is maximal. So when the loop is in the horizontal position, the result would be that the induced current is zero; when it is in a vertical position, the current should be maximal.

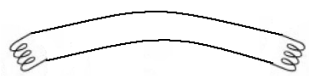
But just the opposite is correct, because *it is not the speed of flux change that is essential, but the speed of motion of the wire towards the magnet or away from it.* In producing the current in the rectangular loop, only two of its sides play a role - in the image the shorter ones - which approach the magnetic poles and move away from them. As we have seen from the earlier examples, this speed is zero when the mentioned sides of the loop are closest to the poles.

In the given example of the loop, however, there is another inconsistency. To make this clearer, we will present an experiment. From a lacquered copper wire we cut off twenty to thirty pieces of about 10 cm. From them we form a bundle of parallel wires and connect the two ends with one more wire each. These two wires are connected to a sensitive analog ammeter. We hold the bundle horizontally and move quickly a strong and broad magnet downwards on its left side. The pointer of the instrument will make a deflection on one side. If we now move the magnet quickly downwards on the right side of the bundle, the pointer will make a deflection on the opposite side. The magnetic flux that we have produced in the wire is now in the opposite direction to the one in the first case, which is why the deflection is in the opposite direction. The moving of the magnet produces current even if we only approach it to the bundle from one side without lowering it below the bundle. In this case, the current is of course somewhat weaker. But if we now move the magnet down to the middle of the bundle, the instrument won't show any current, because the left and right halves of the magnet act on opposite sides of the bundle, canceling each other out. We can do the experiment also with a single wire instead of a bundle, as long as we have a very strong magnet and a very sensitive instrument. We consider this to be the simplest and most impressive experiment that can be carried out with the help of a magnet to produce electricity.

It follows: if a wire moves in a homogeneous magnetic field (to be nearly homogeneous, it must consist of very wide magnets), then no significant current can be produced, therefore such a generator would be very inefficient.

A synchronous motor does not start by itself. In other words, the magnet that should rotate in the middle will remain at rest even if we apply 50Hz current to the coils. From the state of rest, it cannot suddenly fit into the rapidly rotating magnetic field of the current, which comes at an immutable and relatively high frequency. In our basic example, however, we could turn the rotor of the generator by hand, starting slowly and gradually accelerating. As we increase the speed of the generator rotor, so the speed of the motor rotor follows in step with the previous one.

If we attach a pulley on the axle of the rotating magnet of the motor, it can perform physical work. For example, we can fix it on the roof of a building and lift a load with it. Imagine that the motor starts lifting a smaller load and, as it moves up, the load increases. As the load increases, so does the current in the windings of the stator, whereby the magnetic field intensifies to lift the heavier load. The speed of the synchronous motor never decreases. That's why it is called synchronous - it always turns in sync with the generator. If the load increases beyond the prescribed limit for the given motor, it will abruptly stop. Now the question arises as to how the windings of the stator "feel" that the load increases so that the current in them increases, when a material connection between stator and rotor does not exist. What exists is only the invisible magnetic connection. As we have said, the speed of the rotor does not change, and still we should seek in it the cause of the current increase. Let's look at the image on the left. There we see two



coils in a closed circuit, one belonging to a phase of the generator, the other to a phase of the motor. Basically we have here nothing more than a copper wire in a closed loop. The current in this wire increases as the load increases thanks to the, in our opinion, increased eccentricity during the turning of the two rotors, i.e. thanks to their vibrations. When the rotor of the motor rotates at increased load, it does not rotate regularly circularly; rather, as it approaches a coil of the stator, it receives additional attraction from it, causing the rotation to be irregular, eccentric, vibrated. Actually, this happens at whatever load weight, only the phenomenon intensifies at increased load. This additional eccentric speed of approach to the motor stator causes a stronger current in its windings, and this also has an effect on the rotor's rotation of the generator, so that it also experiences an intensification of its vibrations, which in turn leads to an increase in force necessary to maintain the speed of its rotation. The vibrations help i.e. intensify the "swinging" of the magnetism (that is, its constant twisting and untwisting) in the cores of the coils. They do not have to be strong, but at the rate of 50 revolutions per second, or 3000 per minute (3000rpm), even a smaller intensification of vibrations plays a significant role. We can also compare this with the bicycle: if the load in the trailer behind the wheel is increased, then this will lead to intensification of the eccentricity when turning the pedals for the sake of maintaining the speed.

Let us also consider the very moment when the current increases. The moment the load on the motor is increased, it will still slow down for a very short time, but due to its large inertia, it will not stop but continue to turn. After this short delay behind the turning of the stator's magnetic field, the rotor will also for a very brief moment increase its speed beyond the normal, thus catching up with the time previously lost. This brief moment of speed increase will cause a small intensification of the eccentricity, which in turn will cause an increase in the current strength. The increased current as well as the increased eccentricity will be maintained in the further, but now again uniform rotation.

As we have said above, this is just a consideration of the author that is not gained from experience in working with motors and generators, so it needs to be verified.

The magnet between the coils of the motor can be replaced with an aluminum can, say from beer, placed on the same axis in the center, so that this time, viewed from above, we get a turning circle instead of a turning stick. Now we have an asynchronous motor, also called an induction motor.

To explain what is happening here, we will make an experiment. In a flat piece of styrofoam we thrust a copper or aluminum plate so that it can stand vertically. We place the styrofoam plate in a container with water and slowly move a strong magnet close to the copper plate. The "yawl" starts drifting to escape. If we then slowly move the magnet away from the plate, it starts to follow the magnet. When the magnet rests, the plate beside it rests too. In fact, once the magnet has come close to the plate, it cannot get rid of the plate anymore. It constantly floats before or behind the magnet. We could say this: when the magnet approaches the plate, it strives to strengthen its effect, and then the plate moves away; when the magnet moves away from the plate, it strives to weaken or completely cease its effect and then the plate follows it; and when the magnet is still, the plate is still too. In this experiment there is actually nothing new compared to the experiment with the falling magnet through a copper tube.

Here the situation is the opposite of the situation when an iron object was laid on the styrofoam. There, the approaching of the magnet attracted the object, here it repels it; there, the moving away of the magnet repelled the object, here it attracts it. The opposition of what is called ferromagnet (iron, cobalt, nickel) and diamagnet (copper, silver, gold, etc.) is clearly seen here. In physics it is said that ferromagnetic are such elements that are attracted to the magnet, and diamagnetic such that are repelled by it. But this is only a partial perception. The whole truth is that both types are both attracted as well as repelled by the magnet.

Approach and attraction can be characterized as plus-actions, but moving away and repulsion as minus-actions. Since the diamagnet is repelled when approaching the magnet, diamagnetism can be characterized as a minus-property. The ferromagnet is attracted upon approaching; hence, ferromagnetism is a plus-property.

Based on the above said we can say this: when the current in the coil is on the rise, then its magnetic field intensifies, so the can is repelled; when the current in the coil is in decline, then the magnetic field weakens, so the can is attracted; when the current is in the highest point or near it, the magnetic field is invariable (the steepness is zero), so the magnet has no effect on the can. This happens in still another moment: when the current through the coil is zero. So during one cycle of a sinusoid we have two periods of repulsion and two periods of attraction alternately, but also four short moments of no effect.

Since the steepness of the function sin(x) is cos(x), hence the force by which a coil acts on the can depends, besides other things, also on the product of "sin(x)·cos(x)". In the figure below, three 120° shifted sinusoids are shown first, representing the strength of the three-phase alternating current, and below them the products of them and their steepnesses:

$sin(x) \cdot cos(x)$;

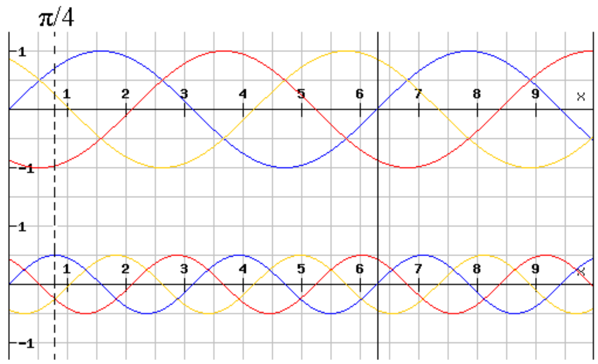
$sin(x+2\pi/3) \cdot cos(x+2\pi/3)$;

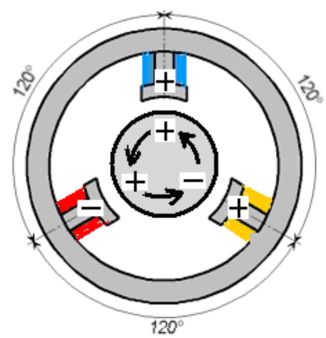
$sin(x+4\pi/3) \cdot cos(x+4\pi/3)$

From the figure we see that these products are also by 120° shifted sinusoids, but with double the frequency (due to the two periods of repulsion and the two of attraction) and with different order compared to the basic ones. Positive half cycles of such a sinusoid are the periods when a coil repels, and the negative half cycles are the periods when the coil attracts the can. But apart from that, that one coil repels or attracts the part of the can in front of it, it also repels or attracts the parts of the can in front of the other two coils.

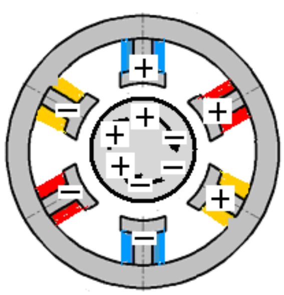
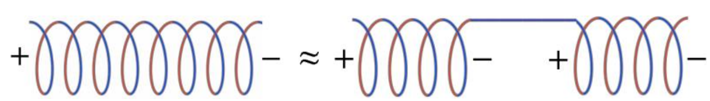
The image at right shows the state at the moment π/4 (since one circle 2π at 50Hz is 20ms, then π=10ms, so π/4=2.5ms). The blue coil has a positive current, so its plus-magnetic pole is turned towards the can. Since its current is increasing, it evokes a whirled current in the can with its plus-magnetic pole facing the coil. At the same time, the red coil has a negative current, so its minus-magnetic pole is turned towards the can. Since its current is in decline, it evokes a whirled current in the can with its plus-magnetic pole facing the coil. In a similar way, we can also derive the state of the orange coil.

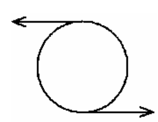
The blue coil not only repels the plus in front of itself, but also attracts the minus before the orange coil, and also it repels the plus before the red coil. The red coil attracts the plus in front of itself, and at the same time it attracts the plus in front of the blue coil (it is important to keep in mind that at this moment this plus is considerably stronger), and it also repels the minus in front of the orange coil. In addition to the corresponding action (i.e. in the direction of turning), the orange coil also appears to have a contrary effect (i.e. opposite to the direction of turning). It has in front of itself a minus which it attracts, and to the left of itself (i.e. before the red coil) it does not have a stronger minus than that in front of itself, but has a plus. This repulsive action towards that plus is canceled by the attractive action towards the minus in front of itself, and what remains is only the corresponding repulsion to, at that moment, the strong plus in front of the blue coil.





Such a motor is not effective because of the very large distance of 120° to which a coil should act to attract or repel the magnetic pole in front of another coil. It can be improved by doubling the number of coils, i.e. by adding to each coil, so to speak, a double exactly opposite it, so that now we will have coils at every 60°. We can imagine each pair as just one coil, which we have stretched exactly in the middle and got a straight line. If at a given moment on the left side of the unstretched coil the magnetic pole is plus, and on the right side minus, then the order of the poles after stretching it will be as in the figure below.



For one thing to be turned around acting with force on two opposing sides, it is necessary for both forces to have opposite directions (figure above). In the interspace of the stretched coil we now have opposite magnetic poles, so the forces that act on the can are exactly in such directions.

If we carefully consider the motor image above (it refers to the same moment π/4), then we will notice two things: firstly, the homonymous magnetic poles, both in the rotor and in the stator, are grouped; secondly, the rotor groupings with respect to those in the stator are displaced by 120°. That's why the rotor turns aiming to clutch its magnetic poles with the opposite of the stator, which are at reach of 60°, but these are continually running away, and the rotor is continually chasing them.

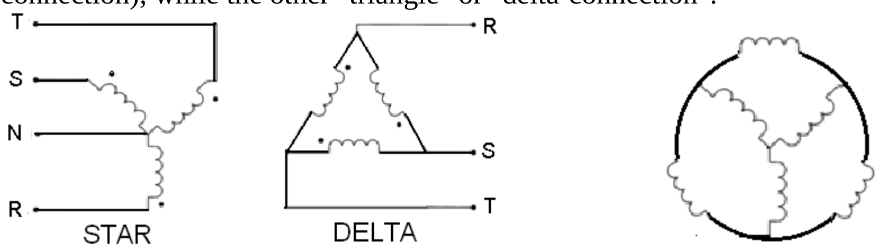
These motors are called asynchronous because they always, even when they are not loaded, turn slower than the magnetic field of the coils. The difference may be smaller or greater and is called slip. It is expressed in percentages. For example, if a motor turns with 10% slip, it means that it makes 2700rpm, because 10% of the 3000rpm is 300.

The slip depends on many things in the motor design, and additionally on the load. This motor is turning slower upon load increase. The principle example of an asynchronous motor just presented is still inefficient and with great slip. It could be improved to a certain degree if a can with a thicker wall is placed, because the six induced looped currents in the can will be stronger, hence the repulsion and attraction. Also, the distance between the can and the stator (so called air gap) should be as small as possible. A great role in efficiency is played by the use of space in the circumference of the circle. In our example with six coils, it is still small. In order to effectively utilize the space of the entire circumference, the stator windings, both in motors and generators (no matter whether they are synchronous or asynchronous), are arranged in a different way than here shown, but we won't speak of that here.

The three-phase asynchronous motors, unlike the synchronous ones, start by themselves. Therefore they are very commonly used in various industrial machines, elevators, pumps, etc. When the load on these motors increases, then the current drawn from the network increases. What we've said about vibrations in synchronous motors can be applied here too.

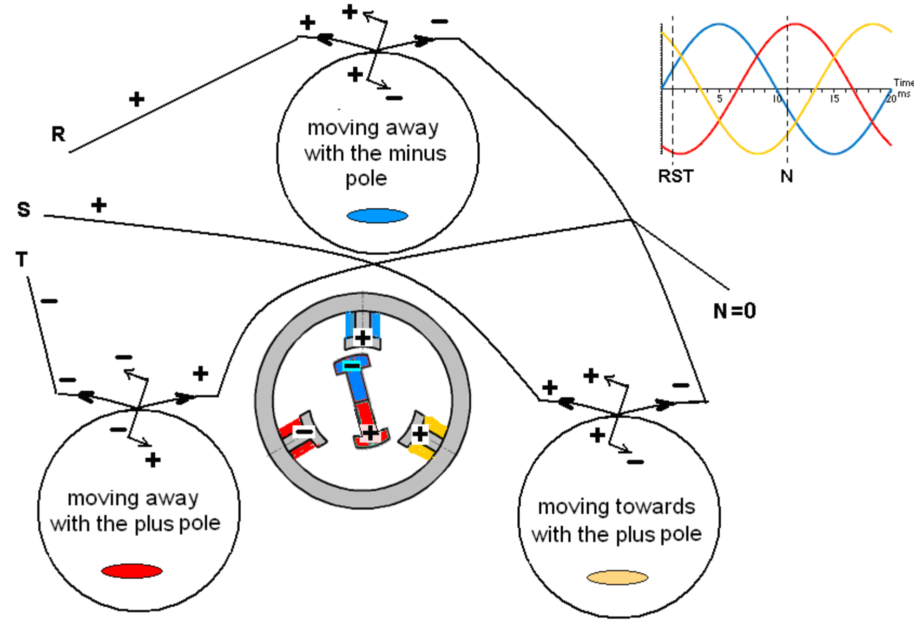
● ● ● ● ●

We have said above that the three coils of the generator are independent current sources. But in practice, however, they are connected. There are two connection modes: the one is called "star" (Y, wye, star-connection), while the other "triangle" or "delta-connection":

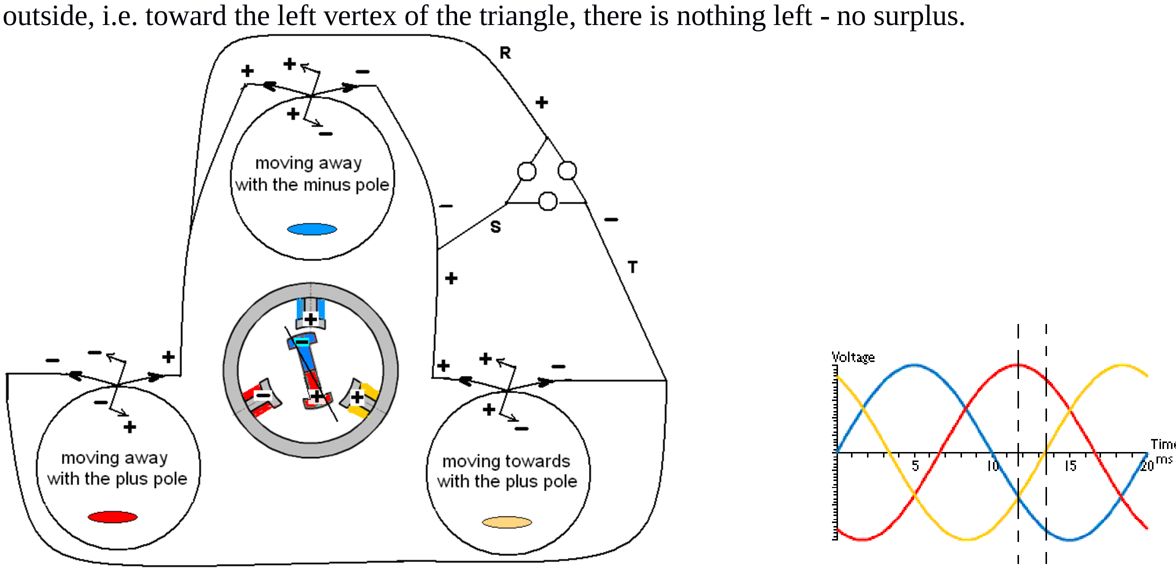


The figure below shows one situation when connecting to a star. As can be seen from it, the right ends of the coils are connected in one point N (neutral conductor, null). In that point at any moment the force of the blowing coming from one (or two) coil(s) is fully met by the suctioning of the other two (or one) coils. When the magnet is positioned as in the figure, then at the right end of the blue coil there is weak suctioning, at the right end of the yellow stronger, and at the right end of the red coil there is so strong blowing that corresponds exactly to the sum of the two suctionings. In the joint point N, they ideally fit, so there is no voltage left toward outside. The left ends of the coils (which in the figure above are marked with dots) are called phases. In Europe they are marked with R, S, T, or more recently with L1, L2, L3, but in the rest of the world there are other markings. The letters RST do not have any special meaning. They are just consecutive letters of the alphabet.

However, it must be emphasized that the aforesaid is valid only if all the three phases are equally loaded (balanced load). What this means, we'll see below.



The figure below shows a situation when connecting in a circle (i.e. delta). Three equal loads are already connected to the generator, e.g. three lamps (the three circles in the triangle). Let's imagine that the magnet is turned slightly further left, that is, it is exactly at an angle of 90° in relation to the red coil (indicated by the thin line). At the right end of this coil there is a maximal plus, and at the left end of the blue coil a mean plus. Only the stronger of these two pluses has an effect to the top vertex of the triangle (this is similar to two parallel-connected cells, one of which gives off a higher voltage). At the left end of the red coil there is a maximal minus, at the right end of the yellow coil a mean minus. Of these two minuses only the stronger minus has an effect to the right vertex of the triangle. At the left end of the yellow coil there is a mean plus and at the right end of the blue there is equally the same minus. These equal plus and minus ideally complement each other, i.e. the minus fully suctions the plus; so, for the outside, i.e. toward the left vertex of the triangle, there is nothing left - no surplus.



But we can also say this: the voltage of the blue coil between R and S and the voltage of the yellow coil between S and T are connected in series and at this moment they are adding up to 1, which is the same

voltage as of the red coil. This resembles three cells - one let’s say of 3V and two of 1.5V, the latter two connected in series and in parallel with the first, i.e. the same configuration as with the loads.
So at this moment in the bulbs’ triangle there is an ordinary parallel connection of two branches, one with one bulb, the other with two series bulbs, hence in the latter branch is flowing two times less current than in the first, while in the summing branch is flowing a current of 1.5 if we treat the strength of the first current as 1.
After 30°, i.e. 1.66ms at 50Hz, the maximal plus of the red coil in R will drop to 0.87, the blue coil’s plus in R will rise to 0.87, and the yellow coil's plus in S will drop to zero. At this moment in the top vertex of the triangle there is a plus of 0.87, while in both lower vertices a minus of 0.87. There is no current flowing through the resistor at the base of the triangle (same minuses at both ends), and a current of 0.87 (more precisely 0.866=sin60°) is flowing through the other two sides of the triangle. This means that a maximum current of 0.866x2=1.73 will flow through the upper summing branch R, and half of it through the lower summing branches (0.87). At this moment there is a parallel connection of only two equal loads.
Each coil we can think of as a battery that gives off a variable voltage. Two coils from the star would represent two batteries connected in series. From two batteries connected in series, each of 1.5V, we can get two different voltages: 1.5V or 3V. It is similar with the coils: from them we can get a voltage of 220V or 380V.
The leads from the coils of a generator go not just to one three-phase motor but thousands of three-phase or single-phase motors and all sorts of other electrical equipment. Electrical engineers design the grid so that consumption is, as much as possible, evenly distributed over each of the three phases. As we saw earlier, increased consumption causes greater resistance to the turning of the generator’s rotor. If a considerably higher load were applied to one phase than to the other two, then the resistance to the rotor’s turning would not be equal everywhere on the perimeter of the circle and it would lead to turbulences.
Because of the often considerable distance between the place of generation and the place of its consumption, the current weakens to some degree. It does not mean that the electricity flow is weaker near the consumer than at the output of the generator, because in a closed circuit the electrical flux is still the same everywhere. The weakening is actually a partial loss of the energy produced. Although the cables of the transmission lines have a small resistance because they are made of copper (more recently of aluminum because of its lower specific mass) and because of their considerable thickness, the losses are still not negligible. The losses cannot be completely avoided, but can be reduced. This is achieved by increasing the voltage of the current produced to high levels by means of transformers (step up transformers).
Thinking of the aforesaid, we can imagine a mechanically stretched wire, several kilometers long. Near one end we strike the wire with a metal object: the tighter the wire is, the more the effect is transferred to the other end - the looser it is, the less is the effect transmitted. However, the longer the wire is, the greater the tension should be in order to obtain the same effect at the other end as with a shorter wire.
What the tension of the wire is in the mechanical sense, that is the voltage in an electrical sense. The farther the power is to be transmitted, the higher the voltage should be in order to maintain approximately the same degree of attenuation of the current as with shorter distances. The high voltages, which nowadays reach up to one million volts, must be reduced by means of transformers before they reach the consumer, because no electrical device operates at such high voltage.
The transformer is an electrical element that serves to increase (step up transformer) or to decrease (step down transformer) the AC voltage to the desired level. Principally it consists of two independent coils wound on a single iron ring. The dependence of what happens in the coils’ windings is established by the twisting and untwisting of the magnetism in the ring.
The coil connected to the source of the current is commonly called “primary”, but we would like to call it an ingoing coil. This causes the twisting and untwisting of the magnetism in the ring, which evokes the current in the other coil connected to the load. It is usually called “secondary” - we will call it an outgoing coil. The ratio of the voltages of the ingoing and outgoing coil depends on the ratio of the wire lengths of their windings. If the two coils have equal windings lengths, there will be no difference between the input and output voltage. When we talked about cells, we have seen that they can only give off gradual voltages

(e.g. 1.5V, 3V, 4.5V, etc) when several of them are connected in a series battery. Regarding the coil, we can think of every centimeter, or millimeter, or micrometer etc. of its length as one cell. In other words, we have here a continuous voltage source and can get any voltage value - but of course only variable and alternating voltage. If the wire of the outgoing coil is longer than the wire of the ingoing coil, then the output voltage is greater than the input voltage; but, if shorter, then the output voltage is lesser than the input voltage. In this respect it holds that $L_1/L_2=U_1/U_2$, that is, the ratio of the input and the output voltage is equal to the ratio of the wires' length of the input and output coil. Since the transformer is only a transducer of energy, it is also true that the power at the input must be equal to the power at the output. In other words, the product of the input voltage and the input current is equal to the product of the output voltage and the output current ($U_1 \cdot I_1=U_2 \cdot I_2$).

In the transformer we encounter an identical picture as in the lever. For the latter it holds that $m_1 \cdot r_1=m_2 \cdot r_2$, where m_1 and m_2 are the masses on the different sides of the lever and r_1 and r_2 are the distances of the masses from the fulcrum. But for the lever one can also say that $m_1 \cdot v_1=m_2 \cdot v_2$, where v_1 and v_2 are the velocities of the masses on the different sides of the lever. In the formula $U_1 \cdot I_1=U_2 \cdot I_2$ we can use Q_1/t instead of I_1 and Q_2/t instead of I_2. The time t can be eluded because it is the same for both coils, so we get $Q_1 \cdot U_1=Q_2 \cdot U_2$. Now we can draw equivalence between m and Q ($m \approx Q$) and between v and U ($v \approx U$), because the velocities of the two masses can be compared with the vibration speeds of the EM-forces in the two coils, i.e. with their voltages.

Let's take a look at some simple experiments:

<u>Experiment No.1:</u> We connect a light bulb of 60W to AC 220-240V. While the bulb is lit, we first touch the phase and then the neutral with the phase tester (PT) (one-contact neon test light). The lamp of the PT lights up only when the phase is touched. In this and all the subsequent experiments it is understood that we touch the free end of the PT with a finger.

<u>Experiment No.2:</u> The end of a wire led out from a lamp socket with a 60W incandescent bulb we connect to the phase, while the end of the other wire remains free. The bulb does not light up. We touch the end of the free wire with the PT. The lamp of the PT lights up.

<u>Experiment No.3:</u> We connect two bulbs in series, one of 60W, the other of 100W, to 220-240V. The 100W bulb lights up belatedly and is weaker than the 60W. The 100W bulb is closer to the phase. While the bulbs are lit, we use the PT to touch three points in succession: first the phase, then the wire between the bulbs and then the neutral. The PT lights up only in the first two cases, in the second case somewhat weaker than in the first.

<u>Experiment No.4:</u> The same arrangement as in Experiment No.3, except that now the positions of the bulbs are changed. Of the three touches, the PT lights up only when the phase is touched.

<u>Experiment No.5:</u> The same arrangement as in Experiment No.1, but instead of the neutral wire, we now use a water tap or a central heating pipe (if there is an unpainted spot) or the third wire of the AC installation, called "earth". It happens the same as in the first experiment. In this experiment, there is no need to fear that someone somewhere may touch the water or heating installation.

<u>Experiment No.6:</u> One lead of a 60W incandescent lamp is connected to the phase; the other we connect once to a larger metal fence, another time to a smaller metal fence, both of which are mounted in concrete, i.e. insulating ground. In the first case, the incandescent bulb lights up noticeably brighter than in the second. The current measurements in our particular arrangements show 0.23 amps in the first and 0.17amps in the second case.

<u>Experiment No.7:</u> We take several different lengths of insulated wire. With one hand we hold the PT to the phase of the socket; with the other we successively touch the pieces of wire at the free end of the PT, starting with the shorter ones, then moving to the longer ones. As we change the pieces of wire, the lamp of the PT lights up more. Our body has no affection the results because the pieces of wire are isolated.

<u>Experiment No.8:</u> We measure the voltage between the phases of two different sockets. If the voltmeter displays a voltage of about 400V, it means that they are different phases (if both sockets are connected to the same phase, the instrument will display 0V). Between two such different phases we connect four 40W light bulbs in series. While the bulbs are lit, we touch with the PT first one of the phases, then the wire

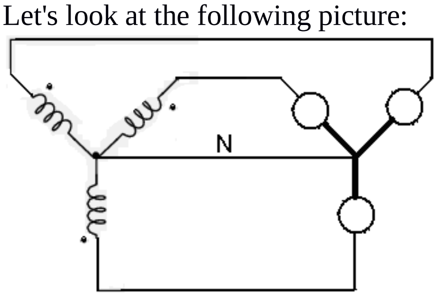
between the first and the second bulb, and then the wire between the second and the third bulb, namely the “middle” of the circuit. In the first point the PT lights up the strongest, in the second weaker and in the middle the weakest. We can do the same from the other side, i.e. from the other phase and observe the same.

Experiment No.9: Similar to the previous experiment, this time we connect two incandescent 60W bulbs to 400V. While the bulbs are lit, we touch the PT to one of the phases and then to the middle of the circuit. Here the PT lights up weaker than with the phase. Then we connect a third bulb of 60W from the third phase of the AC installation to the middle of the circuit. Now all three bulbs are lit. With the PT we check the middle of the thus formed star of wires. The lamp of the PT does not light up at all in the middle. However, it lights up a little if the third bulb instead of 60W is, let’s say, 25W.

Experiment No.10: In the experiment No.9 we made a circuit of only two 60W bulbs between two different phases. Now we measure the current in this circuit. It is about 0.22-0.23A. From its midpoint we lay a wire to the neutral of the power supply, thus giving us three branches. If we measure the current in each of them, we notice that it is the same in all, namely 0.22-0.23A. If we now connect a third 60W bulb from the third phase to the midpoint, the current in the three phase branches is the same (0.22-0.23A), but in the fourth branch, i.e. in the neutral wire there is no current. In this experiment, it is possible that some small differences in the three phase branches may appear, because the incandescent bulbs, although they are all 60W, are not ideally equal.

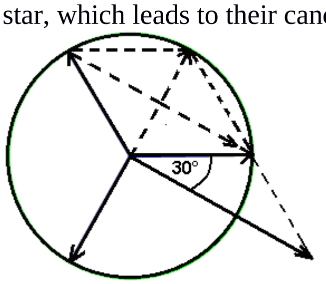
Experiment No.11: Please strictly follow the experimental setup! Standing on dry, isolated ground, we are holding with a bare hand a bare phase wire, being careful not to touch something metallic at the same time with the other hand, or accidentally touch the neutral. We don’t feel anything. This experiment is not dangerous for the following reason: our body is the ending of the open circuit and in such circumstances the voltage of 230V is not high enough that the electrical flux in the body would be strong enough to feel it. But if the voltage of the phase were a few thousand volts, then even if our body would float in the air, an electric shock will happen for sure upon touching the phase wire. Birds sitting on bare power lines carrying thousands of volts are not electrocuted, not because they do not touch anything else, but because their legs are good insulators. In addition, their bodies are small masses (1-2kg).

Let's look at the following picture:



The left part of the figure shows a generator with three coils. Let’s suppose that each coil generates 230V (it refers to the mean voltage). In practice this is the last transformer before the end users. We should not say a transformer, but transformers, namely three of them – one for each phase. Above we have said ”generator”, because the outgoing coils of the transformers can be treated as a three-phase generator, as far as we ignore what's behind them. Now, if we connect one end of each coil in a star and bring out another wire from this midpoint, this wire is called “neutral”. From the generator there are now four wires, three phases and one common wire, neutral. As we can conclude from the experiment No.10, no current will flow through the neutral wire if all the phases are equally loaded, because then they complement each other ideally, so that nothing is left for the neutral. But what do we mean by the neutral conductor? Once this comes out of the transformer, we can call it a trunk. Then it branches off into many buildings and again each branch splits into many apartments of the end consumers. The same applies to the phases. Current still flows in the branches of the neutral (just as in our experiment No.10 it has flowed in the parts from the light bulbs to the midpoint of the star) because these sections of the neutral become parts of the phase lines (in the figure above they are indicated slightly bold). The midpoint of our mini-star is in the larger picture the point where the trunk of the neutral wire begins to branch. This first branching point of the neutral is actually the last point where all

currents cancel each other out as long as the other branching points have left some currents uncancelled. The cancellation occurs to a significant extent because a large number of consumers are connected. If we throw a coin ten times, the result could be 7 to 3, or 70% to 30% in favor of one or the other coin side. But if we throw the coin thousands of times, the result will be very close to 50% to 50%. The same is true for the electricity. The higher the number of consumers is, the greater the likelihood that the load on the phases will be distributed evenly.
The sum or the difference of two sine functions of the same frequency, which have a certain phase difference, is again a sine function. If both functions have equal maxima, then their sum has exactly the same maximum only if their phase difference is 120°. The maximum of the sum of the functions lies exactly in the middle between the maxima of the other two, at 60° from the one and from the other. This results from the following: if we take two segments out of the star of the Mercedes sign as the basis for forming a rhombus, the shorter diagonal of the rhombus is equal to the radius of the circle, since it divides the angle of 120° into two of 60°, giving two equilateral triangles (figure below). This diagonal is exactly in line with the third segment of the star, which leads to their cancellation if we treat them as vectors.



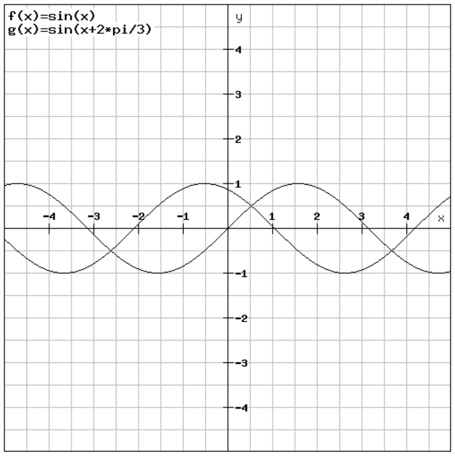
The word "cancel" refers only to the mathematical operations. In the reality of the electrical systems it means something quite different, namely that the system ideally closes in itself so that maximum utilization of the applied energy is achieved, because - as we have already described - the phases take over their currents among each other, that is, they complement each other so that the neutral remains without current or possibly with only small current. If the trunk of the neutral conductor still carries a considerable current, it means that one phase is loaded more than the others, which leads to turbulence in the generator's rotation, which in turn means that some of the applied energy is lost.
If we want to determine the length of the other diagonal of the rhombus, we come to √3, on the condition that the radius is 1. This follows from the theorem of Pythagoras: half of the smaller diagonal is a cathetus, and the side of the rhombus is the hypotenuse of one of the four triangles into which it is divided by the diagonals. For the half of the longer diagonal we get:

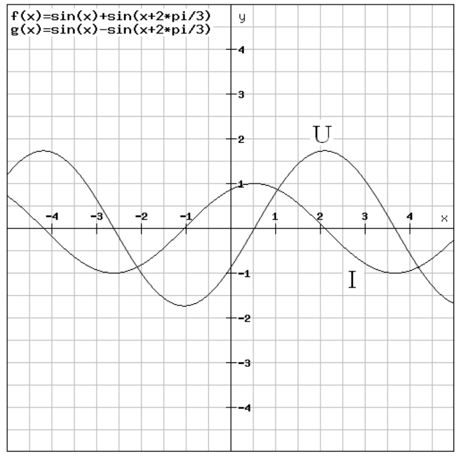
$$(\frac{1}{2})^2 + x^2 = 1^2$$

$$x^2 = 1 - \frac{1}{4}$$

$$x = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$

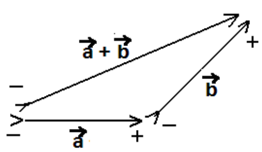
The whole diagonal is thus √3. What does this diagonal represent? It represents the voltage between two phases, the so-called line voltage. If the mean voltage between a phase and the neutral (called phase voltage) is 220V, then the line voltage is 220·√3 = 220·1,73≈380V (usually the phase voltage is 230V, so the line voltage is ≈400V).
The first figure below shows two sine functions that have a phase difference of 120°. They represent the currents or the voltages of two different phases. The second figure also shows two sine functions. The one has the same maximum as those on the first figure. It represents the sum of the currents, i.e. the sum of the functions from the first figure.
$I_1 + I_2 = \sin x + \sin (x+2\pi/3)$
The other sine function has an increased maximum by a factor of √3 (≈1.73) compared to those on the first figure. It represents the difference of the voltages, i.e. the difference of the functions from the first figure.
$U_1 - U_2 = \sin x - \sin (x+2\pi/3)$

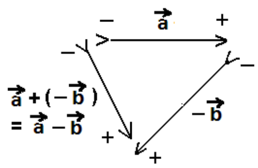




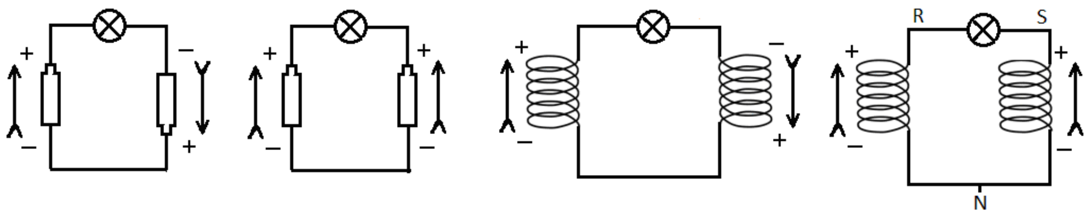
From the second figure we also see that the functions for the current and the voltage differ in phase by 90°. This was to be expected because the diagonals of the rhombus are also at an angle of 90°. What is this phase difference and what are these by 90° shifted functions related to? The function with the larger maximum refers to the voltage between the two different phases; that with the same maximum as the two basic functions refers to the current through the neutral conductor.

When we have to add two vectors (the vectors can be seen as arrows: the front of the vector is plus, and the back is minus − >⟶ +), then we translate one vector so that its minus matches the plus of the other vector (figure left).
When we subtract two vectors ($\vec{a} - \vec{b}$), then again we add two vectors $\vec{a}+(-\vec{b})$, that is, we first change the direction of $\vec{b}$ by 180° and then add it to $\vec{a}$.

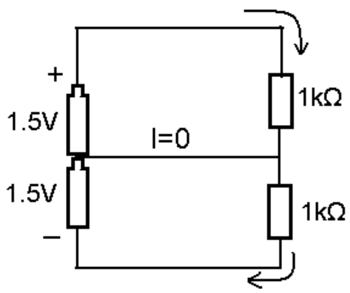




What is the reality behind the addition and the subtraction of two vectors? The first figure below shows a circuit with two 1.5V batteries and a bulb. It will light because the vectors, i.e. the battery voltages add up to 3V total. The bulb in the second circuit will not light because the vectors are in contrary directions, so the voltages are subtracted. In the third circuit, the voltages of the two phases R and S are added because their windings turn in contrary directions. In the fourth circuit, the coil voltages are subtracted because their windings turn in same direction. Now, for the last two circuits, by analogy with the first two we could say that in the third circuit the bulb will light, and in the fourth it will not. This is true only if the magnets, let's say from above, simultaneously enter and exit the coils with the same pole forward; that is, if there is no phase difference between the voltages. But since there is a phase difference of 120° between the phases in a three-phase system, the bulbs will shine both in the third and in the fourth circuit, in the fourth even stronger, because, as we have seen, this voltage is represented by the difference between the two vectors, that is, by the longer diagonal of the rhombus (pay attention to the placement of the dots in the figure further above; they replace the arrows by the coils in the figures below). If in the circle above we draw all the three longer diagonals for the three voltage combinations U_{RS}, U_{ST} and U_{TR}, then we get another Mercedes sign with 1.73 times longer arms and 30° offset in comparison to the arms of the basis sign. It means that in the star connection the line voltages are phase-shifted by 30° in respect to the phase voltages.



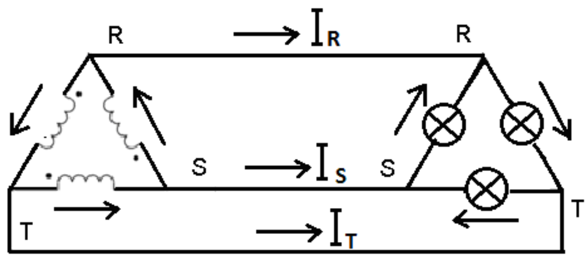
** If two cells of 1.5V are connected to two identical loads (i.e. two resistors) in a circuit as in the figure below, then no current will flow through the middle branch (when a similar configuration is used in electronics, which is often the case with operational amplifiers, the middle branch is called "virtual ground"). If the resistors or the cells don't have equal values, then some current will flow through the middle branch.



Something similar exists also in the alternating current supply in the United States (and more or less also in other countries), where in rural areas only one phase and the neutral enter in the last transformer (in England only two phases) before reaching the final users. The voltage between the ends of the outgoing coil is 240V. From the midpoint of this coil a third wire is pulled out (the so-called center tap wire). There are now three wires for every user. Between any end of the outgoing coil and the center tap wire the voltage is 120V. Bulbs and less powerful home appliances work at this voltage, and more powerful appliances at 240V. If the two "phases" are loaded evenly, then no current flows through the middle, that is, the neutral wire. This wire is additionally grounded in case of an excessive imbalance in the loads.
This, however, is not called a two-phase current, but a "single-phase three-wire system". The two "phases" have a phase difference of 180°.

Let us look on the first graph above the moment in which the basic functions intersect (t= π/6=30°≈0.52), that is, when both functions have equal values. At that moment the voltage function has the value 0 (second graph) and the current function the maximum value (what we explain here corresponds to the first part of experiment No.10, i.e., the part without the third bulb). At that moment, both phase currents have exactly half their positive value, that is to say, both blow with half force to the midpoint and collide, which in turn means that the two summed up will find their way to the neutral conductor. In regard to the voltages, it's as if at that moment we have two cells with equal voltages, facing each other with their positive poles, so their voltages (not individually, but as a group) cancel each other out. Therefore the value of the voltage between the two phases is zero.
Let us consider another moment on the graph t=2π/3 (≈2.1) when both basic functions have the same, rather large value √3/2 (0.87), but with different signs. At that moment, the value of the current through the neutral is zero and the value of the voltage between the two phases is maximal. Since the currents are of equal magnitude but of opposite signs, this means that the blowing of the one phase meets the equally great suction of the other phase at the midpoint, so that the phases ideally complement each other, leaving the neutral without current. The two phase voltages add up. At this moment we can compare it to two cells connected in series, but this time the positive pole of the one joins the negative pole of the other. At this moment the current through the bulbs is the strongest.
The figure below shows a generator whose coils are connected in a triangle, and which is evenly loaded with three lamps, also connected in a triangle. As we saw above when we talked about this connection, each right end of the coils is connected to the left end of another coil. Hence, the arrows of the current

vectors in the coils, and also in the loads, must be closed in a circle, whether to the left or to the right. The arrows of the line currents (I_R, I_S, I_T) must all point either from or to the generator. For I_R, according to the figure shown, we get $I_R = I_{SR} - I_{RT}$ (since $I_{SR}=I_{RT}+I_R$ on the generator side). With the star connection we distinguished line and phase voltages; here, we distinguish line and phase currents. There, the line voltages were differences from the phase voltages; likewise, here the line currents are differences from the phase currents (I_{SR}, I_{RT}, I_{TS}) - $\sqrt{3}$ times greater and 30° shifted in respect to the phase currents. So everything is the same as in the star-connection, only the voltages and currents have exchanged places in the mathematical expressions.



We can characterize the star- and the delta-connection as a plus- and a minus-connection, but on the different sides of the circuit the characterization is reversed. On the generator's side, the star is plus, and the delta is minus-connection. On the load's side the star is minus, while the delta is plus-connection. For comparison: if we connect two identical cells in series, then we get twice as much voltage, and therefore a stronger current through a resistor. If we connect them in parallel, then the voltage remains the same, the current is weaker than when they were connected in series, but it lasts longer. So in the first case there is more power over a shorter time, and in the second case, there is less power over a longer time. Therefore, on the generator's (or the batteries') side, the series connection is plus, and the parallel minus.

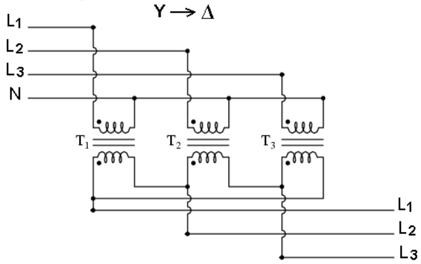
If the outlets of two equal containers with equal amounts of water are connected together in a single outlet, then the pressure in it does not change - this corresponds to the parallel connection of two cells. The series connection of two cells corresponds to the pouring of the water from one container into the other.

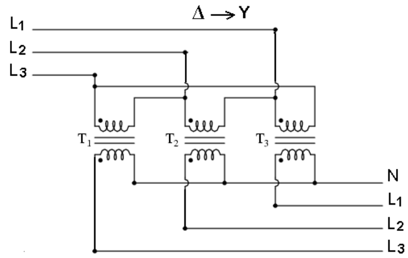
If we have two cells and two resistors, we will obtain the greatest power if we make a series *(+)connection* from the cells on the generator's side, and a parallel *(+)connection* from the resistors on the load side.

In the delta-connection on the generator's side, the line voltages are the same as the phase voltages (220-240V), because, as we have seen, the right end of a coil enters the left end of the next, so these voltages add up. But due to the phase difference of 120°, the sum of two such voltages is the same as they themselves; what changes is only the phase position of the summed voltage. This summed voltage is equal to the voltage of the third coil, and these two are placed in a parallel in respect to the loads.

The delta-connection on the generator's side, more precisely on the transformer's output to the end users is not utilized. Therefore, if we at the household AC make a delta-connection of loads (we will not use the neutral), then on the generator's side we have (+)star-connection (380-400V), and on the load side (+)delta-connection, whereby the power is the greatest. (We cannot make delta-connection with only three lamps because they can withstand up to 240V, but we can with six of them, two in each branch connected in series.)

If the coils' ends of the generator in the power plants are not interconnected, then to the consumers should travel six wires. With the star-connection that number is reduced to four, and with the delta-connection to three. This is why the latter is utilized to transmit the three-phase current at a distance. If the generator is in a star-connection, then the first set of transformers, apart from boosting the voltage, additionally performs the transformation of the star into delta-connection. The figure below shows a three-phase set of transformers that transforms the star- into a delta-connection. As we can see, all the ends of the input windings of the transformer without dots are connected to neutral, while the ends of the outgoing windings are connected in a circle. This is already known to us from above. In the picture on the right we have the same, only the sides are reversed. This set of transformers will lower the voltage and at the same time again produce the neutral for the end consumers.





Due to the often very high voltage at which the current is transferred along the transmission lines, sparks may occur between the phases themselves or between them and the metal pillars, trees, or other things (under high voltage transmission lines there must be no trees, and the grown ones must be regularly cut). In order to reduce the possibility of arcing, the phases are separated at a significant distance; additionally, each phase is transmitted through 2, 3, 4 or even 6 parallel, so-called bundle conductors, set at a distance of 30-40 centimeters as vertices of a segment (in the case of 2 wires) or of a triangle (in the case of 3 wires) or of a square (in the case of 4 wires). The higher the voltage is, but also the higher the current intensities being transmitted, the more wires per phase. Between them, of course, there is no arcing because they belong to the same phase. With this mode of transmission, the current is distributed through several conductors, thus the possibility of arcing is smaller.

Another wire travels along the tops of the transmission towers to protect the phase wires from lightning during storms. It is therefore placed on the top, for the lightning to strike at it. This wire is connected to the ground at every pillar.

In addition to the phase [according to the newer European standards with brown (L1-line1), black (L2-line2) and gray insulation (L3-line3), that is, one color per phase] and the neutral (with blue insulation), there is a third conductor in the sockets. This so-called protective conductor (protective earthing – PE) or popularly called “grounding” (with yellow and green color together) leads to a metal plate buried in moist ground near the transformer or elsewhere. The purpose of this conductor is mainly for electrical devices that have a metal housing (washing machines, cookers, refrigerators etc.) It is possible for a phase wire inside the device to come unhooked or the plastic insulation to be damaged and the wire to touch the metal housing. In that case, the whole housing becomes a phase, that is, potential danger. To avoid this, the protective conductor is attached to the housing. If the phase somehow comes in touch with the housing, then the protective conductor becomes a large mass for the phase’s potential, a strong current begins to flow to the ground, the fuse in the installation blows out and interrupts the current, thus removing the danger.

To protect human life, but also to protect against fire, the so-called RCD switch is used (residual current device). Both the phase and the neutral conductor pass through this device, whereby the currents are compared. Normally they are equal [we said above that the neutral wire becomes part of the phase line up to the first (seen from the load’s side) star point]. If there is a greater difference than what has been established as a threshold between the current in the phase and that in the neutral (e.g. 30mA), then the switch reacts and cuts off the current. For instance, our experiment No.6 would not have been feasible if the electrical installation were equipped with this switch, because a blind current[11] of 170 and 210mA respectively was measured toward the fences mentioned there, which the 30mA wide exceeds, so that the switch would have instantly interrupted the flow, because the current of 170mA passes through the phase, but does not continue through the neutral. We have mentioned the value of 30mA because these devices are often set with this value to protect people from electric shock. Of course, there are also switches that are set for greater differences, but these are not intended to protect human life.

[11] What in the English language is called "residual current" and in the German "Fehlerstrom" (fault current), we call it here "blind current".

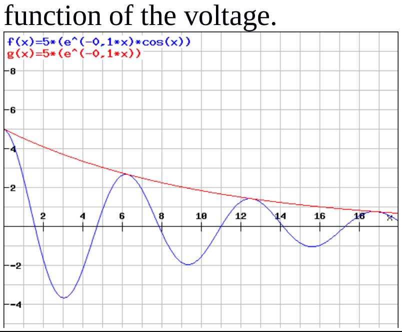
Regarding the experiments on the different intensities of the light of the phase tester (experiments No.3,4,8,9), the reader can explain on his/her own. For comparison, it is enough to imagine a large container with water that has a horizontal outlet in the lower part, i.e. a pipe of considerable length. Although the water flow through the pipe is everywhere the same, the pressure is weaker in the parts of the pipe that are more distant from the container.

●●●●●

If a capacitor is connected to an AC source, then depending on its elastance and the frequency of the voltage source, three situations can occur in a certain sense: 1. the capacitor is not present in the circuit, that is, it behaves like an ordinary copper wire; 2. the capacitor blocks the current, i.e., the circuit is open, and 3. the capacitor acts as an unusual resistor, with greater or lesser resistance.

The voltage at the ends of a capacitor and the current flowing through it are always shifted by 90°. To illustrate this, we will use a comparison. At one end of a thread we attach a small weight, the other end we attach to the ceiling. Then we turn the weight so that the thread is twisted. When the weight is released, it begins to turn in the contrary direction whereby the thread is untwisting. At the moment of complete loosening of the thread, the turning speed is maximal, and then the thread begins to twist in the contrary direction. When it peaks up to a maximum, the process stops, and then it starts over again. If we call the *twist* of the thread "voltage" and the *turning* of the thread "current", then it can be said that when the voltage is maximum, the current is zero, and when the current is maximum, the voltage is zero. The current and the voltage are sine functions phase-shifted by 90°.

Because in the experiment with the thread and the weight there is no added energy, the (co)sine function slowly decreases. Here we have the exponential and the sine function in one place. The figure shows the function of the voltage.



Something similar happens in the capacitor: the current and the voltage are always phase-shifted by 90°.

If a capacitor is connected to an AC source, it will start to swing (here we call 'swinging' its twisting, then its untwisting to null, and then the same, but in the contrary direction). How intensely it will swing (i.e. how large the amplitudes of its oscillations will be) depends on its capacitance and on the frequency of the source. The capacitor receives energy from the source, stores it in itself, and then returns it. By doing so, it opposes the source and hinders it. If it hinders it a lot, then the current in the circuit will be small.

In fact we have here two oscillators, the source and the capacitor, which must be harmonized at some level and find a kind of compromise to oscillate in a stable regime. At what level this will happen (i.e. how strong the current will be) depends, as we have said before, on the physical characteristics of the oscillator (the capacitance) and on the frequency of the source.

If a capacitor with considerably great capacitance is connected to the household AC, then a very strong current will flow through the circuit and soon the fuse will blow out or some other thin spot in the circuit will melt. In this case, the capacitor is swinging very little, with small amplitudes. Since its capacitance is great, it receives a great amount of electricity at a very small increase of its voltage (just as the level, i.e. the pressure of water in a very wide container will rise fairly little even when large amounts of water are added). Since the frequency is not very low, the capacitor does not have enough time to twist considerably.

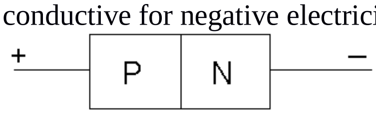
If to the same source we connect a capacitor with a small capacitance, a weak current will flow in the circuit. Due to its low capacitance, this capacitor needs a very small amount of electricity and thus also time to twist to a considerable extent. Hence, it will come to twist to a substantially high voltage even at high frequencies (as a very narrow and high container of water). In this case, the capacitor is swinging a lot. With the great voltage to which it is twisted, it opposes the source and hinders it.
So, when a capacitor is swinging a little, strong current flows in the circuit; and, when it is swinging a lot, small current flows in it. Large oscillations can be done by a low capacitance capacitor. There is only a small amount of electricity entering it, as it is the current in the circuit.
From the foregoing it follows that the capacitor’s resistance, called capacitive resistance X_C, is the smaller, the greater its capacitance and the higher the frequency are. So for the X_C we get:

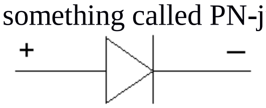
$$X_C=\frac{1}{2\pi\cdot f\cdot C}$$

The capacitive resistance is expressed in ohms. For example, a capacitor of 1μF at a frequency of 50Hz would have a resistance of 3257Ω. Connected to 230V, a current of 70mA will flow through it (230/3257=0.07A).

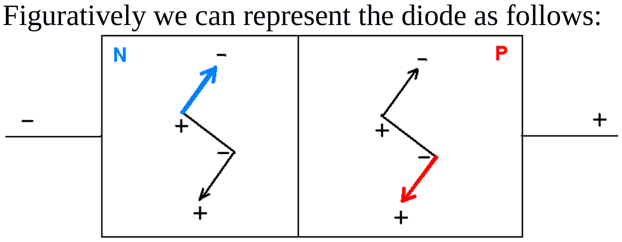
SEMICONDUCTORS

Semiconductors are semimetals (silicon, germanium, selenium, tellurium etc.) in a very pure and very regular crystallized state (intrinsic semiconductor). However, in this state they show much lower conductivity than metals and are practically unusable in the electrical technology. But if they are acted upon with certain substances in a certain way, they become very good conductors (extrinsic / doped semiconductors), yet only for one of the two types of electricity, either plus or minus. In other words, a kind of polarization occurs. The action of certain substances (boron, gallium, indium etc.) makes the semimetal sensitive to plus electricity, and the action of others (phosphorus, antimony, arsenic etc.) sensitive to negative electricity.
Now, if a piece of pure and regularly crystallized semimetal is acted upon on one side with a substance that makes it conductive for plus electricity and on the other side with a substance that makes it conductive for negative electricity, then something called PN-junction or diode is gotten.

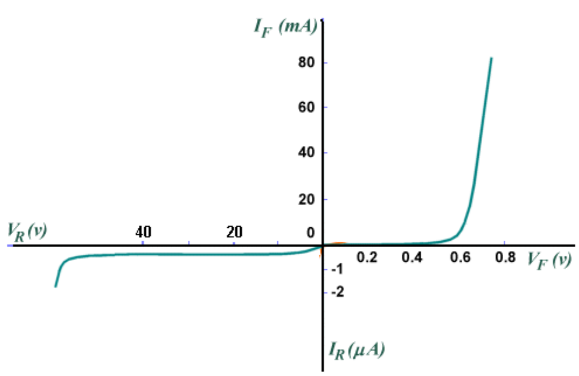




As soon as voltage is applied to the diode in the forward direction (i.e. the plus-pole of the source to the P-region, the minus-pole of the source to the N-region of the diode), then only a negligible, so-called leakage current flows through the diode up to the so-called barrier voltage (that is 0.3V for the germanium and 0.6V for the silicon diode). Above this voltage, the current begins to rise sharply, but the voltage at the ends of the diode changes only slightly. For certain silicon diodes, it can rise to maximum 1.1V, while the current increases many times over. We could therefore say that the EM-forces in the semiconductor have a sort of inertia, so it takes a certain minimum of force (i.e. voltage) to get the process going. If a reverse voltage is applied to the diode (reverse biased), then an insignificant leakage current will flow through the circuit up to the so-called breakdown voltage (depending on the design of the diode, it can be from few volts to thousands of volts). After this voltage, the current begins to increase, but it is now a great number of times smaller than that in the first case.
Figuratively we can represent the diode as follows:

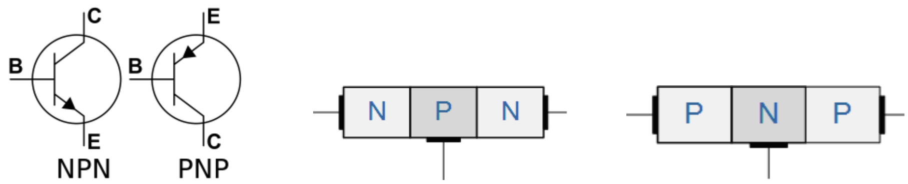


The sensitivity of the right region to the plus-electricity is represented by red coloration of the plus-segment and at the same time by its inclination to the middle of the diode, while the sensitivity of the left region to the minus-electricity by blue coloration of the minus-segment and by its inclination also toward the middle of the diode.
Our suggestion to explain the phenomenon is as follows. Exposure to certain substances makes the crystal sensitive to the one type of electricity and indifferent to the other type. Connecting in the circuit a piece of crystal treated with only one substance would be like connecting a resistor with a small resistance value. But if the piece of crystal has been treated on both sides with different and opposite substances, then depending on its position in the circuit with direct current, it will either block the current or let it pass through. When in the forward direction, the plus-segments of the EM-forces in the P-region of the semiconductor are set in motion by the positive pole of the battery, and these in turn set their neighboring minus-segments in motion. It is likewise for the N-part of the diode. Thus, the plus- and minus-segments of the P-region will connect up to the plus- and minus-segments of the N-region respectively, and the electricity flows.
The increase of the current in the diode will result in an increase in its temperature, which increases the elasticity of the EM-forces in the crystal. In metals, electrical elasticity decreases with temperature rise, in semimetals it increases. We could say that the elasticity of the semiconductor also increases with the rise in voltage (voltage dependent elasticity). That's why it cannot rise a lot at the ends of the diode, for as the voltage rises, the resistance falls. What is the share of the temperature, and what is the share of the voltage in the increase of the elasticity of the EM-forces in the semiconductor, it is a question that requires more investigation.
When we apply a reverse voltage to the diode, only insignificant current flows in the circuit up to the breakdown voltage. After this voltage a polarity reversal takes place, i.e., a reverse spinning of the EM-forces begins; but this is a very small current (usually in the order of microamps) because the resistance of the forces to the spinning, to which they have no inclination, is huge. We can compare this with the battery with zinc and lead plates. The zinc plate had reversed the polarity of the lead plate, but because the first had partly used up its electromotive force to maintain the reverse polarity of the second, it had diminished, i.e. their forces were subtracted. So, we can imagine that the great reverse voltage uses up a great deal of its force to attain and maintain the reverse spinning of the forces in the P- and in the N-region of the diode.
When looking at the U/I graph of the diode (figure below), the reader should pay attention to the fact that the axes don’t use the same scales in their two sections. To the right of the Y-axis, one centimeter on the X-axis represents let’s say 0.2V, but to the left of the Y-axis one centimeter could represent 10V or more. Above the X-axis, one centimeter on the Y-axis could represent 20 milliamps, and below the X-axis one microampere, that is, thousands of times lower current.

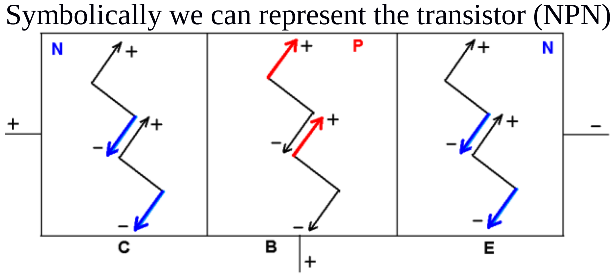


Another important semiconductor element is the transistor. As we saw at the beginning of this book, it is composed of three segments (NPN or PNP) called emitter, base and collector. The base is what we used to call heart. The transistor can be understood as a variable resistor whose resistance value varies as a function of the current at the input of the base (trans-resistor). The greater this current is, the more

conductive the transistor or the lower its resistance is. This varies to a great extent, from nearly zero to the order of megaohms.
In the symbol for the transistor the arrow is always drawn on the emitter terminal and shows the direction of the current from plus to minus.

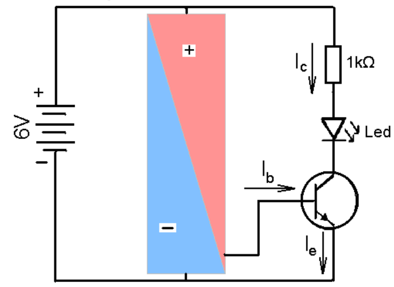


If the emitter is sensitive to negative electricity (NPN), it is connected to the negative terminal of the battery. The reverse applies to the emitter of the PNP transistor. Because the transistor appears symmetrical with respect to its middle part, we can reverse it in the circuit; but, in that case, we won't obtain the same amplification of the current arriving at the base as in the previous case, because the internal symmetry is not exact.
At the beginning we have seen that it is necessary to actuate the base with little corresponding electricity to make the transistor conductive. That is its essence: a small current to the base initiates a larger current through the transistor, which in turn can turn on or set something in motion with its power.
Symbolically we can represent the transistor (NPN) with the following picture:



When there is no positive current to the P-base, then the battery plus-pole cannot penetrate up to the base, because the N-sensitive collector is a barrier preventing it. In this case the transistor is not conductive. But if at least a small plus-current arrives at the base, then the plus-segments of the EM-forces in the base are set in motion. They move their neighboring minus-segments in the base. Once in motion, they can actuate their cognate minus-segments of the N-emitter and N-collector. These in turn set their neighboring plus-segments in motion and so the flow of the current through the transistor is established, also thanks, needless to say, to the force of the battery. Although we are speaking here as if of temporal successive phases, nonetheless, it is a simultaneous and unitary process. It is similar to a system of meshing gears, where we step by step explain which gear moves which other, but when the system is set in motion no gear is late behind another.
In support of the just aforesaid, we will refer to the experiment described when we spoke of the term "mass", where the "bridge" was established not through actuation of the EM-forces in the base, but in the emitter or in the collector. We consider the case with the NPN transistor(s): If we move the electrified vinyl plate toward the free end of the wire (the wire is pulled out from the emitter and its free end is far from the circuit itself), then this minus-flux transmitted through the wire will actuate the minus E-forces in the N-emitter. They will set in motion their neighboring plus E-forces and these in turn their cognate plus E-forces in the base. However, the force of this motion will be insignificantly small unless at least a short metal wire (mass) protrudes from the base, because the actuated plus E-forces of the base encounter a barrier at the collector since it is N-sensitive, whereas the metal wire (the mass) is a neutral terrain where the flux can spread out freely and thus the actuation of the plus E-forces in the base will be much stronger. This now stronger motion of the plus E-forces in the base will be strong enough to actuate their neighboring minus E-forces, which in turn will set in motion their cognate minus E-forces in the collector and these their neighboring plus-forces, of course, with the cooperation of the battery plus-pole that enters

here. So the circuit closes and the lamp lights up. The same applies when the stimulus acts on the wire extending from the collector.

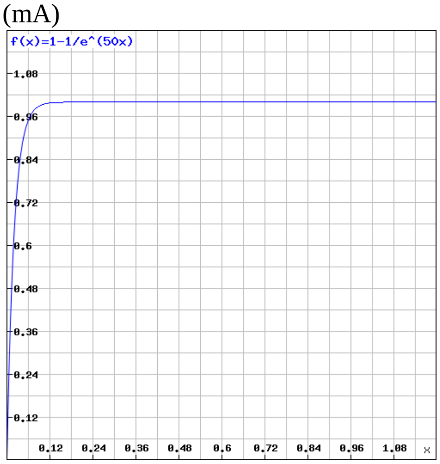


The picture on the left shows a simple circuit with a transistor, an LED-lamp, a resistor and a high-resistance wire. The red-blue rectangle represents the wire. From bottom to top along the length of this wire, a good conductive transverse wire (whose other end is connected to the base of the transistor) slides evenly. At first the lamp does not shine. If the battery is 6V, then when we reach somewhere at 1/10 of the length of the wire, the LED starts to glow dimly, and as we slide the wire up, it shines brighter and brighter. As the wire approaches the top, the light starts to fade again, to disappear completely at the top.

Coming to 1/10 of the wire length at a 6V voltage source, the point is reached where the plus is already strong enough in relation to the pure minus to which the emitter is connected, so that a small current can flow to the base. Translated into numbers, this means that when the sliding wire is in the mentioned place, the threshold voltage of 0.6V between the base and the emitter has been reached. We see that in silicon transistors, just like in silicon diodes, the threshold voltage of the PN base-emitter junction is 0.6V. The current through the transistor is stronger on the base-emitter junction (BE), and somewhat weaker on the collector-base junction (CB); or, we can say that the current in the emitter branch (I_E) is somewhat stronger than the current in the collector branch (I_C), for it is obvious that $I_E=I_C+I_B$. Since the base current is very small, it is usually considered that $I_E=I_C$. As the wire slides up, the plus becomes stronger, hence the current to the base. As a result of the stronger base current, the current through the transistor increases and the lamp shines stronger. Shortly after the 'starting point' (i.e. 1/10 of the wire length), the transistor's current reaches the maximum; the further sliding up does not increase the current through it. It is said that the transistor has entered into saturation. And here also, as in many other cases in the electromagnetism, there is an exponential dependence; in this case, of the transistor's current (which is usually considered the collector current I_C) from the base current I_B. If in the collector's branch there is only a resistor R_C=1kΩ (we will remove the LED) and the battery is 6V, then the maximum current which can flow in that line when the transistor becomes fully conductive is 6/1000, that is, I_{max}=6mA. The dependence of the collector current on the base current can be expressed by the following formula:

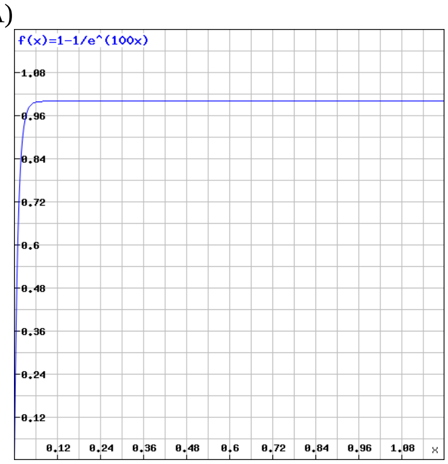
$I_C = I_{max}\,(1-e^{-\beta I_B})$ or $I_C = U/R_c(1-e^{-\beta I_B})$

I_c (mA)



I_b (mA)

I_c (mA)



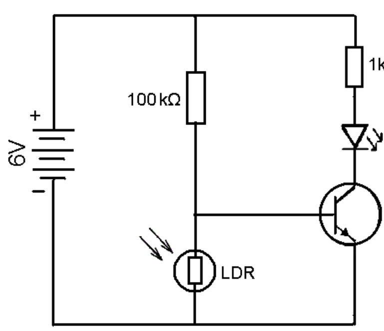
I_b (mA)

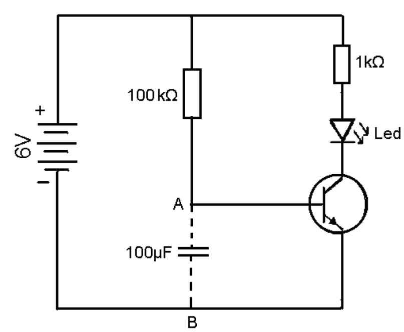
The graphics visually present the I_C-dependence on I_B for two values of β (50 and 100) and a maximum collector current of 1mA (let's say 6V battery and R_c=6kΩ). From the graphics we see that already at a base current of only a few dozen of microamps (μA), the collector current reaches its maximum of 1mA, i.e., it enters saturation. We also see that the function in most of its growth is practically linear; hence, *in* *that part* we can simply replace it with the function $I_C=\beta\cdot I_B$. The number β is called “current gain”. It is often labeled with h_{FE} in cases when the transistor is used as an amplifier in the so-called common emitter configuration.

When we slide the transverse wire to the top (or near it), then we connect the base directly to the (+)battery pole. The lamp stops shining. At that moment, we actually make a short circuit of the battery poles, because the PN base-emitter-junction at a voltage of 6V at its ends has insignificant resistance. Through the circuit flows a strong current, which rapidly uses up the battery, and if this lasts longer, the transistor may suffer damage. The lamp does not shine since very small current flows through its branch because of the huge resistance in it (1kΩ + the resistance of the lamp) compared to the practically zero resistance of the shortened branch.

The transistor can operate as an automatic switch (on-off switch) or as an amplifier. When it is not conductive (cut-off) or when it is fully conductive (saturated), that are the two off-on states. In order to operate as an amplifier, the transistor must not enter any of these states. For example, if it is used as an audio amplifier, then with its occasional entering into those states, a distortion of the sound would occur at those moments. That's why it must constantly oscillate in the linear region of the graph we saw above, from the middle point upwards and from that point downwards, avoiding the extremes. Thence comes the name “linear electronics”, that is, electronics for linear signal amplification. The signal will be amplified, or, we can say, its intensity multiplied by a certain number while fully retaining the original shape.

The figure below (left) shows a circuit where the transistor operates as a switch. The LDR element (light dependent resistor/photo resistor) is a variable resistor whose resistance depends on the brightness of the space around it. As the light level grows, its resistance decreases. The range is very large: from only a few tens of ohms in strong light, to more than 1MΩ in total darkness. When it is dark, the LED in the circuit will light up; and when it is bright, it will go out. If we change the places of the photo-resistor and that of 100kΩ, then the LED will turn on at dawn. By changing the value of the 100kΩ-resistor, we can adjust at what light level the LED will turn on or off.





In this circuit (left) in principle there is nothing different from the one above with the sliding wire. As the light level increases, the resistance of the LDR decreases, so this is equivalent to the sliding of the base wire towards the minus-pole of the battery. The only difference is that the wire of the previous example has a constant resistance, and here the vertical line 100kΩ-LDR has a variable resistance due to the LDR.

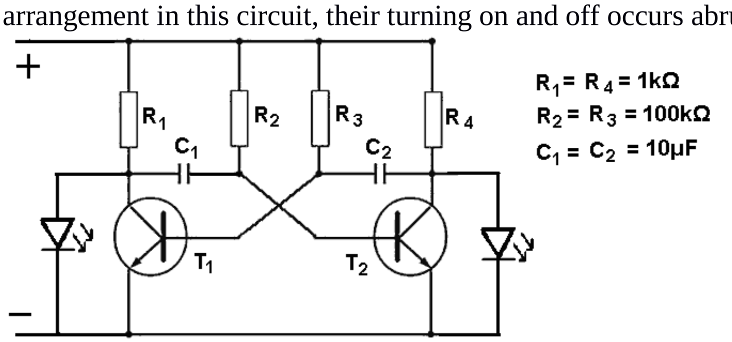
The figure on the right shows a similar circuit to the previous one. At first there is nothing between the points A and B. The LED is lit because a small, but sufficient current flows to the base through the big resistor for the transistor to be fully conductive. While the lamp is lit we add between A and B a great capacitor, say of 100µF. The LED instantly goes out, but after about ten seconds it lights up again. If we put a smaller resistor or a smaller capacitor, or both simultaneously, the LED will light up after a shorter time.

When in the circuit the capacitor is added, then it begins to twist thanks to the current coming through the great resistor. At first there is a greater current flowing through the capacitor, so that in those first moments we can consider as if we have placed a good conductive wire between the points A and B. This means that the base is directly connected to the minus-pole of the battery, and this pure minus turns the

LED off. As the capacitor is twisting, so the voltage at its ends is rising. This again resembles that, as when we were sliding the base wire up. When the capacitor is twisted to 0.6V, then the plus from the left is already strong enough for a small current to start flowing to the right (i.e. to the base), causing the LED to gradually begin to glow, so that at a voltage of 0.65-0.7V there is already enough current towards the base for the transistor to enter into saturation, that is, its resistance to become ≈0Ω, so that the LED shines as strongest as it can at a source of 6V and a series resistance of 1kΩ. Here ends the twisting of the capacitor (0.65-0.7V) and from that moment on we can freely consider as if between the points A and B there is nothing. Without removing the capacitor from the circuit, we connect those points with a piece of wire (we make a short circuit between them). The LED goes out, but after ten seconds it lights up again. When shorted with the wire, the capacitor untwists through it with a considerable, but extremely short-lasting current. Now it is untwisted and the 10-second process starts again.

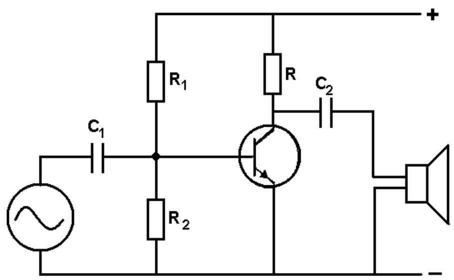
Now we remove the capacitor from the circuit and connect it directly to the 6V battery, so that it is twisting up to this voltage (if it is an electrolytic or tantalum capacitor, we must pay attention to its polarity, i.e. the (+)pole of the capacitor goes to the (+)pole of the battery. These two types of capacitors are polarized, which means that they must not be twisted in the opposite direction. However, it is not a big deal if they are twisted to a low voltage in that direction). Then we return it back to the same place in the circuit, so that the capacitor’s terminal which was attached to the (+)pole of the battery is connected to the point B, i.e., to the (–)battery rail. The LED goes out and it lights up again after considerably more than ten seconds. - When we add the twisted capacitor to the circuit, we add, so to say, another series 6V “battery” to it (thus the battery now becomes 12V). However, in this additional 6V "battery" there is no "inflow", but it is just emptying (i.e. untwisting) through the circle with the 100kΩ resistor. Again, there is a pure minus directly to the base, resulting in the lamp’s going out. Through the 100kΩ resistor, the capacitor needs considerable time to untwist, and when this ends, it begins to twist in the opposite direction, so from now on the already described 10-second process starts again.

Let's take a look at the circuit in the figure below known as the "astable multivibrator", which we like to call a seesaw. It consists of a battery, two transistors, two capacitors, two LEDs and four resistors. The lamps serve only as indicators for an easier understanding of what is happening. When this circuit is connected to the battery, then the lamps turn on and off alternating (at the moment when one lamp turns on, the other turns off). This means that during the period when one lamp is on, then one transistor is conductive, the other unconductive, and then the other way round. As the reader has surely noticed, the lamps this time are not connected with the transistors in series, but in parallel. In this case, they turn on when “their” transistor is unconductive (when T_2 is unconductive, then the right lamp is on). With such an arrangement in this circuit, their turning on and off occurs abruptly.



Let's start with some moment, let's say when T_1 is conducting, and T_2 is blocked. Toward the base of T_1 flows plus-current through two parallel branches, one with the resistor R_3, the other with R_4 and the capacitor C_2, whereby it is twisting. This current is strong enough for T_1 to be fully conductive. This means that its resistance is ≈0Ω, so that while it is conductive, we can freely consider it as an ordinary copper wire which C_1 connects directly to the (–)battery rail. If in this situation (that is, T_1 imagined as a piece of copper wire) we carefully observe the left part of the circuit, we will notice that here we have an identical picture as in the previously described circuit: the line of C_1-R_2 fully matches the upper circuit. So C_1 is twisted up to 0.65-0.7V making T_2 fully conductive. When T_2 is conductive (now it can be

considered as a piece of copper wire), then C_2 is already twisted from before to almost 6V (i.e., the battery voltage), so now we have here an identical situation like when above we have twisted the capacitor separately up to 6V, so twisted we have placed it in the circuit and then we have waited for a significantly longer than 10sec for the LED to light up. At the moment when T_2 becomes conductive, T_1 becomes unconductive. The capacitor C_2 must be first fully untwisted and then twisted in the opposite direction to 0.65-0.7V for T_1 to become fully conductive. Thus the ribbon (∞) closes and the cycle "infinitely" repeats itself.
If we decrease the values of the resistors of 100kΩ or of the capacitors, or of both at the same time, then the turning on and off of the lamps speeds up, that is, they are just twinkling. When the values are significantly reduced, the twinkling changes into seemingly continuous light.
The main application of this circuit is to produce a certain frequency of plus (1) and minus (0) signals (clock) in some electronic circuits.
Let's see the work of the transistor as an amplifier. In the figure below is shown an amplifier with a single transistor.



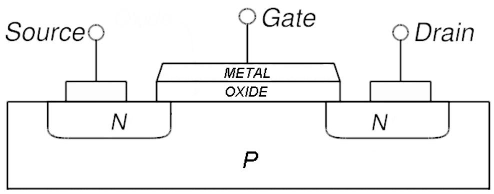
To the left is the source of the signal that needs to be amplified. It can be an ordinary microphone, or a signal from an audio device. The values of the resistors R_1 and R_2 should be chosen so that the so-called operating point of the transistor will be in the middle of the linear part of its above-mentioned curve, so that it will have equal distances to "move" up and down. So even when there is no signal at the input, the transistor is partially conductive and "rests" at the middle point.
The values of R_1 and R_2 can be changed, but the ratio R_1/R_2 should be kept the same. Thereby the operating point of the transistor is not shifted. However, it is not very advisable because the oscillations of the transistor at the arrival of the input signal will be larger if we put lesser resistors, so it will enter both in saturation and in total non-conductivity. If we put greater resistors, the oscillations will be smaller, i.e. the transistor would go up and down only for short distances so that the amplification will be weaker.
The role of C_1 is the following: if we think of the source of the signal as a microphone, then it is a coil wire with a certain resistance. If it were not for the capacitor, then from the plus- to the minus-pole of the battery through the resistor R_1 and this coil would permanently flow a significantly stronger current than the alternating current produced by the sound vibrations in the coil. Moreover, the coil will also lower the operating point of the transistor. Let's take a look at the situation when the capacitor is added. It will twist through the line R_1/C_1/coil and thereby the direct current in its branch will stop. When the microphone starts to vibrate, these vibrations will cause alternating current in the coil, and this in turn will cause oscillations in the twist degree of the capacitor (partial untwisting and twisting again). It is actually an alternating current through the capacitor, which will also go to the base of the transistor. In the periods when a plus-current goes towards the base, then the transistor is more conductive and a stronger current flows through it (from the middle point of the straight part of the curve occurs a "movement" upwards), and in the periods when the base gets a minus-current, then the transistor is less conductive and a weaker current flows through it (from the middle point of the curve occurs a "movement" downwards).
On the output side there is a similar situation. The capacitor is twisted up to a certain degree along the line R/C_2/coil (i.e. loudspeaker) and thereby the current through it dies out if there is no signal at the input. When the transistor's current starts to change the intensity due to the input signal, then during the periods of increased transistor's conductivity the capacitor will untwist through it, more precisely, in the closed loop "loudspeaker/C_2/Tr". However, in the periods of downward "movements", i.e. during decreasing

conductivity, the capacitor will twist again along the line R/C_2/coil. These oscillations of C_2 are more powerful than those of C_1, which means a stronger alternating current through the speaker. The capacitor is always twisted in only one direction, at times more, at times less in respect to the "mean" twist when there is no signal at the input.
So, from the weak input oscillations in the microphone, we've got strong output oscillations in the loudspeaker.

• • • • •

Other important semiconductor elements are the so-called unipolar Field Effect Transistors (FET) (from the point of view of our theory, the designation "unipolar" is a complete misconception). In this type of transistors there are also such, which we would like to call inverse FET. This means: when there is no signal to the heart, right then they are fully conductive; and when a signal arrives, depending on its strength, they become less conductive or fully non-conductive. These transistors are called "Depletion mode FET". Their important representative is the JFET. The 'normal' type of these transistors is called "Enhancement mode FET".
We will examine the most important transistor of the second type: MOSFET (Metal Oxide Semiconductor Field Effect Transistor). Among bipolar transistors we have an NPN-type, which we called a (+)transistor, and a PNP-type, which we called a (–)transistor, according to the type of the base, that is, to the type of electricity it reacts to. Among the MOSFETs there is also a (+) and a (–) type, but because of the current theory which speaks of movement of electrons and holes in the semiconductors, the MOSFET whose heart reacts to (+)current is called N-channel MOSFET, and the one whose heart reacts to (–)current is called P-channel MOSFET.
This transistor also has three segments, hence three leads: source (S), gate (G) and drain (D). The analogy with the previous transistor is the following: the source corresponds to the emitter, the gate to the base and the drain to the collector. The essential difference between these two types of transistors is in the heart. In the first type, the base metal lead is directly attached to the semiconductor, while in the MOSFET there is a layer of silicon oxide between the metal lead (which ends with a metal plate) and the semiconductor (during the production of the MOSFET the silicon is forced to rapidly oxidize, then a metal plate with a wire in the extension comes to the created layer of oxide). Therefore it is called MOS - metal (metal plate), then oxide (oxide layer), then semiconductor. What is the role of this oxide? The oxide is actually a dielectric, that is, a capacitor. Since it is very small, it needs only a very weak and very short-lasting current to twist to the necessary voltage (which is at the same time the voltage between the gate and the source, U_{GS}), so-called *threshold voltage* $U_{GS(th)}$ (which for different types of MOSFETs can be 2, 3, 4, etc. volts), for the transistor to become at least slightly conductive. The smaller is the thickness of the oxide layer (its area cannot vary much), the lower is the threshold voltage. When this capacitor is twisted to this voltage, then its oscillation of continuous small untwisting and twisting again is sufficient to affect the corresponding segments of the EM-forces in the semiconductor, so that it becomes conductive to a certain degree. When we previously connected a capacitor in the base's branch of the bipolar transistor, we saw that the small capacitor caused only a small AC current, which required two transistors, i.e. a great degree of amplification, to be detectable. Here, however, this oxide capacitor is not in the gate's branch, but it is directly 'pasted' to the semiconductor's surface. Hence, only small oscillations of its are sufficient to maintain the MOSFET conductive. A thinner layer of oxide has a greater capacitance, which means that its untwisting and twisting again are powerful enough at a lower voltage, hence the lowering of the threshold voltage with a thinner oxide layer. With the increasing of the U_{GS} above the threshold voltage, say about 0.5-1 Volt, the MOSFET enters saturation, that is, it acts as an ordinary wire with resistance practically equal to zero. The voltage between the source and the drain is then practically zero, too.
As we have said, when maintaining the oxide in twisted and therefore the MOSFET in conductive state, there is a very small alternating current with high frequency to and from the gate, which is practically zero because it is very difficult to detect. In the case of the bipolar transistor, there it flows continually

current to the base, which, although small, is still large compared to the practically zero of the MOSFET. That's why it is said that the MOSFET has a great input resistance (or input impedance), which means that the evoking current to the gate is extremely small. It is also said that the BJT is a current controlled, and the MOSFET is a voltage controlled transistor, again because the current to the gate is practically zero, but there must be a certain voltage, from the threshold voltage upwards. This means that in order a stronger current to flow through the BJT, there should be a stronger current to the base (the voltage between the base and the emitter changes only slightly), and in order a stronger current to flow through the MOSFET, there should be a stronger voltage on the gate (the current to the gate is practically undetectable), which in the oxide causes more powerful oscillations that pass over to the semiconductor.



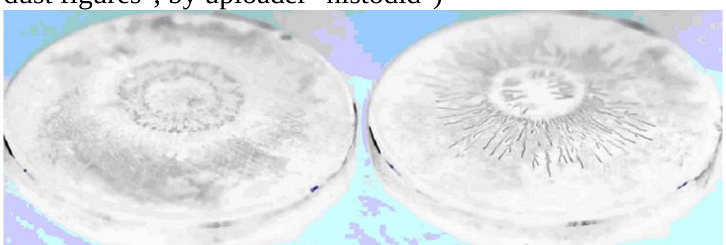
Most of today's digital electronics is based on the MOSFETs. The first MOSFET microprocessors from the early 1970s were composed only of P-channel MOSFETs (PMOS Logic), later they were replaced by microprocessors composed of N-channel MOSFETs (NMOS Logic), for already at the end of the 1970s to appear processors that are composed of both types (CMOS Logic – Complementary MOS).

It is well known that digital electronics is based on the binary system, on ones and zeros. Nevertheless, a machine cannot know of ones and zeros, but only of physical processes. What are these ones and zeros on physical level?

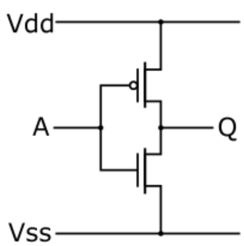
The common interpretation of this question is that ‘one’ means "there is voltage"; or, "there is high level of voltage"; or, "there is current"; and that ‘zero’ means "there is no voltage"; or, "there is low level of voltage"; or, "there is no current". The comparisons with closed and open electric circuit are thereby very dear in the textbooks, where a closed circuit means "1" and an open circuit "0". Hence, the logical conclusion is that ‘one’ is ‘something’, presence of a kind of force, and ‘zero’ is ‘nothing’, i.e. absence of that force. The truth is that both the one and the zero are fully equal forces in intensity, only different in signs. The one is the plus-electricity (the blowing force) and the zero is the minus-electricity (the suction force). Just as when something blows, it must suction somewhere, and just as these forces are equal, only different in signs, so both the one and the zero are equally powerful forces. However, both the plus and the minus can vary in a certain range, so that either the one force or the other can be a nuance stronger. It depends still on the circuit design. But if the minus is somewhat stronger than the plus, it means that somewhere else it is the other way round. In principle they are fully equal. On this (+/–) polarity is built the digital technology.

The misunderstandings on this issue arise from the current theory, which has made the electric current unipolar (motion of negative spheres contrary to the true direction of the current), but also because the man must express the physical processes quantitative when he wants to make use of them. And in order to be able to express them mathematically, he must define a zero reference point. In this case, the battery minus-pole is mostly taken for such a point. If the battery is 5V, then the plus-pole is +5V, and the minus-pole is 0V. When we connect the battery terminals with a wire, the plus is the strongest at the very plus-pole. It is +5V. The minus is the strongest at the minus-pole. It is 0V. Where the suction force is the strongest, there we say it is zero volts, which actually means "no voltage", that is, no force. This paradox is necessary for the mathematical calculations.

Particularly interesting is the congruence that plus-electricity is marked with a straight line (l-one), and minus-electricity with a circle (O-zero). In mechanics, the two opposites are the linear motion (1) and the circular motion (O). The linear motion can be designated as plus, and the circular as minus motion. At the beginning we said that when something expands, it means plus, and when it contracts, it means minus. When a circle is enormously big, then its line becomes straight, and as it narrows, it becomes an increasingly curved line. The combination of the linear and the circular motion gives a spiral motion. If we observe a slow motion video recording of an arrow shot, we will see that the arrow does not fly quite straight, but has a spiral movement. When in the 19th century man began to make rifles with a helical pattern of grooves on the inner wall of the barrel to give the bullet a spiral spinning movement, the range increased several times. So the nature, when it wants to achieve a stronger effect, tends to include both polarities. If we speak not of movements, but of plane figures, then we can say that the combination of the two opposites, the line and the circle, is the ellipse (in fact, this is also true for the movements, as the planets are known to circle in ellipses). When the focuses of an ellipse are moving further apart, then it becomes ever flatter; in the end, when they move apart to infinity, the ellipse becomes a straight line. In the opposite case, when they come ever closer and eventually merge into one point, the ellipse becomes a circle. The combination of a line and a circle is also seen in the very shape of the ellipse: in the middle part its line tends to a straight line, while at its ends to a circle. This polarity of line and circle is interesting to see in the Lichtenberg's figures obtained by means of plus- and minus-electricity (image below) (Georg Christoph Lichtenberg, 1742-1799). The form of the very fine dust from Lycopodium on the left side of the figure is from minus-, on the right side from plus-electricity. (YouTube Video "Lichtenberg's dust figures", by uploader "histodid")



The basic building blocks of digital electronics are the so-called logic gates. The simplest logic gate is the inverter (Inverter, NOT-gate). It performs a polarity reversion of the signal: it turns the plus-signal into minus and vice versa. The figure below shows an inverter in CMOS-technology.

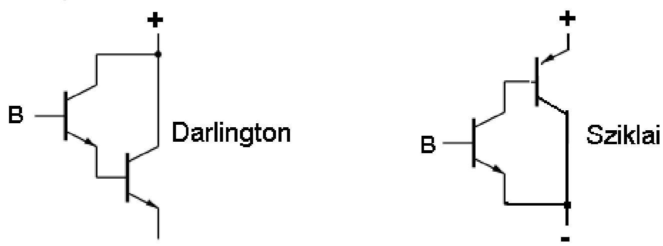


The top line is the battery plus-rail (positive rail / + 5V / Vdd) and the bottom line is the minus-rail (negative rail / 0V / GND / Vss)[12]. The upper MOSFET is a minus (PNP, P-channel), the lower is a plus one (NPN, N-channel). The small circle at the gate lead of the upper MOSFET indicates that it is a minus one. Looking from top downwards there is PNP-NPN, or SGD-DGS. Since the P-Source must be attached to the plus-rail and the N-Source to the minus-rail, it follows that the PNP-MOSFETs are always attached to the plus, and the NPN-MOSFETs to the minus battery rail.

The signal comes through the middle line of the circuit. When at the input A, which is connected to the gates of the both MOSFETs, comes a (+)signal, then only the lower NPN-MOSFET becomes conductive. This means that the output Q becomes connected to the bottom battery rail. In other words, at the output Q there is a (–)signal. Figuratively speaking, the minus from below went up in the middle. When at A comes a (–)signal, then the upper PNP-MOSFET becomes conductive and the plus from above goes down

[12] The notation Vdd for the plus-pole and Vss or GND (Ground) for the minus-pole of the battery do not make much sense, but it is established and often used notation. We consider 'GND' for the minus-pole not appropriate because "ground" means "mass", and the mass is no source of electrical potential, whereas the minus-pole of the battery it is.

in the middle. This output is further an input for the next logic stages up to the last, whose output will no longer be an input for a next one, but through a light bulb, a relay, etc., it will drop to the minus (if it is a plus), or it will climb to the plus (if it is a minus). For illustration, we can connect already to this gate two LEDs: one blue LED from the top line down to Q and one red LED from Q to the bottom line, with one resistor of 400-500Ω in each of the two branches. When we connect the input A to the top rail (whereby we bring to A a plus signal +5V), the upper blue LED lights up (minus at the output = minus color), and when we connect it to the bottom rail (minus signal to A, 0V) the lower red LED lights up (plus at the output = plus color). Why we call the red color ‘plus’, and the blue ‘minus’, we will see later.
In the point Q both transistors are connected with their drains. We can ask the following question: how has the signal from A reached Q, when Q is connected only with one voltage rail? Or, a little more elaborated: one of the two transistors of the inverter is always unconductive; it is a barrier for at the point Q to be ‘present’ the other voltage rail; it is on the one hand; and on the other hand, Q enters gates of transistors of the next stage, and one or some of them should be made conductive right with the signal from Q; therefore, along this path there is also no conductive path to the other voltage rail.
Something reminiscent of this we had at the beginning of this paper when we connected two transistors, that is, when we have made the so-called Darlington pair (the name is in honor of the engineer Sidney Darlington, who first made such a connection of transistors). The same effect can be achieved by connecting two opposite, or, as it is said, complementary transistors. This connection is called Sziklai pair (George Sziklai). The two connections are shown in the figure below.



In these cases too, the corresponding (+)action on the first base reaches the next base, although only one of the other two segments of the first transistor has connection to a voltage rail.
Regarding these two pairs, we can say that the incoming electric flux sets in motion the EM-forces in the first base, which in turn are able to set in motion the EM-forces in the neighboring emitter (in the case of Darlington) or in the collector (in the case of Sziklai) in cooperation with only one battery pole (the plus pole in the first case, the minus in the second) because the EM-forces of the base are the middle part, and therefore powerful. When in an experiment before we directly acted on the emitter or on the collector, there was still necessary enough mass behind the base, because the emitter and the collector are the outer parts, and hence their pure actuation could not have enough force if we have not attached a mass behind the base for its forces (thus the middle part again) to get enough powerful spin for establishing the bridge.

If we now return to the MOSFET’s inverter, then, on the basis of what have just been said, we can say that the actuation of the EM-forces behind one gate (which means *the middle* of the MOSFET) is powerful enough, that in cooperation with the forces in the source and to it connected voltage rail to actuate the forces in the drain, although the drain at that moment is still not connected to the nearby voltage rail. This motion of the forces in the drain means a certain current to the gate(s) of the next stage. One or several transistors of it become conductive, thereby establishing the connection to the other voltage rail.

A part of today’s integrated CMOS logic circuits works with a voltage of 5V, thus we use it in the examples. The MOSFETs in these circuits enter saturation at a voltage greater than 3.5V. This means that the plus signal (HIGH) has a tolerance from 3.5 to 5V, and the minus signal (LOW) from 0 to 1.5V. These values are called *logic levels*. The reader should pay attention to the fact that when at the input comes a minus signal with a strength of let’s say 1V, then it is strong enough to drive the PNP-MOSFET to saturation, but it has still a significantly smaller strength than the case when at the input arrives a minus signal of 0V.

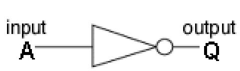
Values between 1.5 and 3.5V for the input signal must not occur, because in that case either one MOSFET in the inverter or both together are only partially conductive, so the output signal, which is the input for the next stage at the same time, will again have some intermediate value, that is, it will be undefined.

However, such an output as above with two lamps, i.e. two loads, in the upper branch as well as in the lower branch, does not exist in practice. The example was only for certain clarification. There is a load only in the lower branch. It will be ‘moved’ only when the output has a plus-signal. This concept is because the operational role has the ‘one’, the plus-electricity. When the plus-electricity is chosen to have the active role, it is called positive logic. There are also electronic circuits where the ‘zero’ has the active role (negative logic), but it is still a rarity. So, since at the end of the circuit one kind of electricity should be given priority, here it makes some sense to talk about that the ‘one’ is a closed circuit and the ‘zero’ is an open circuit. For example, if all the necessary conditions for an elevator to start off, which are reflected in logic circuits, are fulfilled [the door sensor gives a signal that it is well closed (+, that is, 1), the button for the desired floor is pressed (=1), the overload sensor gives a signal that the elevator is not overloaded (=1) (in case of non-overload it gives actually 0, but with an inverter is converted to 1)] then the output will have a plus-signal (since all the signals are joined by an "AND" function, which, as we will see soon, gives plus at the output only if all the inputs are plus) and between it and the minus there will be a relay (switch) which will turn on the electric motor of the elevator. Otherwise, there will be a minus at the output, so that this minus with the one from the negative rail (it is actually the same minus) cannot turn on anything, that is, no circuit is closed.

However, when already at the very beginning of many textbooks, even represented with drawings, it is stated that ‘zero’ means open electric circuit (no current flows), and ‘one’ means closed electric circuit (current flows), or that ‘one’ means "there is voltage" and ‘zero’ means "there is no voltage", then the reader’s puzzlement has already begun, resulting in hampered progress and probably a soon quitting.

In the CMOS circuits we should speak about "there is a (+)voltage" (=1) and "there is a (–)voltage" (=0), and in the BJT-circuits we should speak about "there is a (+)current" (=1) and "there is a (–)current" (=0), because there must be a current flow to the base(s).

The most commonly used inverter’s symbol is shown in the figure below. The circle after the triangle indicates inversion, i.e. negation (the circle can sometimes stand in front of the triangle). The voltage rails for the sake of simpler drawings are usually not drawn. The table is called the truth table.



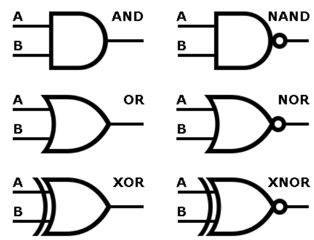
A	Q
0	1
1	0

This logic gate has only one input and one output. The next logic gates which the digital electronics is based on have two inputs and one output[13]. These gates are three in number: two unbalanced and one balanced. One of the unbalanced we will call the "dominant plus" (plus-gate) and the other "dominant minus" (minus-gate). The plus-gate is called ‘OR’ and the minus-gate ‘AND’. The balanced gate is called ‘XOR’ (Exclusive-OR). When after each of these gates we add an inverter, then we get their reversals (NOR, NAND and XNOR); thus the three basic gates are doubled to six.

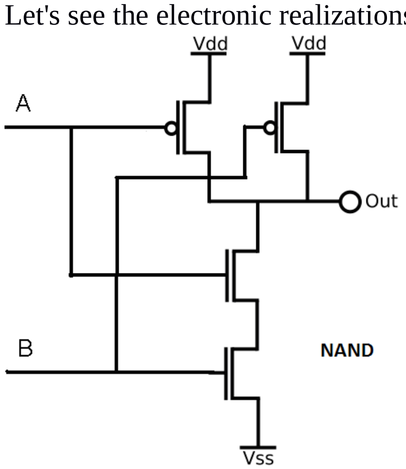
[13] The inputs may be also more than two, but the gate with, let’s say, three inputs can be considered as a two-input gate, whose output is at the same time an input for the next identical gate along with the third input.

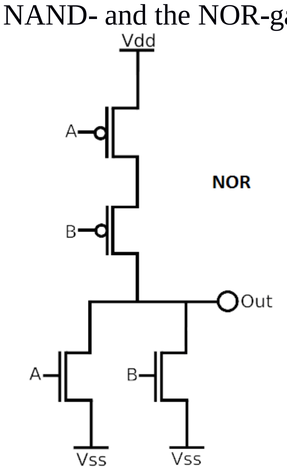
We can describe the plus-gate with the following words: if at least one of the inputs is plus (1), then the output is plus (1).
We describe the minus-gate the same, just the other way round: if at least one of the inputs is minus (0), then the output is minus (0).
We describe the balanced gate with the following words: the different gives a plus (1), the same gives a minus (0). When a plus meets a minus, it means attraction, reinforcement (+); when a plus meets a plus, it means repulsion, attenuation (–). To make it easier to remember what gives 0, we can also say otherwise: the sum 0 + 0 = 0, and also the sum 1 + 1 = 0 with the transfer of one to the next position, something like when in the decimal system we add 7 + 3, which gives zero with transfer of 1 to the next position.
The truth tables and symbols for the described gates are as follows:

INPUT		OUTPUT					
A	B	AND	NAND	OR	NOR	XOR	XNOR
0	0	0	1	0	1	0	1
0	1	0	1	1	0	1	0
1	0	0	1	1	0	1	0
1	1	1	0	1	0	0	1



The starting articles for the realization of the basic AND- and OR-gate in CMOS-technology are actually the NAND- and NOR-gate, to which then an inverter is added (the reasons for this will be discussed below). The essence of the NAND- and NOR-gate is the combination of the parallel plus- and the series minus-connection of transistors.
Let's see the electronic realizations of the NAND- and the NOR-gate.



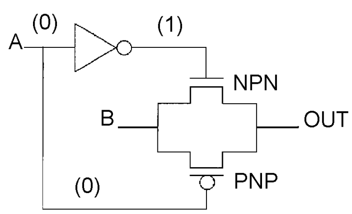


At the NAND-gate (left) we see that in the upper part, which gravitates to the (+)battery pole (Vdd), there is a parallel connection of two PNP-MOSFETs, and in the lower part a series connection of two NPN-MOSFETs. The parallel plus-connection gives dominance to the sign which it gravitates to, that is, an opportunity to expand is given to it; thus in the results we get three pluses and one minus. This means when either in A or in B, or in both simultaneously arrives (–)signal, then at least one upper MOSFET becomes conductive, so that the plus from above will come down to the output. With the series connection the other sign is reduced. For the minus to go up at the middle, A and B must receive (+)signal, so that both lower NPN-MOSFETs become conductive (and none of the upper ones), thus the minus from below climbs to the output.
In the case of the NOR-gate the parallel and the series connection are set vice versa, so that here dominates the minus.

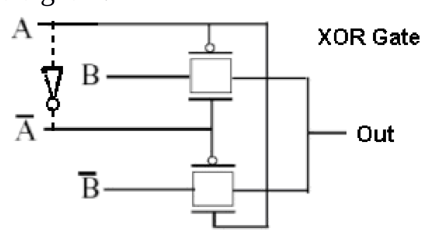
Note that if we either in the NAND- or in the NOR-gate join the inputs A and B together, then the gates turn into inverters. So, in a sense, both NAND- and NOR-gate can be seen as expanded inverters with one transistor up and down, on one side in parallel, on the other side in series connection.

When at the output of each gate one more stage with an inverter is added, then we get the basic AND- and OR-gate. They have to be realized so, because to the battery plus-pole must be attached P-source and to the minus-pole N-source. This means that the upper transistors must be PNP, so that when at their inputs arrives minus, at the output appears plus, that is, we will always get reversed gates, and NAND and NOR are just such: from at least one ‘zero’ at the input, we get ‘one’ at the output (three ones total=NAND), from at least one ‘one’ at the input, we get ‘zero’ at the output (three zeros total=NOR).

To show one of the possible solutions for the realization of the XOR-gate, we will first introduce another significant ‘cell’ of CMOS-integrated circuits. It is called transmission gate (TG). On the left side of the figure below is shown one such. New in this gate is that one of the two signals, in our example it is B, does not arrive at the transistor gates, but at the sources. Whatever signal arrives at B, it will not pass through to the output if there is an ‘one’ at A, and it will do it only if there is a ‘zero’ at A (this case is shown in the figure with (0) at A and at the PNP-MOSFET and with (1) after the inverter). When there is a ‘zero’ at A, then the gates of both transistors have corresponding signals, thus the TG is conductive. If at that moment there is (1) at B, it will pass through the lower PNP-transistor; with (0) at B, through the upper NPN. It would be wrong for this gate to present a truth table as in the middle figure below, because when there is (1) at A, then the output does not have any signal, and the last two zeros in the column ‘OUT’ cannot represent "no signal", but they are a minus signal. Therefore, in the last two places of that column should be an X in each, if we agree X to mean "no signal".[14]



A	B	OUT
0	0	0
0	1	1
1	0	0
1	1	0



In case the inverter is relocated in the branch with the PNP-MOSFET, the signal from B will pass through the TG only if A has (1), that is to say, the opposite from the previous case will happen.

To the right of the figure is shown an XOR-gate realized with transmission gates. A' is connected to A via an inverter, and the same is true for B' to B. These inverters are not drawn for greater simplicity. We have still drawn one with dashed lines. When A=0, then only the upper TG is conductive, so whatever at that moment is present at B, it will be transmitted to OUT. When A=1, then only the lower TG is conductive, and since here enters B', at the output will be ‘copied’ the reversed values of B (*see* the truth table for the XOR-gate).

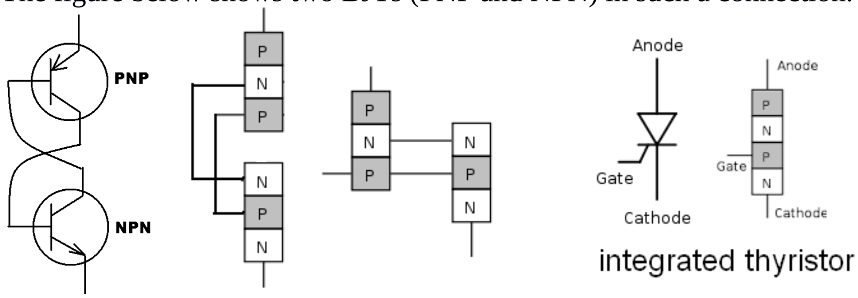
In order to realize an XNOR-gate, we don’t have to add an inverter after the previously described gate; rather, we have only to change the places either of A and A' or of B and B'.

With little thought we can conclude that in this balanced gate we have something very similar to the fans, to the spirals and to the dashed circles from the beginning of this paper.

For some tasks executed by the computer one gets the impression that it performs them simultaneously, but sometimes it is an illusion, because it often jumps from one unfinished operation to another, then to a third, to a fourth etc., and then goes back to the first one and so on in circles until every task is done. However, the electric current does this with tremendous speed, so we have the impression that it all happens simultaneously. In the logic circuits mentioned above the signal will pass through to the last

[14] When in some logic circuits, the possibility of not having any signal at the output (so-called ‘high impedance state’) is purposively implemented, which is not rare, then it is called ‘three-state logic’.

stage without waiting anywhere. But it often has to wait somewhere until it comes its turn to resume. The logic circuit enabling this is called a flip-flop.
At the ‘basis’ of the flip-flop lies something similar to that in an improvised thyristor, i.e. when two complementary BJT-transistors are connected so that something corresponding to a thyristor is obtained.
The figure below shows two BJTs (PNP and NPN) in such a connection.

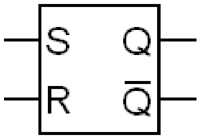


The base of the first transistor is connected to the collector of the second. The base of the second is connected to the collector of the first, thus a kind of closed loop is formed. Only the emitters are free. If in the continuation to one of the emitters we connect a resistor and an LED-lamp and then connect the ends to the plus and minus battery poles correspondingly, then even a very short and weak plus-current to the plus-base, or such minus-current to the minus-base will actuate the transistors, the lamp will turn on and no current is required to any base anymore for the lamp to stay on [for the sake of truth, the lamp in this circuit with the improvised thyristor will turn on even without any current to any base. To disable the conductivity of this improvised thyristor before the arrival of current to one of the bases, we add a 200-500Ω resistor between the (+)base and the (–)battery pole (whereby to the (+)base we bring somewhat attenuated minus signal), or between the (–)base and the (+)battery pole. After the arrival of a brief corresponding current to one of the bases, the conductivity is established, and the current through the resistor has no influence on the “thyristor” anymore because it is negligible compared to the current through the branches with the transistors, where the resistance is practically zero.

To the right of the figure is shown an integrated thyristor. It consists of four segments (PNPN), but only three leads come out of it: cathode (–), anode (+) and gate. The gate’s lead on our figure is taken out from the middle P-region, but it could also be taken out from the middle N-region. In the first case the thyristor becomes conductive with a small and brief plus-current to the gate, in the second it would become conductive with such minus-current. The plus-current to the P-gate in cooperation with the battery minus-pole will actuate the EM-forces in the lower PN-junction. This action will be transferred to the (–)E-segments of the EM-forces in the upper N-region and so on. Once the thyristor becomes conductive, it no longer needs a current to the gate to remain such, because here (unlike in the case of BJT) the two end segments are connected to the corresponding battery-poles (i.e. P to the (+)pole, N to the (–)pole). In the case of BJT, the collector is attached to a non-corresponding battery pole, therefore current must incessantly flow to the base to maintain the conductive state.

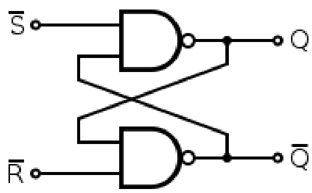
Thyristors are exclusively used as switches.

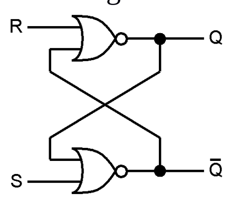
Let's go back to the flip-flop. It has two inputs. The one is called ‘set’, the other ‘reset’, and two outputs, Q and Q' (figure below). The flip-flops are also called bistable devices. This means that they can have two stable, but inverse states at the output: the first state is Q=1, Q'=0, and the second Q=0, Q'=1. Q and Q' must always be different, which is symbolically expressed by the bar over one Q. These building cells are nothing other than holders (i.e. memory storage elements) for some time of one bit (**<u>BI</u>**nary digi**<u>T</u>**), that is, of one ‘one’ or one ‘zero’. The agreement is that the holder of this 1-bit information is the Q-output.



Set / Reset Flip Flop

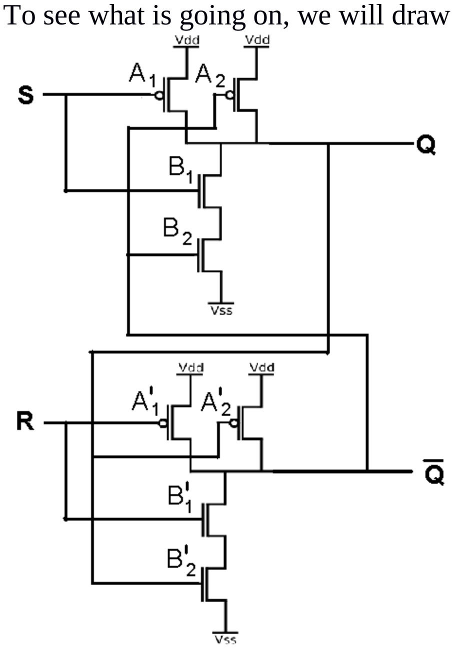
Operational is the (+)current (i.e. the ‘one’) because the positive logic is in force. When at the S-input arrives (1), then the R-input should have (0); then at Q must appear (1), and at Q' (0). This is called SET. When at R-input arrives (1), then the S-input should have (0); then at Q must appear (0) and at Q' (1). This is called RESET. The two inputs S-R must not, that is, should not have (1) at the same time. If there is a zero on both inputs (which is allowed), then it must not affect the previously established state.
The basic bistable elements are built using two NAND- or two NOR-gates.

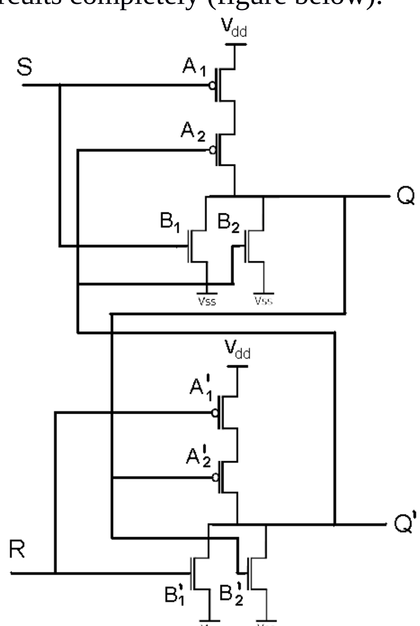




NAND Flip-flop NOR Flip-flop

Note some details on the figures above. The NAND-FF has bars, i.e. negations over S and R, and at the NOR-FF there is no such thing; but here R is up and S is down, which is not in accordance with the usual practice, S to be up and R down.
To see what is going on, we will draw the two circuits completely (figure below).





On both figures we see that the S-input operates two MOSFETs [one PNP (A_1) and one NPN (B_1)] in the upper gate, and also the input R, but in the lower gate. The other two transistors in the upper gate (A_2 and B_2) are operated by the output of the lower gate, and those in the lower gate are operated by the output of the upper gate.

When in the electric schemes two wires cross (+) and the crossover point is not thickened, it means that there is no contact between those wires.

Let's see what's happening with NAND-FF (the figure left). If at the input S arrives minus, then A_1 becomes conductive and the plus from the plus-rail goes down to Q. This plus goes down to A_2' and B_2' too. On A_2' it has no effect, but on B_2' it does. When S=0, then R should be (1). This means that both series MOSFETs in the lower NAND-gate have a (+)signal at their inputs, thus both become conductive, so the minus from below climbs to Q'. This minus goes up to A_2 and B_2 too. It only affects A_2, so the second MOSFET at the top is also conducting, which is not necessary.
Similarly we can derive the opposite situation: when S=1, R=0. Then both Q and Q' will be with inverse values compared to the previous case.
If both inputs S and R get (–), then both outputs become (+).
If both inputs S and R get (+), then it will have no effect on the previously established state because only one (+)transistor from both series connections becomes conductive, which is not enough.
Now let's look at the NOR-FF. If at the input R arrives (+), then B_1' becomes conductive, thus the minus climbs in the Q'. It will climb to B_2 and A_2 too. Only A_2 becomes conductive. When R gets (+), S should get (–). This means that A_1 is also conductive. The two series A_1 and A_2 are conductive, so the plus goes down to Q. It will descend to B_2' also, thus making it conductive, which is not necessary.
If both inputs S and R get (–), then it will have no effect on the previously established state.
If both inputs S and R get (+), then both outputs become (–).
In the tables below we see the summarized states in NAND- and NOR-FF.

NAND flip-flop

S	R	Q	Q’
0	0	1	1
0	1	1	0
1	0	0	1
1	1	Q	Q’

NOR flip-flop

S	R	Q	Q’
0	0	Q	Q’
0	1	1	0
1	0	0	1
1	1	0	0

The repetition of Q and Q' in the results means that the input signals have no effect on the previously established values of Q and Q'.

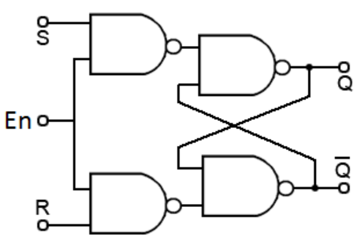
Since we work with positive logic, the ‘one’ has the active role. But two ‘ones’ at the input should not occur. Two ‘zeros’ at the input must not have an effect on the previously established state. When S=1, the agreement is Q to be 1, and when R=1 the agreement is Q' to be 1. (The information carrier is the Q-output.)

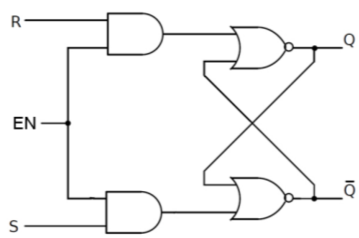
If we look at the tables we see that none of them meets the requirements just said. NAND-FF does not fulfill the requirement when S=1 then Q to be 1, also in relation to R, and neither the requirement that two zeros at the input should have no effect. NOR-FF does not meet only the first two of those about the NAND-FF. Therefore, the circuits must be adjusted so that they will meet the requirements. If we connect one inverter on each of the two inputs of the NAND-FF, and on each of the two outputs of the NOR-FF, then the requirements will be met. The contrast is also seen here: in the first we add inverters to the inputs, in the second to the outputs. But in the case of NOR-FF the problem can be solved more easily, whereby saving one stage in it by simply crossing the output leads. The lower lead will be directed up, and the upper lead down, which is identical to that, Q' to call Q, and Q to call Q'.

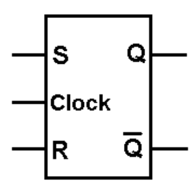
Now it becomes clear why in the symbol for the NAND-FF there are bars over the letters S and R (they represent the added inverters), and in the symbol for the NOR-FF the S-input is down, and the R is up (which is identical to that, Q and Q' to exchange their positions).

There is no need for two simultaneously generated signals to arrive at the flip-flop. It is possible that the S-R inputs through two resistors (say 1-10kΩ each) should be kept on negative potential, that is, through them to be connected to the battery minus-pole (two zeros at the inputs have no effect). Only a brief plus-signal to the S-input (which we can simulate by its brief direct connection to the battery plus-pole) will cause setting of the flip-flop (direct connection means a maximal plus; at that moment the considerably weaker minus that comes through the resistor is canceled), in which condition it will remain even after the termination of this contact. Then only a brief contact of the R-input to the plus-pole will cause resetting of the flip-flop.
For example, when we push the button on the elevator from the outside, then we can imagine that for a short moment we connect S to the (+)pole of the source. At the output Q (which is via an LED-lamp connected to the (–)pole) appears a plus and the lamp in the button lights up. After releasing the button the lamp continues to shine, because now there is a minus on both inputs, which has no effect. When the elevator stops on our floor, then for a brief moment it connects the R-input with the (+)pole of the source. Now Q=0 and Q'=1, thus the lamp goes out. The output Q' is not connected to anything. If we put between it and the (–)pole a resistor and a lamp in another color, then, when the elevator is not called, a different color would shine, which would be more confusing than necessary.
In the case of the so-called Delay- or Data-flip flop (D-flip flop), the sole signal sent to it splits immediately at front of it in two lines, one of which goes directly to S and the other via an inverter to R. Thus, both inputs have always opposite signals.

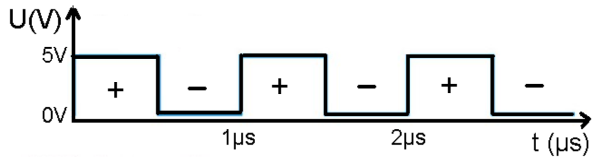
We have described above the effect of a thyristor between two complementary BJTs. Here, we have this effect between a pair of series plus- and a pair of parallel minus-transistors in NAND-FF, whereas a pair of series minus- and a pair of parallel plus-transistors in NOR-FF. In one stable state of the flip-flop, the effect occurs between the pair of the uppermost and the pair of the lowermost transistors and in the other stable state it occurs between the middle pairs of transistors (according to the displayed figures).
Now we will extend the flip-flop with a third input, which we will connect to both S- and R-input via AND-gate each (figure below). This input is called "Enable" or "Control". Since the AND-gate gives (1) only when both or all of the inputs are (1), it follows that when ‘one’ waits either at S or at R, it will not pass through to the output if the Enable-input is (0). Therefore it is called "enable": only when (1) arrives at it, the flow is enabled.





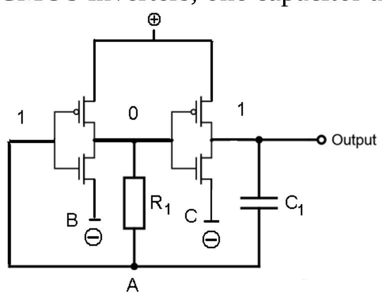


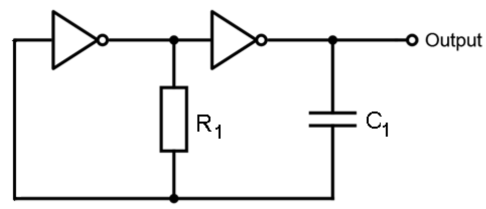
In the first two of the three figures above, we see the same that we’ve discussed before: the left flip-flop from NAND-gates instead of two AND-gates at EN-input has two NAND-gates (two AND-gates and the two inverters in front of NAND-FF, which turns into two NAND-gates), and the flip-flop from NOR-gates (in the middle) has two AND-gates at EN-input, but here S instead of being up, it is down.
The input "EN" in the computer technology is very often a clock. A clock is an alternating change of the plus and the minus with a 50% duty cycle. The figure below shows a graph of a clock signal with a frequency of 1MHz, whose period is 1 microsecond (1μs). Since the duty cycle is 50%, it means that 0.5μs lasts the plus-, and just as much the minus-signal (30% duty cycle at 1MHz would mean that the plus-signal lasts 0.3μs and the minus-signal 0,7μs; the percentage always refers to the plus-signal). The third figure above shows the symbol for this flip-flop.



We could roughly simulate the clock by touching the input lead of the inverter to the plus and minus battery rail alternately at equal intervals.

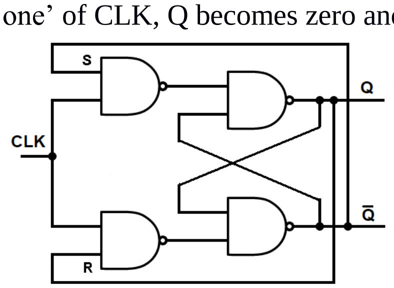
The figure below shows a simple clock generator (also called an astable multivibrator) composed of two CMOS inverters, one capacitor and one resistor:

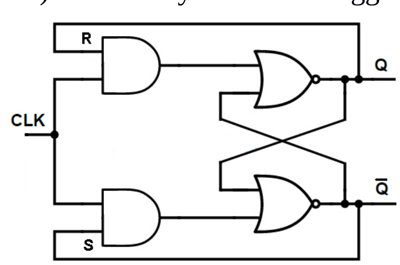


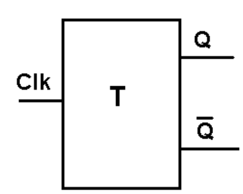


At the output of this circuit we get alternately 1-0 signals, i.e. a "square wave". In the figure on the left by means of the digits 1-0-1 is represented one of the two possible states in the circuit (the other is 0-1-0). When there is (1) at the output, the upper PNP-MOSFET of the second inverter is conducting. Then C_1 twists along the line: (+)pole>the mentioned PNP>C_1>R_1>the lower NPN of the first inverter>(–)pole at the point B. During this twisting there is a (+)current from the point A to the first inverter, making its lower NPN conductive. When the capacitor twists close the maximum, then the (+)current to the first inverter becomes weak. At that moment comes into force the (–)current from the point B through the lower transistor, via R_1, to the input of the first inverter (actually C_1 and R_1 act as a voltage divider). In this inverter the upper transistor now becomes conductive, whereby the capacitor starts to untwist, which further provides the (–)current from the point A to the first inverter. The capacitor untwists to zero and begins to twist in the opposite direction along the line: (+)pole>upper left PNP>R_1>C_1>lower right NPN>(–)pole at the point C, so that this time we have the opposite state (0-1-0). When the capacitor twists close to the maximum, then the same we have just described is valid again (this time the (+)current to the first inverter will come through its upper transistor and the resistor). The output of the second inverter, as well as the output of the first, we can connect through one LED and one resistor at each to the battery (–)pole. They will light up alternately. The frequency depends on the values of the resistor R_1 and the capacitor. With a 500-600nF capacitor and a 1MΩ resistor, we can get a frequency of approximately 1Hz. This clock signal from the output of this circuit can be used as an input for ‘Enable’ of the flip-flop.

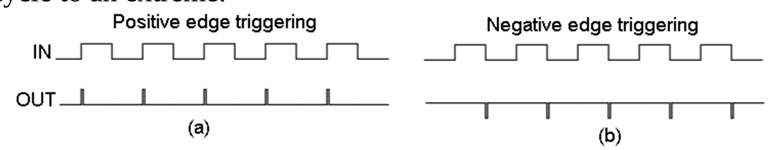
If the output Q of the extended flip-flop is connected back to the R-gate, and the output Q' back to the S-input, then this flip-flop so to say closes in itself - only the clock-input remains free (the left figure below shows this flip-flop made up of NAND- and the middle figure of NOR-gates). This flip-flop will toggle at each positive part of the clock signal (this means: if at a given moment Q=1, Q'=0, then with the first next ‘one’ of CLK, Q becomes zero and Q' one). That's why it is called toggle-flip flop (T-flip flop).



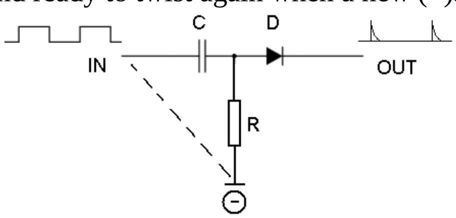




However, in thus constructed T-flip flop there is one problem, the so-called "race around condition". If the clock is 1MHz, then the positive as well as the negative half-cycle is 0.5µs (=500ns). If the passing time of the signal through the flip-flop (propagation delay time), i.e. the time from the moment of arrival of the (+)signal at the input up to the moment of its appearance at the Q-output [simultaneously the (–)signal at Q'], is let's say 50ns, then during only one positive half-cycle of the clock the values at Q and Q' will toggle 10 times. To avoid this, the positive half-cycles of the clock are immediately before entering the flip-flop transformed into brief positive impulses (as we will see, it can be done also with the negative half-cycles, but in negative impulses). This is called "edge triggering". The circuit that realizes this transformation is called "pulse detector circuit". It could also be said that this circuit drives the duty cycle to an extreme.

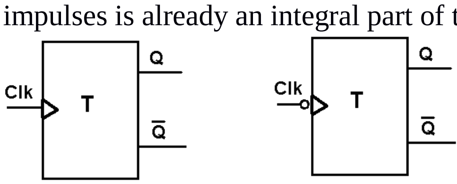


The figure below shows a simple circuit for positive impulses. The capacitor C is small and it will twist quickly along the line C>R>(–)pole when (1) appears at IN. During the twisting the plus-signal passes through the diode. Once the C is twisted, then through R comes a minus-voltage to D, but no current can pass through it. When the input after the (1) becomes (0), then the capacitor untwists through R, because both the left and the bottom point are minuses, i.e. it is a closed loop (the dashed line). During the untwisting there is (–)current towards the diode, but it has no effect on the output. Now C is untwisted and ready to twist again when a new (+)signal appears at the input.

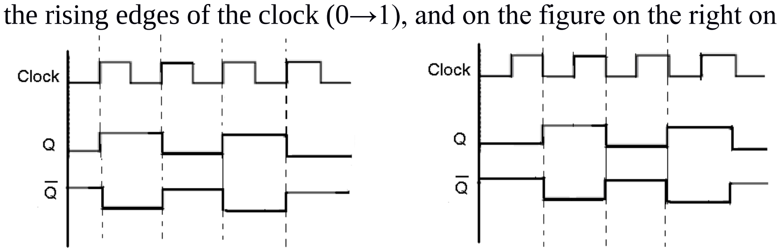


We can easily turn this circuit into one for negative impulses. It is enough to turn the diode in the opposite direction and to move the resistor up, that is, to connect it to the battery plus-pole.

The figure below symbolically shows two T-flip flops, the first is sensitive to positive, the second to negative impulses. The difference is in the small circle at the clock input. The circuit for generating brief impulses is already an integral part of these flip-flops.

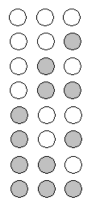


If at the input of any of these two flip-flops we connect a clock generator with a certain frequency, then at both outputs Q and Q' we get signals with twice lower frequency (frequency divider). These two signals are the same, only inverse. The twice lower frequency behind the flip flop is due to the fact that the triggering, which changes its state from plus to minus and vice versa, occurs only once during one full cycle of the clock. The left figure below shows the timing diagram of the clock and of the outputs of the flip-flop which is triggered on the rising edges of the clock (0→1), and on the figure on the right on the falling edges (1→0).

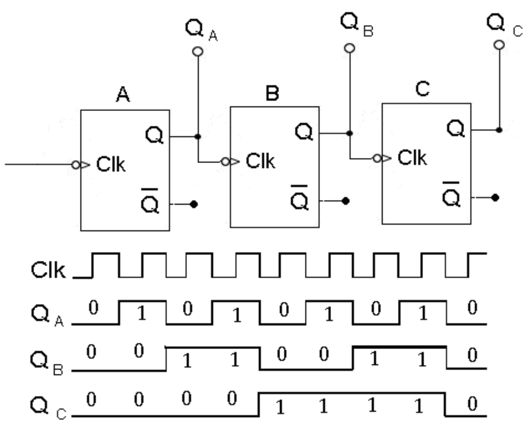


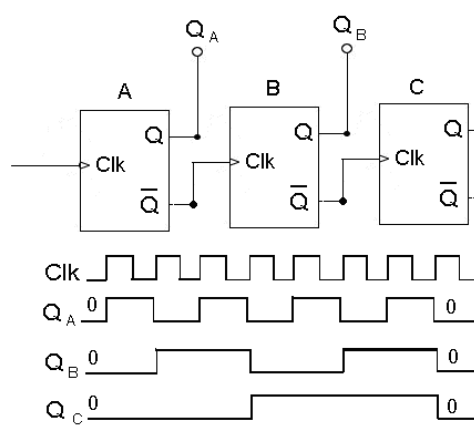
If we now use the output of this flip-flop as a clock signal for a next flip-flop, then at the output of the second we get twice lower frequency compared to the previous one, that is, four times lower frequency than the source clock. This has been used for construction of very important circuits in electronics, that is, of counters. What is a counter? Let's imagine a three-digit time counter that counts and shows seconds from 0 to 999. The first digit from the right changes once in a second (1Hz). The second changes ten times slower, i.e. once in ten seconds (0.1Hz), the third 100 times slower than the first (0.01Hz). So at each digit we have clocks with different frequencies, whose ratio is 1 : 1/10 : 1/100.

In our positional decimal numeral system[15] we have ten digits, or say ten states (0, 1, ... , 9). But if we have only two states available, then analogously to the previously described, the counting will look like the following figure:



If we imagine these 8 rows of circles as different states of three lamps that turn on and off, and each new row at the figure is a new state of the lamps at equal intervals (with the colored circle indicating a lamp turned on), then the frequency in the first column from the right is two times higher than that in the second, and this is also twice as high as that in the third column. This is called 3-bit counter. It can count from 000 to 111 (i.e. from 0 to 7) and is composed of three connected flip-flops:





The figures above show two counters that have the same function. The difference between them is that the first is made up of flop-flops triggered on the falling edges of the clock, while the second is made up of flip-flops triggered on the rising edges. As we see, the Q-outputs in the first counter are inputs for the next flip-flop, while in the second, the just said refers to the Q'-outputs. These counters are called up-counters.

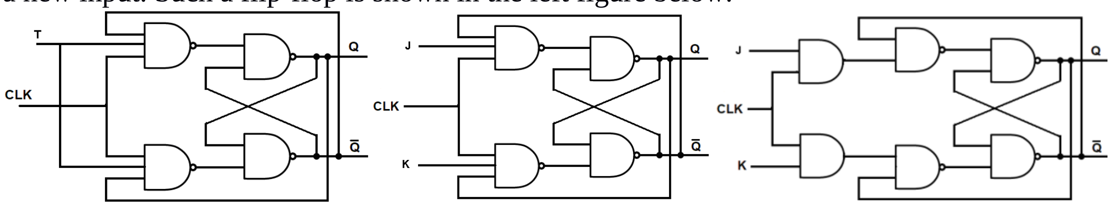
[15] The Hindu-Arabic decimal numeral system, which in the recent history is globally spread, is called a positional system because the value of a digit does not depend solely on itself, but also on the position it is located at (e.g. 2 means two, but 2 in 25 does not mean two, but twenty). The Roman numeral system is an example of a non-positional, that is, III means three, without the position of the signs playing a role. This system was in use in Europe until the 14th century. With its abandonment and the introduction of the Hindu-Arabic system, the calculation was significantly simplified, and at the same time the expressing of decimal point numbers (e.g. 2.75).

Note the moment on the right figure when Q_A falls from 1 to 0 for the first time. Then B-FF is triggered. On the other hand we say that it reacts on rising edges. But at that moment Q'_A changes from 0 to 1, and Q'_A is the output which is connected to the clock input of the next flip-flop.

If each of the three Q-outputs through one LED-lamp and one resistor is connected to the battery minus-pole, then the lamps will turn on and off as described in the figure above with the circles, only we need to put Q_A Q_B Q_C in the order Q_C Q_B Q_A, that is, the figures with the flip-flops have to be drawn from right to left.

What will happen if we flip the connections in these two counters, i.e. in the left counter we connect the Q'-outputs to the clock inputs of the flip-flops, and in the right counter the Q-outputs? Then we get down counters.

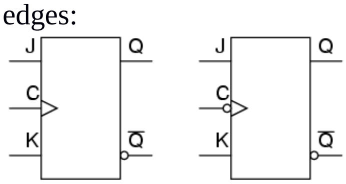
Thus constructed counters will constantly count in a circle without pause. It is often necessary that they be put into operation or be stuck at a particular moment. Therefore, the T-flip flop needs to be extended with a new input. Such a flip-flop is shown in the left figure below.



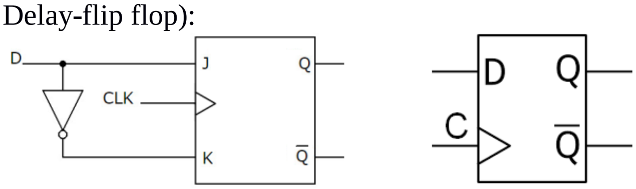
When the new input, called T, is (1), then the flip-flop is active. When it is (0), the flip-flop is stuck, regardless of the fact that the clock continues 'ticking'.

If we tear the T-connection of this flip-flop into two separate inputs, then we get the so-called universal or JK-flip flop, represented in the middle and the right figure (in fact J-K are S-R inputs, but in order to make a distinction from the simple S-R flip-flop, they are marked with two other successive letters of the alphabet). Returning the procedure back, that is, joining the J with the K-input, we get the T-input, and connecting the T-input to the plus battery-pole we get the T-flip flop. We said above that the inputs S-R (now J-K) must not be (1) at the same time. That's just what is done here. But in every rule there is an exception.

These are the symbols for JK-flip flops, the first triggered on the positive, the second on the negative edges:



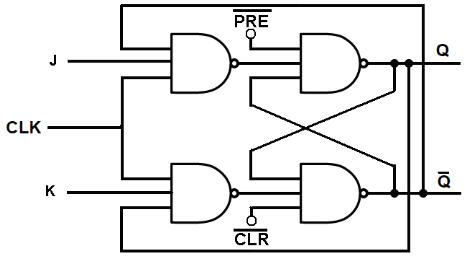
If we add an inverter from the J- to the K-input, then we get the very often used D-flip flop (Data- or Delay-flip flop):



Counters should often run in circle, however, not to their last possible number, but to a smaller one. For example, to display the decimal digits at least a 4-bit counter is needed. It should count from 0000 to 1001 (i.e. 0 to 9) and then return to 0000. These are called modulo η-counters. The counter from 0000 to 1001 is called modulo-10 – MOD-10 counter.

To accomplish this, the JK- or the T-flip flops need another Reset-, also called Clear-input (CLR). In addition to it, they actually have another Set-, also called Preset-input (PRE), which we don't need now.

The figure below shows a JK-flip flop made up of NAND-gates with these two inputs. PRE and CLR (and not only they) may have or may have not a bar over them. What does this bar mean?

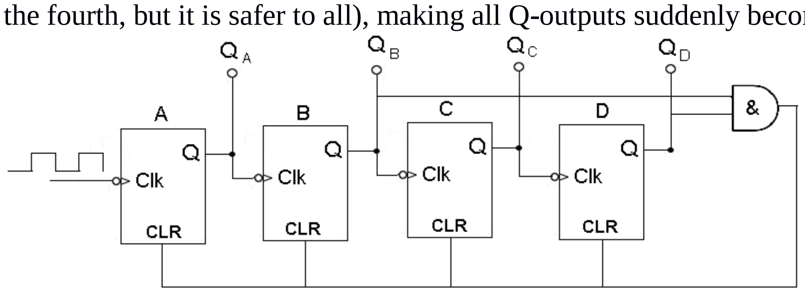


We said above that the digital electronics works mostly with positive logic, i.e. that the (+)current has the active role. But it is only basically. In fact the electronic circuits are constantly working with both positive and negative logic. Imagine that the MOD-10 counter is in the 0000 state and the digit zero should appear on the display. So, in the state of the outputs of only zeros, electronic circuits *should be activated* to turn on some lamps. Or another example: when we talked about the elevator we said that its motor will be actuated when the elevator is not overloaded. But the sensor will produce current, i.e. an (1) signal, just at overload, and zero at normal condition. Thus the elevator should start moving at a zero signal from the sensor, and since the circuits work with positive logic, this zero should be inverted into (1).

The JK-flip flop above, composed of NAND-gates, can be reset with a (0) at the CLR'-input. If it is composed of NOR-gates, then it can be reset with an (1) at the CLR-input, so this input will not have a bar over it. Before the CLR'-input of the NAND-FF we could add an inverter; then the bar over the CLR could be deleted.

So, the bar indicates that here we are working with negative logic, that is, the function 'Clear' will be executed when zero appears at the input. Such inputs are called 'Active LOW'. When there is no bar, it is an "Active HIGH' input.

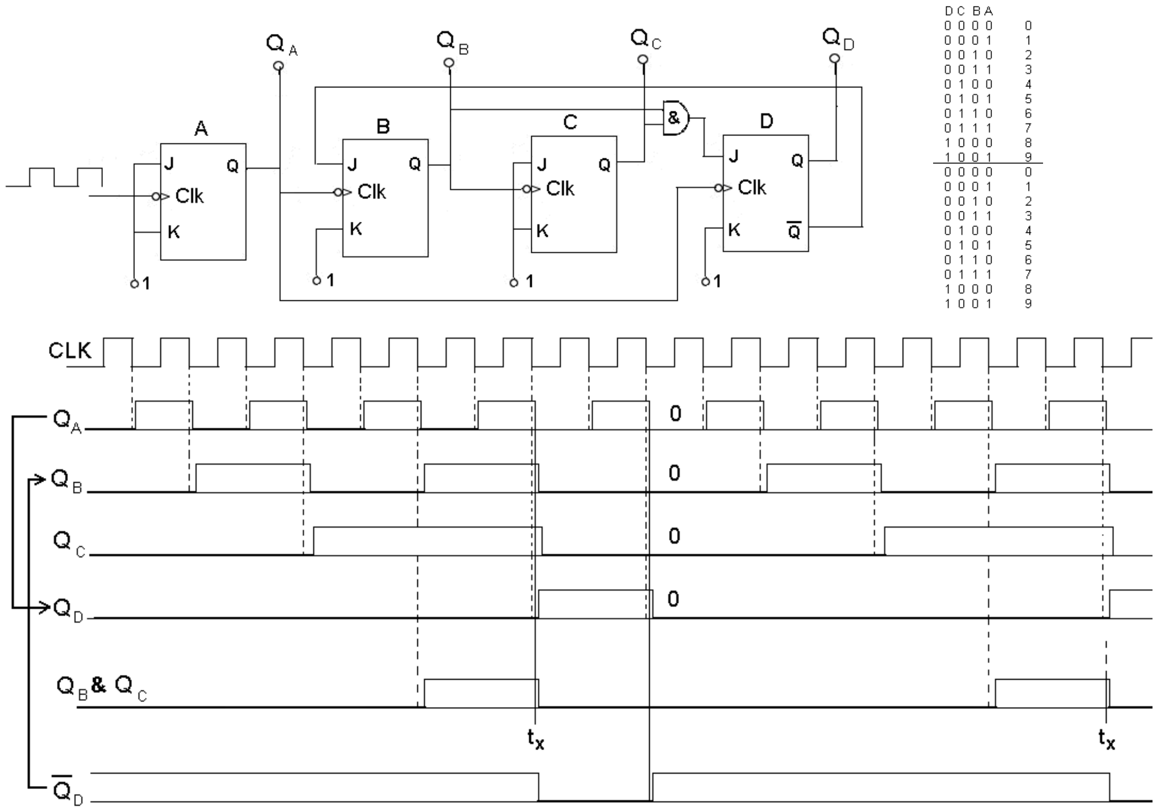
Let's go back to the MOD-10 counter. It should count to 1001. At the moment when the next state occurs (1010), the counter should be zeroed. On the second and fourth position from right to left of this state there are ones. Such state cannot happen before. Therefore the outputs of the second and the fourth flip-flop are connected with AND-gate, whose output becomes (1) at the moment when 1010 occurs. This output is sent to the CLR-inputs of the flip-flops (in fact, not necessarily to all but only to the second and the fourth, but it is safer to all), making all Q-outputs suddenly become zero (figure below).



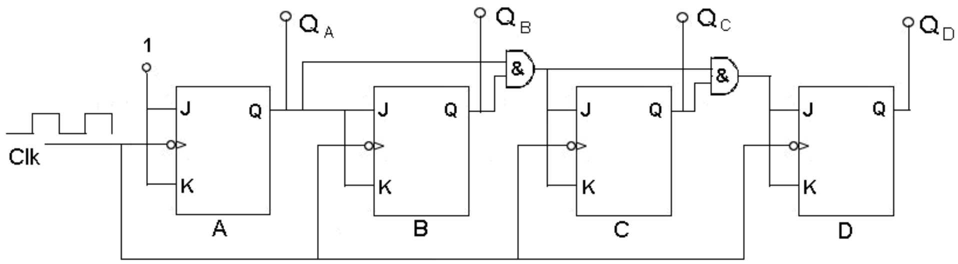
In this counter the state 1010, which is equal to our 10, still occurs for a very short moment. This can cause problems in high-frequency counters. In principle, the counter must be reset at the moment when the clock falls or rises (depending on the type of counter), so that '1010' won't happen at all. Before we describe the counter that solves this problem, let us mention something that should have been mentioned perhaps earlier.

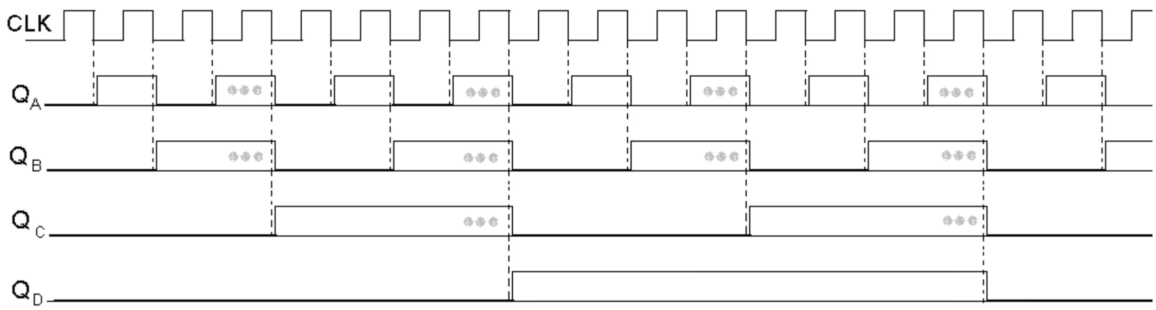
When the first clock signal arrives at the input of the counter, then the appearance of the signal at the output of the first flip-flop will be 50 nanoseconds later if the "propagation delay time" for that particular type of flip-flop is that much. This output, i.e. the clocking of the first flip-flop is the input for the second, so that the output of the second will be 100ns belated relative to the source clock. The more bits the counter has, the longer the output of the last flip-flop is delayed (for a 10-bit counter, the output of the last FF will be delayed 10x50ns=500ns). That's why these counters are called asynchronous. In the case of

fast counters, that is, for those who work with high frequencies of the source clock, the delay causes errors, thus they are constructed differently, i.e. as synchronous counters.
The figure below shows the counter mentioned above and its time diagram. This time the diagram shows the delays in the state changes of the flip-flops. The counter consists of JK-FFs and the "Clear" function is not used. Only those inputs and outputs that are in use are displayed.



All four flip-flops in this counter work somewhat different thanks to the possibilities offered by the JK-flip flops. The first A-FF operates as a T-flip flop with regular frequency (its J and K are attached to 1). The second B-FF also works as T-flip flop, but only while Q'_D=1 (i.e. Q_D=0). The third C-FF works as T-flip flop, but since it takes the clock from its predecessor, its frequency is also not regular. The fourth D-FF works as T-flip flop when it is set, and as SR-flip flop when it is reset.
When the state 1000 (=8) occurs (i.e. the first switching of D), from then on the B and C should no longer switch. Therefore, the switching of D is used as a trigger for stopping of B (the output Q'_D is connected to the J-input of B).
Since the falling of Q_A remains as the only possible trigger for resetting of D, it is necessary Q_A to be the clock-input for D. But to prevent its premature setting by this clock, Q_B and Q_C are via an AND-gate connected to the J-input of D. At the moment t_x, when Q_A falls, Q_B and Q_C are (1), thus D will switch its state. Very shortly afterwards the Q_B&Q_C becomes zero, leaving D simply as an SR-FF at the next fall of Q_A to reset and further to remain indifferent to the switching of Q_A until the next timespan when Q_B&Q_C will be (1).
In synchronous counters, the clock signal enters all the flip-flops so that in a sense they can be called counters in parallel, and the asynchronous – counters in series connection. Synchronous counters must be built from JK-flip flops, for through them to be controlled which flip-flop at a given moment will be unlocked, and which locked. The image below shows a 4-bit synchronous counter.

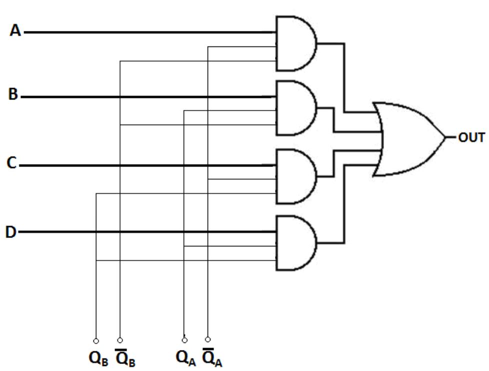


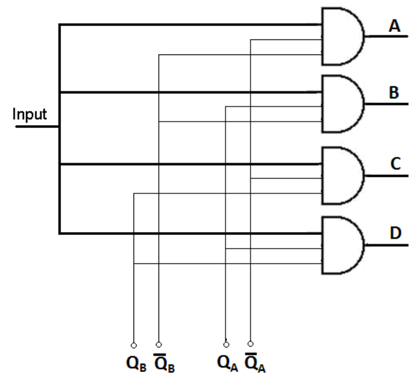


The J-K inputs of the first A-FF are attached to the (+) so that with each fall of the clock (1→0) it will switch. The Q-output is connected to the J-K inputs of the second B-FF so that this one will switch if two conditions are met: one is the clock to fall and the other is Q_A=1. The third C-FF should switch when three conditions are met: the clock falls, Q_A=1, Q_B=1. Therefore, Q_A and $Q_{\mathbf{B}}$ are coupled via an AND-gate, whose output is input for the J-K of the C-FF. Analogously for the fourth D-FF. We can conclude all this just by looking at the table above. When the state 0011 occurs (i.e. Q_A=1, Q_B=1), then the conditions are fulfilled with the next clock pulse the third digit to switch its state (0→1). When the state 0111 occurs, then the conditions are fulfilled for the third digit to switch its state again (this time from 1→0), but also the fourth digit (from 0→1) and all the rest.

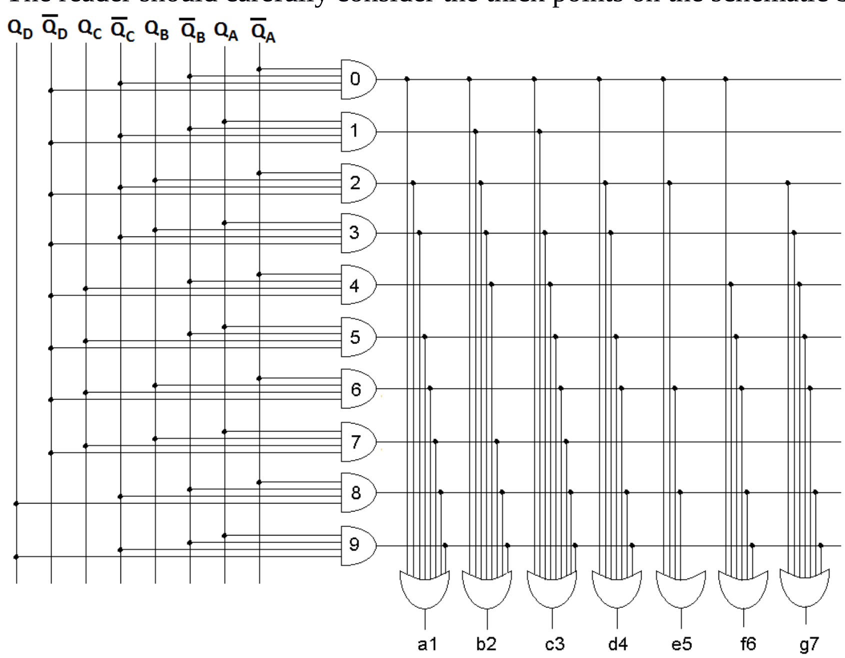
In the timing diagram we notice that all FF are switching simultaneously with equal delay behind the clock signal.

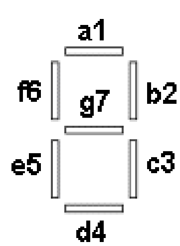
The counter is one of the essential circuits in digital electronics. Here is an example of its use in the multiplexer (MUX). The multiplexer is a circuit that converts the parallel into series connection. Let's say we have 4-bit information that comes in four parallel lines (parallel bus), and then it should continue in a single line (serial bus), something like a four-lane street which narrows into one lane. In the left figure below is shown a "4 to 1-bit" multiplexer. Each of the parallel lines (A,B,C,D) enters into an AND-gate. In these AND-gates enter also the outputs of a 2-bit counter (Q_A, Q'_A, Q_B, Q'_B). When the counter is in the initial state 00, then the A-lane signal will appear at the MUX-output, because the inverted values Q'_A and Q'_B (that is, 11) enter the uppermost AND-gate. At the next state of the counter 01, at the output will appear the signal of the B-lane, because into the second AND-gate enter Q_A and Q'_B, which means 11. The OR-gate at the end "summarizes" the four lanes into one. In the figure on the right is shown a "1 to 4-bit" demultiplexer, which has the reverse function of the multiplexer.





For the “8 to 1-bit” multiplexer we need a 3-bit counter, for a “16 to 1-bit” a 4-bit counter.
At the outputs of the MOD-10 counter we can attach additional gates to display decimal digits (0...9) on a 7-segment display. This display is composed of 7 elongated LEDs (a1, b2, c3, d4, e5, f6, g7) arranged in the form of number 8. In order to display the digit zero, all the LEDs except g7 should be on, for the digit one only the LEDs b2 and c3. For each digit, there is one AND-gate with four inputs. When the counter is in the initial state 0000, on the display should be shown a zero. In that case the set 0000 should be inverted into 1111 (in order to save on inverters the Q'-outputs are used) and should be send to the first AND-gate, whereby only its output turns into (1). This (1) will be directed to the six LEDs: a1, b2, c3, d4, e5 and f6, but through one OR-gate in each line (for this gate we know that only one input needs to be 1 for the output to be 1). The output from the AND-gate must go through OR-gates to the LEDs and not directly for the following reason: to the same LEDs go also the outputs from the other AND-gates for the other digits. All these outputs have a state (0) when the first AND-gate has a state (1). If this output directly goes to the LEDs, then they would be bypassed by a short circuit between the output’s (1) (the +pole) and the zeros (the -pole) from the outputs of the other AND-gates.
Of course, after the OR-gates we add resistors in series with the LEDs, then it all goes to the minus-pole of the battery.
The reader should carefully consider the thick points on the schematic below for better understanding.





By adjusting the clock (capacitor/resistor in the clock generator), this display can also serve as a digital clock that counts 10 seconds. With another 3-bit MOD-6 counter and 7-segment display a clock can be made that counts 60 seconds.

▪ ▪ ▪ ▪ ▪ ▪ ▪

Let's say something about the way of measuring current and voltage. The measuring of the current strength can be carried out through its external manifestations, which are the magnetic field, the heat, the light, the adhesion of matter during galvanization etc. But for everyday practice the measurement through the magnetic field is solely suitable.

Measuring the strength of the magnetic field using a compass is difficult and impractical. The first step towards overcoming this was made when Johann Schweigger (1779-1857) gathered the magnetic field of a wire in one place, that is, he made the so-called Schweigger's multiplier. He actually wound the wire in a coil. Thus the path of deliverance from the clamps of the north-south direction was opened, but also from the constraints of the uniformity of 0-90° deflection of the compass needle, at the same time the possibility for realization of several varieties of what, how and how much will deflect.

When we want to measure the current or voltage, then the measuring instrument becomes a part of the circuit itself. Each inset in the circuit changes the initial intensities of the physical quantities in it, so actually we no longer measure the original, but the changed, which must not be. Therefore, the measuring instrument must influence the current in the circuit as less as possible, and if it still influences it significantly, then it should be calibrated so that the magnitude of influence of the measuring instrument is also taken into account.

Volt- and ammeters work on the principle of interaction of two magnetic fields: the magnetic field of the current in a coil and that of a permanent magnet. If either the coil or the permanent magnet is not absolutely fixed, that is, if it can rotate on an axis, then, when current flows in the coil, the non-fixed will move tending to clamp its magnetic field with the other. Thereby the coil (if it is what moves) compresses a spring, which does not allow it to be fully clamped with the permanent magnet. The stronger the current, the larger the angle of rotation of the coil and the compression of the spring. A pointer is attached to what moves, indicating the strength of the measured quantity on a scale. One variant is called a moving coil ammeter and the other a moving magnet ammeter.

Voltage measurements are realized so that a great resistance is added in series with the coil. For example, when we measure the voltage of a battery, we connect its poles through a series connection of a coil and a resistor of considerable resistance. Due to the high resistance, the current is very small, so it has a negligible influence on the battery voltage. The higher the voltage we measure, the greater the series resistance must be. So when switching the voltmeter with the rotary switch, let's say from the range 0-10V to 0-100V, we actually add one more resistor in the series connection of the coil and the resistor.

Thus we see that when measuring voltage, we actually measure a small current. We can compare this with, let's say, measuring the pressure in a water-filled barrel. In the lower part of the sidewall we make a very small hole and according to the distance the small jet reaches, we can determine the water height in the barrel if we previously have made a scale trying out with the same hole and various, in advance known water levels. Connecting the battery terminals through a great resistor corresponds to the making of the small hole. Adding the second resistor for measuring higher voltages corresponds to the narrowing of the hole when the scale no longer suffice, that is, when the water pressure is so high that the jet exceeds the scale.

When we connect this instrument to the Weston cell (the 1V standard), then for the magnitude of the pointer deflection we will say to be one volt. If the pointer from the utmost left to the utmost right position makes a deflection of 100° and by the connection of the cell we get a deflection of 70°, then in series with the coil we will add even more resistance and adjust it so that the deflection of the coil (i.e. of the pointer) is, let's say, 10°. Here on the scale we engrave a dash and write over it 1V. Then we connect two cells in series and on the scale near the pointer deflection we write 2V etc. If with two cells we get a deflection of 20°, with three a deflection of 30° etc., then there is a linear dependence, and hence with ten cells we will get the maximum deflection of 100°. So, with this so-called input resistance connected in series with the coil we can measure voltages from 0-10V.

When to a stable voltage source of 1V is connected an 1Ω standard resistor, then the current through this circuit will be 1A (1V / 1Ω = 1A). Now in this circuit we cannot simply insert the moving coil to engrave a scale, because the wire of the coil has already a considerable resistance, which can be tens or even hundreds of ohms. That's why a parallel connection from a very small resistor (say tens of milliohms) and a coil is made, thereby increasing the total resistance in the circuit only slightly. The greatest part of the current passes through the small resistor, and the tiny part that passes through the sensitive coil will be enough to cause its deflection. If the deflection is let’s say 30°, then we will increase the small resistor slightly to gain a deflection of 20°. With this so-called shunt resistor, the instrument can measure current from 0 to 5A at maximum deflection.

We notice that when measuring voltage, we make a series (–)connection of a coil and a great resistor, whereas when measuring current, we make a parallel (+)connection of a coil and a small resistor.

In both cases the coil "steals" a very small current which is necessary for the measurement to have an insignificant influence on the electrical quantities in the circuit.

We don’t mention digital instruments here because they are complex and their understanding requires greater knowledge of electronics.

Measurement with digital instruments at last instance comes down to measuring time. They usually through fast counters measure the time for which a certain capacitor is being twisted to a certain voltage, which is still in the practically linear part of its rise (in other words, the capacitor must not come close to "saturation"). It must be then permanently untwisted and twisted again to monitor possible changes of the current during the measurement. The stronger the current is, the shorter the time for twisting to the defined voltage.

If we connect the volt- or ammeter to the mains electricity, we will not get any deflection of the pointer because this current one centisecond is flowing in one way, and the next in the opposite way. Since it changes the direction very quickly, the pointer due to its inertia cannot follow the changes, so it remains stationary.

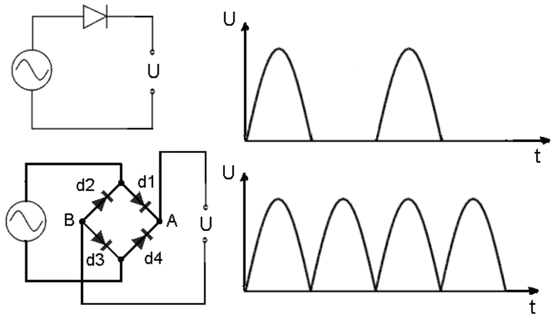
The mains electricity has a frequency of 50Hz. This means it makes 50 cycles in one second. One cycle is when it rises from zero to its maximum, then drops to zero, then grows to the maximum in the opposite direction and falls to zero again. Given that one second has 100 centiseconds and 100/50=2, this means that one cycle lasts 2cs; that half cycle lasts 1cs, and in the first half centisecond the current reaches the first maximum. Since the current grows and drops sinusoidally over time, its intensity can be expressed by the following formula:

$I = I_0 \sin (2\pi f t)$

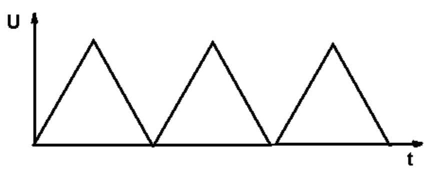
$U = U_0 \sin (2\pi f t)$

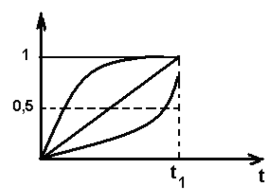
The term 2π expresses the length of a circle with radius one, in fact an angle of 360° expressed through the length of the arc, that is, in radians (2π radians, i.e. 6.28 radians is a full angle of 360°). This is necessary because the radians coincide with the real numbers, while it is not the case with the degrees.

In order to measure the intensity of the alternating current, we need to ‘rectify’ it first, which means we turn it into direct current, and then let it go to the voltmeter/ammeter. The rectification of the current is carried out with one or four diodes. The figures below show an AC source to which once a single diode (half wave rectifier), the other time four diodes are connected (full wave rectifier). The diode of the half-wave rectifier lets the current flow only when it’s positive, thus the output has interruptions, i.e. gaps. To understand how the full wave rectifier works, let's first imagine that instead of an AC source, we have an ordinary battery whose plus is up. In that case, the diodes d_1 and d_3 are conductive, so at point A there is plus, and at B minus. If we turn the battery upside down (the minus up, the plus down), then the diodes d_2 and d_4 are conductive. Again at point A there is plus, at B minus.

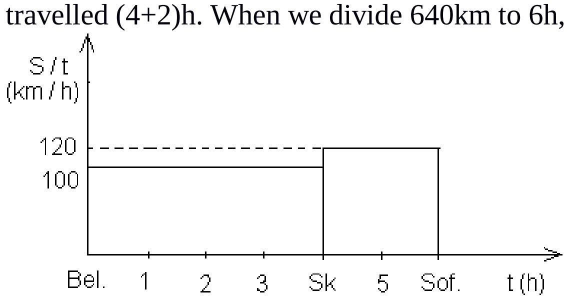


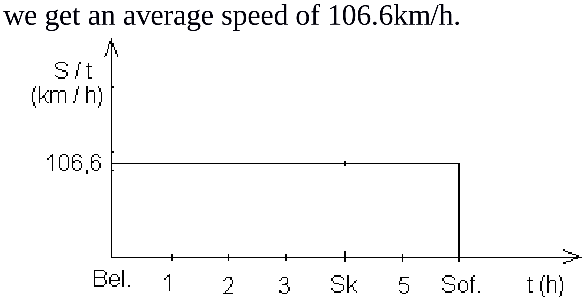
Now this *variable* DC goes to the measuring instrument. What will the pointer do? It will show some value between the maximum and the zero of the voltage or else of the current. The question is what is that value? And will the values of the half-wave rectifier and the full-wave rectifier be the same? With little thought we will realize that on no account they can be the same.
If we draw a triangle-wave voltage (figure below) instead of sinusoidal rectified voltage, then it is easy to guess how much the mean value will be – exactly half between the maximum and the zero. For the full-wave rectified sinusoidal voltage we can guess that it will be greater than the half. Look at the figure on the right on which there are two curves and one line. For the upper curve the mean value is greater than ½; for the line it is equal to ½, and for the lower curve, the mean value is less than ½. At the first curve, the voltage is for a smaller part of the time interval (0-t_1) under ½, and for the greater part it is above ½; in the line there is balance; and at the lower curve for most of the time it is below ½.





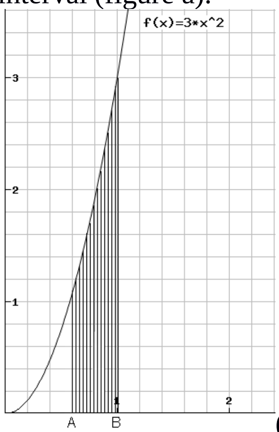
If a car from Belgrade to Skopje drives at an average speed of 100km/h, then from Skopje to Sofia at an average speed of 120km/h, then for the average speed at the total distance we cannot say anything until we know how much time the car has travelled separately on both distances or until we know the distances between the cities. If the distance Belgrade-Skopje is 400km, and Skopje-Sofia 240km, then the car has travelled (4+2)h. When we divide 640km to 6h, then we get an average speed of 106.6km/h.



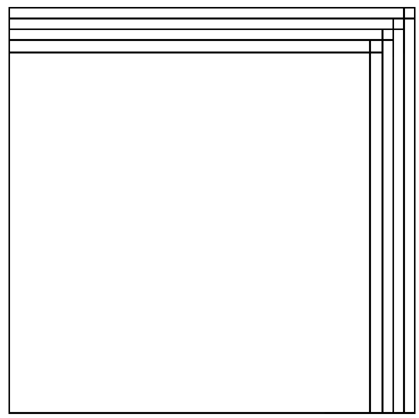


On the left graph we have the speeds successively, and on the right graph the average speed. What do these two graphs have in common? The sum of the areas of the two rectangles on the first graph is equal to the area of the single rectangle on the second graph. The vertical axis is the speed (S/t) and the time t is on the horizontal axis. The area of the rectangles is the distance (S•t/t=S). We could say the following: an irregular area - so we will call the area on the first graph – is turned into a regular area. The height of this area represents the mean value. Hence, in order to find the mean value of a physical quantity whose intensity changes over time, we need to turn the irregular area of its time graph into a regular one.
So, what interests us is the value of the area below the graph of a certain time function in a certain interval. If we find it, we can get to its mean value in that interval. The problem is that the line of the function is curved. The method of calculating such areas is called a definite integral.

When we talked about the steepness of the function x^2, we said that it is 2x because the growth of a square comes down to its widening at two sides, that the steepness of x^3 is $3x^2$ because the growth of a cube comes down to its widening at three square sides. Let us now imagine that we have some physical quantity whose intensity changes over time according to this function $3x^2$. If we want to calculate its mean value in a certain interval, we need to find the area that this curve occupies with the x-axis in that interval (figure a).



(a)


(b)

Look at the figure (b). Let's imagine that the narrow rectangles are narrow cuboids, i.e., that it is the growth of the cube (it is drawn so because it is difficult to present it three-dimensional). Such cuboids are three, and their total volume is $3x^2dx$. The small squares in the upper right corner (in reality cubes), that remain outside the calculation are negligible if we imagine dx very small. What is the sum of all those stacked cuboids? Their sum is what remains when the initial smaller cube is "cut off" from the ultimate enlarged cube, in other words x^3 (of the enlarged cube) minus x^3 (of the initial cube). In figure (a) under the curve $3x^2$ we have very narrow rectangles, which are actually not rectangles due to the upper curved sides. However, if we imagine those irregular figures as extremely narrow, then they transform into rectangles, practically in lines, and each such line has an area of $3x^2$ multiplied by the minute differential on the x-axis (dx). We actually substitute the value of a minimal spatial growth (and it may also be something else), which is so to say scattered on three sides, with the value of the area of just one line. The line can rightly be considered as a rectangle – in reality each line has a certain width. The sum of all those lines is the area under the curve in the given interval, as well as the growth of the cube. The sum is marked with ∫ (**S**um), below and above which the limit values are placed (A and B in figure (a)).

The indefinite integral of the function $3x^2$ is the function x^3, and the steepness of the function x^3 is the function $3x^2$. Steepness and integral are inverse functions.

Let's calculate how much the definite integral of the function $3x^2$ is within the limits from 0.6 to 1:

$$\int_A^B 3x^2dx = \int_{0.6}^1 3x^2dx = x^3(1) - x^3(0.6) = 1 - 0.216 = 0.784$$

So, the hatched area between the points A (=0.6) and B (=1) is 0.784. When we divide this number with 0.4 (B−A=0.4) we get 1.96. The value of the function in the point 0.6 is 1.08, and in the point 1 it is 3. The mean value between 3 and 1.08, if the function was a straight line, is (3+1.08)/2 =2.04, but since it is slightly curved downwards, the mean value is a little less than 2.04, that is, 1.96.

The hatched irregular area, whose lower side is 0.4, we have turned into a regular, i.e., into a rectangle whose horizontal side remains 0.4 and the vertical is 1.96.

Now we can calculate how much the pointer of the ammeter or the voltmeter will deflect if we connect to it a rectified sinusoidal current which changes according to the already mentioned formula $I=I_0\sin(2\pi f\, t)$

or $U=U_0\sin(2\pi f\ t)$. Since the integral of $\sin(x) = -\cos(x)$, and the integral of $\sin(a\cdot x) = -1/a\cdot\cos(ax)$, it follows that:

$$\int_0^{0.01} U_0 sin(2\pi ft) = U_0 \left| -\frac{1}{2\pi f}cos(2\pi ft)\right|_0^{0.01} =$$

$$= U_0(-\frac{1}{2\pi f}\cos 100 * (0.01)\pi - (-\frac{1}{2\pi f}\cos 0)) =$$

$$= U_0(-\frac{1}{100\pi}cos\pi + \frac{1}{100\pi}) = U_0(\frac{1}{100\pi} + \frac{1}{100\pi}) =$$

$$= U_0\frac{2}{100\pi} = U_0\frac{1}{50\pi}$$

So, the area of a half-period of the sinusoid (that is, from t=0 to t=0.01sec) is $U_0/50\pi$. When this area is divided by the lower side, which is 0.01, we get the mean value:

$$U_0\frac{1}{50\pi}/\frac{1}{100} = U_0\frac{100}{50\pi} = U_0\frac{2}{\pi} = U_0 * 0.637$$

We could have solved the task much more easily by computing an integral only for sin(x) within the limits of 0 to π [we can consider U_0 as well as 2πf as 1 (f=0.16Hz)], because in this way we would have also gotten the same, that is, 2/π (=0.636619...≈0.637). However, with the previous more difficult calculation we wanted to show that regardless of the sine wave frequency, the mean value is always 0.637 of the amplitude.

When we talked about the vehicle that traveled from Belgrade to Sofia, we said that its average speed is 106.6 km/h, which means that the vehicle could, ideally speaking, move constantly with this speed and arrive in Sofia in the same time, in 6 hours. Something similar can be found in the case of electric current, when a process depends only on the current strength. In the electrolysis of water, where it breaks down into hydrogen and oxygen, the process depends only on the strength of the current: the stronger the current through the water, the more we get hydrogen and oxygen for a given time. Instead of fully rectified sinusoidal current, we can let through the water a constant direct current from a battery. If the strength of the latter is 0.637 of the amplitude of the former, in both cases we will get the same amounts of hydrogen and oxygen after a given timespan.

However, voltmeters and ammeters, when measuring AC, are calibrated to show a slightly higher than the mean value by 0.07, that is, 0.707 instead of 0.637. Basically it is because the effect that the AC has on a resistor (and therefore the transferred energy) does not depend only on the current strength, but also on the voltage, that is, from their product (I•U)[16]. The mean value of the effect (i.e., of the power) we can see in reality by taking two ideally identical lamps, one of which we connect to an alternating, the other to a direct current source. If we somehow manage to determine that the light given by both is ideally strong, then we can say that same quantities of energy are consumed in both. In the voltage and in the current of the DC circuit we will have a value of 0.707 of the maxima of the voltage as well as of the current of the

[16] Through a small incadescent bulb of 6W which is for 12V flows a current of 0.5A. The same current flows through a bigger bulb of 120W which is for 230-240V. We can imagine the bigger bulb as a longer, and the smaller one as a shorter tungsten wire, both with same thickness. If we connect the bigger one to 12V, then current will certainly flow through it, but it cannot cause the bulb to shine.

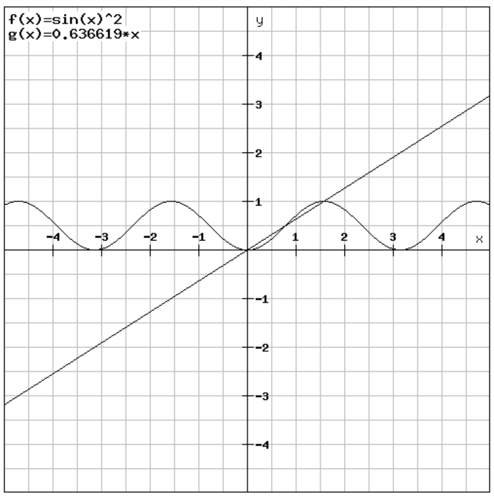
AC circuit, also in the product of the first-mentioned we will have the mean value of the power in the second one.

In an AC circuit the power at any moment is different:

$I \bullet U = I_{max} \sin(2\pi f t) \bullet U_{max} \sin(2\pi f t)$, that is

$I \bullet U = I_{max} U_{max} \sin^2(2\pi f t)$

The transferred energy to the load in a given period can be calculated if we integrate the function $I_{max}U_{max}\sin^2(2\pi ft)$ for that interval (that is, if we calculate the area under the curve in that interval); and the mean value, i.e. the mean power we will get when we divide that area with the length of the interval. But instead of endeavoring to calculate that integral, our work in this particular case could be simplified in the following way: if we look at the two graphs in the figure below (for the functions $\sin^2(x)$ and $0.637x$), it can be noticed that the areas under the both graphs in the interval $0...\pi/2$ are equal, hence it is sufficient to observe only the second function in the mentioned interval.



The mean value in that interval is obviously ½, thus for the mean power of the AC we get $I_{max}U_{max}/2$. Since the number 2 could be treated as a product of two identical numbers, one of which belongs to the current, and the other to the voltage, then we can write the number 2 in a form of two square roots of 2 ($\sqrt{2}\cdot\sqrt{2}$), so for the mean value of the current we get $I_{max}/\sqrt{2}$ and for that of the voltage $U_{max}/\sqrt{2}$. The fraction $1/\sqrt{2}$ is 0.707.

Now we can calculate the maximum value of the voltage when the voltmeter shows 230V during the measuring the household voltage:

$0.707 \cdot U_{max} = 230$,

$U_{max} = 230/0.707 \approx 325V$

So, the voltage in 100 very short moments during one second is considerably higher than that we usually tell: 220-240V.

ON LIGHT AND COLORS

Our account on light and colors we begin with phenomena that everyone can see in daily life. If we look at the flame of a candle or a lighter in a dimly lit room, the flame is blue-violet in the lower part and yellow-orange in the upper part. If the part of the wick running over the wax is very short, the flame is very small and it appears only blue.

The flame of a gas stove is also blue, but as soon as we increase the gas supply, yellow tongues appear in the upper part of the flame.

The smoke of a cigarette, whose smoke columns are not very dense, appears blue against a dark background, whereas against a light background - for example a window in daylight - yellow-orange.

When the sun rises or sets, the sky around it is yellow-orange or even red. The color is also projected onto the opposite horizon or onto the clouds at various points in the sky, while the upper part of the sky retains its blue color. If we look closely at the sunset horizon, we can sometimes also notice that a greenish part appears between the lower yellow-red and the upper blue part.

During larger volcanic eruptions, when the air currents spread the smoke towards the sun, it was noted that the sky turned orange-red (such a case has been described during an eruption of Krakatau).

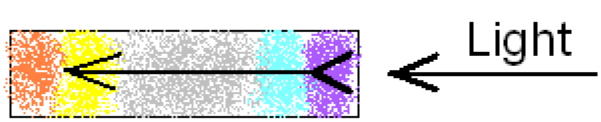
The cloudless day sky appears blue. The higher we look the deeper blue it gets, especially on clear days by low humidity. The lower we look, the brighter the blue of the sky, even turning into whitish. Mountaineers climbing mountains of thousands of meters tell us that the sky's color is the deeper blue the higher they get.

The mountains often appear blue, darker blue the nearer, lighter blue the more distant ones.

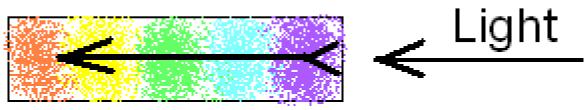
Dirty windows appear orange-yellow. If we apply a thin layer of sour milk to a pure glass surface and look through it toward a window with daylight, the surface appears yellow-orange. The same surface against a dark background appears faintly blue.

This phenomenon can be observed much more intensively on opalite stones, which can be found in jewelry stores. The opalite is a mineral, a sort of glass with certain additives. When we look through this stone toward a light source, it appears yellow, orange or red, depending on the concentration of additives in it (we need several stones with various concentrations to see this). If there are no additives in the opalite, it is transparent and colorless. At low concentration the stone turns yellow, with the increase it turns orange, to change to red and become almost opaque at very high concentration. The same stone against dark background appears violet at low concentration, blue at greater, to turn to bright blue (i.e. cyan) and then to whitish at very high concentration. If a small white LED lamp is neared the stone in the darkened room, then the yellow (orange or red) color is projected onto the opposite wall, while the stone itself is violet (blue, cyan or white).

If a longish stone with moderate and uniform concentration of additives is illuminated with a white directional LED lamp along its long axis, then it gives the following picture:



If a longish stone with a quick uniform *increase* of the concentration of additives along the long axis is illuminated with a white directional LED lamp along that axis from the side with the lower concentration, then it gives the following picture:



All the aforementioned phenomena and experiments, but also others whose description will follow, come down to a single principle, that of plus and minus. In this particular case we can call it “principle of an arrow”. The front part of the arrow is plus, the rear minus. We call the front part ‘plus’ because it has a penetrating effect, the back part ‘minus’, because it has a suction effect (+ <—< −) (the sign '<' appears both at the front and at the back of the arrow, so that already in it there are the plus and the minus (+<−).

In the flame of the candle and that of the lighter we recognize the same arrow pattern. In the front of this ‘fire arrow’ appears the plus color (yellow), in the back the minus color (blue-violet). As we will see later, the yellow and the violet are the two fundamental plus/minus colors of the world. The red (including the orange) is nothing else but weakening of the yellow and the cyan (including the blue) is nothing other than weakening of the violet. When the strong plus (yellow) meets the weakened minus (cyan), their overlapping bears the green. In this color dominates of course the plus, which can be seen, among other things, on the leaves of the trees. In autumn they turn from green to yellow, then to orange, and even to red on some trees, but never to cyan, blue or violet.

If the strong minus (violet) overlaps with the weakened plus (red), this gives birth to magenta. From the mixing of the strong plus (the yellow) and strong minus (the violet) we get nothing, that is, we get their cancellation, which means more or less dark gray.

When we observed the small opalite with moderate concentration of additives against a dark background, it appeared blue. What does that actually mean, to observe the stone against a dark background? It means that we are positioned sideways to the stone (*considering the direction of the light*), seeing the ‘tail’ of the light permeating through it. Therefore the stone appeared with minus color. Since it is small and with moderate concentration of additives, it cannot hold up the light, that is, the light just passes through it and makes a bright yellow spot on the opposite wall (of course we couldn’t see that by diffused daylight, but we have perceived it with the help of the directional white light of the LED lamp. To see the yellow color by daylight, we hold the stone close to the eye looking through it in the direction where the light comes from). But when the stone was longish with moderate and uniform concentration of additives, then, illuminated along the longitudinal axis, it absorbs i.e. withholds the light to a great degree, thus the yellow and red color project already in it on its far side. Therefore it gives the image of a candle or a lighter. When the concentration increases rapidly along the longitudinal axis, then the cyan catches up the yellow, thus the green color appears here as well.

From what has been said so far, we can conclude that the colors are born when the light encounters matter as an obstacle on its way of propagation, or, in other words, when it comes to friction between the light and the matter. For the eye which looks toward the light source, the light is white when there is no or little matter between the eye and the source (that is, when there is no or little friction). As the concentration of matter increases, so the light becomes yellow, then orange, then red; and at the end, at very high concentration, the light is completely attenuated and it turns black. In order to increase the resistance it is not always necessary to increase the concentration of the matter, but the thickness of the material layer, through which the light passes, can be increased as well. When the sun is rising or setting, the thickness of the atmosphere is largest, which is why the light is red. The upper part of the sky (i.e. the atmosphere) is blue because in relation to this part we are positioned sideways and see the rear of the light permeating the atmosphere.

As the sun rises further, so the thickness of the material layer through which we directly see the sunlight diminishes, thus the light becomes orange, then yellow, to finally become to a great degree white because the thickness of the layer, that is, the resistance is small and insufficient to color the light considerably. However, the rest of the sky is blue for the same reason we have already stated.

When we observe a small opalite stone with low and uniform concentration of additives sideways, it is blue-violet, with greater concentration blue and even greater light blue (i.e. cyan), and then turns into whitish. As the concentration increases, the light can permeate the stone ever less, thus the permeation gradually transitions into reflection of the light from the additives. And the white light, when reflected, remains white. For the same reason the part of the sky in the zenith is blue-violet (the thinnest layer to the dark, immaterial universe, especially at higher altitudes), the lower part is light blue, to end in whitish on the horizon. For the same reason we see often the near mountains blue-violet (especially if they are dark, that is, forested), the farther ones blue, and the very distant whitish. This largely depends also on the current degree of air humidity. The more transparent the air is (i.e. with less vapor), the darker blue the near mountains are (this does not apply of course to the very close ones).

Hence, the gradations in the two cases when we see the plus-side and the minus-side of the light, starting from low concentration of matter and going to higher, or starting from a thin, going to a thicker material layer by unchanging concentration, are the following:

White → yellow → orange → red → black (plus-side)

Black → blue-violet → blue → cyan → white (minus-side)

The images below show a kind of spiral, which we will call a double plane spiral. ‘Double’ because there are two spirals that intertwine. But which spirals are we talking about, the white or the black? The question makes no sense, because the one cannot exist without the other.

If we place the first spiral on a spinning top and turn it to the left, then it will expand (the circles will move outwards), that is, it will push us. If we observe carefully what is happening when the spiral is turning, we will notice that thin red strips appear along *the outer edges of the white circles* as long as the spiral is spinning fast, and as the spinning slows down, the strips turn into yellow (the experiment should be carried out near a window through which daylight comes in, but not directly from the sun). Along *the inner edges of the white circles* appear thin light blue strips as long as the spiral is turning fast, and as the speed decreases, the strips turn into blue-violet.

If we turn the spiral to the right, then the circles are moving inwards, so that now the red or the yellow strips are visible at the inner edges of the white circles and the light blue or the blue-violet strips at the outer edges.

Regarding the second spiral (on the right), it happens the same: when the circles are moving outwards, the red and the yellow strips appear on the outer edges of the white circles, and when they are moving inwards, these colors appear on their inner edges.

We see that here recurs the same pattern as in the previous phenomena: the yellow or the red color appear on the front and the blue-violet or the light blue on the rear side. For, when the circles are moving outwards (we say circles because the white and black spirals appear as circles as the top spins), then the outer edges of the white circles are the front and the inner edges are the rear of the light.

Since the light is reflected only by the white circles, we can imagine light and dark tubes, or, let's say toroids above the spiral. These toroids are static when the spiral stands still, however, they move outwards or inwards when it spins. To clarify what follows, we will make a comparison with a tennis ball bouncing off a horizontal surface at an acute angle. Thereby the angle of reflection is equal to the angle of incidence. However, this applies only when the surface is still. If it moves in the same direction as the projection of the ball's trajectory onto the surface, then the ball receives an additional movement, thus the angle of reflection is no longer equal to the angle of incidence, but greater.

Something similar can be assumed with light. When it is reflected by the white circles which move let's say outwards, the light gets additional movement towards the dark parts, in other words, it penetrates sideways into the matter (i.e. the air) over the dark circles which are in front of it. The sideways penetration means more intense friction than the frontal penetration, so this friction will produce plus colors on the front (that is, on the outer edges of the white circles) and minus colors on the rear side of the light (that is, on the inner edges of the white circles). During faster turning of the disk, the white circles move faster, the reflection angle is greater and therefore the friction intensity too, hence red color on the front and blue on the rear. As the rotation slows down, so the red turns into yellow and the blue into violet due to the diminishing of the reflection angle and therefore the intensity of the friction too.

It should be kept in mind that we deal here with diffuse light. This is mentioned for the reason that the direct light in some parts of the disk would receive additional motion, but in other parts retardation.

From the aforesaid the reader can conclude on his/her own, how we would explain the appearance of colors on Benham's disk when it turns.

Benham's disk

Let us consider the following simple experiment. We take a sheet of paper and hold it about 15-20 cm in front of one eye (the other eye is closed) opposite to an uncurtained window through which daylight comes in. If we observe the edge of the sheet carefully, we will notice a thin yellow strip along the edge. We can also pierce with a needle a small hole in the sheet and look through it towards the window. Then we will notice yellow circle along the edge of the hole.

If we now move the sheet in front of our eye so that its edge and the dark edge of the window, which of course is farther away, form a narrow gap, then we will notice that a blue strip appears at the farther edge. If this arrangement is fixed and the gap is quite narrow, then green emerges in the middle between the plus- and minus-colors.

Now we go over to experiments with a triangular glass prism (hereinafter: prism). The experiments that will be carried out we can divide into subjective and objective ones. Subjective experiments are called those when we hold the prism in front of our eyes and look through it, and objective ones are called those when we let directed light pass through the prism and then, at a greater or lesser distance behind the prism, the path of the light is blocked with a white wall which may also be a white sheet of paper that we hold behind the prism.

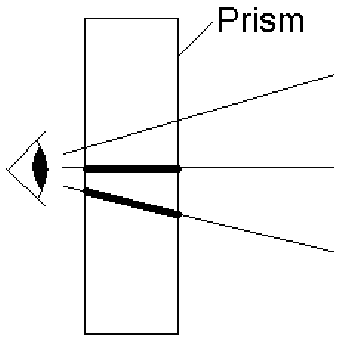
The prisms are usually made equilateral with three equal angles of 60°. One of them, in this case no matter which, is called the (refractive) angle of the prism (also called apex angle). It can be directed upwards (▲) or downwards (▼). In essence, there is no difference between the two possibilities. We will choose the second variant as more practical. Therefore, unless explicitly stated otherwise, the variant of refractive angle being directed downwards is assumed.

If we hold the prism horizontally in front of our eyes and see through it the objects around us, we notice that they take on additional impressive colors. On closer inspection, we perceive that these colors appear where abrupt horizontal or oblique dividing lines (boundaries) between lighter and darker surfaces can be seen, whereas on the monochrome surfaces and on those with gradual transition from light to dark as well as on vertical abrupt ones, it is not the case. In the last variant, however, we will notice additional colors if we hold the prism obliquely or vertically in front of our eyes. The additional colors are paler if the transitions are milder, i.e., when the boundary contrast is weaker, and they are stronger when the transition is abrupt. They are actually the strongest on boundaries where the contrast is maximal, that is, between deep black and bright white.

Therefore, we will begin our experiments with detailed observations of boundaries between black and white areas. We can do it by preparing white papers with black areas on them or vice versa, or we can attach on a window, through which daylight comes in, opaque cardboards into which certain shapes are cut. The attached cardboard represents a black area, and where there is none, it is a white area. If we attach a cardboard in the shape of the image below (left) and the rectangle is wide (it is meant the smaller side), and if we look at this shape through the prism from a few steps away, then we see the image in the middle. A yellow-red strip appears at the upper formerly black-white boundary and a violet-cyan one at the lower boundary. If we turn the prism so that the refractive angle is upwards, then we see exactly the same image, only turned 180°. We notice that in the yellow-red strip the yellow stripe is noticeably wider than the red one, and in the violet-cyan strip the violet stripe is noticeably wider than the cyan one. If we approach the rectangle without losing sight of it through the prism, the yellow-red and the violet-cyan strip become ever narrower, and as we get very close, they practically disappear. If we move away from the rectangle, then the colored strips spread. As they spread, so the yellow and the cyan stripe begin to approach each other, the white space between them is getting narrower, and at the moment when the two colors touch, a green color appears.

As we move farther away, so the cyan disappears, more precisely, it is completely covered by the yellow and converted into green, so now we see four colors: red, yellow, green and violet (the right image above). As we move even farther, a dark field begins to appear between the green and the violet, and even farther, between the red and the green. The gap between the green and the violet can be clearly seen when we observe through the prism a distant, not very bright white street light.

If we think about where we actually see the image through the prism, we realize that it is shifted and that we see it significantly below its true position when the refractive angle is turned downwards, and significantly above its true position when the angle is turned upwards (the eye always projects the image in a straight line; for example, when we see an image in a mirror at an acute angle, then we see it as if it were behind the mirror. Moreover, when we look through a window the image is also shifted, except in the case when we look through it at right angle.). We see the colored rectangle also a bit curved, namely, curved upwards in the first case (▼), and curved downwards in the second (▲) (here again we get a kind of arrow shape with the plus-colors at the front and the minus-colors at the back) . This is because when we look through the prism, we see both in straight and oblique direction at the same time. In the straight direction there is smaller thickness of the prism in front of our eyes, in the oblique direction larger thickness (figure below). With larger thickness of the prism, the image is shifted more.



This curvature does not appear in objective experiments when the sunlight falls on the prism, but it can be seen when the light source is a small white LED lamp that we hold close to the prism. The reason is that the sun is a broad and distant source, thus its rays can be regarded as parallel, whereas the LED lamp is a small and near source and therefore to it the same applies as to our eyes.

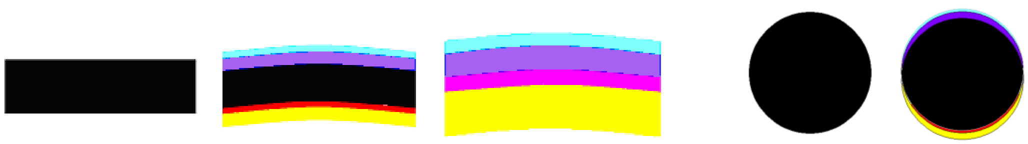
If we observe the same image through a prism with a smaller refractive angle, say 15°, the only difference is that the image is shifted less and the width of the colored strips is smaller, of course at unchanged distance of the observer from the image. In order to get the same width of the strips with this prism as with the equilateral one, we have to move farther away from the window.

If we look at the image through prisms with equal refractive angles, from the same distance, but made of different types of glass, let's say from flint glass, crown glass and acrylic glass, then here again we will notice differences in the displacement of the image and the width of the colored strips. With the flint-glass prism the image is mostly shifted and the thickest strips can be seen, because the flint-glass has the largest refractive index of these three types of glass, that is, the light is mostly refracted by it.

Now on the window we attach two square cardboards, horizontally placed, one above, the other below, which touch each other only with their vertices (left image below). When we look at this image through the prism we will see the image on the right. We have brought the yellow-red and violet-cyan strips on the same line. We notice that there is equality in width of both groups as wholes, but also separately: the width of the yellow stripe is the same as that of the violet, and the width of the red stripe is the same as that of the cyan. What is also very important here is that the red stripe appears at the expense of the black field, while the yellow one at the expense of the white field. In other words, a part of the black field has turned into red and a part of the white field has turned into yellow. Consequently, on the other side a part of the black field has turned into violet, and a part of the white field into cyan. Where previously was the boundary between the white and the black, now that line is the border between the yellow and the red, i.e. between the violet and the cyan.

Let's move on to observing a circle. For that purpose we cut a circle in a cardboard and attach it to the window (left image below). When we look at it through the prism we see the image on the right. The circle has taken on an elliptical shape with the well-known yellow-red belt above, and the violet-cyan one below. They are widest at the top and the bottom edge and become narrower to the left and the right side, to completely disappear at the extreme left and right points. If we fix a small circular cardboard in the center of the circle, in other words, if we place a central point in it, and although this small circle observed through the prism becomes colored and therefore not clearly visible, we can still notice that it appears in the upper part of the ellipse.

Instead of observing white rectangles and circles on a black background, we can do the opposite - observe black rectangles and circles on a white background. For this purpose, we attach on the window an opaque cardboard, once in a shape of rectangle, another time in a shape of circle. We will get the images below. In this case, if we cut the rectangle or the circle smaller or if we move away from the window, then the violet and red color will approach each other and at the moment of their overlapping, a magenta color is born. In the first case the wider yellow(+) and the narrower cyan(–) overlap to produce green. This time the wider violet(–) and the narrower red(+) overlap producing magenta.



In the case of the black rectangle, it is as if we had two light sources, one up and one down (the rectangle can be imagined very long). The “head” of the lower light with the plus-colors manages to catch up with the “tail” of the upper light with the minus-colors and their overlapping gives birth to magenta.

If on a white sheet of paper we draw horizontal narrow black stripes so that they are of same width to each other, but also to the white space between them (i.e. to the white stripes) and if we look at this image through the prism from a certain distance, then only two colors remain in the image: the green and the magenta. In other words, instead of black and white, we now see magenta and green stripes. In the white stripes the yellow has merged with the cyan becoming green, in the black stripes the violet has merged with the red becoming magenta (white stripe → green; black stripe → magenta).

Instead of fixing cardboards on the window, we can carry these experiments out in such a way that we observe the mentioned shapes as black and white areas on paper. We will see through the prism exactly the same images as before. But in this case we can make the black and white areas in different shades of gray, from black to white. We will notice that absolutely nothing changes in the results, except that the colors get paler the less the contrast is between the darker and the brighter shades of the gray areas. In order to accomplish this grading on the window, we need papers with different transparency. This can be achieved by using tracing paper in different number of layers.

The simplest objective experiment we can make with the prism is the following: in a room in which direct sunlight enters we hold with one hand the prism with the refractive angle downwards, of course in the field of the direct light, and with the other hand behind the prism we hold a white sheet of paper. If we hold the sheet very close to the prism, then we will see only a regular bright rectangle without any colors.

As we move the sheet away from the prism backwards and upwards at once, then a violet-cyan strip appears on the upper edge of the rectangle, and a yellow-red strip on the bottom edge. With the further withdrawal of the sheet, the strips widen, the middle white field of the rectangle narrows, and in the moment when the yellow and cyan stripe overlap, green appears. Now we have a continuous colored spectrum.

All that has been described so far for the subjective experiments with the prism fully recurs in the objective experiments; therefore we don't feel it is necessary to repeat it again. We will only describe the way these experiments are carried out.

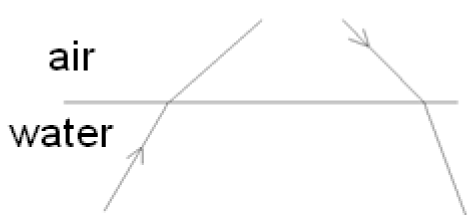
On the incident or emerging surface of the prism we stick an opaque paper and once we cut a rectangle, another time a circle in it. Then we hold the prism in the sunlight as we did in the previous experiment. This corresponds to the observing of the rectangle and the circle in the subjective experiments. Moving the sheet of paper behind the prism corresponds to the changing of our distance from the window with the attached cardboards.

Another time on the prism we stick two opaque squares that touch each other with the vertices. In order to get a magenta color we stick only an opaque strip horizontally in the middle of the prism. To carry out the experiments with darker or brighter gray surfaces, we need to stick on the prism more or less transparent papers.

Glass prisms that can be purchased have not large incident surface, so that no significant amount of sunlight can fall on them. By covering it partly with opaque papers we reduce the amount of light further. Therefore, the colored phenomena at a not large distance behind the prism already weaken. Glass prisms with bigger surface would be quite heavy. Johann Wolfgang Goethe for this purpose used a water prism in his experiments. It was a prism-shaped 'aquarium' fixed in a wooden construction with the refractive angle directed of course downwards. The prism had a large incident surface so that the phenomena at longer distances didn't weaken significantly. Its drawback was that after every experimenting the water had to be exhausted because the lime would have blurred the glass.

What we have said earlier, that there is no difference between the subjective and the objective experiments is not entirely true. There is still one small and baffling difference. When we observed a white field on a black background through the prism (▼) in a subjective experiment the plus-colors were up, the minus-colors down, whereas in an objective experiment the colors are placed reversed.

The colors obtained with the help of the prism are called refractive colors, because the Latin term "refraction" refers to the change of direction of light when it falls at an acute angle on its way from one medium to another which have different densities. The refraction is always inclined to the denser medium, that is, the made "curve" encompasses larger part of the denser medium (figure below).

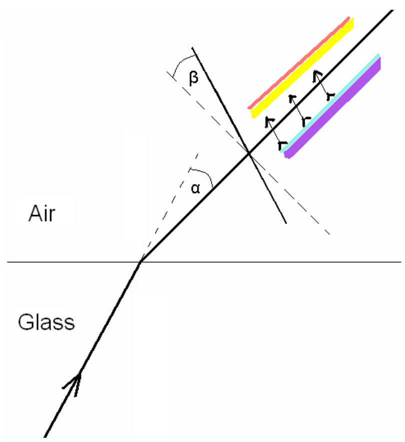


We can compare the change in the direction of light with a vehicle passing over at considerable speed at an acute angle from solid asphalt into soft muddy ground or vice versa. The vehicle then inevitably changes the direction of movement.

When a vehicle makes a curve as it moves, then the inner wheel travels a shorter distance than the outer. Once the vehicle enters the soft earth, one wheel will collide with the greater resistance of the soft earth earlier than the other, causing the vehicle to turn to the side where the first wheel travels a shorter distance. The same principle applies in the opposite case.

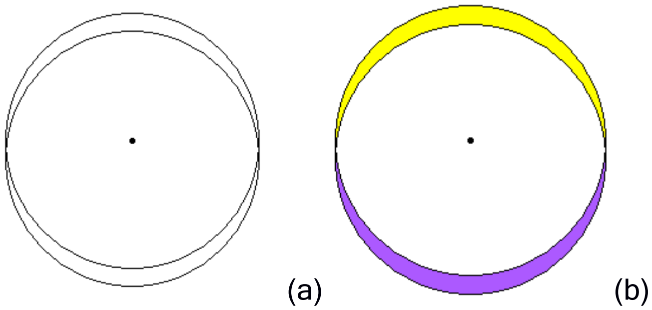
The light undergoes two refractions on the prism, one at the entering and the other at the emerging surface. For the birth of the refractive colors there is no need of a double, but only of a single refraction. For example, when a circular beam of light falls on water surface at an acute angle, then it refracts, and on the bottom of the water, if it is white, the same can be noticed as on the screen in the objective experiment, in which an opaque paper with a circular hole was stuck on the prism surface.

When a beam of light propagates in space, its frontal surface is at right angle to the direction of propagation. We call this frontal propagation of light. But when the beam makes an ‘arc’, that is, when it is refracted, then it propagates sideways, meaning that its frontal surface is no longer perpendicular to the direction of propagation. We call this sideward propagation of light. This mode of propagation entails an increased friction of the light with the matter, and the sharper the ‘arc’ described, the more intense the friction. The friction causes birth of colors, and the plus-colors in the objective experiment are always on the outer side - we would like to say on the outer front of the ‘arc’ - and the minus-colors on the inner front. We see that the same pattern recurs here as in the previous cases: the plus-colors hurry ahead, while the minus colors run after. (The straight lines of the incident and the refracted beam can also be considered as two arms of the arrow sign (<). The plus-colors are on the front, and the minus-colors on the rear of the arrow, but of course only on the refracted arm.) Thereupon it can be assumed that the angle for which the light is slanted, i.e., the angle (β) between the new frontal surface and the normal of the new direction is equal to the angle (α) by which the light is refracted, i.e. that β = α.



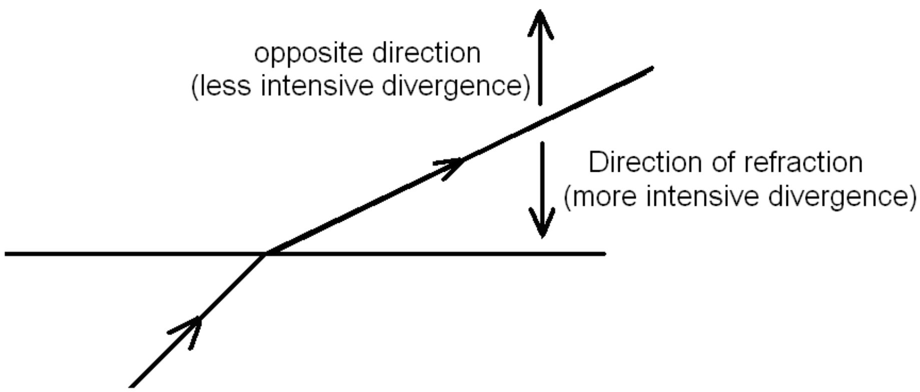
It follows that if we mount a prism at one end of a vacuum tube and a white screen at the other end, then from the falling light on this prism no colors on the screen will appear, because although the light is slanted and spreads sideways, it has still nothing to rub against behind the prism.

When we observed through the prism the white circle on black background with a point in the center, we noticed that the point appears in the upper part of the elliptical shape. In the figure below two identical circles are drawn, one slightly below the other, with a central point belonging to the upper circle. This image actually represents the slanting of the light beam after the refraction. Although the circle resembled an ellipse, it is not an ellipse, because two circles one below another cannot form an ellipse. The figure (b) shows where the two basic and - in the case of the prism - wider colors appear, the yellow and the violet. For greater clarity the narrower colors, red and cyan, are not shown. The red appears outside the circles over the yellow, and the cyan inside the circles over the violet.



An important term that we need in order to understand the further details of the refractive colors is the divergence of light, that is, the cone-shaped propagation of a light beam. Every light diverges more or less, even the laser light. For example, if we direct a laser beam from the roof of a tall building onto the street at night, we will notice that the bright circle on the street is considerably larger than that on a nearby surface.

When a light beam refracts, its heretofore uniform divergence to all sides becomes non-uniform, namely more intensive in the direction of refraction and less intensive in the opposite direction (figure below). The greater the refraction, the more intensive the divergence.

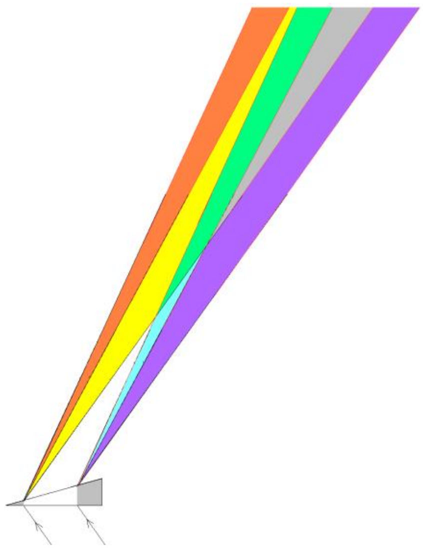


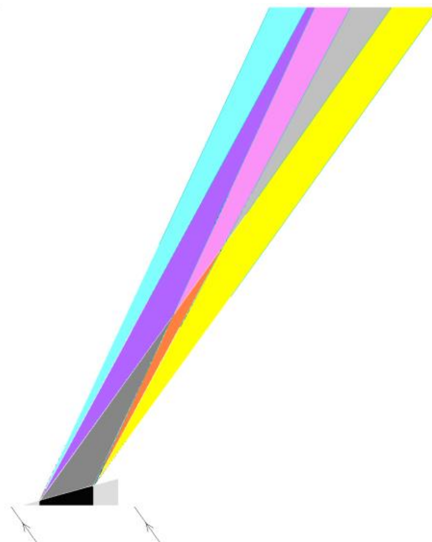
The yellow stripe is wider. In its width manifests itself the larger divergence in the direction of refraction. The red stripe is less wide. In its width manifests itself the smaller divergence in the opposite direction[17]. Since here the light spreads counter the direction of refraction, its attenuation is the greatest, so it appears red in color. The same happens to the other side, where what was said for the yellow applies to the violet, and that for the red applies to the cyan. When the yellow and the cyan meet at a certain distance behind the prism, their overlapping gives birth to green. A little farther the cyan is completely overlapped by the yellow, so now there are only four colors: red, yellow, green and violet. However, the overlapping does not mean that the colors are mixed as chemical colors so that they no longer exist as separate, but they continue to spread out as separate entities, thus a little farther the yellow begins to overlap with the violet. What's going on here? When something moves circularly or spirally, then in one point of the circle it has one direction of motion, but in the diametrically opposite point it moves in the contrary direction. Yellow, overlapping with the violet, encounters something that has a 180° contrary movement; therefore, the two cancel each other out, resulting in the appearance of a more or less dark gray part that comes at the

[17] It is interesting that chemical additives in the glass can change the intensity of divergence. With certain additives that are more on the acid plus-side (e.g. metal oxides) and with certain additives that are more on the alkaline/base minus-side, two different glasses can be obtained which have the same refractive index, i.e. they “bend” the light equally, but the width of their colored strips are different. They are wider in the glass with oxide (+) additives and narrower in the glass with alkaline (–) additives.

expense of the violet. This happens because both have the same intensity, which was not the case with yellow and the cyan. Although they have contrary motion too, the intensity of the cyan is less than that of the yellow, so the resulting green is actually a slightly stifled yellow. Even farther, the cyan will penetrate into the red field resulting also in cancellation, that is, in a dark part that goes at the expense of the red stripe, for here applies the aforesaid about the intensities.

The appearance of gray parts between the prismatic colors in combination with something else, which will be left undiscussed for now, is the basis for understanding of what in science is called a discrete spectrum.





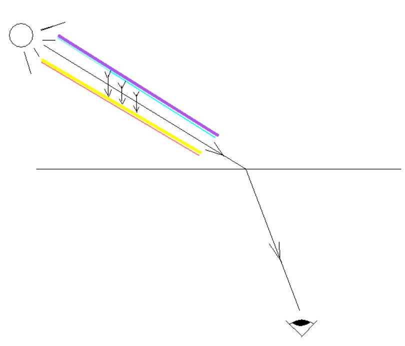
All that has been said above applies also in the case when instead of opaque paper with a circular hole we stick on the prism a circle of opaque paper. But this time instead of overlapping of the wide yellow and the narrow cyan, an overlapping of the narrow red and wide violet takes place, resulting in magenta.

The following pairs: yellow-violet, red-cyan and green-magenta are the plus-minus pairs that cancel each other because of their equal intensity. These pairs are also called complementary colors.

If we draw in a computer program a yellow field on a white background and stare at it for a while, then we close the eyes and wait for a few seconds ‘looking’ into the blackness, then we will ‘see’ a violet field with the same shape forming out of the blackness. Instead of closing the eyes, we can point them to a nearby white wall. The same effect happens. If we repeat the experiment while looking at a violet field on the screen, then after closing the eyes we ‘see’ a yellow field. The experiment can be repeated with the other pairs: red-cyan and green-magenta. Trying out with other ‘inter-colors’ we can easily determine which color is complementary to which.

If we stare at the double plane spiral for ten seconds or more while it is spinning and then point our eyes to a nearby surface, we will notice that now we see the opposite of what we were looking at before. If the spiral was contracting (–), then the surface we look at thereafter is expanding (+); and if the spiral was expanding (+), then the surface we look at thereafter is contracting (–).

From the previous two groups of experiments we can conclude that after a certain stimulus the eye does the opposite to relax: if the stimulus was plus, then minus follows for relaxation and vice versa, similar to a rolled up paper which afterwards we roll up in the opposite direction to better flatten it.



We said that in the subjective experiments with prism the plus and minus colors switch their places in comparison to the objective ones. It comes therefrom, that in the objective experiments the outer side of the made ‘arc’ is its front, thus the plus colors appear there, while in the subjective experiments the front is the inner side of the ‘arc’. We can explain this on a single refraction experiment. In the image on the left the eye is in water, while the light source is in the air above the water. I believe the figure speaks for itself.

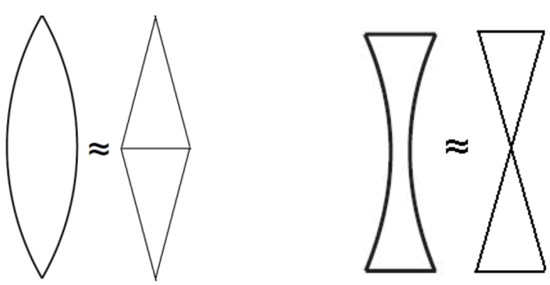
If we place a coin on the bottom of a bucket filled with water and look at it from above, we notice that the coin seems closer than it actually is, that is, closer compared to when the bucket is empty. We can say that the light in the water, as a medium where it is confronted with greater resistance, experiences greater compactness of its spirality than in air. Here we can make an analogy between light and electricity. Electricity in wires with greater electrical resistivity (ρ) also experiences a greater compression of its spirality.

If we hold a convex lens (i.e. a magnifying glass) before one eye - the other is closed - and look through it at a distant white circular street light at night, then we see it surrounded by a violet-cyan ring. As we move the lens slowly away from the eye, the light circle along with the ring enlarges so that at one point everything dissolves in whiteness on the whole lens. With the farther moving of the lens, a distinct white circle begins to form again, which is now shrinking. But this time the violet-cyan ring has turned into a yellow-red[18]. However, not only this can be noticed. Everything that we see now is upside down. If we carry out this experiment as an objective one, by sending through the lens white light and catching the image on a white sheet of paper behind the lens, then we get the opposite situation. When the sheet is close to the lens, the bright circle has a yellow-red ring. With the moving of the sheet farther away, the circle shrinks to a point; thereafter it begins to grow, but this time with a violet-cyan ring. Now the image is turned upside down.

At this point, on the example of this objective experiment, I would like to call on the reader to think dynamically in moving pictures: the sheet moves away from the lens, the image is upright, the colors are plus; the image shrinks to a point, then it turns upside down (what we can call a minus image), the colors also turn into minus.

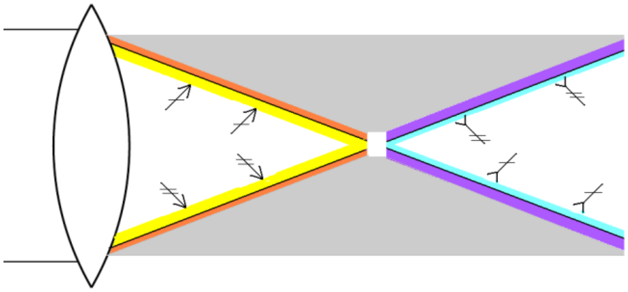
The experiments with a concave lens do not have the diversity of the previous ones. In a subjective experiment with this lens, we see the street light surrounded by a yellow-red ring and the directed light through the lens in an objective experiment with a violet-cyan ring. There are no inverted images here, therefore no transitions from minus to plus colors or vice versa.

A convex lens can be regarded as a very large number of small prisms arranged in a circle with refractive angles directed outwards. The surfaces of these prisms are not flat, but slightly curved. The same applies to the concave lens, except that here the refractive angles of the prisms are directed inwards, that is, towards the center.



[18] If the lens has a large focal length (which happens when the curvature of the lens is very mild), then the length of our arm will not be enough to see the yellow-red ring. In this case the magnifying glass should be stuck to the window and we should move away from it looking constantly at the light through the lens.

The image below shows the objective experiment with a convex lens. Before the light focuses in a point, its front rubs with matter; after the focal point it occurs with its rear. Hence, yellow-red color in the first section and violet-cyan in the second.



If we cut two equal squares from a polarizing film and put one on top of the other and look through them, then we can see quite well. If we now slowly rotate one of the squares, the picture gradually darkens and when we have rotated it by 90°, that is, when the square exactly covers the other again, the picture is completely dark. In this position the squares become opaque. By rotating it further, the picture brightens, so that in the next 90° position the squares are well transparent again. In few words, with every 90° turn we get alternately full brightening and full darkening, and a gradual transition in between.

Every polarizing film, though transparent, it is still considerably dark. We can notice this well if we lay a piece of crystal clear glass and a piece of polarizing film next to each other. Therefore, we can assume an existence of regularly ‘woven’ dark ‘threads’ throughout the film.

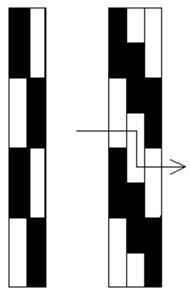
If we print two identical copies of the double plane spiral on tracing paper, then put the one on top of the other and look through them, we will notice that the same happens: at every 90° rotation we get brightening and darkening. This cannot happen with another type of spiral. With a single spiral it happens at every 180° rotation, with a triple spiral at every 60°, with a fourfold at every 45°, etc.

What does the double spiral actually represent? In it we have the same principle that we have already encountered in the EM-element. And in this element there is at the same time the principle of the arrow.

Sign or symbol from an ancient people

If we cut a third equal square from the polarizing film and place it between the other two (when they are arranged in a state of opacity) either in the 0° or 90° position, then nothing changes, i.e. the opacity remains. But if we begin to slowly rotate the middle square, the picture gradually brightens. When we reach the 45° position with it, the picture is completely brightened and if we continue to rotate it further, the picture darkens again and is completely dark at 90°. The same recurs with the further rotation of the middle square.

When it is in the 45° position, the middle square serves as a bridge between the other two. For the first square we could say that it is at 12 o'clock position, the middle one at 1.30, and the third at 3 o'clock, so the spirality can be established through the three, that is, the light can pass through the foil squares.

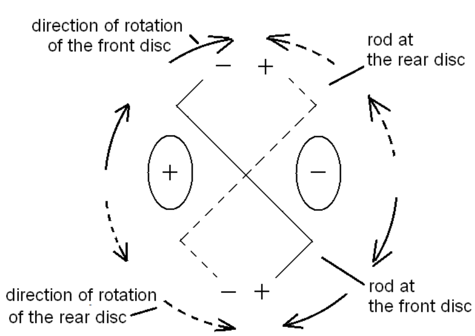


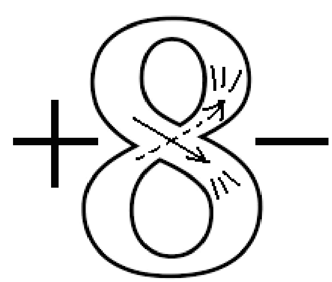
The aforesaid can be described also a little differently: the first figure on the left schematically and principally shows the situation with only two polarizer foils when they are in arrangement of opacity. The second figure shows the situation when between the two is inserted the third polarizer foil at an angle of 45° relative to the others. At this arrangement of the foils the light can pass through, which is symbolically shown with the broken arrow. This is possible thanks to the inherent spirality of light.

Supplement about the Wimshurst-Generator

Whether the electricity will appear in the horizontal or the vertical quadrants depends on the direction in which the disks turn, but also on whether the metal rod at the front disk is set like this (\), and the one at the rear disk like this (/), or vice versa. When randomly choosing one of the four possible combinations, we can guess in which quadrants the two electricities will appear on the basis of the following observation: the electricities appears in the pair of quadrants approaching the metal rods. In other words, if the disk in front of us is turning to the right and its rod is set like this (\), then the horizontal quadrants are those which are approaching the rod.

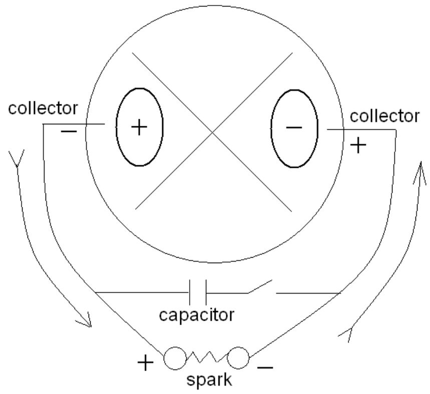
Our preliminary explanation is the following: if the disks rotate without the metal rods, then the electricity generated in them by influenza (if at least one disk is initially minimally electrified) chaotically appears and simultaneously neutralizes, resulting in zero electrification. The rods are those that play the role of separators (differentiators) of the two electricity types. The picture below shows a situation where the front rod is set like this "\" and the direction of rotation of the front disk is clockwise. The electricities appear in the horizontal quadrants. The upper brush of the front rod due to the rotation of the disk bends to the right, the lower brush to the left, whereas at the back rod it is reversed (interesting thing is that the rods get the form of an EM-force element). In the upper and lower quadrant the electricities neutralize themselves in a kind of closed electric circuit in a form of a lemniscate (or Mobius strip), i.e. of number 8 (figure on the right), thereby in the left and the right quadrant they get the opportunity to freely develop.

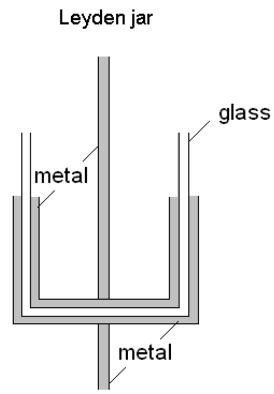




We deem the metal sectors stuck on the disks to be of no key importance because in another variant of this generator, the Bonetti-machine, there are none of them. Along the metal rods there are metal combs facing the disks, but they don't touch them by just a little. Although the combs do not bend like the brushes of the Wimshurst, the principle does not change.

When in the close vicinity of the disks we add two more metal rods (so-called collectors) which at their other ends end with metal balls, then weak and frequent sparks begin to jump across them when the disks are spinning. Adding a capacitor or capacitors (Leyden jars) between collectors' conductors, the sparks appear less frequently, but they significantly gain power. The accumulated energy in the twisted capacitor(s) multiplies the power of the spark many times. Thereby the capacitor(s) abruptly untwists and the process begins anew.





⬩⬩⬩⬩⬩⬩⬩⬩⬩

We will finish here this paper for now. I am convinced that when the principle put forth in it will be applied to the other branches of natural science, then they too will experience such deepening as they do not know in the past. For, I believe that everyone is crystal clear that the principle is universally applicable. Look at the chemistry for instance. The two opposites there are acids and alkalis (bases). The acids color the blue litmus in red; alkalis the red in blue. What does that mean? Nothing else than that the acids are plus-, the alkalis minus-compounds. These at first sight not very significant designations will lead to such an extension of knowledge, the scale of which can be predicted not by many. Recall the beginning of this paper. From a simple principle, visible in everyday life, we have come to an explanation of the essence of electromagnetism and of light, about which for three hundred years, despite a series of generations, hundreds of institutes, expensive technical aids, etc. etc., there exist huge misconceptions. And those who even after reading this paper still believe in the claims of the current theories, I invite them to ask themselves, what their belief is based on? How many experiments can they adduce and describe to support the theories?! And not only that, but much more importantly, do they know some details about those experiments (if they know of any experiments at all)? For, every single phenomenon in nature has many details and a significant number of them are for sure not in accord with the current theories.

The more details one knows about a phenomenon, the more pebbles he has ordered in the mosaic, thus the picture begin to reveal by itself alone. Goethe deemed that man should not set up theories, but as much as possible should manifold the experiments[19], and when there are enough known details about the phenomenon (that is, enough pebbles in the mosaic), its being comes to light by itself.

Take the example of the prism. How many details people know nowadays about the phenomena thereof? My assumption is that the vast majority of the mankind has not seen a real prism, let alone knows some detail of the experiments with it. And the thing costs couple of dollars and everyone can afford it. In the

[19] By "manifolding an experiment" Goethe meant that one should carry out all possible variants of it, i.e. one by one to change all the circumstances that may come to his mind. First, let's say, we change only the distance of the screen behind the prism but all other circumstances are static. Then we change the refractive angle of the prism at an unchanged distance of the screen. Then we change the type of glass which the prism is made of. Then we observe various black-and-white, but also colored shapes through the prism. Then we make a large water prism to clearly see the behavior of the phenomenon as far away from the prism as possible, etc., etc. As the observations from the experiments are multiplied, so the phenomenon becomes part of the observer's being and it is only a question of time and patience when it would reveal itself to him.
Of course, this is often a slow and painful process, because sometimes changing only one circumstance of the experiment requires a lot of effort, accompanied by a lot of frustration due to the frequent lack of appropriate requisites to carry out the imagined variation of the experiment.

school textbooks you will see only a drawing of a triangle from which colors emerge (which drawing, by the way, is a flagrant deception, because at the very exit of the prism there are no colors yet) and here ends the whole story. So: we have a prism, a light falls on it, colors appear on the other side. Hence, we conclude: the light consists of colors and the prism divides it into its constituent parts. End of the story. We have come to the truth. We have learned how the light produces the colors. That the nature is not so vulgarly trivial and that there is a multitude of details about the phenomenon, the young student does not even anticipate it, because he naturally has trust in the common sense, in the scientific methods and the sincerity of the people - that is, that they will tell all the details they know about the phenomenon, even if they are not in favor of the theory.

However, the colors are born not only on the prism, but in countless other ways. And now the ‘truth’ that a man has fantasized out about the development of the prism’s colors begins to graft onto all other phenomena, and thereby to invent some most incredible stories, without any basis, without any experimental fact that goes in that direction[20]. For example, the sky were blue because the air molecules scatter the blue color from the sunlight?! What about the other colors? What happened to them? And where are the experiments this explanation is based on? Probably the air molecules are especially angry only with the blue color of the light. If it were not tragic, it would be . And this is just one of a great many other examples.

⋄⋄⋄⋄⋄⋄⋄⋄⋄

Today's human is little aware of the scope of the impact of science on the overall society and on the general social conditions. A natural science, different from this nowadays, will undoubtedly lead to profoundly positive changes not only in terms of understanding nature, but also in terms of social conditions in general.

[20] And this science is constantly boasting that it, as an antipode of religion, is based not on beliefs, but on experimental facts.