A Qualitative Approach to Electricity

A Guide to Visualising Eletrodynamics using symbols and images

Hermann Haertel,
Guest Scientist at ITAP , University Kiel
(Institute for Theoretical Physics and Astrophysics)

The following paper was written during a sabbatical at the Palo-Alto Research Center at Palo Alto and published in September 1987 as IRL-Report 87.0001. (ED 287 730,SE 049 690)

It is now made electronically available in a slightly shortened and revised form.

ABSTRACT

In the teaching of physics, the study of electricity and magnetism typically follows the introduction of the basic concepts of mechanics. However, there are some new concepts associated with electromagnetic fields that seem at first to the student to be unrelated to, or even incompatible with, Newton's third law as learned in mechanics. Furthermore, the transition from electrostatics to studies of moving charges and associated magnetic phenomena seems to many thoughtful students not to be consistent with concepts learned earlier in the course. This report describes approaches to electrostatics, to elementary circuits, and to the effects of moving charges in a way designed to be fully consistent throughout, so that the student is not left with quandaries about the relationship of each set of basic concepts to the other sets in the course. Included are chapters on: (1) qualitative reasoning; (2) basic electricity; and (3) electromagnetic induction and relativity.

PREFACE

In Chapter I some basic principles about qualitative concepts and qualitative reasoning are presented. In Chapter II, the subject matter of Basic Electricity, including Voltage, Current, Resistance, Ohm's Law and Kirchhoff's Laws, is reconstructed stressing the importance of the systems aspect of the electric circuit and the relation between microscopic and macroscopic effects, especially in respect to voltage.

In Chapter III, the magnetic interaction, the electromagnetic induction and wave propagation are described. This description is based on the relativistic change of the Coulomb field due to the constant velocity of charge carriers and the existence of circular fields, connected with accelerated charge carriers.

The material presented depends on the fact that the presentation will be supported by interactive and animated computer graphics. The development of material for teaching and instruction on the basis of this approach as well as a further development of the conceptual framework will be part of the research agenda of the Institute for Research on Learning, Palo Alto, and the Institute for Science Education, Kiel, Germany. This text is therefore not addressed to students as newcomers to this field. It is presumed that the reader has a good knowledge about the basic facts of electromagnetism as they are described in traditional textbooks.

Acknowledgment

I want to thank all of my friends and colleagues at Xerox PARC Intelligent Systems Laboratory (ISL) and at the Institute for Research on Learning (IRL) for their stimulating and constructive support. My discussions with Zvonko Fazarinc from Stanford were of great help for writing the last chapter.

Especially, I wish to thank John Seely Brown with whom I began to discuss these ideas at the conference on Intelligent Tutorial Systems in Tübingen, June 1985. He has made my stay at PARC possible and has supported this work with numerous, valuable ideas, constructive criticism, and continuous encouragement.

List of Contents

Chapter I QUALITATIVE REASONING

Chapter II Basic Electricity

Chapter III Electromagnetic Induction and Relativity

Chapter I QUALITATIVE REASONING

1.1. INTRODUCTION

The dominance of quantitative procedures and the importance of mathematical formalism in teaching physics is well-known. What seems to be missing is an understanding of the role of qualitative reasoning. How does insight emerge from qualitative models, and what is the importance of constructing explicit bridges between the formalism and these conceptual models? This deficit may be one of the reasons why so many people fail to learn successfully in this field.

The current method of teaching physics is good in handling the abstractions in mathematical notation, but it is weak in supporting qualitative thinking in a consistent way. There are reasons for this deficiency. In the history of physics, many qualitative models can be identified with generations of physicists, who have strongly (and sometimes desperately) believed them to be valid. The theories about phlogiston and about the ether are two famous examples which misled physicists for a long time, and much scientific effort was used for the incorrect purpose. All these models had finally to be given up; and when quantum physics, wave/particle dualism, and the Theory of Special and General Relativity arrived, it seemed to be clear that only the quantitative approach - the system of differential equations, and the correct handling of this formalism -could guarantee progress and success. Qualitative models or concepts could only be used for special cases with a limited range of validity and without underlying common principles.

The body of physics knowledge condensed in textbooks and simplified for the different school levels is usually described as a consistent quantitative system, carefully prepared and standardized according to units, syntax, methods, etc. This quantitative system is surrounded by isolated qualitative models, meant to help students to understand isolated phenomena, to give a background for causal relations, and to model certain processes. In contrast to the quantitative side, the qualitative one is not treated with equal care. For many aspects, the underlying essence is only presented implicitly or is not presented at all. The statement of Hertz, "The physics of electromagnetism is Maxwell's equations." expresses clearly the attitude of overemphasizing the quantitative side and even denying the existence of qualitative models and questions about the underlying ontology as part of physics. If qualitative models are presented, their limits and questions about these limits are, in most cases, left aside. Inconsistencies are overlooked or hidden under shallow explanations with the excuse that, as a rule, students would never detect these inconsistencies and would only be confused by any further and more detailed explanation. Worse yet, there may even be a belief that students should not worry about inconsistencies; truth is within the equations themselves.

This point of view and the established practice are to be challenged for the following reasons:

1. There are many hints that physicists who are working in new fields, and other experts who have to apply their knowledge in a creative way, are permanently using qualitative models or concepts in an implicit or explicit way. Einstein, who said about himself that he was weak in calculus but strong in handling pictures and using spatial imagery, is a famous example to support this statement. It is an open question to what extent this qualitative expert knowledge could be made explicit and be supportive for better teaching.

2. It is unknown and difficult to detect to what extent students are, in fact, aware of inconsistencies and missing relations between qualitative models or have the capability to see these discrepancies. It is unknown to what extent this may cause them to experience failure and frustration in the learning process and inhibit interest and motivation for further learning in this field.

3. It is unknown to what extent the teaching of the underlying qualitative models and the application of consistent qualitative reasoning can support physical intuition and the capability of problem-solving.

4. There are indications that better models can be developed and presented to learners - especially with regard to modern computers and their ability for animated and interactive graphical representation. The introduction of the computer as a new medium for teaching can therefore have a strong influence on the way physics is taught. Such a medium is asking for a revision of all qualitative ideas, models, concepts..

The project described in the following paper concentrates on electricity. This field was chosen because of its importance as a basis for modern technology and because it is traditionally known to be very abstract and difficult to understand. Even small progress in this field would therefore have a positive influence on a broad field of applications. The objectives of this project are to find the answers to the following questions:

1. What are the most common discrepancies and inconsistencies among qualitative concepts offered in traditional teaching?

2. Is it possible to develop a consistent qualitative concept for electromagnetism which could be used in a generic form to support an introductory course and which could be developed in a consistent way when more facts and phenomena are presented and higher levels of abstraction are addressed?

3. What are the underlying ideas and principles which are helpful in finding these discrepancies and inconsistencies and which could be used in a constructive way for revision and further development?

The answers to the first and second question required the most time and effort. They are described in Chapter II and III of this report. The answers to the third question evolved at the very end when a series of examples for discrepancies and inconsistencies were visible in comparison with the new model and the common structure could be found. Once formulated, these basic characteristics can now be presented at the beginning to clarify the underlying model and to indicate the level of abstraction. This is done in the following part of this chapter.

1.2. QUALITATIVE CONCEPTS IN ELECTRICITY

Almost any physics curriculum or major textbook starts with mechanics which is regarded as a more fundamental subject, not only historically but also by virtue of its basic principles such as force, acceleration, energy, momentum, e.t.c.

lt can, therefore, be assumed that most of the students who start learning about electricity do have a more-or-less well-structured knowledge about Newtonian mechanics which may sometimes be in conflict with, or overruled by, common sense interpretations of mechanical problems due to daily life experiences.

lt is a well-established fact that this knowledge about mechanics is normally limited to certain standard problems and cannot be easily applied to unknown tasks by most of the students. Nevertheless, it can be assumed that most students have developed some kind of Newtonian view when they analyze interactive particles and that they come to the following conclusions:

- Materialistic objects with well-defined surfaces interact when they come too close to each other (with the exception of gravitational interaction). Within the region of contact, action and reaction forces are created due to Newton's Third Principle and these forces are the cause of elastic deformation, deceleration and acceleration. In addition to Newton's Third Principle, it is postulated that there is no action at a distance.

The important characteristics for this kind of analysis are:

- independent objects and systems of objects
- interaction due to action and reaction forces at the same point in space
- causality relations between force and acceleration
- changes develop continuously in space and time

Moving to the electrical world where Coulomb's Law describes the basic phenomenon of interacting charges, a new kind of interaction has to be analysed where action and reaction forces are created at a distance. This task is accomplished by introducing the concept of charges and fields a concept which is strongly supported by quantitative methods.

The important point is that qualitative reasoning about this new kind of Coulomb interaction, using the concept of charge and field, cannot develop harmoniously out of the mechanical approach. There are not only some fundamental differences but also some fundamental discrepancies between these two qualitative approaches. These discrepancies (which are described in the following chapter) are connected to some basic procedures which are believed to be of general importance for qualitative reasoning:

- - the definitions or constructions of basic objects
  - the way in which causality is used to structure the interaction between these objects

In mechanics, the application of Newton's Third Principle provides a basis for such a causal relation by introducing action and reaction forces for any locality of subsystems which are involved in the interaction. It is mainly this aspect of locality (the fact that action and reaction forces are created at the same place) which is missing in the electrical case. The electric or magnetic forces are described to exist at different places in space, and no mechanism is offered to construct a connection between them. The question is how this fact limits the possibility for students to reason about causality and, more seriously, to learn with understanding.

In the following chapter, some implications for qualitative reasoning are outlined which connect to object definition and causality.

1.3. OBJECT DEFINITION

When a mechanical system with two colliding objects or point masses is to be analysed, the definition of the isolated objects and the structuring of the interaction phase is straightforward. There are many different ways and methods to establish these definitions independently and to agree about the underlying principles such as isolated objects in space, the process of interaction including elastic deformation, and forces due to "action equals reaction", deceleration, acceleration, and, finally again, separation into isolated objects.

The analogous case of two colliding charged objects presents a new and difficult problem. The fundamental objects involved in this process are charges and fields, and these objects cannot be defined independently but only in mutual dependence. Charges and fields can only be detected and defined in the presence of other charges and fields and only through the result of interaction From the inception, charges and fields are at the same time representations of interaction and of isolated objects. This is a basic difference in comparison to the mechanical world, and it imposes a basic difficulty for qualitative reasoning. In the electrical world, the basic primitives are more the result of construction by the human mind than determination by nature. This implies a difficulty for all following steps when new facts or knowledge have to be acquired. If charges and fields are seen as individual, isolated objects (and this reflects the common practice), then the mechanism of interaction is hidden behind a formal product of charge and field ($F = q \times E$). Because there is no way to explain how, and within which volume, this interaction takes place, qualitative reasoning may be misled or totally blocked.

When moving charges and, therefore, magnetic effects are involved, the definition of the magnetic field as on object arises again in an even more complicated form. The magnetic field as an object is created by the movement of charges and can exist as a wave in space a long time after the cause of its creation (the movement of electrons) has already disappeared. This constitutes a very peculiar way to create an object which has no analogy in mechanical problems. In mechanics, a constant movement is just a matter of the frame of reference; and the laws in physics should be invariant with respect to them. The magnetic field, however, is said to be caused by such a movement. On the other hand, the magnetic field can disappear and be transformed to an electric field when changing the frame of reference. Electric and magnetic fields become components of a 4-dimensional tensor with hardly any connection to a qualitative concept.

When looking for quantitative solutions, these arguments are of no real concern because the system of differential equations always involves integration of the whole system and does not depend on the interpretation of objects and interactions.

The work presented here is, however, concerned with teaching physics and the importance of qualitative reasoning. In the light of the new computer tools for animation and interactive graphical representation, the following questions arise:

- What kind of basic primitives should be selected as a starting point for qualitative reasoning within the curriculum of electricity?
- What are the basic objects and mechanisms and how could they be represented, reasoned over and displayed so that further learning is supported and not hindered or blocked by misleading derivations?
- What kind of basic principles can be offered as a ground for causal argumentation which connects qualitative reasoning in the mechanical world with the new electric phenomena and which still leads to correct physics?

Preliminary answers to these questions are proposed in Chapters II and III of this report. In this approach, the basic objects for qualitative reasoning in electricity are defined in a broader and more general sense than charges and fields. The notion of distortion in space is proposed to characterize charge and field occupying a certain volume and expressing a certain symmetry. Magnetic effects are derived from a change in symmetry due to the velocity of the charge carriers together with some basic assumptions derived from the theory of special relativity.

These proposals and the derived representations have to be regarded as a first step. They are constructed as a tool to improve learning with the focus on qualitative reasoning. They are not meant to represent correct physics from the very beginning. The problem does not lie in the development of learning tools which avoid any risk of misinterpretation in the light of the established theories of physics. The problem is to minimize and to control these risks while giving support to a step-wise development from the mechanical world towards the electrical one and offering a concept without basic discrepancies or confusing open ends. Studies about the reaction of students who work with these materials will help to determine the direction of further development.

1.4. QUALITATIVE REASONING AND CAUSALITY

1.4.1. NEWTON´S THIRD PRINCIPLE

In the mechanical example of two colliding objects or point masses, the implication of Newton's Third Principle as a basis for causal explanation represents a standard approach to analyse such a collision and to derive such a fundamental law as conservation of momentum. To make this analysis and to apply Newton's Third Principle is not a trivial task; causality is not built into nature but is dependent on the applied theory and the underlying principles. In mechanics courses, this principle, however, is introduced as a fundamental one; and no limits are presented - especially when systems of interacting particles are to be studied.

When Newton's Third Principle is applied in a straightforward way to the interaction between charges and fields and especially to the interaction between magnetic fields and moving charges, qualitative reasoning is faced with a new and basic discrepancy. The locality of Newton's Third Principle gets lost. Equal and opposite action and reaction forces can no longer be defined at each single point in space but only for the complete system, and no mechanism is presented to explain the interrelation between these single forces at different places.

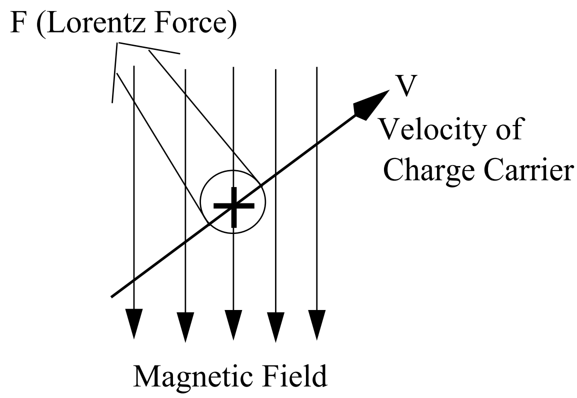


Fig. 1.1.Lorentz Force

In the case of the magnetic interaction, a single force acts on a moving charge carrier within a magnetic field which results in an acceleration of the charge carrier perpendicular to its velocity. There is no local reaction back onto the magnetic field.

Such a single force or a single acceleration is a contradiction to Newton's Third Principle. For the magnetic interaction, this principle is only fulfilled when the reaction of the moving particle with the complete system is taken into account. The principle is, however, not valid on a local scale.

Also, in the case of the Coulomb interaction between two particles, Newton's Third Principle cannot be applied in a completely analogous way to the mechanical case of colliding objects. The Coulomb forces between two charged carriers are equal and opposite, but they do not act at the same point in space. Each charge carrier experiences a single force due to its interaction with the field; but locally, there is no reaction back onto the field. This reaction force is found when the force on the opposite charge carrier is taken into account. This reaction force, again, is the result of an interaction of field and charge with no reaction back onto the field.

There is no qualitative explanation for this new kind of mechanism between two action and reaction forces acting at different points in space. This missing information severely limits the potential of qualitative reasoning learned so far from mechanical colliding processes.

1.4.2. CONTINUOS CHANGE IN SPACE AND TIME OR ACTION AT A DISTANCE

The majority of the problems treated in electrostatics, and in dc and ac currents, are static or quasi-static cases. The distribution of the charge carriers is either static or stationary, and Coulomb's Law and Kirchhoff's Laws only describe these states of equilibrium. These laws do not tell how the system is changing from one state to the other or how the new state of equilibrium is reached. They are even invalid during the time interval of change. Qualitative reasoning based on these equations is, therefore, limited to static cases and cannot cope with changes in time. Knowledge about change has to be handled in the form of knowledge about a series of equilibrium states, and this may seriously limit the development of physical intuition and the capability to accomplish tasks such as trouble-shooting.

A simple example for such a limitation is the change of the electric current within an electric circuit due to a change of voltage or resistance. In the absence of action at a distance, the change of the current has to start at a certain point within the circuit, and this change has to spread out around the whole system.

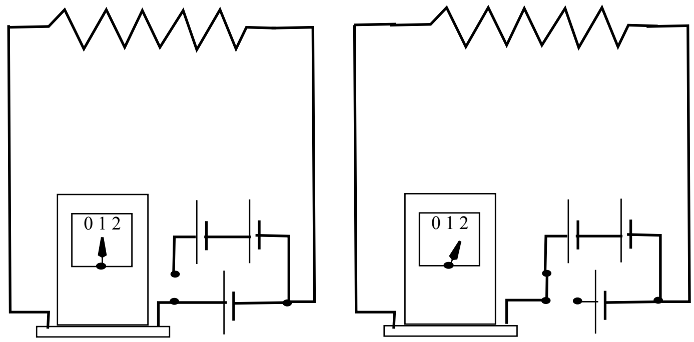


Fig. 1.2. Transition Between States of Equilibrium when Doubling the Voltage

The time for this process is rather short and can be neglected for a quantitative analysis. For qualitative reasoning, however, it seems to be important to know how this change of the current is started, how it is transmitted and what mechanism determines the new state of equilibrium. Certainly this change cannot occur simultaneously for all points of the system because this would include action at a distance. Ohm's law and Kirchhoff's Laws are even counterproductive for qualitative reasoning about this process. The question is whether a student, who can treat this process explicitly and who can qualitatively reason about the connection between different states of equilibrium, reaches a better understanding and whether this activity supports the integration of knowledge about different parts of the system.

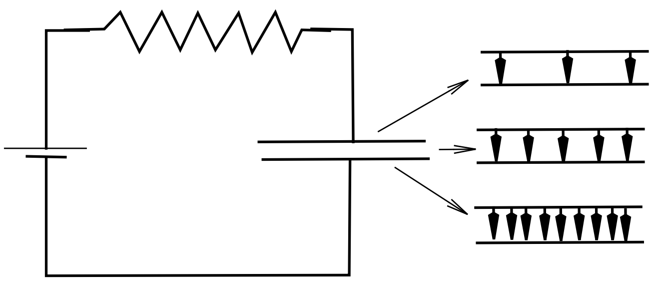


Fig. 1.3.Charging a Capacitor

Another example for the importance of change is the charging of a capacitor. The equations can describe the starting and the final state and can also describe the time dependence of the charge flow into the capacitor. The change of the electric field within the capacitor can, however, only be represented by a series of static pictures with different densities of parallel lines indicating the change of the field strength.

The problem is that the actual, continuous change of the field strength cannot be shown. The field strength cannot at the same time be changing and be identical all across the capacitor because this would imply action at a distance. The increase of the field strength can also not start from one or from both sides and travel out in space towards the other side. This would imply - at least for a short moment in time - an open field line in space which can be excluded due to Maxwell's equations.

This is a fundamental problem which has led Maxwell to the invention of the displacement current and the formulation of his equations, solving the problem in a quantitative way. For qualitative reasoning, this problem can, however, arise much earlier within the learning process. The question arises, how much is gained for the support of physical intuition and problem solving when this information about change is provided and integrated into the learning process. Also, this knowledge could lead to different types of questions about the underlying principles of the system and to a selection of different types of problems to focus on.

1.4.3. RELATION BETWEEN THE MICROSCOPIC AND MACROSCOPIC LEVEL

The kinetic gas theory is a classical example to demonstrate the power of causal explanations on a microscopic level for a large number of macroscopic effects. Statements on a macroscopic level, like

- "In thermal equilibrium, each part of the system has the same temperature."

or

- "Heat always flows from higher to lower temperature "

do more to describe than to explain. The change from the macroscopic to the microscopic level and the use of the kinetic gas theory to describe heat flow is richer in explanation than description. This should be true for any macroscopic quantity which evolves from an underlying microscopic model.

Electricity is a subject with a distinct microscopic level based on electrons, their movement, and their interaction with matter. For this area of physics, it should therefore be possible to make a close connection between macroscopic and microscopic effects.

In the area of electrostatics and electrodynamics, the concept of current and voltage are treated in a different and inconsistent way according to the macroscopic and microscopic level. From the beginning, the current is explained on a microscopic level. Kirchhoffs first law is, for instance, explained by the "Principle of Conservation of Charge" and the movement of electrons within a closed circuit. For electrostatics, the voltage also is connected to a microscopic level. The work done by separate charges explains the electric energy which is then used to define the voltage. In electrodynamics, however, the voltage is only connected to potential difference which is directly related to energy considerations. The main problem is not that the microscopic explanation for voltage is missing. The real problem is that a student searching for such an explanation would be confronted with a contradiction. The contradiction arises from the following facts:

- In electrostatics, the concept of voltage is connected to separated charges.
- For the dynamic case of a constant current flow, the microscopic model predicts an identical situation between two cross-sections before and after a resistor. The same number of electrons is drifting through these cross-sections, and also the field is the same.
- Between these two cross-sections, there is a voltage drop but no explanation on a microscopic level.

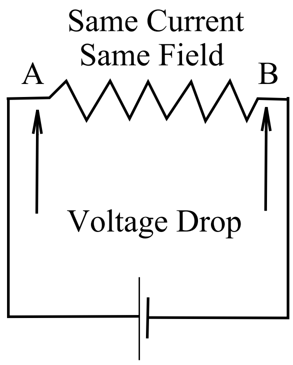


Fig. 1.4.What corresponds to the voltage drop,

The conclusion that there is no microscopic difference between the cross-sections A and B according to the model for the current would lead to the contradictory result that a voltage drop can exist between two points which are identical on a microscopic scale. This is contradictory because one could change the scale continuously from the macroscopic to the microscopic scale and ask when the voltage would disappear.

There would be no answer. A solution out of this problem would be to assume a difference between the voltage in electrostatics and electrodynamics but this assumption is also not supported by Maxwell's theory.

For qualitative reasoning, such considerations could act as a block and could prevent further progress. The question is whether students detect this discrepancy on their own and how they react in case they do. Another question is, what amount of qualitative reasoning and physical intuition about voltage and current can be improved by explicitly teaching a microscopic model for voltage. A proposal for such an approach is presented in Chapter II based on the concept of surface charges which are present when a current is flowing through a conductor. This concept of surface charges relates to a rather small effect which is normally neglected in any quantitative approach. For qualitative reasoning, however, the size of an effect is relatively unimportant; what is important is the fact that it exists at all and that it plays a causal role in the ensuing behaviour.

Chapter II
BASIC ELECTRICITY

2.1. INTRODUCTION

In the following chapter, a course in Basic Electricity is designed covering the concept of current, voltage and resistance, the subject area described by Ohm's Law and the two Kirchhoff's Laws. It is a primary concern to develop a set of consistent qualitative concepts which can serve as a frame or base for quantitative analysis.

Knowledge about quantitative physical expressions and practice in solving the adequate problems may lead to a tacit and implicit understanding of the underlying ontology, but often it will not. It is our belief that such knowledge can be made explicit, can lead to deeper understanding, can support physical intuition, and can enhance the capability of problem-solving.

For the development of such qualitative concepts, the underlying primitives and the mechanism of the interaction have to be treated in detail and have to be related to macroscopic effects. This involves, for instance, the question: What microscopic differences at two measurement points are responsible for a voltage drop between these two points? Further, it is necessary to treat changes between states of equilibrium explicitly. This involves questions such as:

- How is information about a voltage change transmitted to all parts of the system?
- How is the information of a change of a resistor transmitted to a branching point, and how is the current redistributed according to the new state of equilibrium?
- How are all parts of a system informed about the fact that a switch has been opened or closed?
- How fast does a current increase if anywhere within a circuit a short occurs?

Finally, it seems to be necessary to explicitly mark the unique features of the electric current in comparison to a mechanical flow of matter so that the risk of incorrect analogies is minimized. This includes, for instance, the question of how the electric energy is transported in comparison to mechanical systems.

Within the frame of a qualitative concept, these and similar questions (which could also include semiconductors) should be answered in a consistent way. It is hoped that this approach, when integrated with quantitative methods, will help the student to develop physical intuition and to improve his or her other abilities to solve problems within new areas.

It is assumed that at the beginning of this course the students have built some simple circuits on their own and that they know the following facts:

- A closed loop built of conducting material is necessary for a continuous electric current to flow.

- Each element of an electric circuit (like a bulb, a resistor, or a motor) has two terminals which have to be connected to the two terminals of the voltage source.
- Each circuit needs at least one voltage or energy source to function.
- A switch is used to open or close the loop of conducting material.

2.2. THE ELECTRIC CIRCUIT AS A TRANSPORT SYSTEM

2.2.1. MECHANICAL ANALOGIES AS A STARTING POINT

In general, the electric circuit can be interpreted as a system to either transport energy or information. Both functions are of fundamental importance for technology and have a strong influence on our daily life. It is hard to imagine what society would be like without them.

Teaching electricity in connection with these basic technologies has to deal with the fact that both, energy and information, are rather abstract notions. Especially with energy, most people have developed some ideas about this concept and the transportation mechanisms which are strongly connected to mechanical objects like charcoal or oil and mechanical transportation systems like a pipeline, a water heating system, etc.

Understanding basic principles of the electric circuit requires that this knowledge about mechanical systems or models be changed or reconstructed and that some new aspects of electrical phenomena with no analogy to mechanical systems have to be learned and understood.

In order to support qualitative reasoning, it seems to be crucial to prepare this reconstruction of mechanical knowledge and the transition from the mechanical towards the electrical world with great care and to offer a variety of material to support individual exploration and exercise. Otherwise, there is the constant threat that incorrect mechanical analogies structure implicitly the knowledge about electricity which then can lead to a series of incorrect expectations, misunderstandings, and failures.

An example of such a misleading mechanical analogy can be found in modern textbooks (see fig.2.2. below) where the electric current is compared with cars on a highway or with some strange creatures carrying energy around. In these models, the single particles which symbolize the current flow, possess their own transportation mechanisms. Within a certain range, they can act in an isolated way. At a branching point, these particles have to "decide" where to go; and the speed of energy flow is bound to the speed of these moving objects.

Within an electric circuit, the interrelation between the moving electrons, constituting the current, and the atoms of the conducting wire is of much larger influence than single isolated electrons. The electrons can only move in a closed circuit as an ensemble. There is no place for individual movement. This last statement is not true in an absolute sense. Even an electron gas within metal or within a resistor is not absolutely rigid but can rearrange itself due to changing external forces like a changing EMF or a changing magnetic field. This rearrangement of the electron gas is, however, much smaller than can be predicted from any mechanical analogy.

Because of this strong interrelation between the electrons of the electric current, the flow of energy which is transported by the electric current is completely disconnected from the movement of the charge carriers. The energy can be interpreted to flow very fast (nearly with the speed of light) while the electrons move only very slowly.

A mechanical model, which could serve as a good starting point for the development of such knowledge, is a transmission belt or a "stiff" (only slightly compressible) ring while any system with individual or quasi-isolated elements is strongly misleading. As a first step in teaching electricity, different mechanical systems should be compared which have circular symmetry and which can transport energy. The objective of such an introduction would be to filter out the model which has the largest number of attributes in common with the electric circuit. Some of these attributes are:

- A single source of energy is responsible for the whole system to move or to function.
- The functioning of the system depends on the circular movement of all elements as an ensemble and not on the movement of single, isolated particles.
- The energy is not transported in close connection with the moving particles but in an independent way, connected to the transmission of force and movement by the complete system.

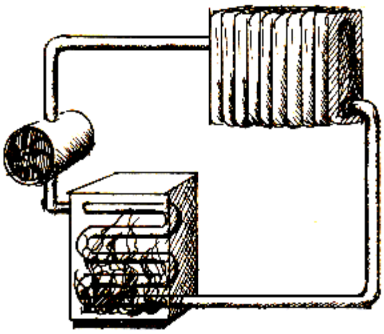
The use of well-chosen mechanical analogies can support the task to rearrange mechanical knowledge and to understand the importance of certain features of the system. It can, however, also mislead further learning and block full insight if it is taken too literally and if the basic differences between the mechanical world and the electrical one are not explicitly taught. Some of these basic differences are:

- The range of interaction between mechanical and electrical objects is quite different. Electric interaction reaches out in space. Mechanical interaction, besides gravity, is limited to immediate contact.
- The nature of the interaction is different. Mechanical interaction is connected to the fact that the same space cannot be occupied by two material objects at the same time. Different electric fields, however, superimpose at the same place.
- The order of magnitude of the forces between electrical and mechanical objects (the ratio of force over mass in the mechanical and the electrical world) is completely different (some 40 orders of magnitude).

2.2.2. SIMILARITIES AND DIFFERENCES BETWEEN MECHANICAL SYSTEMS AND THE ELECTRIC CIRCUIT

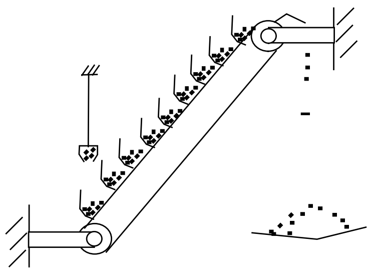
There are quite a variety of mechanical systems which, in common with the electric circuit, show some kind of circular flow or movement of matter and some kind of energy transportation.

Some examples for such mechanical systems are:

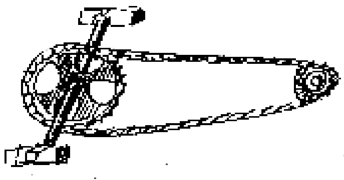


- a warm water heating system

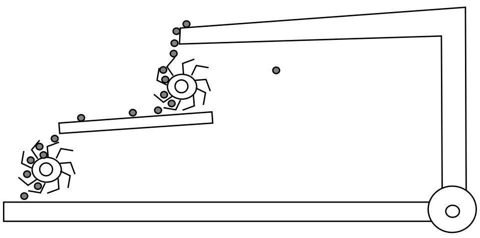
-a transmission belt



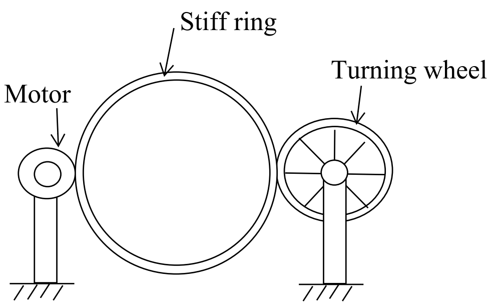
a transportation belt



- a bicycle chain



an open water system

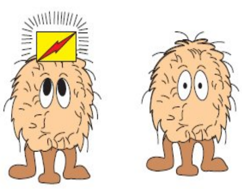


a stiff ring

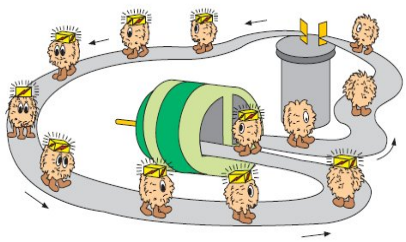
Fig. 2.1 Mechanical circular systems for transport of matter and energy

2.2.3. ISOLATED ELEMENTS AND SYSTEM COHERENCE

As mentioned earlier, a chain of trucks is found in some textbooks as an analogy for the electric circuit. The same idea has led other authors to invent an artificial system with special features to carry the energy around.



Electric particles
with without
energ



Electric particles,
carrying energy from
the battery to the motor

Fig. 2.2. Misleading Models for the Electric Current (from a Textbook and a Curriculum)

In these two systems there is a striking and fundamental difference with respect to the electric circuit which is bound to the fact that the moving particles have their own transportation mechanism. They can stop individually, their interrelation is very weak, and the energy is moving exactly at the same speed as the moving particles. All of this is opposite when looking at the electric circuit. Although it can be assumed that there are single electrons moving around in the circuit, they cannot move individually but only as an ensemble. The demand for neutrality within the conductor imposes the same density of free electrons all around the entire circuit. In addition to that, the movement of the energy is completely de-coupled from the movement of the electrons.

There are a series of consequences in connection with these different system attributes which render these mechanical systems useless or even counter-productive when taken as analogies for the electric circuit. In this course, they will be used as a negative example in order to point out the importance of the system coherence in comparison with isolated particles and to demonstrate the negative consequences of poor analogies.

2.2.4. TRANSPORTATION OF MATTER INSTEAD OF ENERGY

It is common practice for most people not to differentiate very precisely between matter and energy when, for instance, charcoal or oil is transported by any kind of transportation mechanism. It is clear that in these cases it would be more precise not to talk about energy transportation at all. The amount of energy which is connected with oil, for instance, is strongly dependent on the amount of oxygen available at the burning place; and this oxygen has not been transported at all. In this course, these mechanical systems (like a transportation belt) will not be used as the starting point to learn about basic electrical facts but only to point out basic differences between mechanical and electric transportation systems.

2.2.5. TRANSPORTATION OF MATTER TOGETHER WITH ENERGY

In the case of a warm-water heating system or a water pipe system within a hydraulic power plant, there is energy transported together with the water (or the steam) either as thermal energy or kinetic energy of the moving medium. These models are found in some textbooks, and they seem to be quite easy for students to understand. A discrepancy arises when information is presented about the drifting speed of free electrons and the speed of the energy flow. In this respect, the two systems differ completely; and any conclusion by analogy fails.

2.2.6. AN OPEN WATER CIRCUIT UNDER THE INFLUENCE OF GRAVITY

A final negative example is a water circuit where the main driving force for the water flow is due to gravity while at the lowest place a pump lifts the water up again to close the circuit. Such a model is used in many textbooks when the notion of voltage is introduced. It is quite tempting to use this model because the similarity between the gravitational field and the electric field helps to "explain" some basic features about the voltage in series and parallel circuits.

There are, however, also some basic differences - especially when the water system is open (like a river). In this case, the coherence of the system is very weak. Changing a resistor will lead to overflow and not to a change of the current within the whole system.

2.2.7. TRANSMISSION OF ENERGY BY FORCE AND MOVEMENT

Mechanical systems like a bicycle chain, a transmission belt, a "stiff" ring, or a slow-moving hydraulic system have in common that the energy is transmitted through force and motion or, in more physical terms, in the form of physical work. This is a basic structure which also holds for the electric circuit and which is one of the basic factors to explain the coherence of these systems. The mechanical example of the "stiff" ring has the advantage that both, a pushing and a pulling force can be applied in analogy to the attractive and repulsive forces between positive and negative charge carriers. One of the main drawbacks of these solid mechanical systems lies in the fact that, first, the "stiff" ring is an unusual device and, second, parallel circuits are hardly possible to simulate.

With hydraulic systems, parallel branching comes in a natural way; but these systems allow only pressure and no pulling forces. You cannot pull on water but only take the air out and rely on the fact that somewhere air pressure will push onto the water. Moreover, a water circuit for energy transport normally allows the water to flow at high speed. Energy is then transported as kinetic energy of the water which leads to inconsistencies when compared with the electric current. A more correct analogy is a hydraulic system with slow-moving water under high pressure. Such a system, however, has never been built to transmit energy because it would be rather inefficient and the fitting and leaking problems would be hard to overcome. The only practical solution for a slow-moving water circuit is found in a hydraulic system where the fluid is pressed through some kind of a resistor or valve. Such a system, however, cannot transmit energy on any reasonable scale.

There are more basic differences between a hydraulic system and the electric circuit. The mechanical flow is caused by contact forces acting through the whole volume. In the electric circuit, the electrons are interacting via field forces with distant charges. In the mechanical systems, energy can only be transformed by force and motion. In the electric circuit, there are many different kinds of energy transformation - especially the magnetic coupling to the outside world.

2.2.8. CONCLUSION

The electrical system has some attributes of the "stiff" ring as well as some attributes of a slow-moving hydraulic system. It allows pushing and pulling forces and parallel branching; and, in addition, it can transmit energy on a very wide scale. Furthermore, it has a series of new and unique features which have no analogy within the mechanical world. In this course, the development of knowledge about electricity, therefore, starts with proper mechanical models like a "stiff" ring and/or a slow-moving hydraulic system to point out some important features in comparison with other mechanical systems, thus serving as a guide to structure the understanding of electricity phenomena. The main task will be to isolate these features which have value within the electrical world and to integrate them with new knowledge.

2.3. REPRESENTATION OF FORCES

In traditional mechanics courses, forces are described as vectors and are represented as arrows. This representation is very useful for graphical solutions, and it supports vector calculus. There are, however, some drawbacks in respect to learning and understanding. The representation of a force as an arrow and the fact that an arrow is in itself a complete figure or symbol can lead to the assumption that a single force can exist and be represented by a static symbol which is, of course, incorrect. There is either an acceleration, or the action and reaction forces add up to zero. Nevertheless, there are many cases where the reaction forces are neglected, suppressed, or merely overlooked; and this may be tempting due to the fact that an arrow is a complete figure and that this kind of representation does not ask on its own for the representation of its counterpart.

Another drawback is connected with the fact that an arrow always has only a single point of attack and that this arrow representation does not support the qualitative concept of surface or volume forces or tensions. However, only volume forces can be qualitatively connected to real phenomena, and forces acting at a point are an abstraction just as point masses are.

This question has a specific importance for the electric interaction where the actual place of interaction is not immediately clear. The use of a one-sided symbol could, therefore, work against a crucial understanding of the underlying mechanism. It is an open question whether, and to what extent, students are confused or misled by this abstract representation of an abstract concept.

For the following course, it is proposed that, as a first step, the ellipse be used as representation of a force and that the vector representation be introduced in parallel or at a later time. The ellipse represents a volume of interaction rather than interaction at a point and along a single line. Moreover, half of an ellipse is immediately seen as incomplete; and the representation itself leads to the search for the reaction forces.The following pictures demonstrate how ellipses could be used parallel to arrows to represent interaction between colliding objects.

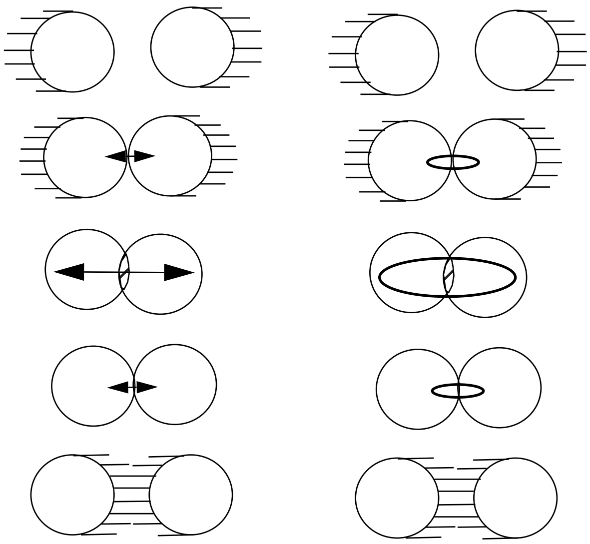


Fig. 2.3.Different Representations of Forces

The representation of forces as ellipses has, of course, also its drawbacks. An ellipse does not reinforce the centre of interaction as an arrow does, and vector addition is not supported. These two representations have dual properties and when used alternatively, could possibly help the student to differentiate between representations and concepts or, more generally, between the name and the object. At the end of each course, when quantitative considerations are dominant and when forces have to be superimposed and decomposed, only the vector representation is suitable.

2.4. THE STRUCTURE OF COULOMB INTERACTION

2.4.1. COULOMB FORCES BETWEEN CHARGED PARTICLES

In order to explain the cause for current flow, it is necessary to start with the Coulomb interaction between charged objects and, as a thought experiment, between two isolated electrons.

A system with separated charges can be created by rather different mechanisms. For instance:

- chemical interaction
- mechanical surface interaction
- electromagnetic interaction
- photo-chemical interaction

As a result of any such separation of charges, repelling or attracting forces occur depending on the sign of the charges. In this course, the Coulomb interaction is used only in its qualitative form:

- more charges lead to a stronger interaction
- larger distances lead to a weaker interaction

In the following picture, this interaction is represented in two different ways. The representation with arrows focuses on the centre of the charges, and the ellipses emphasize the space between the two objects.

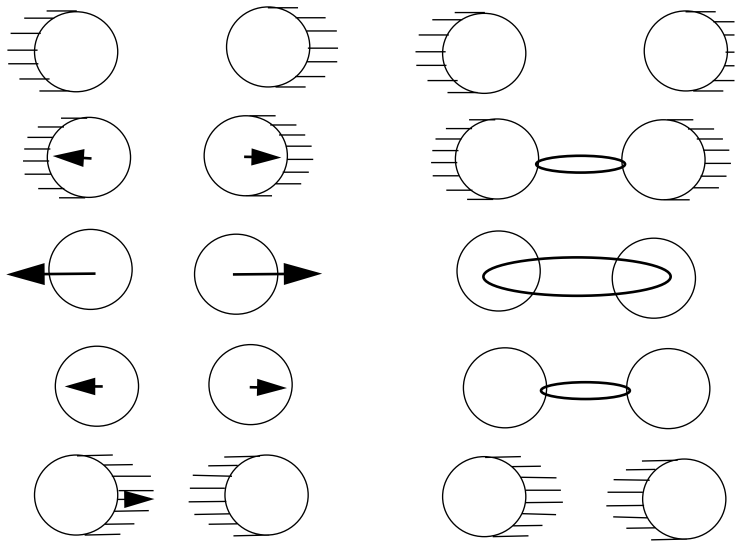


Fig. 2.4. Different Representation of Coulomb Interaction

This immediately raises the important question of where the interaction actually starts and in what qualitative way this interaction could be structured.

In traditional physics, charges and fields are defined as separated objects; and the interaction is constructed as a local event at the centre of the charges, expressed as the product qxE. This construction implies that the locality of Newton's Third Principle is being replaced by two action and reaction forces at different points in space. In this construction, a charge can be accelerated by a field, but there is no reaction directly back onto the field.

As argued earlier, it is proposed here to introduce different kinds of objects and interactions as learning tools which do not originally rely on definitions for measurements but which support some need for qualitative thinking. Some of these needs are:

In the following, it is briefly described how such an approach which fulfils these needs could be developed by introducing the concept of "distortion of space" or "change in space" and "overlap of distorted or changed space". These notions, however, have to be taken as a first proposal which yet have to undergo a process of discussion and criticism. The objectives of such a process should lead to correct physics and, at the same time, to the development of more powerful tools for qualitative reasoning which could then lead to learning with understanding and the development of physical intuition.

In order to develop a new qualitative concept about electricity, it will be necessary to revise the concept of empty space which seems so natural in organizing our daily life experiences. The idea that a transverse wave can travel in empty space was, however, one of the most puzzling ones for physicists from the time, when Hertz made his first detection. The existence of an ether was at that time an unquestioned axiom, and the history of physics is full of efforts to defend this concept.

It is assumed here that a student has to undergo to a certain degree a similar process of reconstruction and development of a new concept of space before the concept and the interpretation of electromagnetic fields and waves can be understood.

The following questions could guide such necessary considerations about space:

- What is the result of two equal but opposite electric fields which could be constructed around any material body? Is the result an empty space or a space with two equal but opposite distortions?

- Could such an "undistorted space" have properties like some kind of elasticity so that disturbances can travel from one place to another?
- Could an electron or proton be some fundamental unit of distortion in space with opposite symmetries (like "left and right" or "in and out") which has the potential to couple with other objects/distortions?

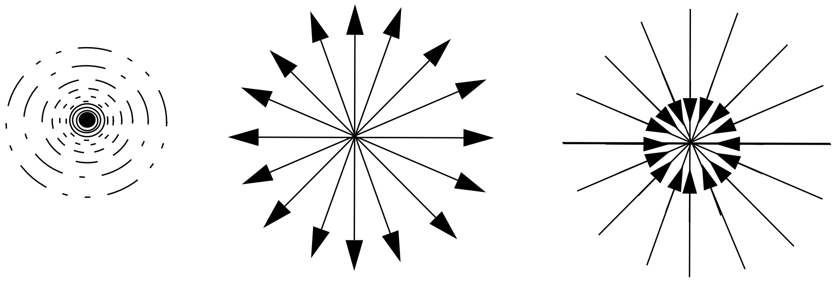


Fig. 2.5. Alternative visualisations of charges and fields

- Could the undistorted space have a certain level of noise which hides or limits the influence of an isolated electron beyond a certain distance?
- Could the overlap of such distortions lead to repulsion or attraction according to equal or opposite symmetries of the distortions?

The following text does not give answers to all these questions. As a first step, however, the representation of Coulomb interaction does not follow the usual path. The traditional method to represent a force as vector applied at the centre of charge carriers is expanded. Ellipses are used instead to emphasize the importance of the space between the charges. Especially in Chapter III, the overlap of charges and their surrounding space is emphasized and visualized in the following way:

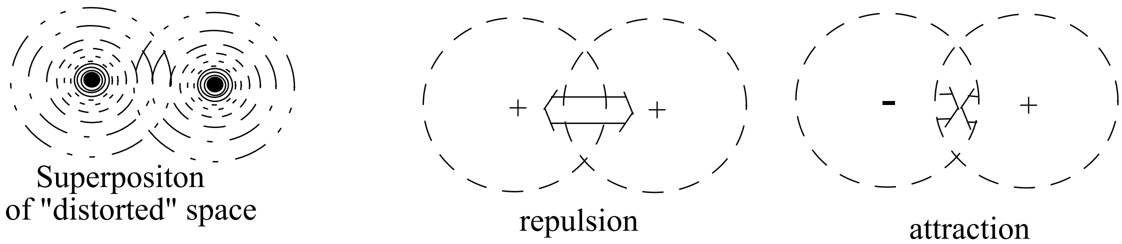


Fig. 2.6. Visualisation of Coulomb interaction

Such a visualization has, however, no consequences for the derived conclusions of physics but serves primarily intuition and to support better understanding. It can always be replaced by the usual vector representation of Coulomb forces.

2.4.2. MAGNITUDE OF COULOMB FORCES

The interaction between charged particles is extremely strong compared with normal mechanical processes or gravity. In the book about his lectures, Feynman describes the magnitude of the Coulomb forces with the example of two persons standing 1 meter apart from each other and who increase the number of electrons in their bodies by 1%. These extra electrons would, of course, immediately fly apart; they could never be accumulated in the first place. But if they could, the force resulting from this interaction between the two persons would be comparable to the weight of the entire earth.

Another example: If one could separate the electrons and protons of about 1 liter of Hydrogen gas, keep these separated charges close together, and put one part on the other side of the earth, the attracting forces between both parts would be something like 1 t (10000 N). It takes, therefore, only a few electrons to exert large forces; but it is, therefore, also very difficult to separate charges. The interactions which result in such a separation are different from Coulomb interactions; e.g., chemical processes, collisions, or molecular forces.

2.4.3. STATIC DISTRIBUTION OF CHARGES ON MACROSCOPIC OBJECTS

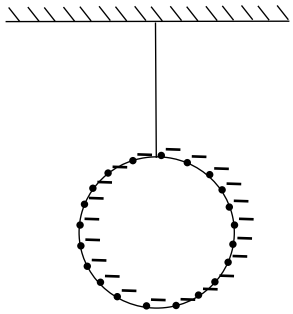


Fig. 2.7.Distribution of Charges on an Isolated Conductor

A conductor is characterized by a lattice of fixed positive charges and mobile electrons. Because of the large Coulomb forces, each single atom has a balanced number of positive and negative charges. An insulator is characterized by the fact that negative charge carriers are more or less fixed to the positive ions and cannot move freely. If a metallic object is isolated from other objects (e.g. the earth) and is charged (which means that extra electrons are added), then these electrons have stable positions only at the surface.

This is an experimental fact and is also grounded on theory. It is, however, not immediately understandable. One could ask why the electrons do not leave the surface to join their original partners outside or why they do not distribute throughout the whole volume to increase the distance from their nearest neighbours from which they are repelled. To answer the first question, it has to be accepted that the surface of a piece of metal acts like a rather high barrier for the mobile electrons. Only at very high temperatures can this barrier be penetrated by the electrons. This capability is used in electronic tubes to get the electrons into the vacuum in the outside. At normal temperatures, however, the barrier at the surface is too high for the electrons to overcome.

A uniform distribution all over the volume would, indeed, increase the distance to the nearest neighbours and decrease the resulting repelling forces. In our mechanical world where the interaction with the next neighbour is normally the most important, this expectation is very reasonable. The interaction between charged particles does, however, not depend on close contact. The interaction reaches far out into space; and, therefore, all interactions with all other charge carriers have to be taken into account. A uniform distribution all over the volume would reduce the distance to the nearest neighbours but would at the same time decrease the distance to nearly all other neighbours and, therefore, increase all the other repelling forces. It turns out that for a charged particle within the volume of a charged body, the sum of all Coulomb interaction always results in a force towards the surface. The only equilibrium state is at the surface. Here the electrons are kept in place by two opposite forces - the repelling forces resulting from all other charge carriers directed towards the outside and the forces due to the surface barrier directed inside. This situation is represented in the following picture for only one electron where either the vector or the ellipse representation for forces is used.

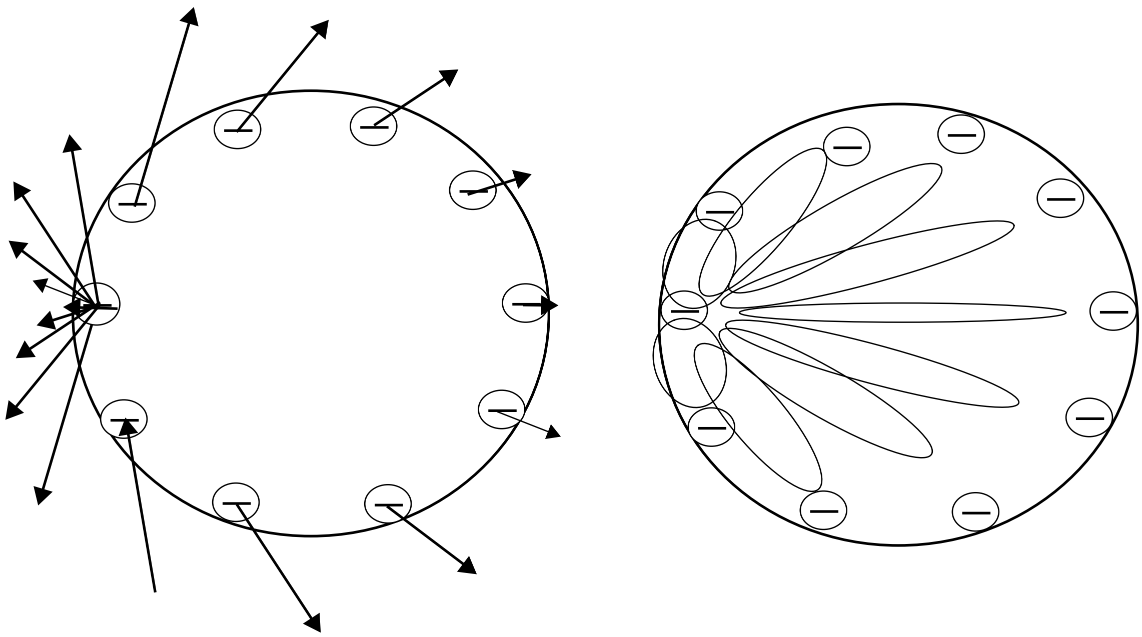


Fig. 2.8. Interaction between a selected electron and the other surface charges (in two different kinds of representation)

2.5. CONSTANT ELECTRIC CURRENT AND THE DISTRIBUTION OF SURFACE CHARGES

2.5.1. DEVELOPMENT OF SURFACE CHARGE

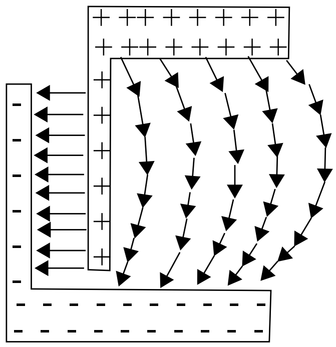


Fig. 2.9.Charged Battery

In the following diagram, a battery is shown as a device to separate charges. This separation is performed by a chemical process resulting in a state of equilibrium with a certain number of electrons on the surface of the negative metallic outlet and the same number of electrons missing on the surface of the other side.

A dynamic equilibrium is reached when the attracting Coulomb forces between these separated charges are equal to the separating chemical forces.

If two conductors and a resistor are connected to the battery, an electric current starts to flow. There are also surface charges spreading out all over the circuit within very short time. The way in which these surface charges are distributed is represented in the next picture (in a rather crude and exaggerated form).

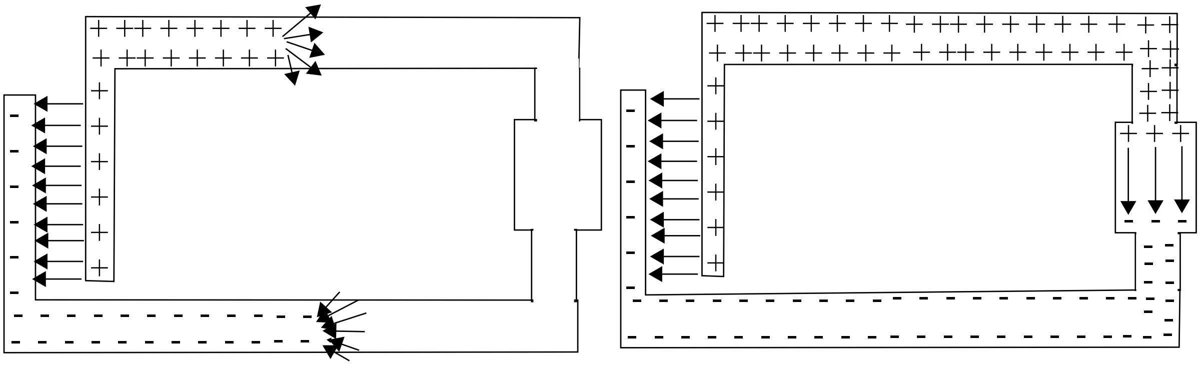


Fig. 2.10.Development of Surface Charges

The existence of these surface charges and their interaction through space can be demonstrated with the following experiment.

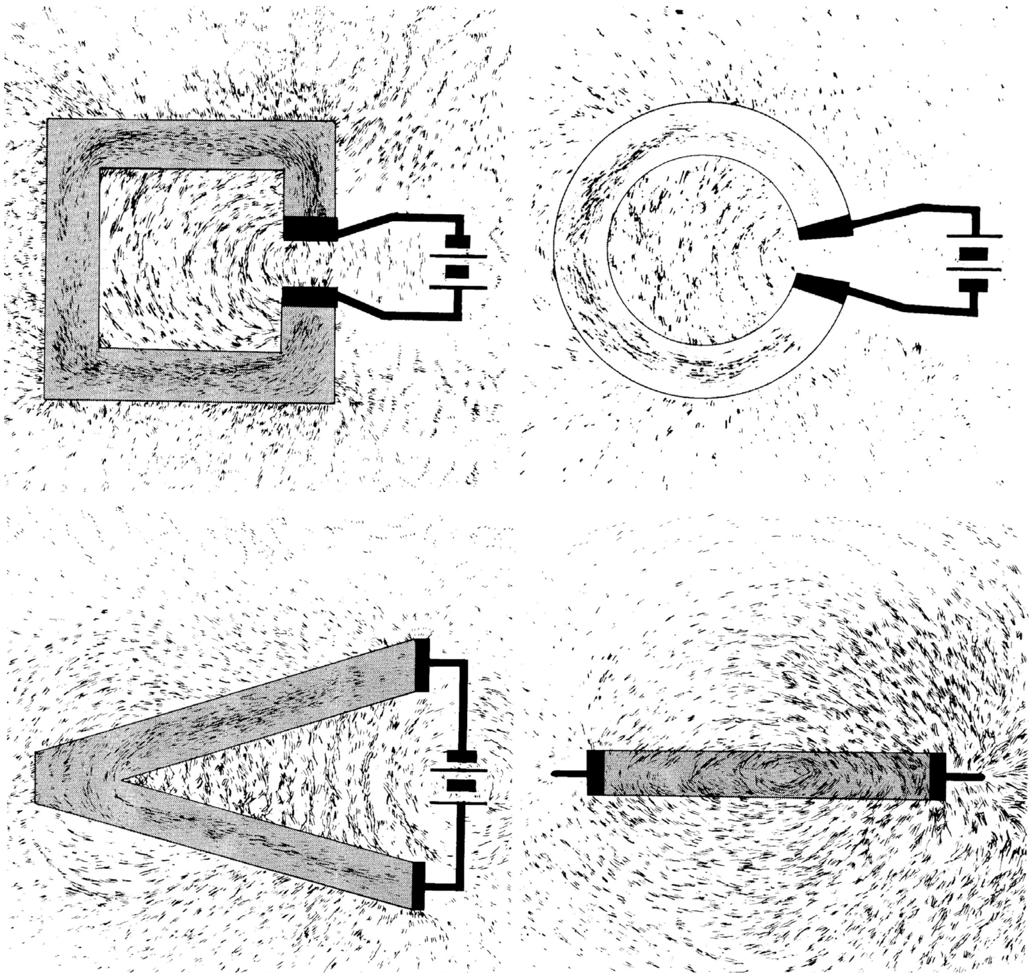


Fig. 2.11. Electric Field Lines Around Conductors

For this demonstration, the circuits are made out of conductive ink and placed on glass plates (Jefimenko,1963). The change in space around the circuit due to the surface charges is demonstrated with the aid of grass seeds strewn upon the glass plate. Grass seeds orient themselves in the neighbourhood of charges because they act like electric dipoles (similar to magnetic compass needles).

2.5.2. GRADIENT IN SURFACE CHARGE DISTRIBUTION

In order for the student to understand the importance of these surface charges in respect to current flow and voltage, the distribution of these charges has to be studied in detail. The following points are of interest:

- order of magnitude of these surface charges
- distribution along a linear conducting wire
- influence of symmetry and distance

Order of Magnitude

If a current of 1 A is flowing through a wire, there are about 10^{19} electrons passing through each cross section within one second. This number is so large that it has no qualitative meaning in comparison with other values known from daily life. Because of the enormous strength of the Coulomb interaction and the very high mobility of electrons in metals, it takes only a few electrons at the surface of the wire to push 10^{19} electrons around in a circle and to overcome the resistance of a metallic wire.

Distribution of Electrons on the Surface

For reasons of symmetry, it can be concluded that the interaction of equally distributed electrons on the surface of a round wire with free electrons inside the wire cancels to zero. For every small area A containing a certain number of electrons, there exists another area A' at the opposite side with the same number of electrons which will balance the interaction with any electron inside the conductor (for example at point P.

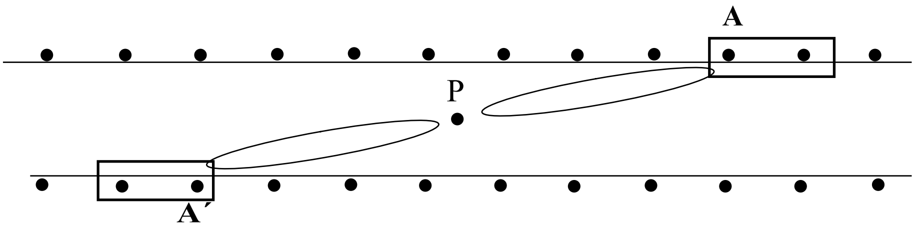


Fig. 2.12. Influence of a Constant Surface Charge for Points inside the Conductor (E.g. Point P

The forces perpendicular to the axis have to be zero because otherwise the electrons would move to the surface and redistribute until there is no longer a force towards the surface.

To break this equilibrium between equally distributed surface charges and to produce a driving force for the electrons within the wire, it is necessary to change this symmetric distribution. On one side, there have to be more surface electrons than on the other side, so that the two interactions of opposite areas A and A' do not balance.

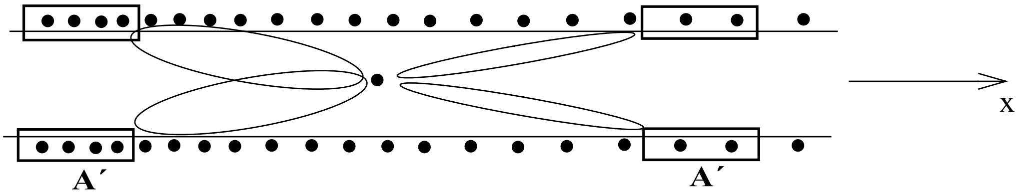
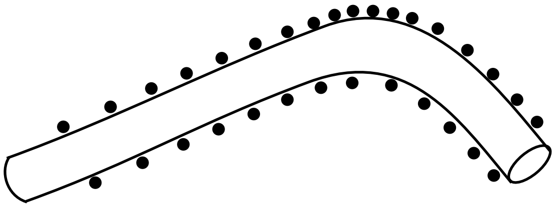


Fig. 2.13. Gradient in Surface Charge Distribution

For a long and thin wire, it can be shown by calculation that a uniform decrease of surface charges in the direction of the axis (a surface charge with a constant gradient in x-direction) results in a uniform and constant force for electrons in the direction parallel to the axis of the wire. (Walz 1984). If the wire is curved or if other wires with surface charges are close together, the gradient is no longer constant and the circular symmetry of the charge distribution around the wire will be broken.



In the case shown on the left, there have to be more electrons at the outer part of the curve than at the inner side to change the direction of the moving electrons inside the wire. Surface charges are, therefore, only a qualitative indication for voltage. They could not serve for a quantitative definition as the expression "energy per charge" does.

Fig. 2.14.Non-linear Distribution of Surface Charges

2.5.3. CONDUCTION MECHANISM AND RESISTORS

The mobility of electrons in metals like copper is many orders of magnitude larger than in a resistor. Within the copper wire, the electrons experience a very small force; and it is possible to think of a mechanism of conduction where the single electrons are accelerated under the influence of the interaction with the surface charges and collide with some atoms within the lattice. The number of collisions per second is high (about 10^{14}/s), and the mean free path between two collisions is about 10 diameters of a copper atom. In this theory about the mechanism of conduction in metals, the electrons can, therefore, be regarded as independent particles which are accelerated for a short moment and than stopped by collision. In this approximation, the interaction between the free electrons is neglected.

The function of a resistor is, in general, to hinder the flow of charge carriers and to exchange energy. There are different ways to accomplish this function. One possibility is to reduce the diameter of the conducting wire without changing the material.

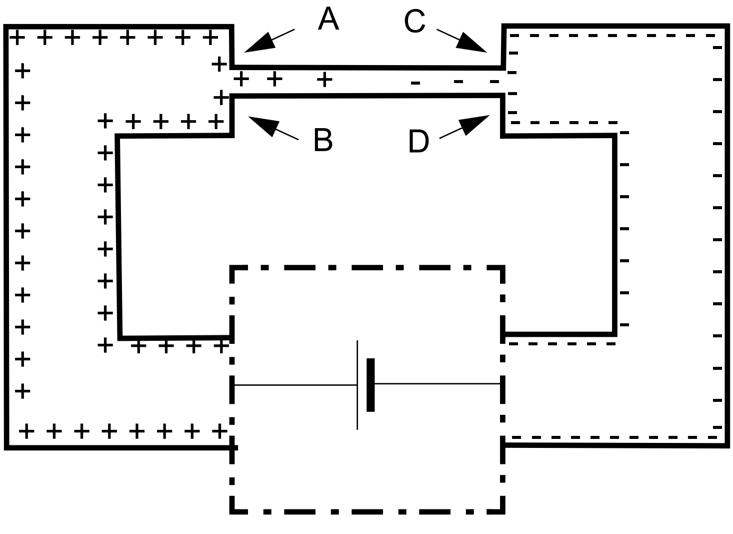


Fig. 2.15. Resistance Through Reduction of Cross-section

The effect of resistance in respect to electron flow can easily be explained through a geometric arrangement. The surface charges on the areas A-B-C-D which always exist in the presence of a changed cross-section, will oppose the effect of the original charges on the battery terminals and will, thus, reduce the intensity of the current flow in the conductor.

The gradient of the surface charge distribution is larger along the thin wire than along the thick ones. Therefore, a stronger force is exerted on the free electrons within the resistor leading to a higher drifting speed. The longer the thin wire, the smaller this gradient will be and this will lead to a smaller current.

When equilibrium is reached, the same number of electrons will pass through any cross-section of the system in any given line. The drifting speed, however, will be different at different cross-sections.

In most practical cases, a resistor is introduced by choosing a different material. Such a material possesses a higher density of obstacles to electron flow, and/or it offers a smaller concentration of free electrons. Both effects normally appear together within a resistor and can cause a dramatic change in mobility for the electrons.

If a uniform conductor and a uniform resistor are soldered together, there exists two thin areas or better volumes within which the change of conductivity occurs. These areas are represented in the next picture on an exaggerated scale.

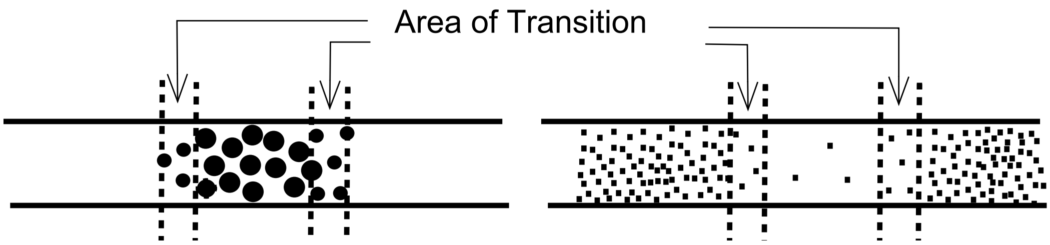


Fig. 2.16. **Increased Number of Imperfections** **Reduced Number of Free Electrons**

In the left picture, the filling represents the density of obstacles for the moving electrons. In the right picture, the density of the free electrons is represented. In both cases, the function will be the same. Within the area of transition from high to low values for the mobility, there will be on one side a slightly higher concentration of electrons than neutral and a slightly lower one on the other side. These extra charges which are spread out within two thin volumes will have two effects: On one hand, they oppose the flow in the conductor; and on the other, they increase the force on the free electrons within the resistor. In dynamic equilibrium the same number of electrons will flow through any cross-section in any given time. Due to the extra charges and the different gradient of the surface charge distribution, the driving force within the conductor and the resistor are, however, quite different.

To treat the effect resulting from these transition areas seems to be important for two reasons. First, the principle that nature changes continuously in space and time is basic for qualitative reasoning. Therefore, the dramatic change in mobility should not be treated as a step function with no further consequences. Second, this area of transition plays an important role when the conduction mechanism in semiconductors and the function of diodes and transistors have to be explained.

A last possibility to resist an electron flow is through magnetic interaction from the outside. In this case, it can be assumed that an additional force from outside is holding back parts of the electron distribution within the circuit. This leads to a redistribution of the surface charges within the other parts just as if an ordinary resistor would have been introduced.

For any of these different resistors, a similar surface charge distribution will occur which is represented in the following picture.

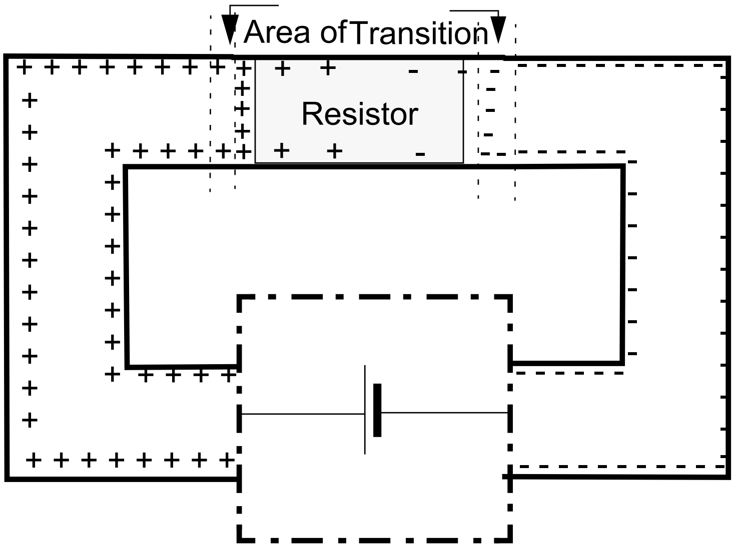


Fig. 2.17. Surface Charge Distribution (Circuit with one Resistor)

This picture indicates the principal ideas, it does not represent reality in all respects. The scale of the different parts is chosen to optimise visibility, not to represent reality on a correct scale. Moreover, the gradient of the surface charge distribution on the copper wires is not shown because it is too small compared with the gradient along the resistor.

A representation of the surface charge distribution in series and parallel circuits is shown in the following figures.

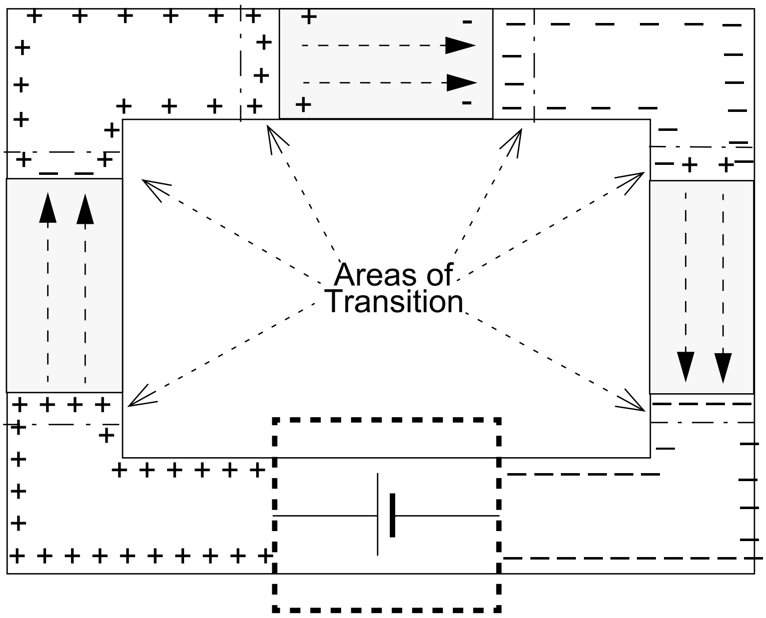


Fig. 2.18. Surface Charge Distribution in a Series Circuit

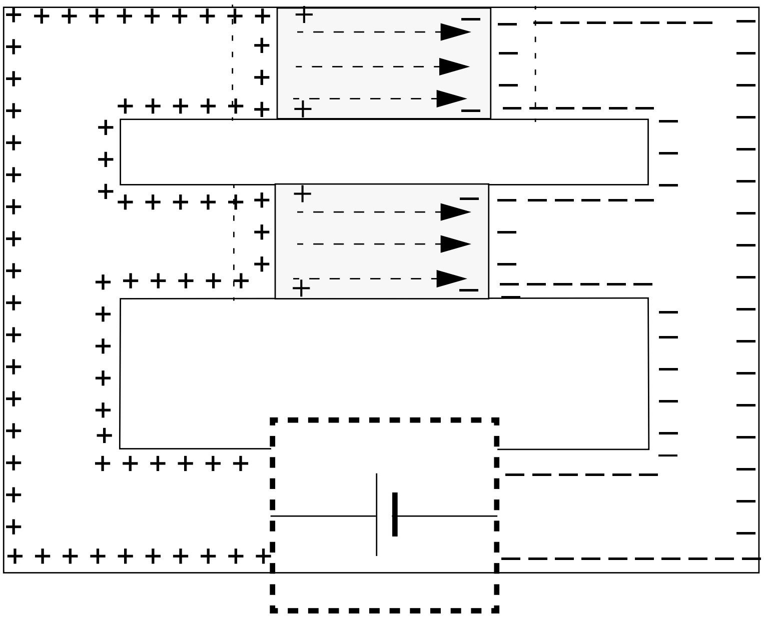


Fig. 2.19. Surface Charge Distribution in a Parallel Circuit

2.5.4. REPRESENTATION OF VOLTAGE

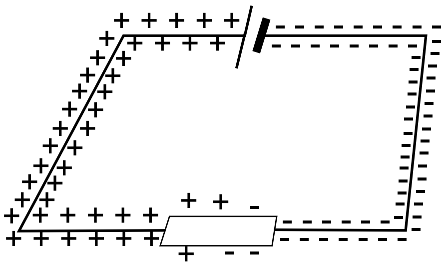
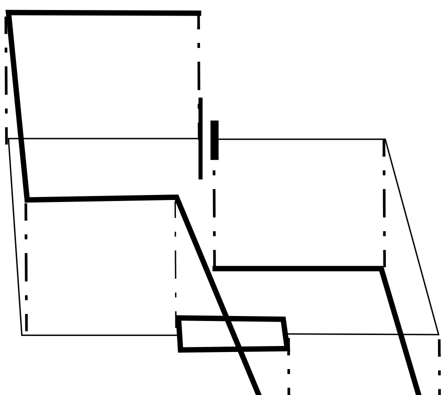


Fig. 2.20.Circuit with Surface Charges

On the left a complete circuit with an indication of all surface charges is shown.

Once the principle of the distribution of the surface charges is understood, a different and more practical representation can be offered by representing different values of surface charges by vertical lines of different length. In the following figure, negative surface charges are indicated as lines in the upper direction and positive charges as lines downwards.

This representation is more abstract but better to read, and the gradient of the surface charge distribution across resistors is immediately visible. The same information can be presented by choosing one side as zero (same height as the circuit).



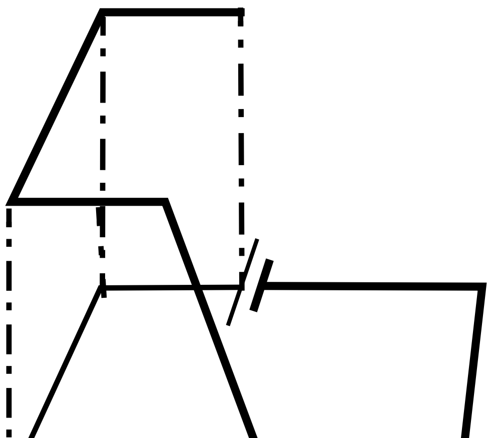


Fig. 2.21. Representation of Voltage **Change of Reference Point**

lt is not the absolute value of the surface charges but only the difference between two points that causes a net driving force on the mobile electrons within the conductor and the resistor. Any gradient in surface charge density along conductors is too small to be shown.

2.5.5. TRANSITION PHASE BETWEEN STATES OF EQUILIBRIUM

When a voltage source is applied to a circuit, the change in charge distribution travels like a wave-front through the circuit. This conclusion follows from the principle that there is no action at a distance but that a change can only be transmitted continuously in space and time.

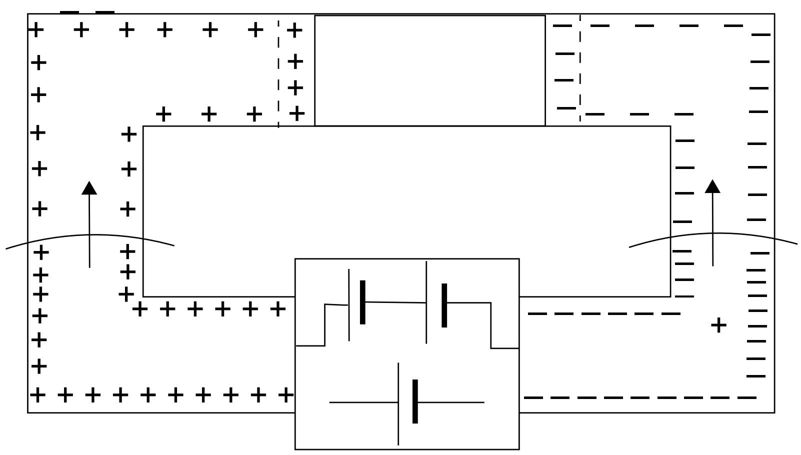


Fig. 2.22. Change of Voltage

Knowing the speed of light, it can be estimated that for a circuit of a length of 1 meter, the time to reach a new state of equilibrium is something like 10^{-8} seconds. To come to a better and deeper understanding of the underlying mechanism which controls or causes equilibrium, it may be useful to study in more details this short interval of time between two states of equilibrium.

Changes take place as a result of a voltage change (a change in the distribution of surface charges) or a change of resistance. The final result of such a change is a new equilibrium with a changed intensity of the current.

In the following, the change from one state of equilibrium to another is studied for the simple case of a uniform system with one voltage source and one resistor where the voltage in suddenly doubled :

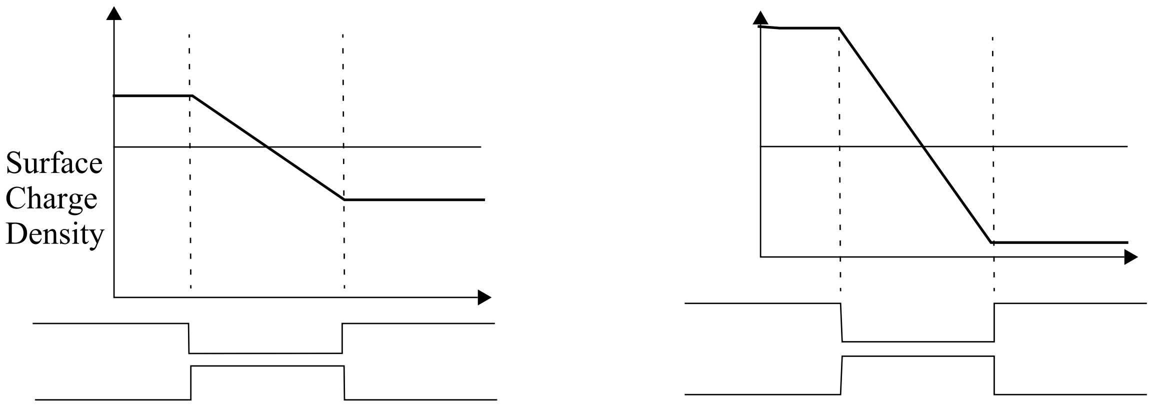


Fig. 2.23.Doubling the voltage Initial State Final State

Since the time for the transition process between these two states is rather short, it is not easy to be experimentally investigated. One can exclude the possibility that the same change occurs at all points along the circuit at the same time. This would imply action at a distance and this is not possible.

A simple case where a closed solution of the underlying differential equations can be found belongs to a linear and long (theoretically infinitely long) transmission line, where the distance between the 2 lines is much smaller than the length of the line. For such a system the corresponding algorithm can be used for a simulation program to visualize the transition process in real time.[1]

Such a simulation shows how a wave of higher voltage - and this means also of higher surface charge density - is starting from the voltage source on both lines and moves on to the two connections of the resistor. If more than 1 resistor is in place several reflection may occur, depending on certain boundary conditions. In any case, at the end a final state will be reached which is in accordance with Ohm´s law and Kirchhoff´s laws.

As another example the change of a resistor in a parallel circuit with 2 different resistors is analysed. The next picture represents the distribution of the surface charge with an interruption in the upper resistor.

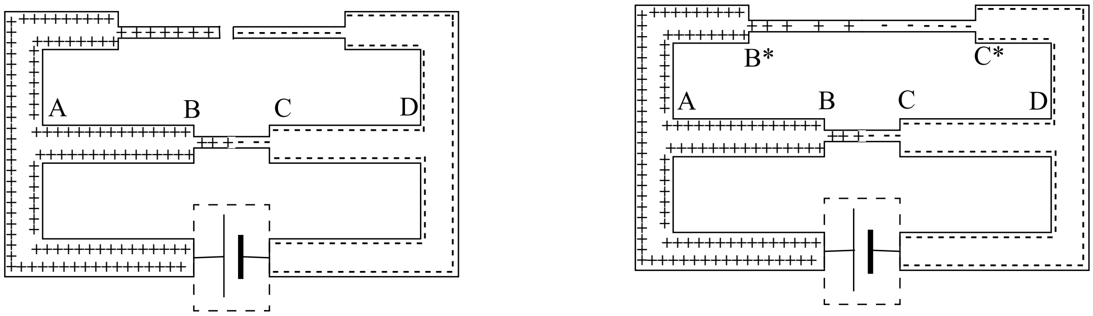


Fig. 2.24. Change of Surface Charge when a Parallel Resistor is Added

When the connection is closed (instantaneously), the large difference in surface charge exerts a driving force on the free electrons which reduces the gradient of the surface charge distribution in both directions. If the voltage source is strong enough, the distribution at A and D will not change, and the current in the lower branch will not be altered.

Due to the larger distance between B* and C*, the gradient in the surface charge distribution will be smaller than along the short resistor; and this results in a smaller current for the new state of equilibrium.

For practical purpose one can neglect the resistance of the conductors and claim the same constant voltage for every parallel branch. However, to answer the question, why the current is split according to the different resistors at the branching point, a more careful look to the surface charge gradients is helpful. To drive this smaller current from A to B* and from C* to D within the conducting wire, a smaller difference in surface charge gradient is necessary than for the larger current from A to B and C to D. These differences correspond to the difference in voltage drop along the conducting wires in the parallel branches. Due to the high conductivity of conductors these voltage drops can be neglected in normal quantitative analysis. Though these values are small, they are not zero; and this is an important, or even decisive, factor for qualitative reasoning.

At the branching point, there is, therefore, a different gradient of the surface charge distribution in the two branches when equilibrium is reached. According to this difference, the current is redistributed at the branching point. Any further change of the resistor will lead to a change of the surface charge gradient within the parallel branches and, therefore, to a redistribution of the cur-

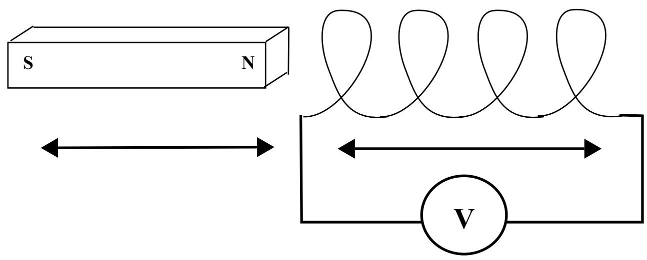
1.For a simulation program for teaching/learning purposes see http://http://www.astrophysik.uni-kiel.de/~hhaertel/TL/TL.zip
For a tutorial about transmission lines and transition processes see: http://http://www.astrophysik.uni-kiel.de/~hhaertel/TL/TL-tutorial/index.html

rent at the branching point.

Chapter III
ELECTROMAGNETIC INDUCTION AND RELATIVITY

3.1. TRADITIONAL DESCRIPTION OF ELECTROMAGNETIC INDUCTION

In its simplest form, the law of electromagnetic induction can be formulated as follows:

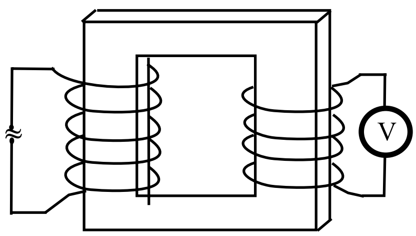


During the time when a magnetic field is changing within a conducting coil, a voltage is induced between the ends of this coil. The validity of this law is normally demonstrated with the help of a permanent magnet, a coil, and a volt or ampere meter.

Fig. 3.1.Relative Motion Between Magnet and Coil

While either the magnet or the coil (or both) are moving in relation to each other, it is stated that the common underlying process is the change of the magnetic field in time within the coil which is inducing the voltage. This process is often "explained" by the Lorentz Force, which is observable when a charge carrier is moving relative to a magnetic field.

In either case, when the magnet or the coil is moving and the magnetic field is changing within the coil, field lines are "crossing" over the filament which contains the free electrons.



The direction of the Lorentz Force which follows from the right-hand rule can then "explain" the induced voltage and current. The quotation marks around "crossing" and "explain" indicate that this picture of crossing field lines is questionable. The problem becomes visible when the permanent magnet is replaced by an electromagnet and the change of the magnetic field is controlled by changing the current.

Fig. 3.2. Transformer

In this case the process cannot be explained with the Lorentz Force because the field lines are not crossing over the filaments. The induced EMF can only be described by the change of the magnetic field within the coil. Normally, no explanation is given to make the connection between the change of the magnetic field in the interior of the iron core and the coil at the outside of the magnet. There is also no explanation why, in this case, the Lorentz Force cannot be used as a cause for the induced EMF. This kind of contradiction is, of course, hidden behind abstract notions like change of a magnetic field in time and is probably in most cases, not detected by students.

Flux Law and Lorentz Force - As a universal law of induction, it is derived from many different experiments that the time derivative of the magnetic flux causes a voltage. The relation between voltage and electric field, known from electrostatics, is used to introduce circular fields.

$$U_{ind} = \int E ds = -\frac{d\phi}{dt}$$

This equation expresses Faraday's Law, stating that the change of magnetic flux causes or induces a circular electric field. It states that there is an electric field even in the absence of a conductor.

Normally it is shown in the textbooks that the motion of the conductor leads to the same change of the magnetic flux as expressed in the equation above. This could lead to the assumption that the flux law expresses a general law, while the Lorentz force is an additional method for a certain subset of cases. This is, however, not correct. In Feynman's book about his lectures (Feynman et al. Pg 17-3), he states explicitly that there are exceptions to the flux law which can only be described with the Lorentz Force and that the correct physics is only given by the two basic equations:

$$F = q(E + vxB)$$

$$\mathrm{curl}E = -\frac{1}{c} \cdot \frac{dB}{dt}$$

3.2. CONTRADICTIONS IN TRADITIONAL TEACHING

3.2.1. TWO DIFFERENT EXPLANATIONS FOR THE ELECTROMAGNETIC INDUCTION

To describe the effect of an induced voltage, either the change of the magnetic flux or the effect of the Lorentz Force on a moving charge is used.

$$U_{ind} = \int E ds = -\frac{d\Phi}{dt} \qquad F = qE + \frac{q}{c} v \times B$$

Feynman states in his book about his lectures: "We know of no other place in physics where such a simple and accurate general principle requires for its real understanding an analysis in terms of two different phenomena." In most textbooks, however, this problematic situation is not mentioned; and, worse, it is hidden by incomplete and even incorrect "explanations." To prove this statement, it is analysed how the following experiments are treated in the textbooks.

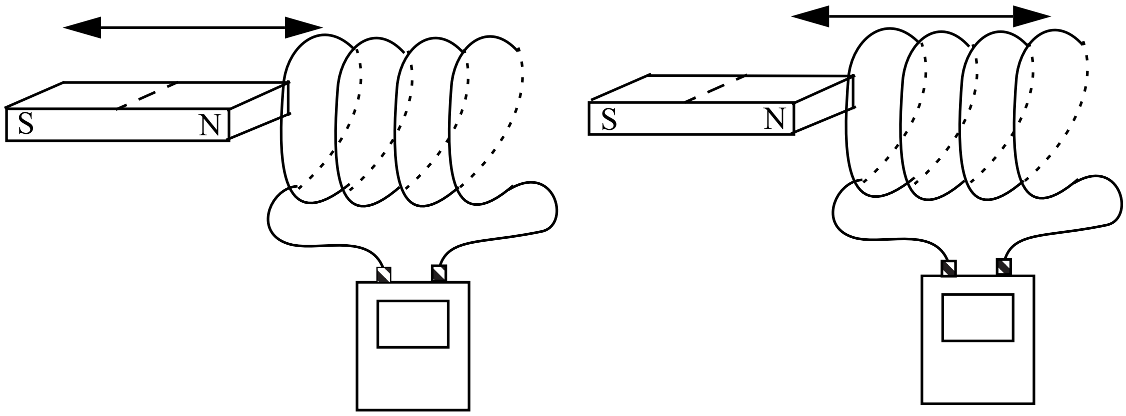
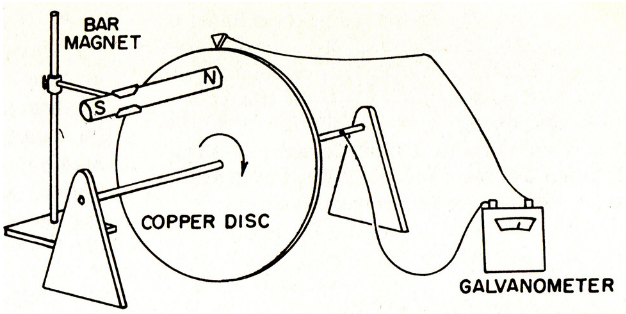


Fig. 3.3. Induction by Moving the Magnet Induction by Moving the Coil

In both cases, when either the permanent magnet or the coil (or both) are moving relative to each other, the same thing happens. The magnetic field lines and the electrons in the coil are crossing each other. If the relative velocity is the same, the induced voltage is the same. The physical description, however, has to use two different phenomena. The experiment on the right can be described and "explained" with the Lorentz force, because the electrons in the conductor are actually moving. The experiment on the left (which is practically the same) can only be described by Faraday's Law. Otherwise, one would have to introduce the velocity of a moving magnetic field line to find a v for the expression (v x B). The velocity of a magnetic field line is, however, not defined in physics.

It is, of course, possible to hide this contradiction by using only Faraday's flux law to describe both experiments.

A student cannot, however, understand why the Lorentz Force can be applied only in one case and why it is better not to apply it at all.



In some textbooks, it is stated explicitly that the mechanism underlying Faraday's Flux Law is the fundamental one and that the application of the Lorentz Force is only possible in special cases. This statement is incorrect. There are cases when the flux law cannot be applied as, for instance, in the following experiment presented by Feynman.

Fig. 3.4.Induction Without Change of Flux

While the copper disc is turning between the poles of the permanent magnet, there is an induced voltage without a change of the magnetic flux through the area of the circuit. The same argument holds for any experiment where the Hall voltage is measured.

3.2.2. FARADAY'S LAW AS A CAUSAL RELATION

$$U_{ind} = \int E ds = -\frac{d\Phi}{dt}$$

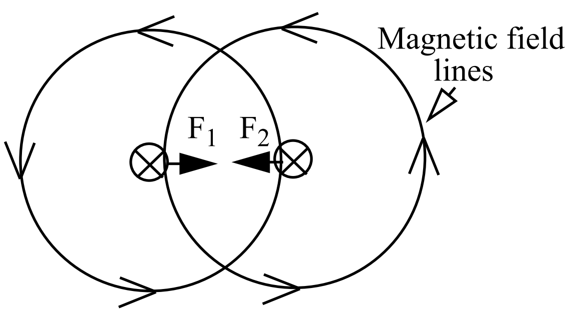
Faradays' Law is in most cases interpreted as a causal relation. The change in time of the magnetic flux within an area causes a circular electric field around this area. If there is a conductor around this area with a small gap, a voltage can be measured across this gap.

The interpretation of this equation as a causal relation is problematic in two ways: First, there is no mechanism offered to the student to connect the flux change in the interior of the area with the induced electric field at the border of this area. This is a specific difficulty when the wire of the conduction coil is always within an area where there is no field at all, and therefore no field lines can ever cross the wire of the induction coil. Such a case is found in any transformer with a closed iron core or around a perfect solenoid.

Second, there are strong arguments stating that this causal relation between change of flux and induced voltage does not exist (Rosser, 1968). It can be shown that these two phenomena always coexist and that the cause of any change is the movement or acceleration of some charges which originally caused the magnetic field.

3.2.3. LORENTZ FORCE AND ACTION AND REACTION

The Lorentz Force is normally taken as the cause for the change of the velocity of an electron moving through a magnetic field. In such an interpretation, the two partners of interaction are the magnetic field, represented by lines, and the moving particle. The result of this interaction is a force perpendicular to the direction of the movement and the direction of the field.



As a result of this interaction, one partner is accelerated while the field is not altered at all. The reaction to this action takes place at some distance where other moving particles in other parts of the magnetic field are accelerated in opposite directions. There is no mechanism presented to the student which could explain the connection between these different places of action and reaction.

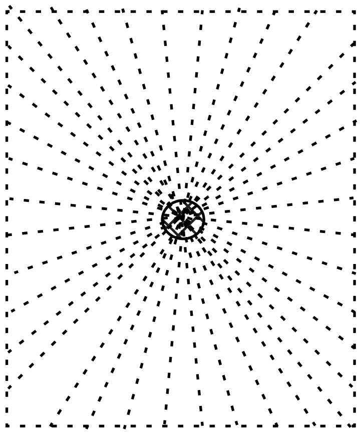
Fig. 3.5.Interaction Between Parallel Conductors (Perpendicular to the Paper Surface)

Again, there are arguments stating, that there is no causal relation between a moving charge carrier and a magnetic field (Rosser, 1968). It can be shown that a moving particle interacts not with the magnetic field, but with the current, which was connected to the magnetic field. The magnetic interactions can be explained by a change of the Coulomb interaction due to relativistic effects.

3.3. BASIC ASSUMPTIONS FOR A NEW APPROACH

The basic assumptions described in Chapter II about distortions in space, when charges and electric fields are present, can be developed further by including relativistic changes due to the movement of charge carriers. As described in the Berkeley Physics Course, Vol. II, these relativistic changes of the Coulomb field can be used to explain the magnetic interaction between currents.

The Theory of Relativity states as a basic principle that length, time, and mass depend on the velocity of the system for which these values are defined. All this is a consequence of the postulate that the Speed of Light can only be reached approximately by any material object. No object can ever move faster than the Speed of Light. As a consequence of this postulate, it can be shown by calculation that the field around a charged particle changes in a specific way. To date, there is no mechanism to make this change understandable or plausible in a qualitative way. This change, which is described in the following paragraph, has to be taken as a basic fact connected with electric interaction between moving particles. When a particle is moving with constant velocity relative to a frame of reference, the symmetry of the distorted space around the centre is changing from spherical to cylindrical in the following way:



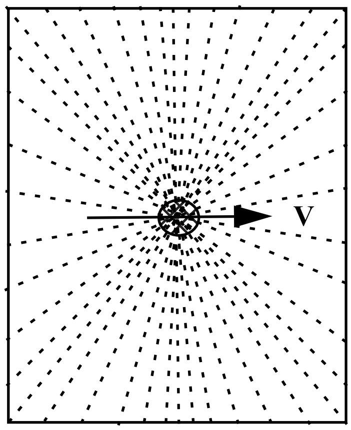


Fig. 3.6. Charged Particle at Rest (spherical symmetry **Charged Particle with Velocity V (cylindrical symmetry)**

There is an enhancement of the field perpendicular to the direction of the velocity and a reduction in the forward and backward direction. As a consequence of this change, the surrounding of a positive charge at rest and a negative charge moving with constant speed do not cancel completely even when they are close together. There would be an enhanced negative interaction perpendicular to the velocity and an enhanced positive interaction in the forward and backward direction.

The electric Coulomb fields, surrounding the centre of charge carriers, inerpreted as distortions or changes of space (Chapter II) do have spherical symmetry for particles at rest and cylindrical symmetry for particles moving with constant velocity. For the following arguments, this change in symmetry is basic for a description of the interaction between moving charge carriers and for an explanation of electromagnetic induction. This symmetry and its change in connection with the movement of a charged particle is represented in the following way:

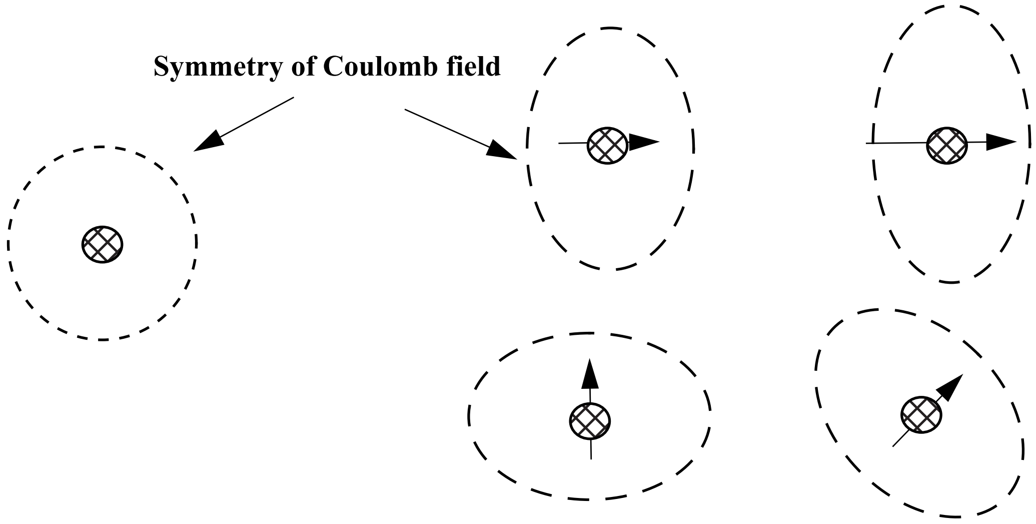


Fig. 3.7. Charged Particle at Rest **Charged Particle with different Velocities**

This representation indicates that there will be an increase of interaction perpendicular to the direction of the velocity and a decrease in the forward and backward direction.

In this chapter it is shown how this approach can be used to describe magnetic interaction and the electromagnetic induction. This is done by first analysing the interaction of the following cases:

- - two single particles moving with constant speed.
- - a single particle moving with constant speed and a linear current
- - a single particle moving with constant speed and a single closed loop
- - interaction between two parallel currents

The analysis is limited to movements in parallel and perpendicular directions. After that the change of symmetry due to changing currents or accelerated charge carriers and the effects connected with these changes are described.

3.4. INTERACTION BETWEEN CHARGE CARRIERS MOVING WITH CONSTANT VELOCITY

3.4.1. INTERACTION BETWEEN SINGLE CHARGE CARRIERS

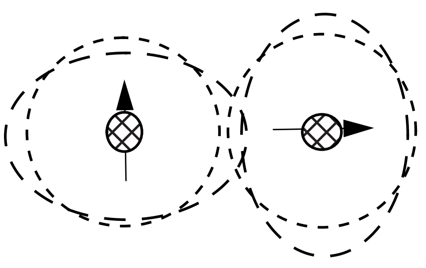


Fig. 3.8.Two Electrons Moving at Right Angles

It follows from the Special Theory of Relativity that one cannot derive the interaction between moving particles by relating every velocity to the frame of reference of the laboratory. In the very unrealistic but simple case of two electrons moving at right angles, it can be seen that such a method would lead to contradictory results.

These two electrons would, of course, repel each other due to the normal Coulomb interaction; and this interaction would be many orders of magnitude larger than any influence due to relativistic effects. But if one could imagine that this Coulomb interaction is balanced by some positive charges, then the relativistic effects give two different results for the two electrons. The interaction is enhanced for one electron and reduced for the other one. This would imply that Newton's Third Principle of "action = reaction" would not hold for electrodynamic interaction. To get the correct result, it is necessary to change to a frame of reference of one of the involved particles and to derive the interaction for this particle from this frame of reference. This leads to one of the two following pictures and indicates equal and opposite interaction for the two electrons.

Fig. 3.9.Two Electrons Moving at Right Angles (Represented in Different Frames of Reference)

3.4.2. SINGLE CHARGE CARRIER AND LINEAR CURRENT

MOVEMENT IN PARALLEL AND ANTI PARALLEL DIRECTION

A single electron, when moving in the same direction as a linear current will be attracted attracted and will be repelled when moving in the opposite direction.

This effect is normally explained with the existence of the Lorentz Force and the expression:
F = q (v x B)
In the frame of reference of the laboratory, the current and a moving electron can be represented as follows:

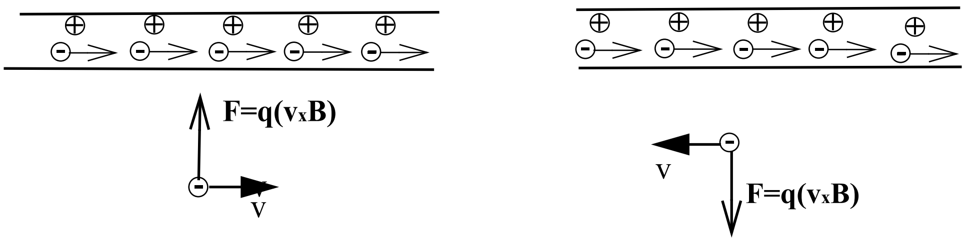


Fig. 3.10. Interaction between a Current and an Electron
Moving Parallel Moving Anti parallel

To simplify the representation, it is assumed that the single electron and the electrons in the conductor do have the same speed.

The concept of distorted Coulomb interaction gives the same result. To find the correct interaction for the single electron, the frame of reference has to be changed to this particle. In this frame of reference, the left part of the picture shows that only the positive charges seem to move, resulting in an enhanced interaction perpendicular to the direction of their velocity. For the single electron remains an interaction with a positive distortion and, therefore, an attractive force towards the conductor.

To simplify the following representations and explanations, the current is represented in a symmetric form. It is assumed that the free charge carriers are half positive and half negative and are moving in opposite directions. Such a current, which could be realized in semiconductors, will create the same interaction as a current where only the electrons are moving (either twice as many or twice as fast).

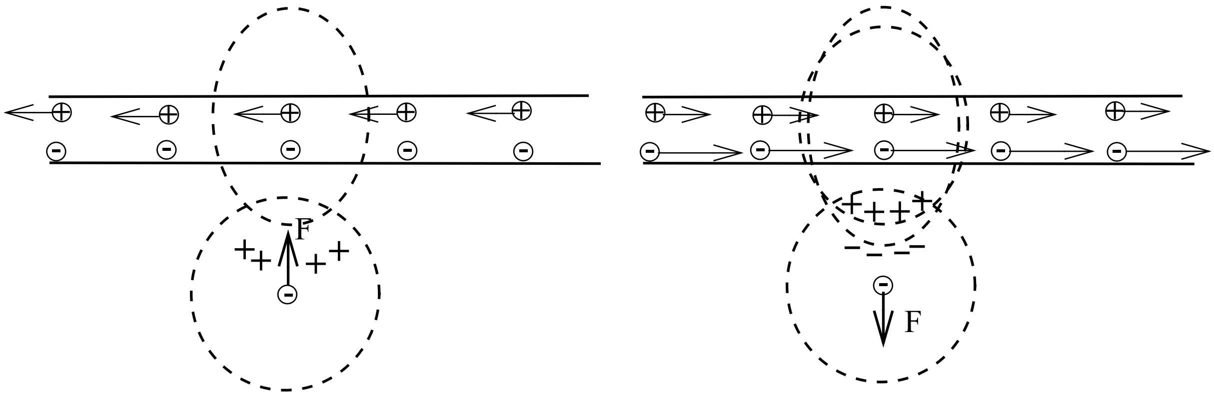


Fig. 3.11. Current and Electron (seen from Reference Frame of the Moving Electron)
Electron Moving Parallel Electron Moving Anti Parallel

The result for the frame of reference of the single particle moving in the anti parallel direction is shown in the right part of the above picture. Due to the larger speed of the electrons in the conductor relative to the single electron, the "negative" distortion is larger than the "positive" one, resulting in a repelling force.

MOVEMENT AT RIGHT ANGLE

An electron moving at a right angle in respect to a straight current and within the plane formed by the straight current and the particle is deflected due to the Lorentz Force in the following way:

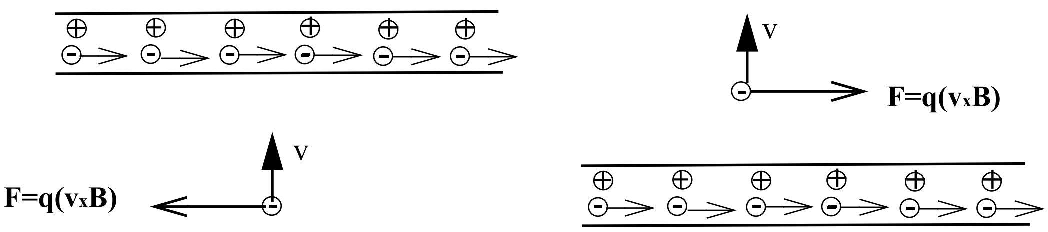


Fig. 3.12. Lorentz Force for an Electron
MovingTowards a Current at Right Angle Moving Away from A Current at Right Angle

To derive the result, based on the concept of the distorted Coulomb interaction, the current, composed from positive and negative moving charge carriers, have to be looked at from the frame of reference of the single electron. The result looks like the following:

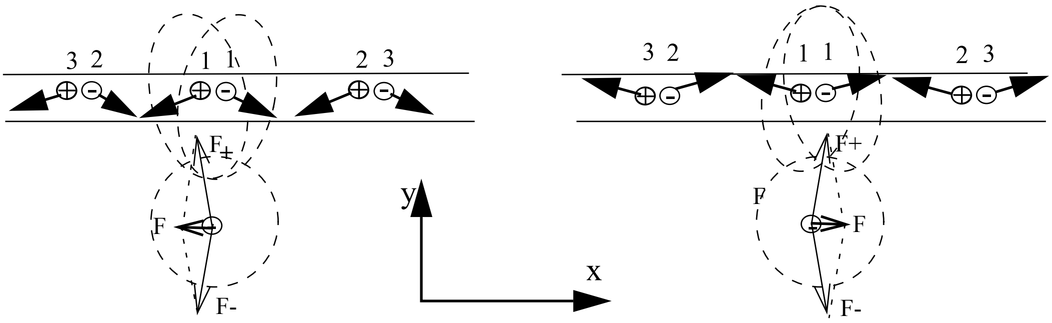


Fig. 3.13. Electron Moving Towards a Current (At Right Angle) Electron Moving Away from a Current (At Right Angle)

Pairs of positive and negative particles at symmetric positions in respect to the single particles (1-1, 2-2, 3-3,) have to be related to the single electron. Due to the different directions of their velocities in respect to the frame of reference of the single electron, two forces exist which have equal components in the x-direction and equal but opposite components in the y-direction. As result, there exists a force parallel to the current just as it is indicated by the "Right-Hand-Rule".

INTERACTION BETWEEN A MOVING CHARGE AND A CONDUCTING LOOP.

To explain the phenomenon of induction to newcomers as interaction between conductors with varying currents or localities, in most cases the explanation starts with a single moving charge carrier and the Lorentz force in combination with the so-called "Right-Hand-Rule". The following cases are quite easily described by pointing to the fact that the moving charge carrier is cutting magnetic field lines and thereby inducing the Lorentz force.To simplify the graphical representation of this case, the conductor loop is designed as a rectangle.

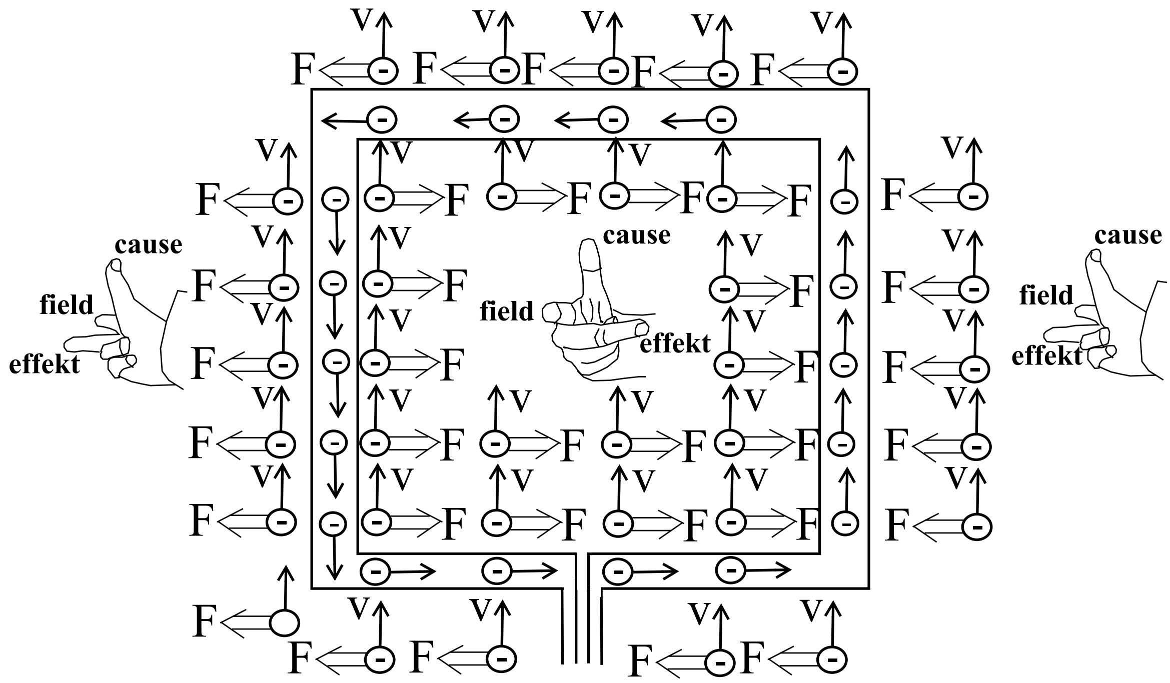


Fig. 3.14. Interaction between a moving charge carrier and a conducting loop Applying the Lorentz Force

From this picture the following rules can be derived:

- 1. A moving electron is attracted by a parallel current and repelled by an anti parallel one.
- 2. A moving electron approaching a current at a right angle is moved to the opposite direction as the current flow
- 3. An electron moving away from a current at a right angle is moved to the same direction as the current flow.

The Symbol F⇐⊖ is representing the result of the interaction of a single moving negative charge carrier and the current within the loop. This interaction is composed of two interactions with the two opposite parts of the loop acting always in the same direction. The parallel movement for one part is anti parallel to the opposite part. While moving towards a current at a right angle inside of the loop, the distance to the opposite part is increased. The same final result is obtained, of course, by applying the Lorentz Force and building the cross product of v and B. This law expresses the underlying symmetry of the system and can be interpreted as a convenient shorthand to connect the starting conditions and the final outcome.

The same result can be achieved based on the concept of disturbed Coulomb interaction due to relative velocities of the interacting charge carriers. To simplify the graphical representation the circuit is again designed in form of a rectangle with only 4 different relative positions of the interacting partners. The current is again seen as composed of moving positive and negative charge carrier in opposite direction. Each time the frame of reference has to be taken from the single charged interacting particle.

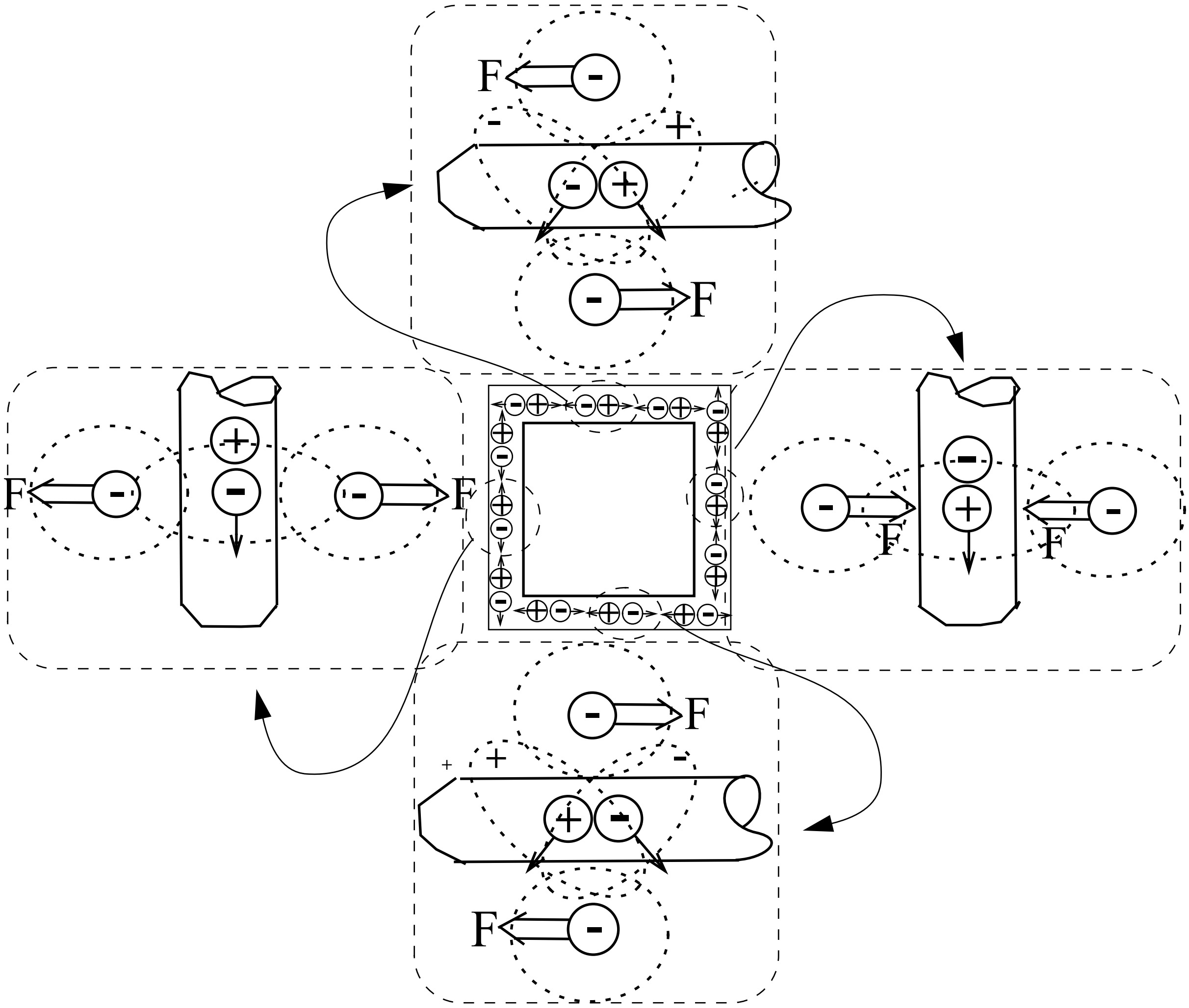


Fig. 3.15. Interaction between a moving charge carrier and a conducting loop Applying relativistic change of Coulomb interaction

The Symbol F⟸⊖ is again representing the result of the interaction between a single moving charge carrier and the positive and negative charge carriers within the wire due to the relativistic distorted Coulomb interaction.

3.4.3. CURRENT/CURRENT INTERACTION

The description of the interaction between two linear currents is straightforward. The effects for a single electron have just to be integrated to give the macroscopic result for the conductor carrying the current. For closed circuits and coils, symmetry arguments have to be added in the same way as has been done for the interaction of single particles with a conducting loop . In each case, the same result is derived that follows from the application of the force law: F=q(vxB)

3.5. INTERACTION DUE TO ACCELERATING CHARGE CARRIERS

3.5.1. SINGLE CHARGE CARRIERS MOVING WITH CONSTANT SPEED

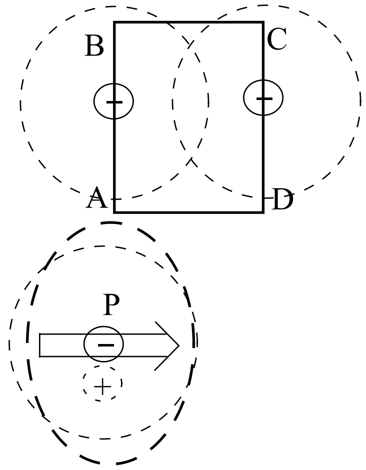


Fig. 3.16.Moving Charge Carrier and a nearby Loop

If a single charge carrier is travelling through space with constant velocity relative to other charge carriers, the interaction with charge carriers within the loop A-B-C-D is continually changing. During the time when the particle is moving from the left towards the point P, it can be argued that the interaction with charge carriers along the line A-B will always be stronger than with those along the line C-D.

This statement is true due to the fact that the relativistic change in space is increasing the interaction in the direction perpendicular to the velocity and decreasing in the forward and backward direction. This difference in interaction would cause a current to flow in the loop A-B-C-D, if a conductor were present. In classical terms, one would argue that the approaching charge carrier is producing an increasing magnetic flux through the loop A-B-C-D; and according to the flux law, this would induce a circular E-field.

For the more realistic case of a complete current, the situation is changed.

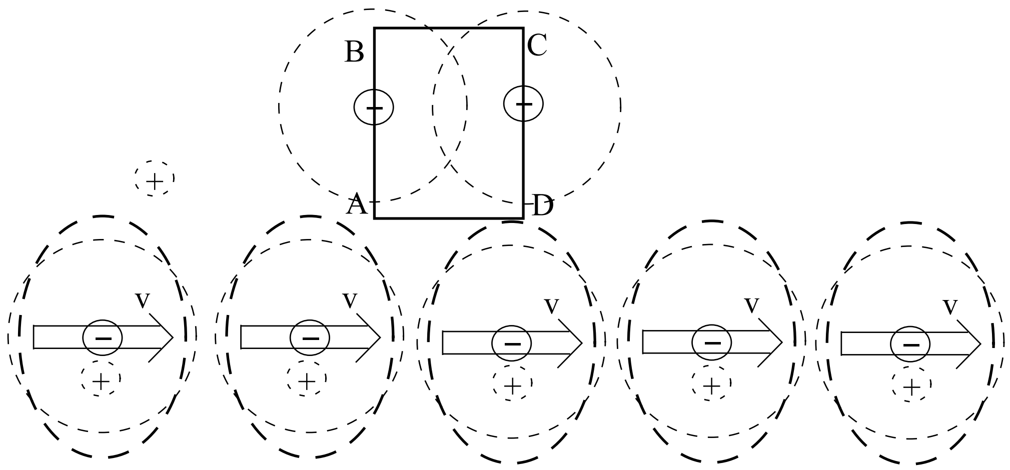


Fig. 3.17. Symmetry Around a Linear Current

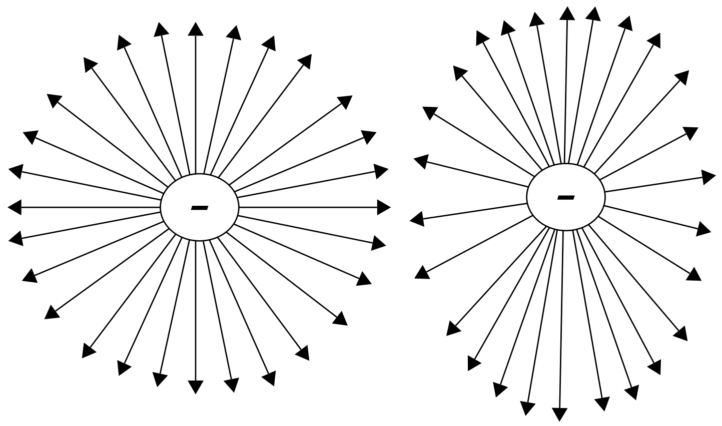
Due to symmetry, there can be no difference between the line A-B and C-D; and, therefore, there is no induced circular E-field around the loop A-B-C-D for a constant current flowing nearby. Such an effect, a circular E-field around this loop, occurs only when the current is changing or, in other words, when an ensemble of charge carriers is being accelerated or decelerated. If charge carriers are accelerated or decelerated during a certain time, a change occurs which travels out in space with the speed of light and which will be analysed in greater detail in the next paragraph.

Effects caused by changes in time should be represented in an appropriate manner, which means with the help of animated graphics. If static pictures in print media are used to represent changes in time, the reader continually has to transform this sequence of pictures into a representation of the same object or the same place changing in time. It is an open question how well this task can be done by the majority of newcomers to this field and how much help can be pro

vided when animated graphics in an interactive environment will be available.

3.5.2. ACCELERATION OF SINGLE CHARGE CARRIERS

Until now, only the relativistic distortion of the Coulomb field of a particle moving with constant speed has been analyzed.



To understand the effect of acceleration, one has to consider the change in the electric field - visualized as lines of force - due to this change in symmetry and in addition due to a changing position of the charge carrier - the centre of the field. Such a change cannot happen everywhere at the same moment in time. Action at a distance is not possible

Fig. 3.18. Field of a Charge at Rest and Moving with Constant Velocity

Such a change (of position and symmetry) has to start at the charge carrier and will develop out in the environment with the speed of light. Such a phenomenon is known in Physics under the term "Retarded Potential".

An analysis for accelerating charge carriers has been carried out in Berkeley Physics Course for two cases. First, the surrounding of an electron which has stayed at x = 0 for a while was shown which then had been accelerated for a rather short time and moved on with constant speed. In the second case, an electron which had moved with constant velocity had been stopped in a rather short time and has been at rest for a certain time.

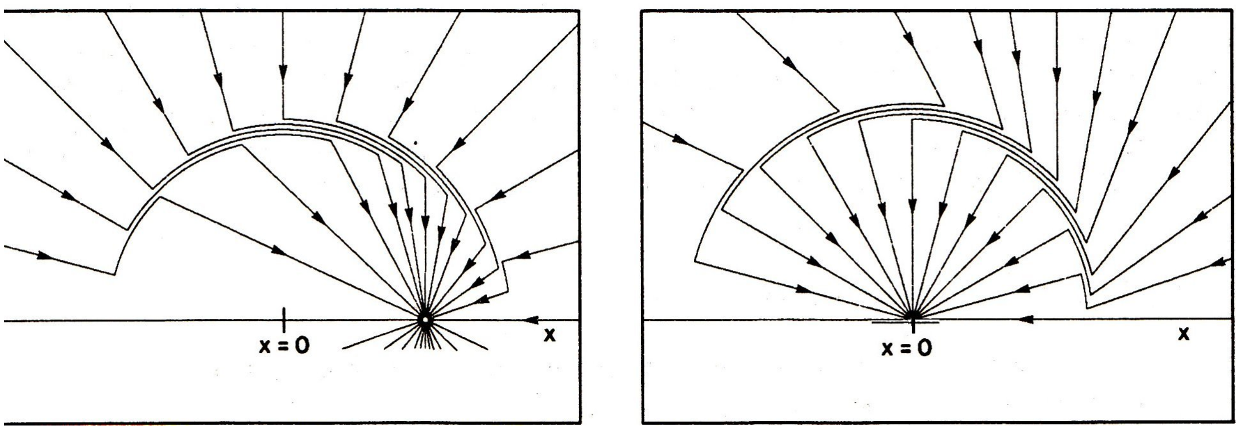


Fig. 3.19.A charge carrier initially at rest at x= 0, is abruptly accelerated and is moving and with constant velocity thereafter.

A charge carrier that has been moving with constant velocity in x-direction has been abruptly stopped at x = 0.

(Berkeley Physics Course, 1963, pages 164-165)

From figure 3.19 it can be seen that the field in the outer environment has not yet been "informed about the new position of the charge carrier as centre of the new field in the inner part. The transition between these two parts results in a circular wave, moving to the outside with the speed of light.

In 1924 Leigh Page published a so-called Emission Theory of Electromagnetism. The basic idea is that for a single charge carrier socalled "moving elements" are leaving in all direction with the speed of light. These particles are purely kinematical in nature, meaning, that they do not carry any energy or momentum. The path of those virtual particles, which are emitted under the same angle, correspond with the well known lines of force, normally used to indicate the strength and symmetry of the Coulomb field around a charge carrier. The basic ideas of this emission theory, which include relativity effects can be used as algorithm for a simulation program to visualize processes like those in the Berkeley Physics Course and many other similar processes[1]. The following picture shows two examples.

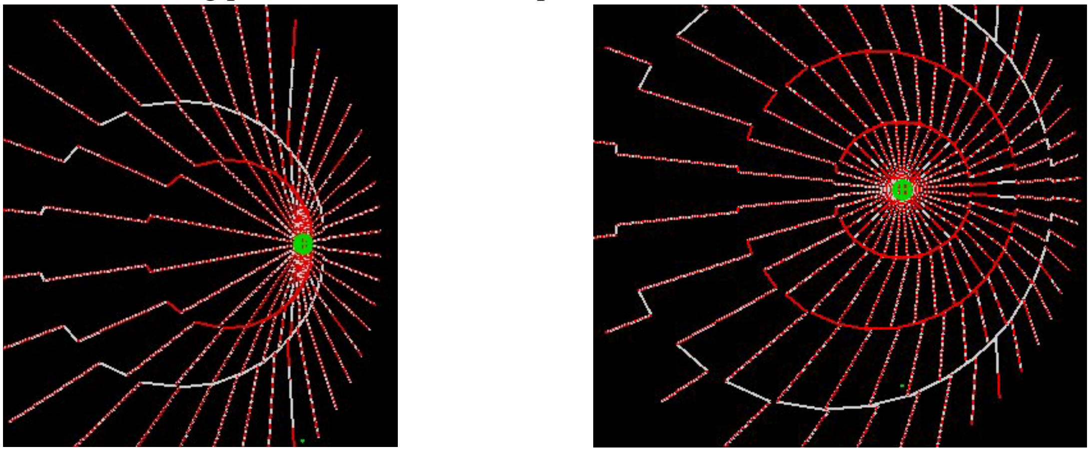
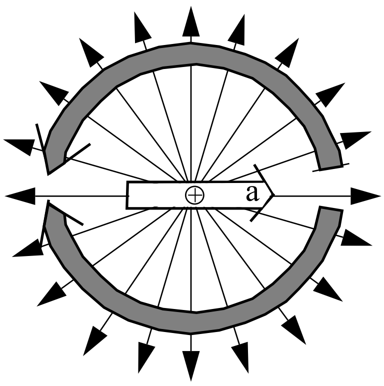


Fig. 3.20. A point charge being , accelerated twice **a point charge being decelerated twice**

The conclusion is: when ever a charge carrier is accelerated, the associated Coulomb field is affected. This change in the field, represented traditionally by field lines, shows up as a circular wave, moving to the outside with the speed of light. This wave adds a circular component to the normal radial component of the Coulomb field.



A charge carrier of equal sign close to an accelerated one would therefore not only be repelled to the outside, but would in addition to that be accelerated in the opposite direction as the original acceleration.The general symmetry of such an arrangement can be represented as shown on the left:

Fig. 3.21. Change of Symmetry due to Acceleration a

Following the convention that the field is pointing from positive to negative charges, the results for opposite charges and opposite directions for the acceleration are represented in the following

1.Under http://astrophysik.uni-kiel.de/~hhaertel different screen shots and videos are found visualizing the field around an accelerated point charge figure:

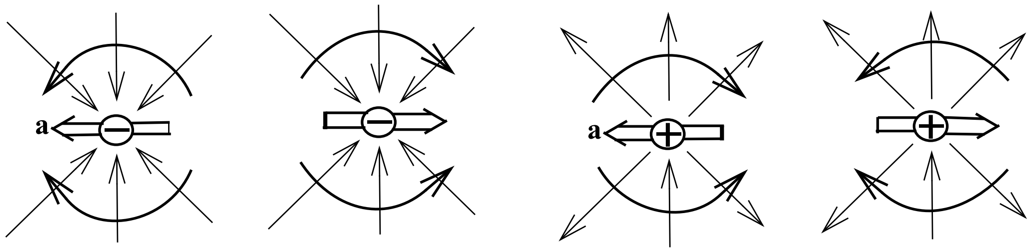


Fig. 3.22. Direction of Circular Waves for Different Charges and Different Direction of Acceleration

A negative charge accelerated to the right and a positive charge, accelerated to the left thus cause the same circular wave. In the following pictures the symmetry of space around charges will be indicated in the traditional way: the arrows (field lines) are pointing from plus to minus and negative charge carriers are being accelerated against this direction.

3.6. INTERACTION DUE TO CHANGING CURRENTS

3.6.1. CHANGE AROUND A SINGLE CIRCUIT

The symmetry around an accelerated single particle changes when many particles move together to form a current within a conductor. The radial component cancels because the current can always be thought of as a flow of positive and negative charge carriers of equal magnitude and opposite direction. The sum of all circular parts results in a cylindrical symmetry around the current during the time period when the current is changing. This change, which appears during the time of acceleration, has a direction parallel but opposite to the current flow and travels out in space with the speed of light.

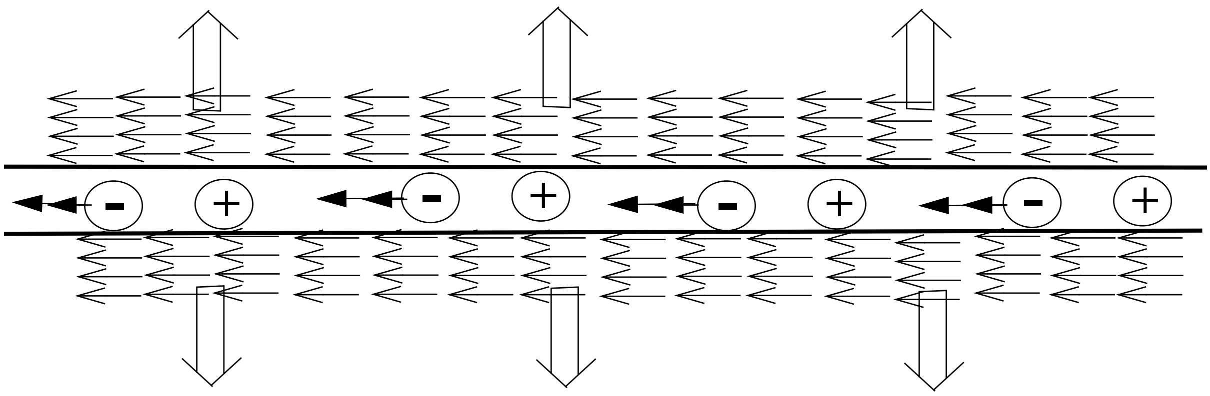


Fig. 3.23. Symmetry of Space Around a Changing Linear Current

When the acceleration has ended and the current has come to a steady state, this change disappears again with the speed of light to the outside

3.6.2. INTERACTION BETWEEN ELEMENTS OF THE SAME CIRCUIT - SELF-INDUCTION

Within a closed loop, there is inevitably an interaction between different parts of the circuit giving rise to the effect of self-induction. This self-induction can be interpreted as a kind of inertia or resistance to a current when its charge carriers are either accelerated or decelerated.

The strongest and unavoidable interaction will occur between charge carriers and their nearest neighbours. This interaction will give rise to a kind of inertia either during acceleration or during deceleration.

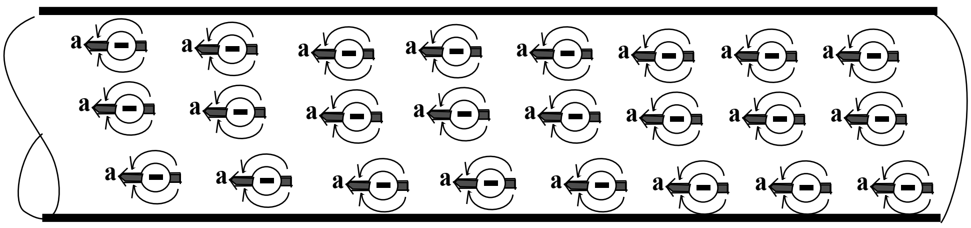


Fig. 3.24. Interaction Between Parallel Accelerated Charge Carriers

There is another interaction with the geometrically opposite part of the conductor, which will, however, not resist but support the original movement.

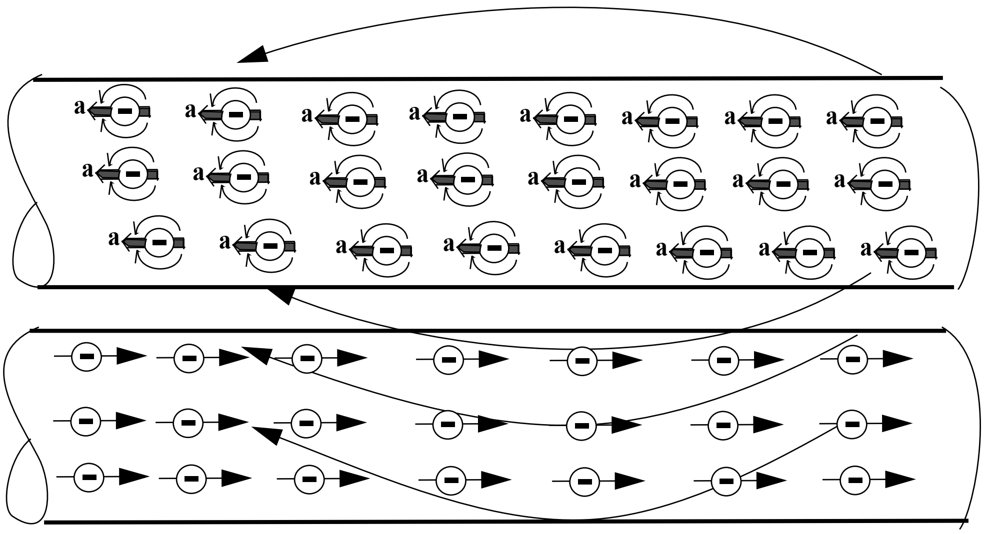


Fig. 3.25. Interaction Between Opposite Parts of the Circuit

If the conductors are very close to each other, this interaction will be at its maximum and will nearly cancel the first described interaction between parallel accelerated charge carriers. This description corresponds to the fact that a circuit, where the wires are very close to each other (or twisted around each other), will show only a small self-inductance. If, however, the circuit encloses a large area and the opposite parts of the circuit are far apart, this second interaction is rather weak, and the self-inductance, due to the interaction between parallel charge carriers has a maximum value. This again corresponds to the traditional description, where the induced voltage, due to selfinduction, is proportional to the change of flux within the circuit and therefore proportional to the area encircled by the loop.

3.7. INTERACTION BETWEEN SEPARATED CIRCUITS - MUTUAL INDUCTION

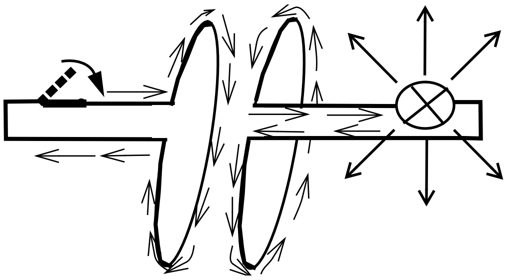


Fig. 3.26.Interaction Between Parallel Circuits (Electromagnetic Induction)

Charge carriers within a separated circuit will be accelerated when a circuit nearby is connected to a voltage source. If the two circuits are parallel, the induced current will flow in opposite direction.

In order to fully understand this process of electromagnetic induction it seems to be helpful to further analyse this interaction between two circuits and to look at the additional interaction due to the induced current back onto the original one.

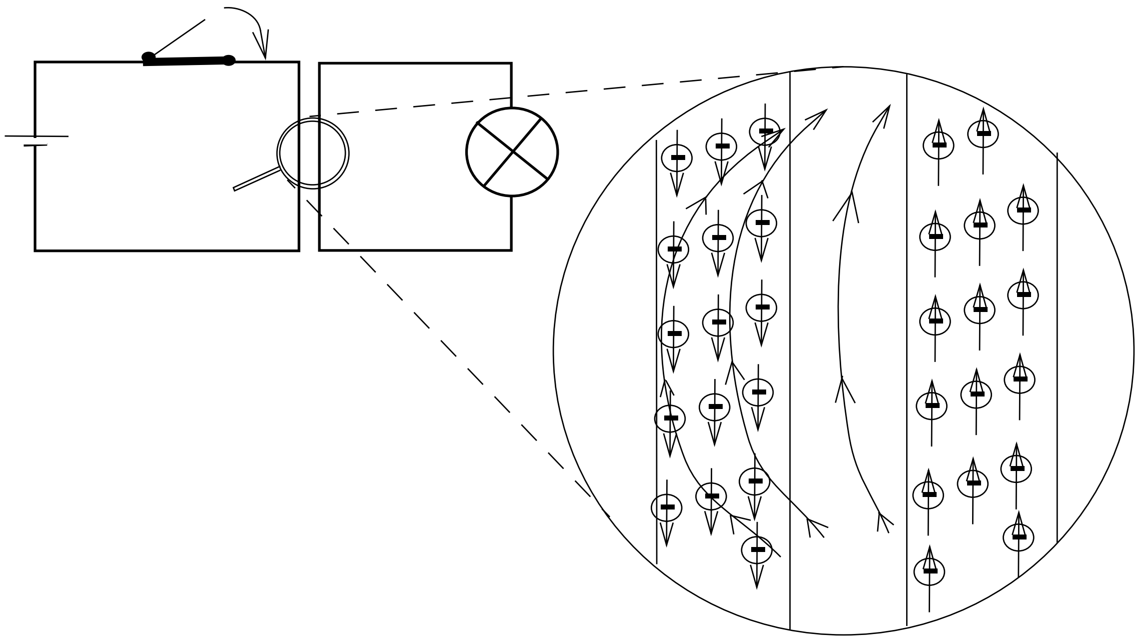


Fig. 3.27. Feedback from the Secondary to the Primary Coil

When the charge carriers in the second circuit are accelerated and start to form a current flow, they also, as all accelerated charge carriers, sent out circular waves.

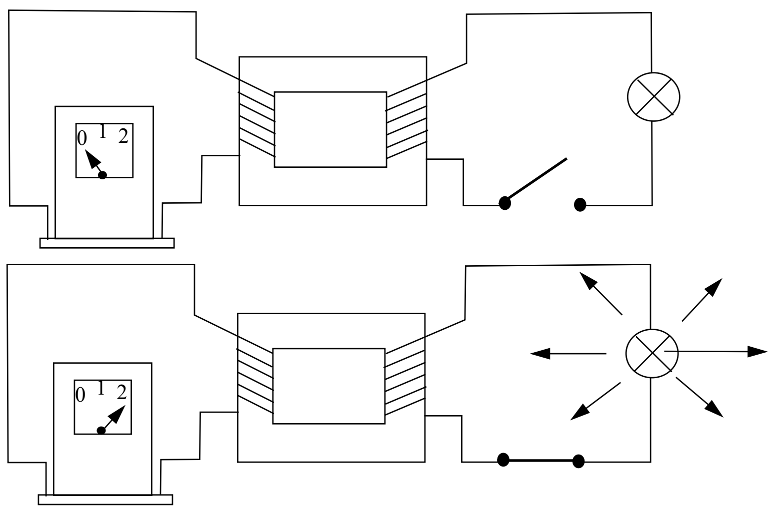


Fig. 3.28.The Current in the Primary Coils is Controlled by the Current in the Secondary Coil

These will interact with the charge carriers within the first circuit. This interaction, however, does not resist but it supports the original current flow. This fact is not a violation of the law of conservation of energy, as one could think at first sight. It is an explanation for the fact that a transformer only consumes energy, when the secondary coil draws some current.

The interaction caused by the current in the second wire decreases the effect of the self-induction within the first circuit. This is the same argument, which was used to describe the interaction between geometrically opposed parts of the same circuit. In traditional language one would say that the second current produces a magnetic field, which is opposed to, or out of phase with the original one. The total field is then reduced and therefore also the induced voltage in the first coil, which itself was opposite to the original voltage source.

ACCELERATING CHARGES AND WAVES

ACCELERATING CHARGES, CIRCULAR WAVES AND FORWARD PROPAGATION

As it has been described in the preceding chapter, an accelerating charge will send out circular waves with the following pattern and symmetry.

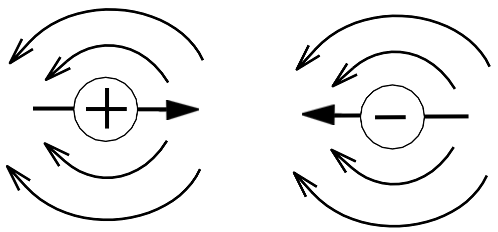


Fig. 3.29.Representation of Circular Waves around an accelerated Charge Carrier

If a voltage is applied to an electric circuit (or a transmission line) the process starts with the acceleration of charges in opposite direction at both ends of the voltage source.

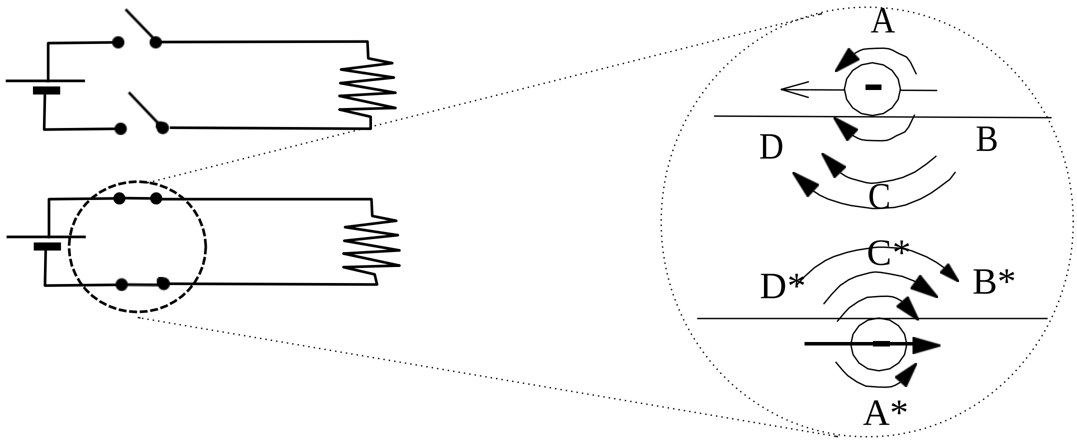


Fig. 3.30. Start of a Wave Propagation

The part of the waves within the conductors (part A and part A*) will be absorbed by other electrons, thus giving rise to the impedance of the system. Part C and C* are absorbed by the charges within the opposite parts of the circuit, thus reducing the impedance of the line. Part B and B* form the front of a wave, travelling along the two conductors to the right and Part D and D* form a wave front, travelling in opposite direction.

Let us first consider the wave front B-B*, which is of main interest for the transmission of energy or information. In visualizing and representing this wave front it should be emphasized that such a wave is not created by few or even a single electron, but that it has macroscopic dimensions and therefore involves a huge amount of electrons.

Huygen's principle can be applied to structure this problem and to give a qualitative description for the observed behaviour. According to this principle the front of a travelling wave at the time t is formed as the sum of circular waves, starting from all those points, which formed a wave front at the time t - Δt

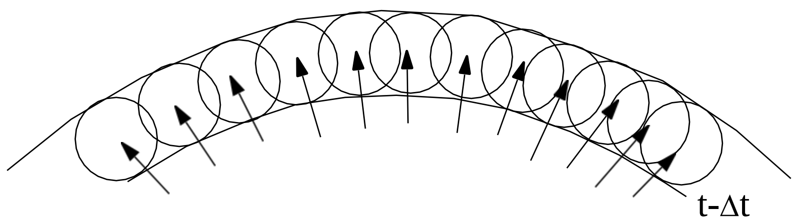


Fig. 3.31.Huygens' Principle

Following this principle the observed wave front in the direction of propagation can be interpreted as caused by the sum of the single circular wave fronts which are sent out by the accelerating electrons.

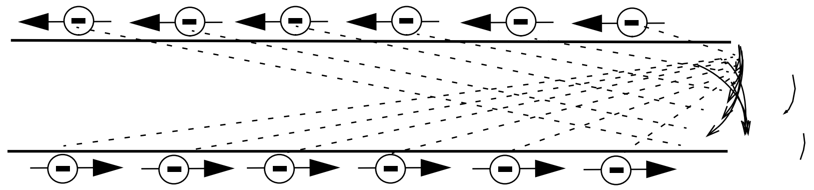


Fig. 3.32. Propagation of a Wave Front Applying Huygen's Principle

For this superposition only the vertical components of the circular wave fronts and only the parts in the direction of the propagation are taken into consideration. The horizontal components are absorbed by the charge carriers in the opposite conductors. The part of the circular wave front travelling backward is discussed later.

In order to create a wave front, it has to be assumed that all the circular waves, caused by the accelerating charge carriers, are send out in phase and that this wave front will accelerate further charge carriers, when it propagates, which then will sent out further circular waves and so forth.

3.8.2. ELECTRIC WAVE PROPAGATION

When the free electrons in metal are accelerated, they will send out circular waves, which will add up to a wave front propagating with the speed of light parallel to the conductors. In the following picture this propagation is represented at three sequential moments in time.

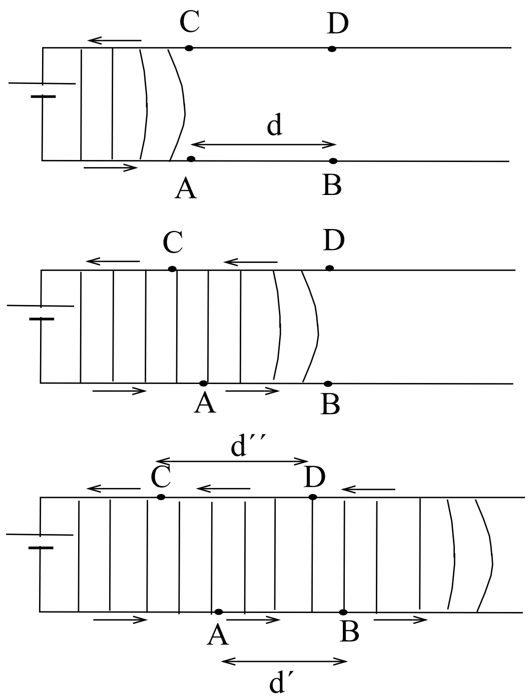


Fig. 3.33.Wave Propagation within a Transmission Line at Three Sequential Moments in Time

The front of the wave is curved so that a (very small) part of the energy is entering the conductor, which is necessary to accelerate the electrons. This acceleration is due to the component of the field in the direction of the propagation. Behind this wave front, the electrons have constant speed and no energy has to enter the conductor, if a loss free conductor is assumed. Lines, indicating the symmetry of the propagation, (the field lines), can be drawn perpendicular to the surface of the conductors.

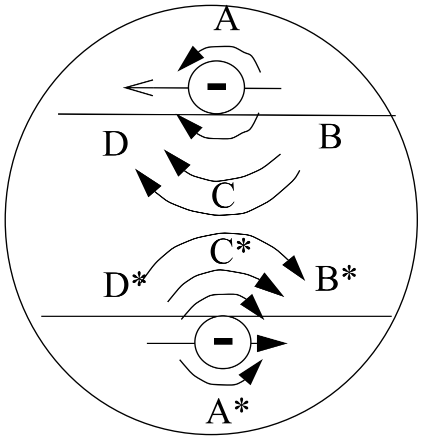
In the upper figure the wave front has not yet reached the electrons A-C , and B-D, which are a distance d apart, corresponding to normal density of free electrons, neutralised by corresponding positive charges of the lattice. When the wave front passes the electrons A and C, electron at A start moving to the right and electron at C to the left.

In a truly loss free conductor the electrons do not need further energy to keep up a drift motion along the wire.

During the time when the wave passes from A-C to B-D, the distance between the electrons A and B decreases to d´ while the distance between C and D increases t d´´. This corresponds to an increase of negative charge on the lower conductor and a decrease of negative charge (an increase in positive charge) on the upper wire. This detailed description explains the cause for the appearance of extra charges on the surface of a conductor, when a circuit is connected to a voltage source.

3.8.3. BACKWARD PROPAGATION AND REFLECTION

The question is still open what happens with the part D - D* of the circular waves that has been sent out in the backward direction.



Looking at a single electron, there should be no difference between the forward and backward direction and the question can be raised, why a wave travels only one way and why there is not something traveling backwards.

It is well known that a wave is reflected backwards when the geometry of the transmission line changes and especially at the end of the line. There will be no reflection only, when this line is accurately terminated with a resistor equal to the specific impedance of the line.

Fig. 3.34.Circular Waves and Wave Propagation

An argument therefore has to be found, why on a parallel line without change in geometry there is no reflection. The answer can be found by pointing to the fact that there is never only one single electron involved and that the sum of a large number of circular waves has to be considered. The mathematics of this problem is governed by Huygens' principle. This principle states that each point of a wave front acts as a point source emitting a spherical wave which travels with a velocity c. The development of this principle into a complete scalar theory was done by Kirchhoff and has been expanded into a vector theory by Sommerfeld. It involves a large amount of calculus and cannot be demonstrated here. Stated in simple words, the result is that all single circular waves in the backward direction are not in phase at a certain point. All their singular waves interfere in a destructive way and add up to zero. But as soon as the geometry changes, the interference is no longer completely destructive and a reflection wave occurs.

3.9. ELECTRIC CIRCUITS AND TRANSMISSION LINES

In the following chapter some arguments are listed why the topic "transmission line "could play a more explicit role within the traditional curriculum about electricity.

3.9.1. TRANSMISSION LINES IN TRADITIONAL TEACHING

In most of the traditional textbooks the phenomena on transmission lines are treated as special solutions of Maxwell's equations. As the name indicates, the transmission of waves of electromagnetic fields, guided by conducting wires or plates is of major interest.

Pictures like the following are typical to illustrate the different solutions of the transmission line equations:

$d^2 V/ dx^2 = U_o L_o d^2 V/ dt^2$

$d^2I dx^2 = U_o L_o d^2 I/ dt^2$

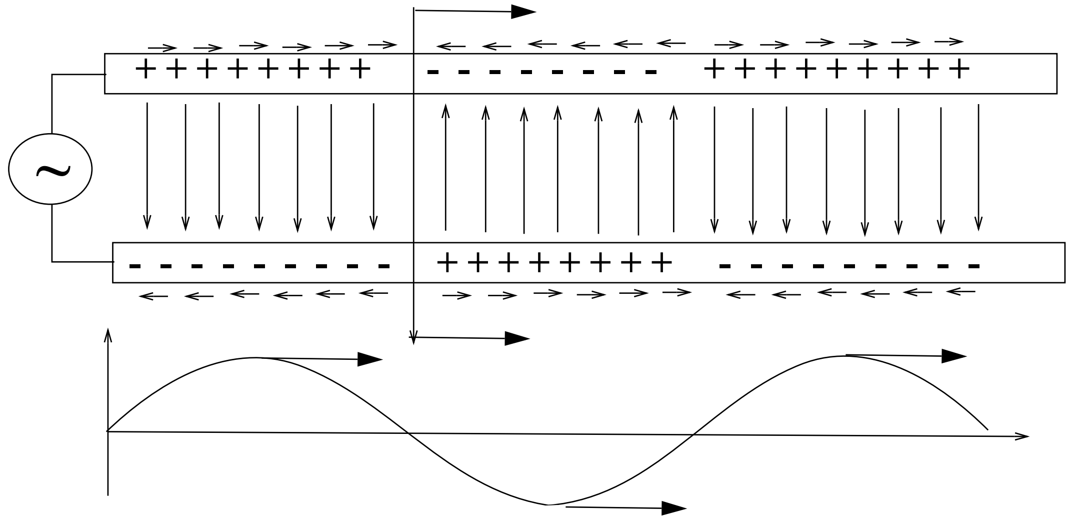


Fig. 3.35. Distribution of Voltage and Current on a Transmission Line

As a rule it can be said that in most textbooks only sinusoidal waves, either propagating or standing, are discussed. Other kinds of waves, for instance the front of a single voltage step is normally not mentioned in an explicit way. With the formalism of Fourier analysis at hand, the propagation of such a step is normally treated as the sum of sine and cosine functions, leaving the sine wave again as the basic primitive for this subject matter.

The subject matter of the electric circuit with stationary states and of the transmission line with propagating and standing waves is traditionally bridged only by the subject matter of ac-currents, containing active elements like impedances and capacitors and the capability to produce oscillations.

The treatment of transmission lines then leads consequently to antenna theory, radiation and wireless propagation of electromagnetic field energy. Maxwell's equations are stated to be the unifying core of this linear chain of more and more developed fields of electromagnetism

- - dc-current
- - ac-current
- - wave propagation along conductors
- - wave propagation in free space

From these equations all these different solutions can be derived. In its most elegant form, including the time as a fourth dimension, this system of equations reduces to two four-dimensional vector equations, which are often regarded as one of the great successes of classical physics.

3.9.2. COMPARISON BETWEEN DC-CURCUITS AND TRANSMISSION LINES

If one compares a simple electric circuit with a properly terminated transmission line, one can see that these two devices are identical expect for some geometrical proportions. These two devices are, however, treated in very different ways.

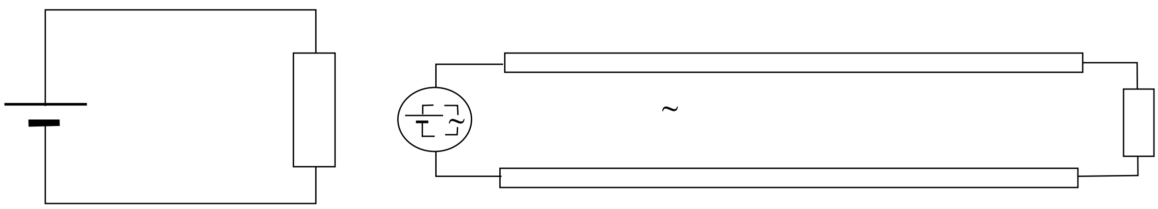


Fig. 3.36. Similarity Between a Circuit and a Transmission Line

In the case of the electric circuit it is common practice to say that the transient state after applying the voltage is too short to be of interest and that therefore only the steady state is treated by applying Ohm's law and Kirchhoff´s laws.

In the case of the transmission line one would explicitly treat the propagation of the voltage step, travelling down with the Speed of Light, after the connection with the voltage source has been made. Reflection of this voltage step is expected at the end of the transmission line, which only does not occur, when the line is properly terminated (when the resistor is equal to the specific impedance of the line). In steady state it is said that a stationary wave is propagating between the two conductors which is then absorbed by the terminating resistor.

There is another difference in the way, dc-currents and transmission lines are normally analysed. In the case of a transmission line it is well known that with higher frequencies the current is pushed to the surface of the conductors, leading to the so-called skin-effect. During the propagation of a voltage step along a transmission line the current therefore is thought to flow only along the surface of the conductors. The question, after what time interval the current will fill the whole cross section and how this process will develop, is hardly ever mentioned.

In the tradition of analysing a dc-current, the transient state is normally overlooked and therefore any kind of skin effect during this transient state cannot come into consideration. The current intensity is therefore assumed to be constant all across the conductor at any time which however is only true for steady state.

3.9.3. COMPARISON BETWEEN A CAPACITOR AND A TRANSMISSION LINE

As mentioned in chapter II, the capacitor and the process of charging and de-charging has been difficult to deal with, concerning the theoretical knowledge of the last century. Maxwell invented the "displacement current" and introduced this term into the law of Biot-Savart in order to avoid an open circuit and therefore manyfold solutions when applying Stoke's theorem. This invention gave rise to the prediction of electromagnetic waves in space which some years later were detected by Hertz.

There have been many disputes about the question, if the term "displacement current" is related to some real physical entity or if it is only a theoretical term, necessary to make a simple mathematical description possible. An interesting debate about this subject matter can be found in Wireless World 1979 to 1980, following a paper from Catt, Dawidson and Walton about "Displacement Current - and how to get rid of it." (Wireless World, Dec 78).

Being mostly concerned with learning and understanding, the important question is, which interpretation or explanation about causal relations is easier to understand for a newcomer and will give more insight. In this respect it can he assumed that the experience of practicing teachers about the displacement current is rather disappointing. This subject matter is hard to grasp for most of the students. The question, what kind of current can be displaced in vacuum and how this current can produce a magnetic field, are difficult to answer, if at all.

There is however a very simple question about the charging of a capacitor, which seems to point to a basic contradiction, implemented in the concept of displacement current. If this statement is true, or can be demonstrated in a broadly acceptable manner, changes in the traditional way of teaching and learning should follow.

The question to be raised was already described in the introduction of this paper an page 10. How does the change of the electric field within the gap between the capacitor plates actually happen?

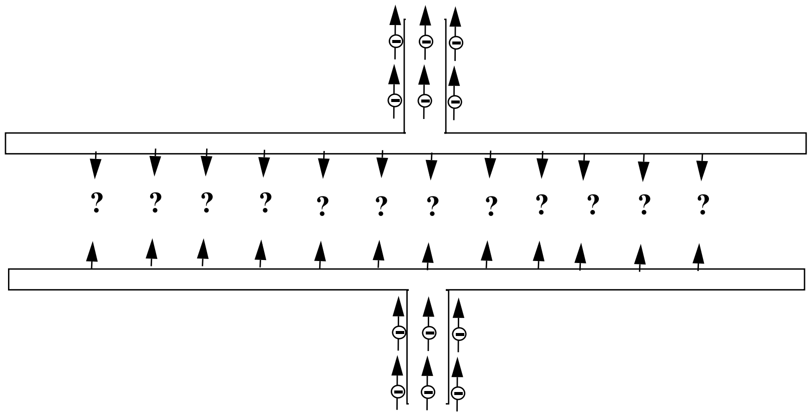


Fig. 3.37. Displacement Current in a Capacitor ? ?

The change of a current, set up by electrons, can in principle be followed point by point, starting at both sides of the voltage source and travelling along the two conductors. What is travelling is assumed to be a region of higher charge density, travelling with the speed of light. Such a construct, however, is not possible for dE/dt, starting either from one side of the capacitor or from both sides and meeting in the middle or at the other side. This is impossible, because this would violate the principle that there never are any open field line in space. A field line always has to end on a charge carrier and cannot start from one point and reach out to the other side. On the other side a change of the displacement current all across the gap of the capacitor at the same time can also be excluded because of the principle that there is no action at a distance.

A surprisingly simple solution to this problem is the idea that a capacitor is in principle nothing else than a transmission line and that during the charging process a wave, starting at the voltage source, enters the capacitor from the side. This idea has been described and discussed by Catt, Davidson and Walton in Wireless World, Dec 1978, p.5 1

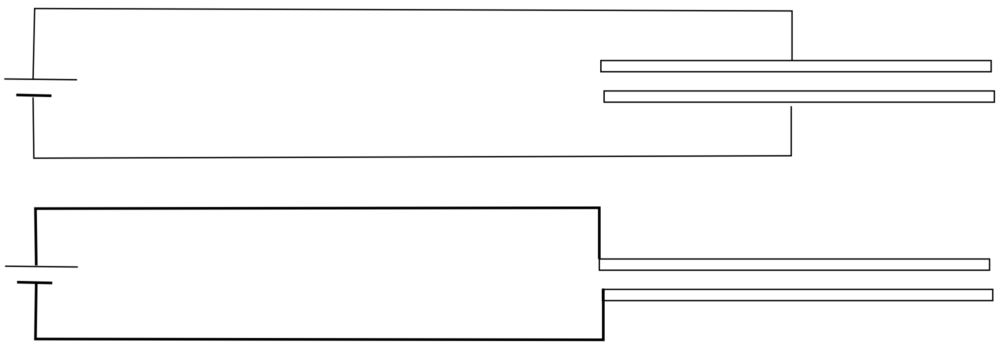


Fig. 3.38. The Capacitor as a Transmission Line

A capacitor, seen as a transmission line, has just a different geometry, but the electric wave enters sideways, after having travelled through the space between the conducting wires. The charging of a capacitor can therefore be seen as the result of a travelling wave, with an increasing part being reflected at the open end of the capacitor..

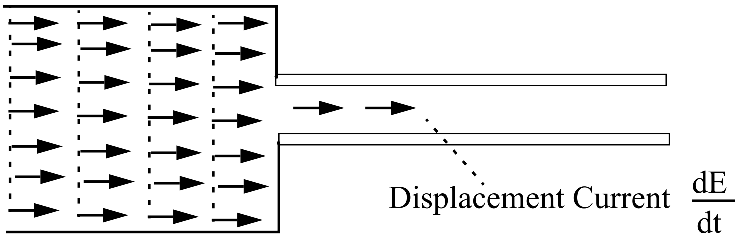


Fig. 3.39. Displacement Current in a Capacitor

3.10. SUMMARY OF BASIC ELEMENTS

In connection with the concept of surface charges, described earlier in chapter II, it is possible to unify the way of analysing electric devices. The point, necessary to be added to this concept, is the fact that with charges on the surface of two conductors, the space between these two conductors will be changed too. Traditionally speaking there will be an electric field between these surface charges. Separation of a neutral state and a change of the space in the surrounding occurs always simultaneously. Both aspects together represent the full picture of nature as we can see it. The separation in charge and field and the concentration on one of these constructs may be convenient but can lead to a one-sided description and can create additional difficulties for learning and understanding.

It has to be stated further that any change of a static state, the creation of separated charges or the reverse, will cause a change in space, some kind of disturbance or distortion, which will travel to the outside with the velocity of light as some kind of a transversal wave. With this concept an electric circuit with either dc- or ac-current, the charging of a capacitor, a processes on a transmission line and wave propagation in space can be qualitatively analysed in the same way, using the same underlying basic ideas or primitives. The basic elements of this concept can be formulated in the following way:

- Electric phenomena occur when a neutral state is changed by separation of opposite charges, where the space between these opposite charges is changed in a specific way.
- The unchanged space around a neutral object can be thought of as either neutral or the sum of two opposite distortions which are exactly opposite, cancelling every measurable effect.
- A single charge carrier can only be created by separation from another charge carrier of opposite sign and by changing the space between these charge carriers in a specific way.

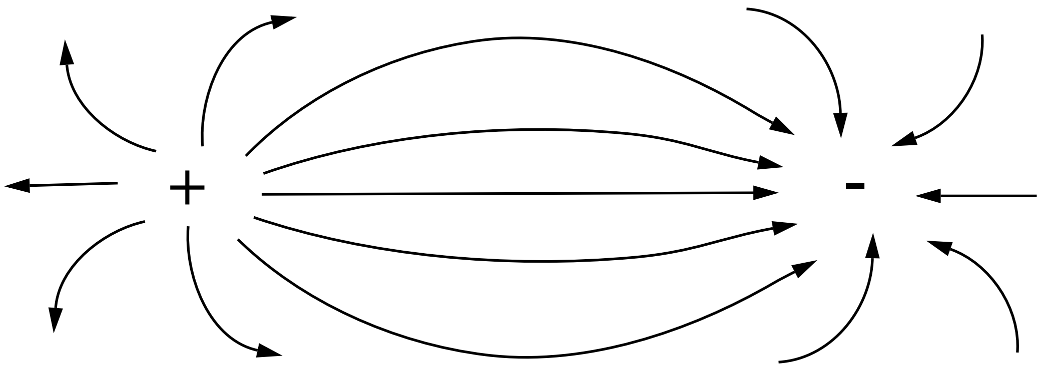


Fig. 3.40. Separation of the Neutral State and Change in Space

- The symmetry of the space around a single charge carrier is spherical when the opposite charge carrier is far away

.

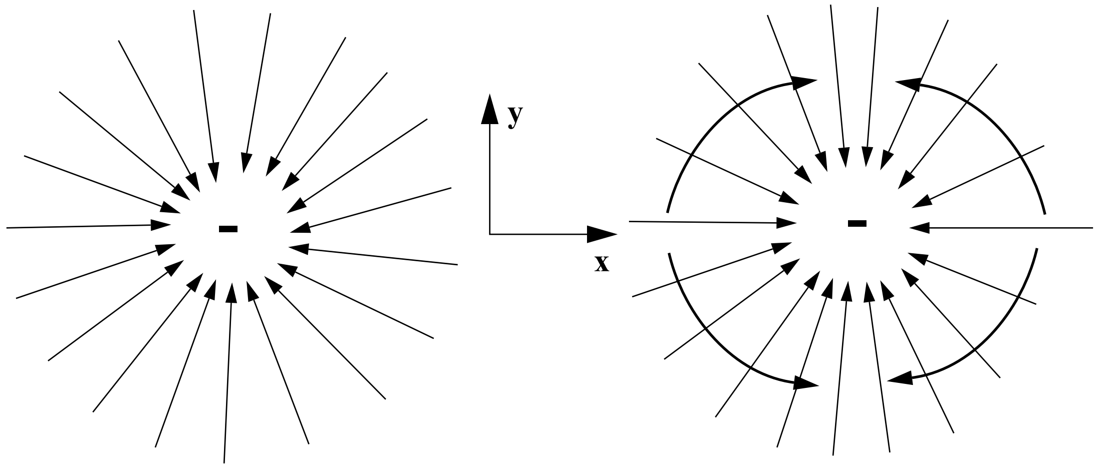


Fig. 3.41. **Isolated Charge Carrier**
at Rest **with Constant Velocity (in x direction)**

This symmetry is no longer spherical, when the charge carrier is moving with constant velocity. It changes to cylindrical symmetry with the axis given by the direction of the velocity in accordance with the theory of special relativity.

Fig. 3.42. Circular Waves Around an Accelerated Charge Carrier

If a charge carrier is accelerated, circular waves are radiated, which move to the outside with the speed of light.

When charge carriers are accelerated within a circuit by a voltage source, surface charges are created along the two wires and the front is propagating with the speed of light. At the same time the space between the wires is changing.

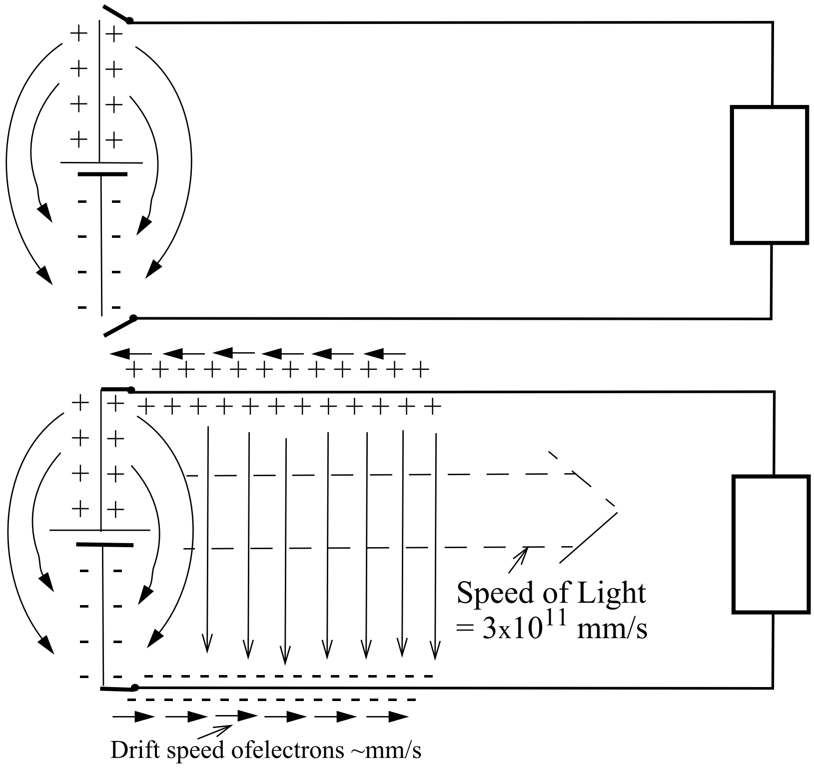


Fig. 3.43. Development of Surface Charges and Fields in a Circuit During a Transient State

Reflection occurs at the "meeting point" of the surface charges (or at an open end) due to deceleration of charge carriers. The reflected wave travels back with the speed of light and establishes equilibrium or steady state.

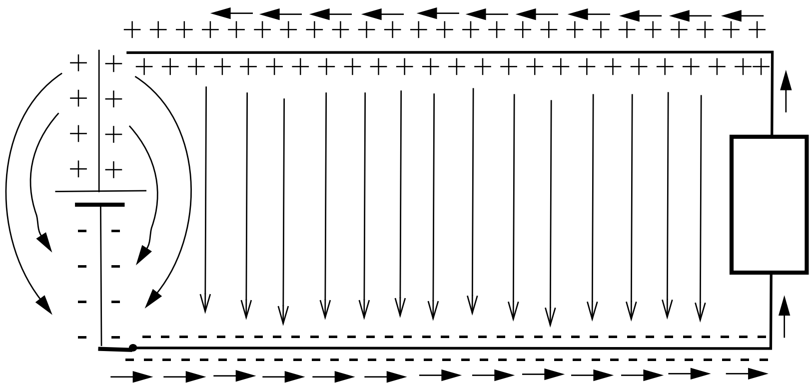


Fig. 3.44. Representation of Steady State in a Simple Electric Circuit

In steady state there is a drift of charge carriers in a circle and a transfer of energy from the battery to the resistor.

For further details about the development of waves, about superposition and reflection on transmission lines see:

http://www.astrophysik.uni-kiel.de/~hhaertel/TL/TL-tutorial/index.htm

BIBLIOGRAPHY

Page, L.,Connecticut Academy of Arts and Science, Vol.26, The Emission Theory of Electromagnetism, pages 213-243

Catt, Davidson and Walton, Wireless World, Displacement Current, Dec 1978, p.51

Coombes,C.A. and Laue, H., Electric Fields and Charge Distribution associated with steady Currents; Amer. J.Phys. 49, 1981, pages 450-451

Feynman, Leighton, Sand, Lectures on Physics, Addison-Westley Pub.1963-65

Haertel, H., IPN Curriculum Physik, UE "Stromstaerke, Spannung, Widerstand. Der elektrische Stromkreis als System", Klett, Stuttgart, 1981

Haertel, H., The Electric Voltage in: Duit, Jung, Rhoeneck (Ed.), Understanding Electricity, IPN-Arbeitsbericht, Schmidt und Klaunig, Keil, 1984

Jefimenko, 0., Amer.J.Phys.30, 1962, pages 19-21

Purcell, E.M., Berkeley Physics Course, Mcgraw-Hil1,1963

Rosser ,W.G.V.,Classical Electromagnetism via Relativity, Butterworth, London, 1968

Walz, A., Elektrische Felder um Stationäre Ströme, PU 2, 1984,pages 61-68