



Practical Industrial Electronics

An Overview:

- Basic Concepts
- Discrete Components



Technology Training that Works

Practical Industrial Electronics – An Overview

THIS BOOK WAS DEVELOPED BY IDC TECHNOLOGIES

WHO ARE WE?

IDC Technologies is internationally acknowledged as the premier provider of practical, technical training for engineers and technicians.

We specialize in the fields of electrical systems, industrial data communications, telecommunications, automation and control, mechanical engineering, chemical and civil engineering, and are continually adding to our portfolio of over 60 different workshops. Our instructors are highly respected in their fields of expertise and in the last ten years have trained over 200,000 engineers, scientists and technicians.

With offices conveniently located worldwide, IDC Technologies has an enthusiastic team of professional engineers, technicians and support staff who are committed to providing the highest level of training and consultancy.

TECHNICAL WORKSHOPS TRAINING THAT WORKS

We deliver engineering and technology training that will maximize your business goals. In today's competitive environment, you require training that will help you and your organization to achieve its goals and produce a large return on investment. With our 'training that works' objective you and your organization will:

- Get job-related skills that you need to achieve your business goals
- Improve the operation and design of your equipment and plant
- Improve your troubleshooting abilities
- Sharpen your competitive edge
- Boost morale and retain valuable staff
- Save time and money

EXPERT INSTRUCTORS

We search the world for good quality instructors who have three outstanding attributes:

1. Expert knowledge and experience – of the course topic
2. Superb training abilities – to ensure the know-how is transferred effectively and quickly to you in a practical, hands-on way
3. Listening skills – they listen carefully to the needs of the participants and want to ensure that you benefit from the experience.

Each and every instructor is evaluated by the delegates and we assess the presentation after every class to ensure that the instructor stays on track in presenting outstanding courses.

HANDS-ON APPROACH TO TRAINING

All IDC Technologies workshops include practical, hands-on sessions where the delegates are given the opportunity to apply in practice the theory they have learnt.

REFERENCE MATERIALS

A fully illustrated workshop book with hundreds of pages of tables, charts, figures and handy hints, plus considerable reference material is provided FREE of charge to each delegate.

ACCREDITATION AND CONTINUING EDUCATION

Satisfactory completion of all IDC workshops satisfies the requirements of the International Association for Continuing Education and Training for the award of 1.4 Continuing Education Units.

IDC workshops also satisfy criteria for Continuing Professional Development according to the requirements of the Institution of Electrical Engineers and Institution of Measurement and Control in the UK, Institution of Engineers in Australia, Institution of Engineers New Zealand, and others.

CERTIFICATE OF ATTENDANCE

Each delegate receives a Certificate of Attendance documenting their experience.

100% MONEY BACK GUARANTEE

IDC Technologies' engineers have put considerable time and experience into ensuring that you gain maximum value from each workshop. If by lunchtime on the first day you decide that the workshop is not appropriate for your requirements, please let us know so that we can arrange a 100% refund of your fee.

ONSITE WORKSHOPS

All IDC Technologies Training Workshops are available on an on-site basis, presented at the venue of your choice, saving delegates travel time and expenses, thus providing your company with even greater savings.

OFFICE LOCATIONS

AUSTRALIA • CANADA • INDIA • IRELAND • MALAYSIA • NEW ZEALAND • POLAND •
SINGAPORE • SOUTH AFRICA • UNITED KINGDOM • UNITED STATES

idc@idc-online.com

www.idc-online.com

Visit our website for **FREE** Pocket Guides

**IDC Technologies produce a set of 6 Pocket Guides used by
thousands of engineers and technicians worldwide.**

Vol. 1 – **ELECTRONICS**

Vol. 4 – **INSTRUMENTATION**

Vol. 2 – **ELECTRICAL**

Vol. 5 – **FORMULAE & CONVERSIONS**

Vol. 3 – **COMMUNICATIONS**

Vol. 6 – **INDUSTRIAL AUTOMATION**

To download a **FREE** copy of these internationally best selling pocket guides go to:
www.idc-online.com/downloads/

On-Site Training

SAVE MORE
THAN 50% OFF
the per person
cost



CUSTOMISE the
training to YOUR
WORKPLACE!



Have the training
delivered WHEN
AND WHERE you
need it!

All IDC Technologies Training Workshops are available on an on-site basis, presented at the venue of your choice, saving delegates travel time and expenses, thus providing your company with even greater savings.

For more information or a **FREE** detailed proposal contact Kevin Baker by e-mailing:
training@idc-online.com

IDC TECHNOLOGIES

Worldwide Offices

AUSTRALIA

Telephone: 1300 138 522 • Facsimile: 1300 138 533

West Coast Office

1031 Wellington Street, West Perth, WA 6005
PO Box 1093, West Perth, WA 6872

East Coast Office

PO Box 1750, North Sydney, NSW 2059

CANADA

Toll Free Telephone: 1800 324 4244 • Toll Free Facsimile: 1800 434 4045
Suite 402, 814 Richards Street, Vancouver, BC V6B 3A7

INDIA

Telephone : +91 444 208 9353
35 4th Street, Kumaran Colony, Vadapalani, Chennai 600026

IRELAND

Telephone : +353 1 473 3190 • Facsimile: +353 1 473 3191
PO Box 8069, Shankill Co Dublin

MALAYSIA

Telephone: +60 3 5192 3800 • Facsimile: +60 3 5192 3801
26 Jalan Kota Raja E27/E, Hicom Town Center
Seksyen 27, 40400 Shah Alam, Selangor

NEW ZEALAND

Telephone: +64 9 263 4759 • Facsimile: +64 9 262 2304
Parkview Towers, 28 Davies Avenue, Manukau City
PO Box 76-142, Manukau City

POLAND

Telephone: +48 12 6304 746 • Facsimile: +48 12 6304 750
ul. Krakowska 50, 30-083 Balice, Krakow

SINGAPORE

Telephone: +65 6224 6298 • Facsimile: + 65 6224 7922
100 Eu Tong Sen Street, #04-11 Pearl's Centre, Singapore 059812

SOUTH AFRICA

Telephone: +27 87 751 4294 or +27 79 629 5706 • Facsimile: +27 11 312 2150
68 Pretorius Street, President Park, Midrand
PO Box 389, Halfway House 1685

UNITED KINGDOM

Telephone: +44 20 8335 4014 • Facsimile: +44 20 8335 4120
Suite 18, Fitzroy House, Lynwood Drive, Worcester Park, Surrey KT4 7AT

UNITED STATES

Toll Free Telephone: 1800 324 4244 • Toll Free Facsimile: 1800 434 4045
7101 Highway 71 West #200, Austin TX 78735

Website: www.idc-online.com
Email: idc@idc-online.com



Technology Training that Works

Presents

Industrial Electronics- An Overview

Lubo Dimitrov

Revision 1.3

Website: www.idc-online.com

E-mail: idc@idc-online.com

IDC Technologies Pty Ltd
PO Box 1093, West Perth, Western Australia 6872
Offices in Australia, New Zealand, Singapore, United Kingdom, Ireland, Malaysia, Poland, United States of America, Canada, South Africa and India

Copyright © IDC Technologies 1999. All rights reserved.

First published 1999

All rights to this publication, associated software and workshop are reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher. All enquiries should be made to the publisher at the address above.

Disclaimer

Whilst all reasonable care has been taken to ensure that the descriptions, opinions, programs, listings, software and diagrams are accurate and workable, IDC Technologies do not accept any legal responsibility or liability to any person, organization or other entity for any direct loss, consequential loss or damage, however caused, that may be suffered as a result of the use of this publication or the associated workshop and software.

In case of any uncertainty, we recommend that you contact IDC Technologies for clarification or assistance.

Trademarks

All logos and trademarks belong to, and are copyrighted to, their companies respectively.

Acknowledgements

IDC Technologies expresses its sincere thanks to all those engineers and technicians on our training workshops who freely made available their expertise in preparing this manual.

Contents

1	Basic Concepts	1
1.1	Atomic Structure	1
1.2	Conductors, Insulators and Semiconductors	4
1.3	Current, Voltage, Resistance, Power	10
1.4	Direct and Alternating Current	15
1.5	Units and Abbreviations	25
1.6	Quiz	26
2	Discrete Components	27
2.1	Resistors, Inductors and Capacitors	27
2.2	Transformers, Diodes and Bridge Rectifiers	49
2.3	Bipolar Junction Transistors, JFETs and MOSFETs	59
2.4	SCRs, DIACs, TRIACs and LEDs	72
2.5	Quiz questions	79

1

Basic Concepts

1.1 Atomic Structure

When you have completed study of this section you should be able to:

- Describe atomic structure
- Define nucleus, electron, proton and neutron
- Define atomic number
- Define energy levels (shells)
- Identify valence electrons
- Describe ionization

1.1.1 Atoms

The smallest particle of an element that retains the characteristics of that element is called an **atom**. Each element has a different atomic structure. The most common model of atomic structure is the classical Bohr model of an atom. According to it, the atoms have a planetary type of structure. This is illustrated in Figure 1.1. The **nucleus** consists of uncharged particles called **neutrons** and positively charged particles called **protons**. Negatively charged particles called **electrons** orbit around the nucleus.

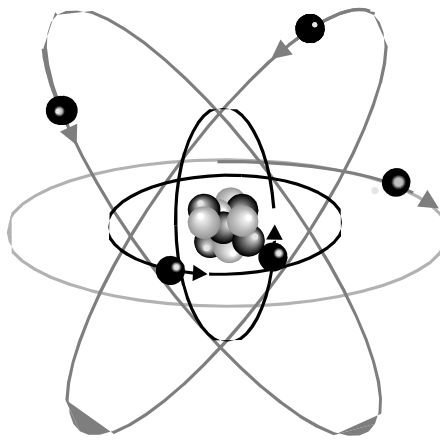


Figure 1.1
The Bohr Model of an Atom

Each of the known 109 elements has a different atomic structure. Every atom has a certain number of electrons and protons that distinguish it from the atoms of other elements. For example, the simplest atom is that of hydrogen. It consists from one electron and one proton, as shown in Figure 1.2 (a). As another example, the nucleus of the helium atom, shown in Figure 1.2 (b) consists from two protons and two neutrons. Two electrons orbit around the nucleus.

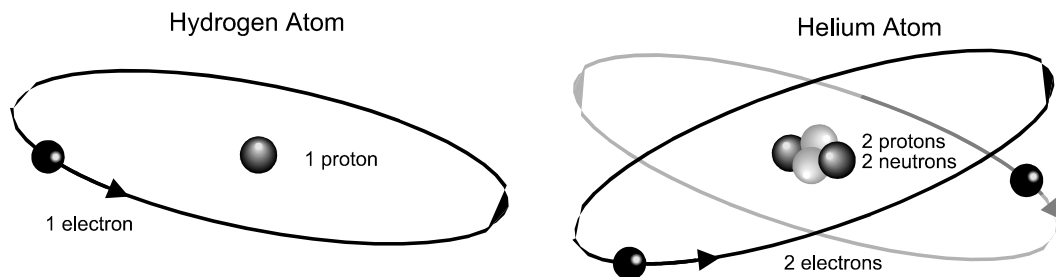


Figure 1.2
The Two Simplest Atoms

1.1.2 Atomic Number

According to their **atomic number**, all elements are arranged in a periodic table. The atomic number equals the number of protons in the nucleus. In an electrically balanced (or neutral) atom, the number of electrons equals the number of protons. For example, hydrogen has an atomic number of 1, and helium has an atomic number of 2. When the atoms of a certain element are in their normal (or neutral) state, they have the same number of electrons and protons. The negative charges cancel the positive charges, and the atom has a net charge of zero.

1.1.3 Electron Shells

Electrons orbit around the nucleus of an atom at distances that are precisely defined and grouped into energy bands called **shells**. Only distinct values of electron energies exist within the atomic structures. Therefore, the electrons orbit only at distinct (discrete) distances from the nucleus. Each one of these discrete distances (orbits) from the nucleus corresponds to a certain energy level. Electrons near the nucleus have less energy than those in most distant orbits. A given atom has a fixed

number of shells. The shells are designated K, L, M, N and so on with K being closest to the nucleus. The energy band concept is illustrated in Figure 1.3.

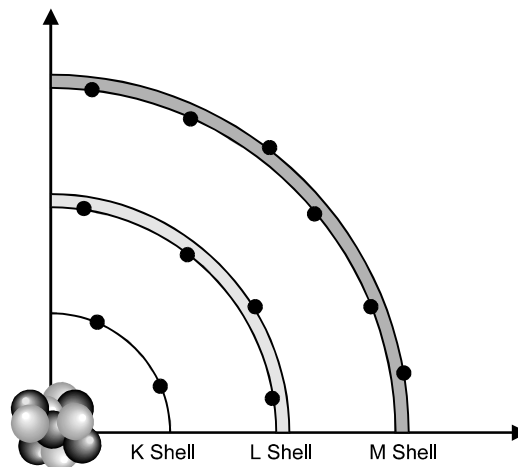


Figure 1.3
Electron Shells

The maximum number of electrons in each shell is fixed and can be derived from the formula:

$$N_e = 2 n^2$$

where:

n is the number of the shell.

Therefore the maximum number of electrons that can exist in the first innermost shell (K) is 2, the second (L) - 8, the third (M) - 18 and so on.

1.1.4 Valence

The force of attraction between the electrons and nucleus decreases when the distance between them increases. Therefore, the electrons in the outermost shell have the highest energy levels and are relatively loosely bound to the atom.

The outermost shell or energy level that contains electrons is known as the **valence level** of that atom. The number of electrons in the outer level is known as the **valence number** of that atom and the electrons in this level are called **valence electrons**. If an electron is dislodged from an atom, the balance of the atom is upset. With more protons than electrons, the atom becomes positively charged. The energy required to free an electron may be obtained from an external source of heat or voltage.

1.1.5 Ionization

The relatively loosely bound to the atom valence electrons can easily jump to the higher energy levels, even when a minimal amount of external energy is absorbed.

When a valence electron absorbs a sufficient amount of energy and dislodges from the atom, it becomes a **free electron**.

The process of **losing** a valence electron is called **ionization**, and the atom with one or more disassociated electrons is called a **positive ion**. At the same time, the opposite process may occur. Free electrons may join positive ions, which turns them back into neutral atoms. This process is called **recombination**.

For example, a neutral hydrogen atom, which consists of one positively charged proton and one negatively charged electron is designated with the letter **H**. If this atom loses its valence electron, it becomes a positive ion and is designated as **H⁺**. On the other hand, if the disassociated free electron joins the outer shell of another hydrogen atom, the atom becomes negatively charged and it is called a negative ion, designated **H⁻**.

1.1.6 Summary

The atom is the smallest item of matter. All substances are made up of atoms. The atom consists of a nucleus, with electrons orbiting around it. The nucleus consists of protons and neutrons. Protons have a positive charge, electrons have a negative charge and neutrons have no charge. A balanced atom contains an equal number of protons and electrons. The number of protons in the nucleus equals the atomic number of the element. Electrons orbit the nucleus in defined energy levels or shells. The outermost energy level of an atom is called valence shell. Electrons in the valence shell are called valence electrons. The process of dislodging valence electrons is called ionization.

1.2 Conductors, Insulators and Semiconductors

When you have completed study of this section you should be able to:

- Describe the atomic structure of conductors, insulators and semiconductors
- List some of them
- Discuss the difference between them
- Explain how covalent bonds are formed in a crystal structure
- Define doping
- Explain what n-type and p-type semiconductors are

1.2.1 Conductors

To simplify the theory, it can be assumed that the **electric current** is made up of many free electrons, moving in the same direction. According to the availability of free electrons, all materials can be divided into conductors and insulators.

Conductors are materials that easily conduct electrical current. Materials, consisting of atoms with a low valence number (1,2 and 3) make very good conductors of electricity. Because of their low valence, very little energy is required to dislodge the outer electrons from their orbits.

The best conductors are characterized by atoms with one valence electron only, such as copper (Cu), silver (Ag) and aluminum (Al). Other metals, such as gold (Au) and platinum (Pt) may have even better conducting properties, but due to their high cost are used only in some specific applications. Other conductors used in electronics are brass, tin-foil and solder.

Most metals have their atoms arranged in a **crystal structure**. This is made up of positively charged ions that are oscillating from specified stationary positions. Many free electrons are moving randomly between the ions and are constantly colliding with them. The electrons also repel from each other. This concept is illustrated in Figure 1.4.

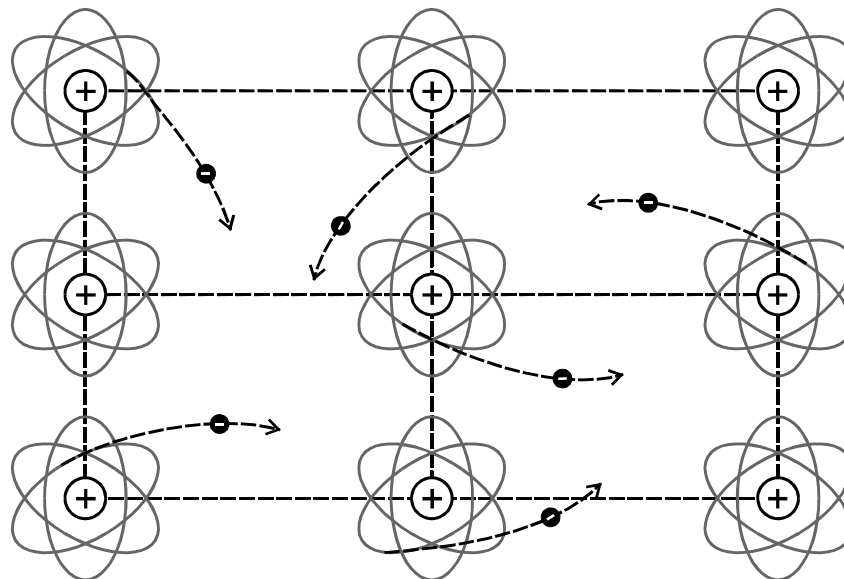


Figure1.4
Conductor Structure

1.2.2 Insulators

An **insulator** is a material that under normal conditions does not conduct electrical current. Atoms of insulating materials consist of at least 5 valence electrons. Valence electrons are tightly bound to the atoms and they have to absorb a considerable amount of energy in order to leave their orbits. The more valence electrons in an atom, the more energy is needed to dislodge them.

Most good insulators do not consist of single-elements. Usually they are compounds, and contain only a very limited number of free electrons. However, if an extremely high voltage is applied to the insulator, it will be forced into the conductive state, and will then break down and fail. This breakdown causes many

of the component failures. Figure 1.5 shows energy diagrams for insulators, semiconductors and conductors, which helps understand this process.

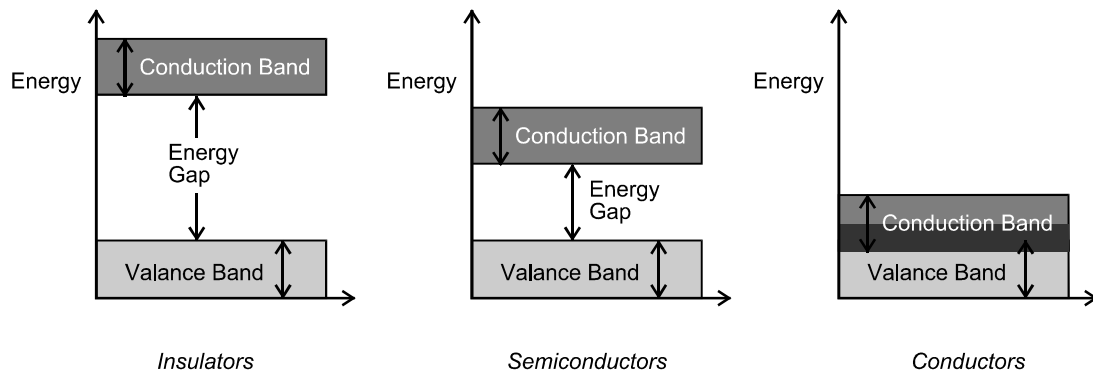


Figure 1.5
Energy Diagrams for Insulators, Semiconductors and Conductors

Remember that each electron is confined to a certain energy band or shell (see Section 1.1.3). For example, when a valence electron acquires sufficient additional energy from an external source to overcome the energy gap, it can jump from the valence band to the conduction band, becoming a free electron. It is important to remember that an electron can jump from one band to another, for example from K to L or from L to M, but it cannot orbit at an energy level, situated between two distinct bands.

The energy required for the electron to leave the band (the shell), where it is orbiting, and to jump into the next one is called **energy gap**. Insulators are characterized by very wide energy gaps. For this reason, valence electrons jump into the conductor band only when extremely high voltages are applied across the material, in which case a breakdown occurs. The wider the energy gap, the better quality of the insulator. Some commonly used insulators in electronics are plastic, ceramics, mica, wood, neon, argon and glass.

In contrast, conductors have a very narrow energy gap and their valence band and conduction band even overlap. This means that the valence electrons do not need any external energy to become free electrons. In fact, even under normal conditions most of their electrons are disassociated from their atoms and *are* free electrons.

1.2.3 Semiconductors

The semiconductors are materials, with conductive properties somewhere between conductors and insulators. Strictly speaking, they are neither good conductors, nor good insulators. Their valence number is always 4. The most popular semiconductors are silicon and germanium. Gallium arsenide is also common.

Both silicon and germanium atoms have four valence electrons. However the silicon atom has 14 electrons, while the germanium atom has 32. Therefore, the germanium valence electrons are at higher energy levels, which makes them unstable at high

temperatures. For this reason, silicon is the most widely used material in diodes, transistors, integrated circuits and other semiconductor devices.

The atoms of semiconductors are arranged in fixed pattern called a **crystal**. Each atom in this structure shares its four valence electrons with another four adjacent atoms. This holds the atoms together, and forms the basis of the **crystal structure**. When two atoms share common valence electrons they form a **covalent bond** between them. The semiconductors have four valence electrons. Therefore, each atom forms four covalent bonds with the neighboring atoms. Each neighboring atom is in turn bonded with another four atoms and forms another four covalent bonds and so on. This is illustrated in Figure 1.6.

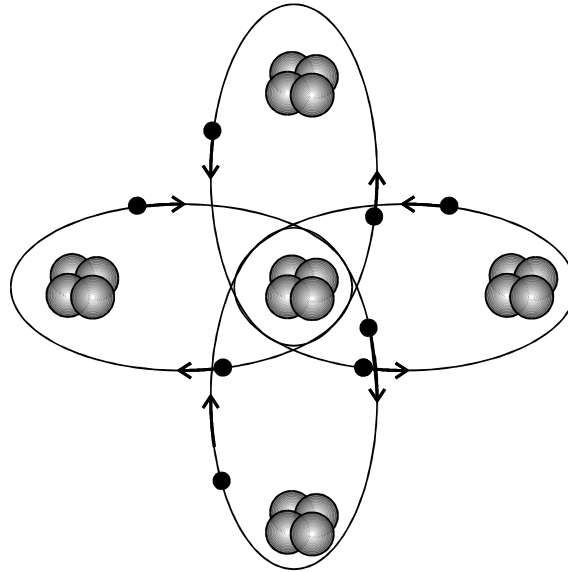


Figure 1.6
Bonding Diagram

Normally 14 electrons orbit around the silicon nucleus, and only 4 of these are in the valence shell. The electrons in the inner shells are tightly bonded to the atom. These electrons do not play any part in the electrical conductivity of the crystal. For this reason, only four valence electrons and four protons are shown in the figures.

A crystal that contains no impurities is called an **intrinsic crystal**. The covalent bonding in an intrinsic silicon crystal at 0°K (absolute zero) is shown in Figure 1.7. At this temperature all valence electrons are in the valence shell. In this case, even if a voltage is applied to the semiconductor, there is no current, as there are no free electrons. Therefore, semiconductors behave as ideal insulators at 0°K.

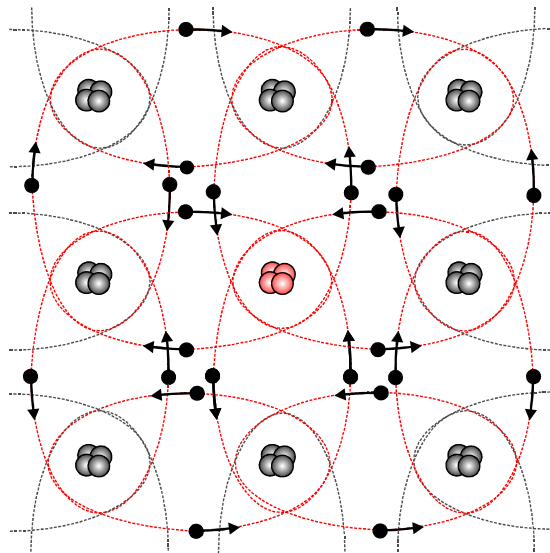


Figure 1.7
Covalent Bonding in an Intrinsic Silicon Crystal at 0 K

However, at room temperature some of the electrons will overcome the energy gap and will jump into the conduction band, becoming free electrons. They move randomly in the crystal structure. If a voltage is applied, they are attracted to the positive end, thus creating an **electron current**.

At the same time, when an electron is freed from the valence band, it leaves a vacancy. This vacancy is called a **hole** and has a positive charge. An intrinsic crystal has an equal number of free electrons and holes. A valence electron may jump from its atom to fill a hole in an adjacent atom with little change in its energy level. In this case, the electron will leave a hole in the atom, where it was previously situated. Effectively the hole is moving from one location to another and this is called a **hole current**. Figure 1.8 illustrates this process.

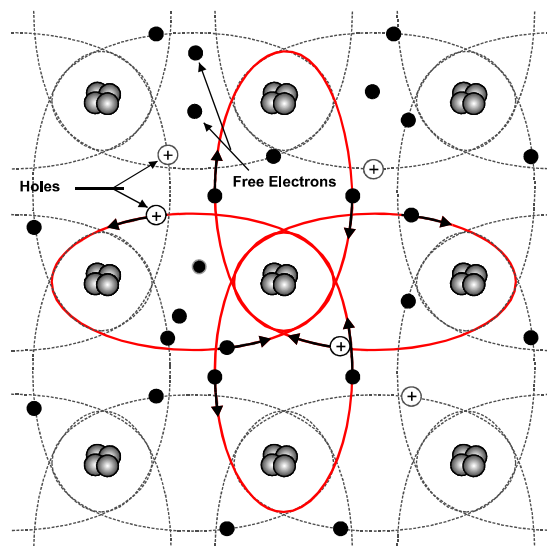


Figure 1.8
Intrinsic Silicon Crystal at Room Temperature

The conductivity of semiconductors can be drastically increased through doping. **Doping** is the process of controlled addition of impurities to the intrinsic

semiconductive crystal. The conductivity of semiconductors could be increased hundreds and thousands of times even with minimal impurities added.

The impurity atoms could be **pentavalent** or **trivalent**. For example, if an atom with five valence electrons (a pentavalent atom) is added to intrinsic silicon, four of its valence electrons will form covalent bonds with the adjacent silicon atoms and one electron will remain free. Thus, the doping process creates an additional free electron, without leaving a hole. The material produced in this way is called **n-type** material, as most of its carriers are electrons. This is illustrated in Figure 1.9. Most commonly used pentavalent impurity atoms are those of arsenic (As), bismuth (Bi) and phosphorus (P).

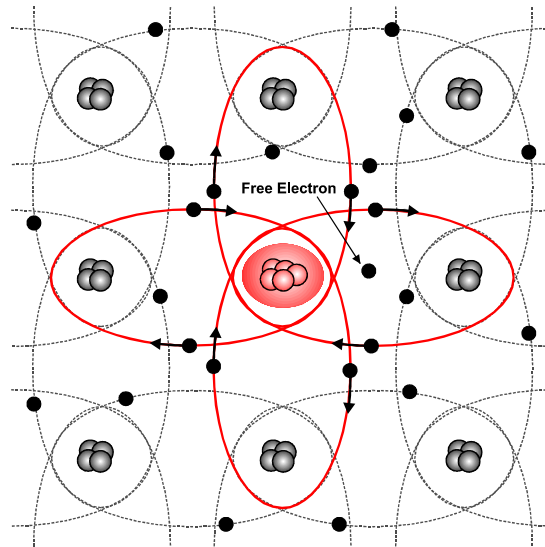


Figure 1.9
N-type Semiconductor

Similarly, if an atom with three valence electrons is added to intrinsic silicon, all of them will form covalent bonds with the adjacent silicon atoms. Since four electrons are required, a hole will appear in the semiconductor crystal. Thus, the doping process will create an additional hole, without leaving a free electron. Such a material is called **p-type**, as most of its carriers are holes. To create p-type semiconductors typically aluminum (Al), indium (In), boron (B) and gallium (Ga) are used. A p-type semiconductor is shown in Figure. 1.10.

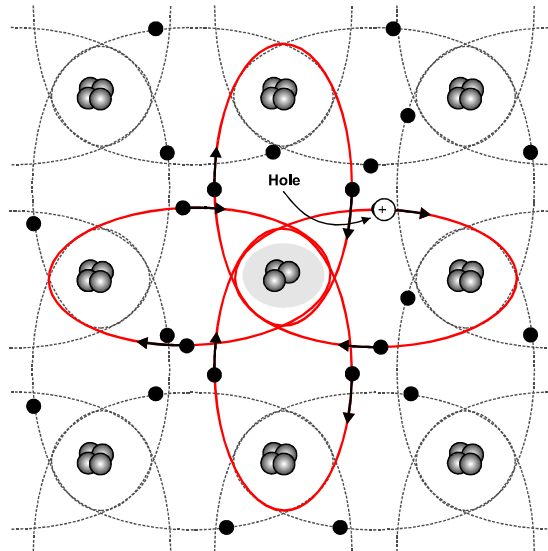


Figure 1.10
P-type Semiconductor

1.2.4 Summary

There are three types of materials: conductors, semiconductors and insulators. Materials with 1,2 or 3 valence electrons are conductors. Materials with 4 valence electrons are semiconductors. Materials with more than 5 valence electrons are insulators. The atoms of the semiconductors are arranged in a crystal structure. When two adjacent atoms share a common pair of valence electrons, they form a covalent bond. Doping is the process of controlled addition of impurities to intrinsic semiconductive crystals. The addition of pentavalent impurity atoms produces n-type semiconductors. The addition of trivalent impurity atoms produces p-type semiconductors.

1.3 Current, Voltage, Resistance, Power

When you have completed study of this section you should be able to:

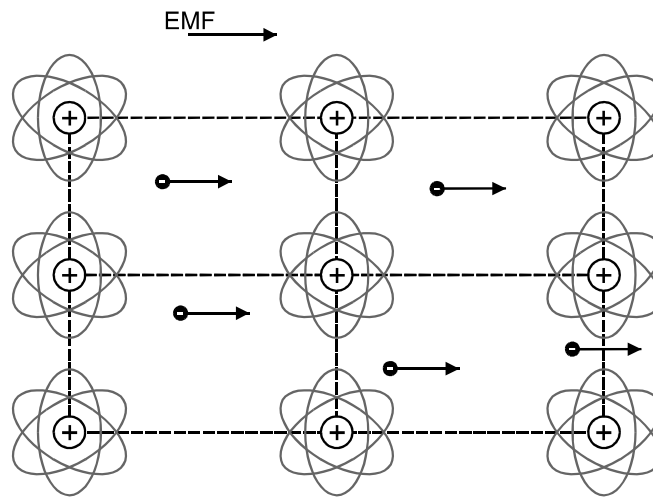
- Explain the terms current, voltage, resistance and power
- Define the relationships between them
- List conditions that are necessary for continuous current flow

1.3.1 Electromotive Force and Current Flow

The structure of conductors resembles a net of positively charged ions, linked together as shown in Figure 1.11. A large number of free electrons move randomly between the ions. In their movement they are constantly colliding with each other and with the ions. The speed of electrons and the number of collisions are proportional to the ambient temperature.

Current only flows if a voltage is applied to the conductor. The voltage source is referred to as the **electromotive force (emf)**. This word comes from the combination of words “electron” and “motion”. This force makes the electrons move in a certain direction. Electromotive force is created between two points, when one of them has more positively charged particles, and the other has more negatively charged particles.

As you can see in Figure 1.11 when the electromotive force is applied to the conductor, all electrons start moving in the same direction following the direction of the emf. At this stage, they still collide with each other, but their energy is also passed as impulses in the same direction in which they are moving. The energy is distributed at a very high speed, close to that of the speed of light. Thus, the **current flow** in the conductor is a complex process, involving electrons moving in one direction and simultaneously transferring their energy through impulses in the same direction.



*Figure 1.11
Conductor with Current Flow*

When electricity was initially discovered, it was decided that the direction of the current flow is from the positive terminal to the negative terminal of the voltage source. This belief was wide spread for many years, until the electron was discovered. Then it was realized, that all theories were wrong, and that the electrons actually flow from the negative terminal to the positive terminal of the voltage source. It was then too late to change all books, which have been written at that time. Moreover, the actual direction of the current is not so important, as long as we accept that the current *has* a direction and we are *consistent* in using one or the other direction.

The current flowing from the positive to the negative terminal is referred to as a **conventional current flow**. At the same time we have to keep in mind that actually electrons flow in the opposite direction, forming an **electrons current flow**.

1.3.2 Current, Voltage and Power Relationships

As was discussed in an earlier section, all electrons can start moving in a certain direction only if there is a potential difference, applied between the two ends of a conductor. In other words the electrical current is the effect, caused from an external source of emf. The emf source converts one type of energy into another, namely its internal energy into that of electrical. This creates at its terminals a potential difference. Because electrons are negatively charged particles, they are attracted to the positively charged terminal of the external emf source. This brings them into motion, orientated towards the positive terminal. The current then starts flowing.

However, the current will flow only if the emf expends some energy. When energy is expended, work is done. The following formula describes this relationship:

$$V = \frac{W}{Q}$$

where:

- V is the emf in volts (V),
- W is the work done in joules (J),
- Q is the charge in coulombs (C).

In other words, the electromotive force will be equal to the work done by the external voltage source in order to transfer a certain quantity of charge via the conductor. Most countries in the world use the **international system of units (SI)**. In this manual only SI units are used. According to SI, if an external voltage source of 1 volt emf, transfers 1 coulomb of charge, 1 joule of work has been done.

As was mentioned earlier, the electron current flows from the negative side of the voltage source, through the circuit and back to the positive side of the voltage source. This means that directions of the emf inside the voltage source and actual electron flow will be the same. The conventional current flow, however, has an opposite direction. Unless it is specifically written otherwise, it is a common practice to analyze electronic circuits, with regard the *conventional* current flow. This manual is no exception. Figure 1.12 depicts a simple electrical circuit. The direction of the conventional current flow is shown with an arrow.

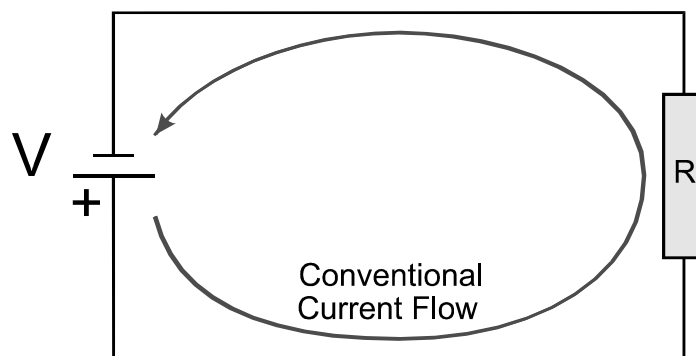


Figure 1.12
Simple Electrical Circuit

Electric current can flow continuously in the circuit if two conditions are satisfied:

- **External source of emf must be available.**

This source creates on its terminals a differential potential. In order to do this it has to transform some other type of energy into electrical. This other type of energy can be a mechanical rotation, a chemical reaction, a thermal energy or a source of light. For example batteries use a chemical reaction, based on the process called *electrolytic dissociation* to produce an emf.

- **The electrical circuit must be closed.**

This means that the electrical current that starts flowing from one of the terminals of the voltage source and goes all the way round the continuous circuit must come back to its other terminal. As it is shown in Figure 1.13, if for some reason the circuit is broken at any given point along its way, the current flow will cease.

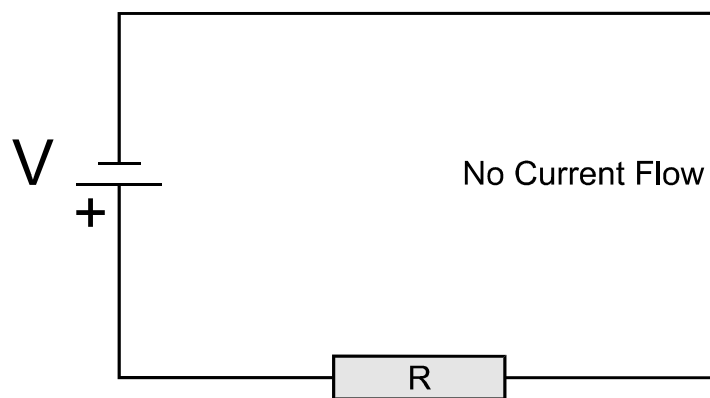


Figure 1.13
Open Electrical Circuit

It becomes clear that the two basic components necessary to maintain continuous current flow are an external source of emf and a closed circuit. However, in order to make use of the electrical energy, an additional component is included in the circuit, and this is the **load**. The load could be for example a bulb, an electrical motor, a speaker and so on. With some thought you can contribute considerably to this list. For example, in Figure 1.12, a resistor is connected as a load.

1.3.3 Resistance and Power

In an electrical circuit, the energy is constantly changing from one form to another. For example chemical energy in the battery is changed into potential electrical energy. Then in the load it is changed again, for example into mechanical energy. Some of it is lost in the conducting wires and dissipated as heat. All this is done according to the law of conservation of energy.

All materials, (including conductors) resist the current flow. Obviously, conductors offer much less resistance to the current flow, compared to insulators or semiconductors. The quality that describes how much the material counteracts the electrical current flowing through it, is called **electrical resistance**. It is measured in **Ohms (Ω)**. If a potential difference of 1 V produces current of 1 A, the resistance of the load is 1 Ω .

The resistance of materials depends on their specific properties, called **resistivity**. This parameter represents the electrical resistance of a certain amount of the given substance (namely one cubic centimeter). It is measured in **Ohm.m**. The electrical properties of different materials can be easily compared by their resistivity. For example, the resistivity of copper is $1.75 \times 10^{-8} \Omega \text{ m}$, while the resistivity of nickel is $1.25 \times 10^{-7} \Omega \text{ m}$. Thus, a conductor, made of nickel, exhibits nearly ten times higher resistance to the current, compared to an identical conductor, made of copper. The resistivity of materials will be discussed further in Chapter 2.

The electrical resistance of metals depends also on the ambient temperature. Higher temperature will increase the energy of positively charged ions. This will increase their rate of oscillation. Therefore, when moving forward, the free electrons will collide more frequently into the ions, losing in this way more of their energy. Consequently, a higher ambient temperature will increase the conductor's resistance.

And finally the electrical resistance of a conductor depends on its dimensions. The longer and thinner the conductor, the higher its resistance.

Electrical power is equal to the work done for a certain time and it can be expressed with the formula:

$$P = \frac{W}{t}$$

where:

P is the electrical power in watts (W)

W is the work done in joules (J)

t is the time in seconds (s)

The electrical power is one of the most important parameters of the circuits, as it shows how quickly the energy is transformed from one form to another. A potential difference of 1 V and a current flow of 1 A result in 1 W power consumption.

1.3.4 Summary

The current flow in the conductor is a complex process, involving electrons moving in one direction and simultaneously transferring their energy through impulses. Current can flow continuously in a circuit only if an external electromotive force is applied and the circuit is closed. The direction of the conventional current flow is from the positive to the negative side of the voltage source. The actual electron

current flow is in the opposite direction. In a closed circuit, the energy is constantly changing from one form to another. Power is the rate at which work is done.

1.4 Direct and Alternating Current

When you have completed study of this section you should be able to:

- Define direct and alternating current
- Recognize primary and secondary cells
- Describe the basic structure of a battery
- Describe the period and frequency of an ac waveform
- Describe sine wave values

1.4.1 Direct Current

An electric current can be either direct or alternating. If the voltage source is a constant polarity supply then the current always flows in one direction. For this reason it is called **dc (direct current)**. The voltage from such a source is called a **dc voltage**. It can be produced from batteries and dc generators. Batteries are much more common and will be discussed here.

A battery is a device that creates a potential difference between its two terminals by means of a chemical reaction. Each battery consists of one or more units, called **cells**. The cell contains electrolytes that take part in the chemical reaction. Most people simply classify batteries into wet and dry. Strictly speaking **dry batteries** do not exist, for the simple reason that if the electrolytes were totally dry they could not operate! However, electrolytes may be in a jelly form, or semi-solid, which makes the battery appear to be dry.

Batteries are made from either primary or secondary cells. Primary cells use a non-reversible chemical process and once their chemicals are used up, they are discarded. Every “dry” primary battery has two structures called electrodes. They are made from different kinds of chemically active materials. An electrolyte placed between them promotes a chemical reaction, known as electrolytic dissociation. The result of the electrolytic dissociation is that one of the electrodes, called an anode becomes negatively charged, and the other one, called a cathode becomes positively charged. The construction of the dry cell is shown in Figure 1.14.

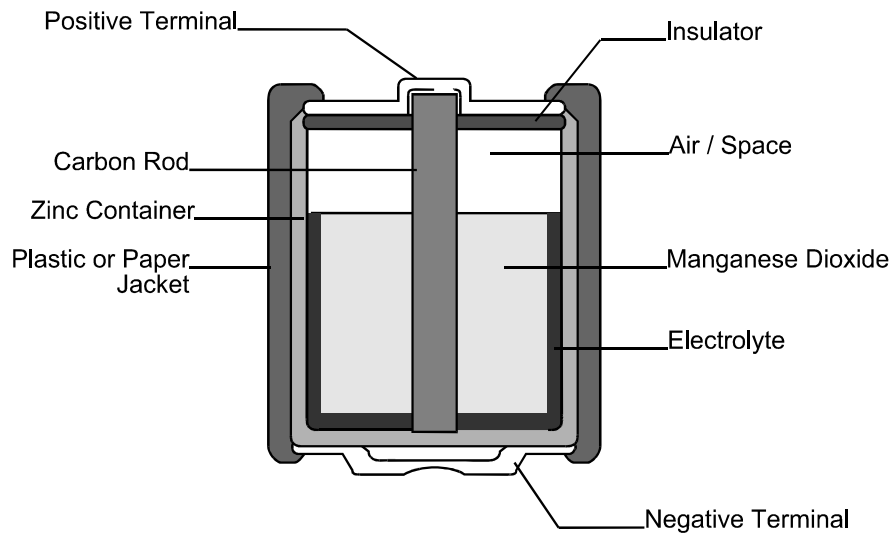


Figure 1.14
Structure of a Dry Cell

Depending on what material the anode and the cathode are made from, dry batteries are classified into three main groups:

- **Carbon/zinc cells.** These are primarily used as general-purpose batteries in radios, torches, toys etc. A carbon rod in the center of the cell functions as a cathode. A sheet of porous material, soaked in the electrolyte separates the anode and the cathode from each other. The electrolyte is a paste, comprising zinc chloride, ammonium chloride and water.
- **Alkaline cells.** The construction of alkaline cells is similar to that of the carbon/zinc cells. The basic difference is that the electrolyte is made from a strong alkaline solution, called *potassium hydroxide*. This compound conducts electricity better, than that of the carbon/zinc cells. This feature enables the alkaline cell to deliver sustained higher level of current more efficiently. For this reason, alkaline batteries are an ideal source of energy for portable TVs, cassette players and electrical toys that require high level of current. They are more economical than the carbon/zinc batteries and last up to eight times longer.
- **Mercury cells.** They have a cathode of mercuric oxide, an anode of zinc and an electrolyte of potassium hydroxide. They produce voltage of 1.4 V, instead of 1.5 V as with the other type of cells. The advantage of the mercury cell is that their voltage remains stable, while that of the other primary cells declines, during use. This feature makes them suitable for use in sensitive devices, such as scientific instruments or hearing aids.

Secondary cells have a chemical structure, which can change between two states. Passing a current through a secondary cell in the opposite direction to its normal discharge current reconstitutes the electrolyte back to its original form. The most common types of secondary batteries are:

- **Lead/acid storage batteries.** They comprise of a hard rubber or plastic container that holds three or six cells. Each cell has two sets of electrodes. The frames of the electrodes, called grids are made from antimony-lead alloy. The electrolyte is sulfuric acid and water. Lead/acid batteries are widely used to produce

current in cars and trucks. They also provide emergency electricity for vital facilities, such as hospitals or sewage-treatment plants.

- **Nickel/cadmium (NiCd) storage batteries.** They operate on the same principle as the lead/acid batteries. The main difference is that the electrolyte is a solution of potassium hydroxide and that the anode and the cathode are made respectively from cadmium and nickel oxide. The chemical composition of nickel/cadmium batteries allows them to be sealed. This prevents the corrosive solution from leaking and makes these batteries suitable for use in portable electronic equipment and many other applications. Nicad batteries are reliable, inexpensive and can deliver plenty of electrical current. On average, they can be re-charged about 1,000 times. Their main drawback is that they suffer from an effect called the **memory effect**. It occurs when the battery is not fully discharged every time when it is used. Then the battery “forgets” its full capacity rate, and “remembers” only the capacity rate, for which it was lastly used. For example if a Nicad battery is repeatedly used only at 20% of its maximum capacity, then it is unable to deliver its maximum capacity current over the rated time. When the 20% discharge point is reached, its voltage drops sharply. This makes 80% of battery’s capacity unusable. The process of draining over 50% of the rated capacity from the battery, and then recharging it in a normal way is called **deep cycling** and is used to overcome the memory effect. Another disadvantage of Nicad batteries is that they discharge internally at the rate of 15% per month.
- **Nickel Metal Hydride (NiMH) storage batteries.** These batteries were first introduced to the market in 1991. They also can be re-charged about 1,000 times, but are much lighter than Nicad batteries and are capable of storing larger charge. NiMH batteries also suffer from the memory effect and discharge internally at the rate of 25% per month.
- **Lithium-ion (Li-ion) storage batteries.** They were first introduced to the market in 1994. These are the most powerful of all. They can be re-charged over 1,200 times, do not suffer from memory effect, and their rate of internal discharging is only 8% per month. Li-ion batteries are four times more expensive than Nicad batteries and twice as expensive than NiMH batteries. At present, their main application is as a backup power supply for portable computers.

A constant-voltage charger or a constant-current charger may be used to charge batteries. For alkaline batteries, a constant-current charger is preferred, as it is more effective in correcting the memory effect. There are three main methods to charge a battery:

- **Fast charge** is used for emergencies only. It lasts about 1 hour. The battery is charged at its full rate. For example, a battery rated at 20 ampere/hours will be charged at 20 amperes for 1 hour.
- **Normal charge** is used for most cases and lasts 5 hours. For example, the same battery, mentioned above will be charged at 4 amperes for 5 hours.
- **Slow rate** is done for 10 hours. Therefore, the same battery, mentioned at the previous two examples will be charged at 2 amperes for 10 hours.

1.4.2 Alternating Current

The direct current flows in one direction only. Conversely, alternating current (ac) constantly reverses its direction. The same rule applies for the ac voltage. To generate and to transmit ac is much easier and cheaper than dc. Furthermore, it is much easier to transform ac into dc, than vice versa. For these reasons, dc voltages are more widely used. The domestic supply voltage in most countries, including Australia is 240 volts ac. Some other countries use mains of 220 V or 110 V, but it is always an ac.

To learn more about the alternating current first we have to know more about the sine wave, as this is the basic waveform, from which all other waveforms can be derived. It is known that if a magnetic flux cuts a conductor, an emf is induced in it. Figure 1.15 illustrates a conductor, which is rotating in a homogenous magnetic field at a constant speed.

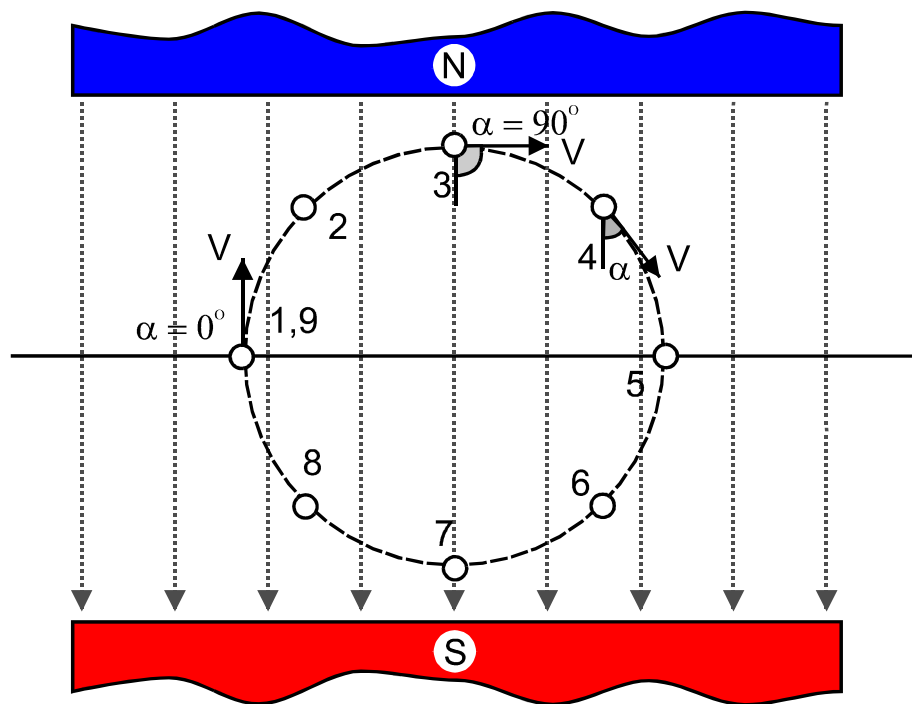


Figure 1.15
Generating Sine emf

Because the magnetic flux cuts the conductor, an emf is induced in it. Its magnitude is:

$$e = B l v$$

where:

- e is the emf, induced across the conductor,
- B is the flux density,
- l is the conductor's length,
- v is the conductor's speed.

On the other hand:

$$v = v_r \sin \alpha$$

where:

v_r is conductor's speed of rotation,

$\sin \alpha$ is the angle between the speed (v) and the magnetic flux (B).

Substituting this expression in the original formula gives us the equation for the sine wave:

$$e = B l v_r \sin \alpha$$

The conductor is rotated with a constant speed, and the angle α is changing gradually from 0° to 360° . In fact, the angle α is the only parameter in the formula that is changing with time. The flux density, the conductor's length and the speed of rotation remain constant. For this reason the induced emf is also changing constantly and gradually, following the changes in the angle α . It is not difficult to see that the emf reaches its peak value when $\alpha = 90^\circ$ (at positions 3 and 7) and has a value of zero when $\alpha = 0^\circ$ (at positions 1, 5 and 9). Every time the conductor crosses the "neutral line" (the line connecting positions 1 and 5), the direction of induced emf is changed. In Figure 1.16, the values of the emf for one full rotation are plotted on a diagram against the time.

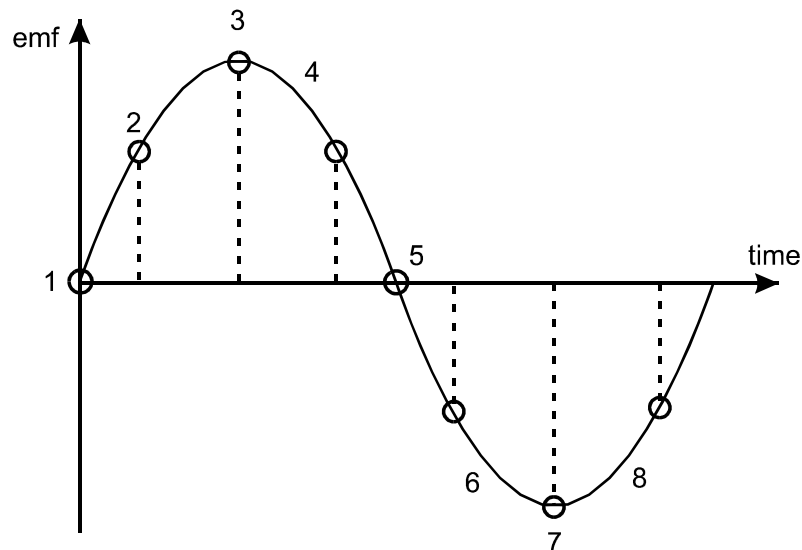


Figure 1.16
Sine Wave

In practice, it is better to use not one but two conductors, connected in series and forming a **loop**. This is illustrated in Figure 1.17. When the loop is in a horizontal position the motion of each conductor in this instant is parallel to the direction of the magnetic field. Therefore no flux lines are being cut by both conductors, and no

emf is induced in them. The total emf in the loop is zero volts, and no current flows through it.

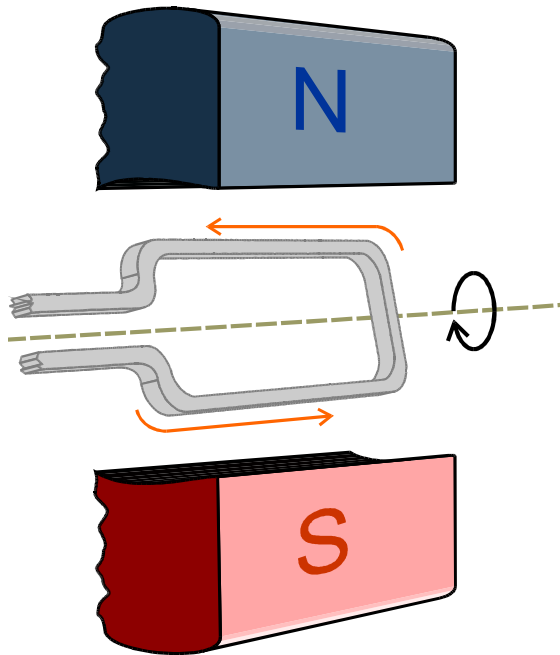


Figure 1.17
Generating Sine emf in a Loop

As the loop starts rotating, it starts cutting some flux lines. An emf is being induced in both conductors. However, because they are influenced by opposite poles (one is closer to the north pole, and the other is closer to the south pole), the emf induced in them are with opposite polarities. The loop represents two conductors, connected in series. When two conductors are connected in series, the total induced emf is the *sum* of the electromotive forces induced in each one of them. This can be expressed with the formula:

$$e_L = 2 B l v_r \sin \alpha$$

where:

e_L is the total emf, induced in the loop.

In continuing its rotation, the loop reaches the vertical position. At this instant, the motion of the conductors is perpendicular to the magnetic field. Therefore each conductor cuts a maximum number of flux lines. The emf, induced in each conductor reaches its maximum value. The total emf at this instant reaches its maximum value, and subsequently the current in the loop is also in its maximum.

From here onwards, the total emf decreases in the same manner, until the loop reaches the horizontal position. In this instant, the current stops flowing. The loop continues its rotation and the emf is induced again. This time the conductor that was previously closer to the north pole is now closer to the south pole. The opposite is true for the other conductor. This means that the total emf now has a reversed

polarity and the resulting current will flow in the opposite direction through the loop.

This reversed polarity emf reaches its peak value, when the loop comes again into a vertical position. After that, it decreases and becomes zero, when the loop reaches again a horizontal position. This process can be constantly repeated, and results in producing an emf, with a waveform, called a **sine wave** (Figure 1.16).

1.4.3 Period and Frequency of Sine Waves

From the processes, described in the previous section, it becomes clear that the value and the direction of the alternating current are changing continuously. Here are some characteristics that describe these changes:

- **A period** is the smallest interval of time taken for the sine wave to complete one full cycle. When the cycle is completed, the sine wave begins to repeat the quantities of the current in the same sequence. For example, let us take the maximum positive value of the sine wave as a starting point. The smallest time taken for the sine wave to revert back to the maximum positive value, is called the **period** of this sine wave. After that, the sine wave begins to repeat the same values in the same sequence. A period is designated with the letter **T**, and is measured in seconds. Figure 1.18 shows a waveform with a period of 20 ms.

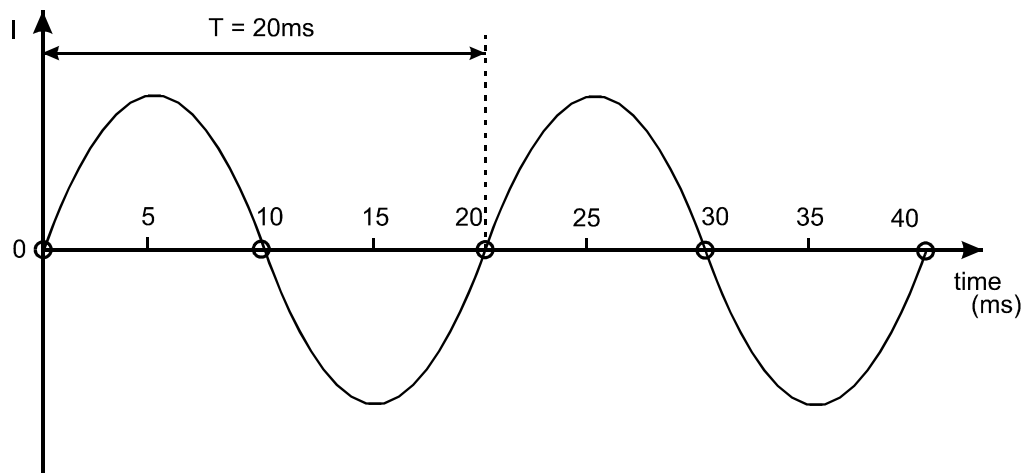


Figure 1.18
A Period of a Sine Wave

- **The frequency** is designated with the letter **f** and depends on the number of periods, contained in one second. Their reciprocal relationship can be described with the formula:

$$f = \frac{1}{T}$$

where:

f is the frequency of the sine wave,

T is the period of the sine wave.

Frequency shows the number of full cycles that occur in one second. The unit of frequency is cycles per second or more commonly known as Hertz (Hz). If a sine wave makes only one full cycle per second its frequency is 1 Hz.

It is not difficult to calculate, using the formula above, that the sine wave in Figure 1.18 has a frequency of 50 Hz. This is the frequency of the mains voltage in most countries around the world, including Australia. In the USA the standard is 60 Hz.

1.4.4 Sine Wave Values

The period and the frequency of a sine wave are related to time. The voltage and current measurements are also important. They are measured on the vertical graph and can be expressed in a number of ways:

- **Peak value.** The peak value of a sine wave is the maximum value that the current or the voltage reaches with respect to zero. The polarity is not important, because the negative and the positive peak should have identical values. Peak values are designated V_p and I_p for peak voltage and peak current respectively. Figure 1.19 shows peak values.

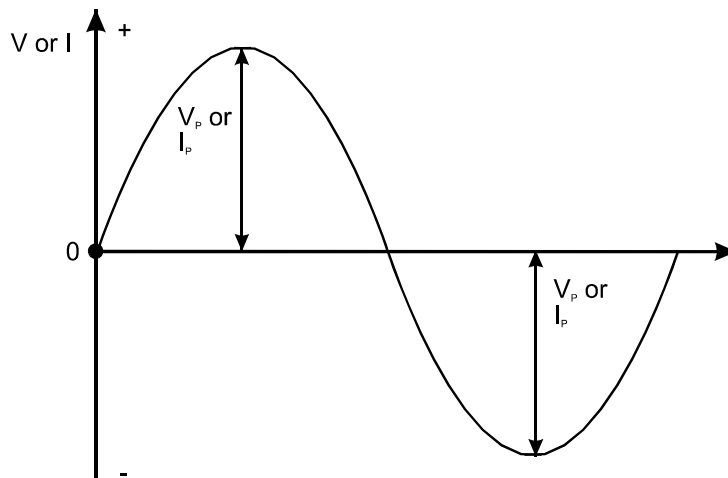


Figure 1.19
Peak Values

- **Peak-to-peak value.** This is the voltage or the current, measured between the positive and negative peak of the waveform. Obviously, the peak-to-peak value is two times greater than the peak value. The peak-to-peak value is shown in Figure 1.20. These values are most commonly used when taking measurements with oscilloscopes. This is discussed further in Section 6.3.7.

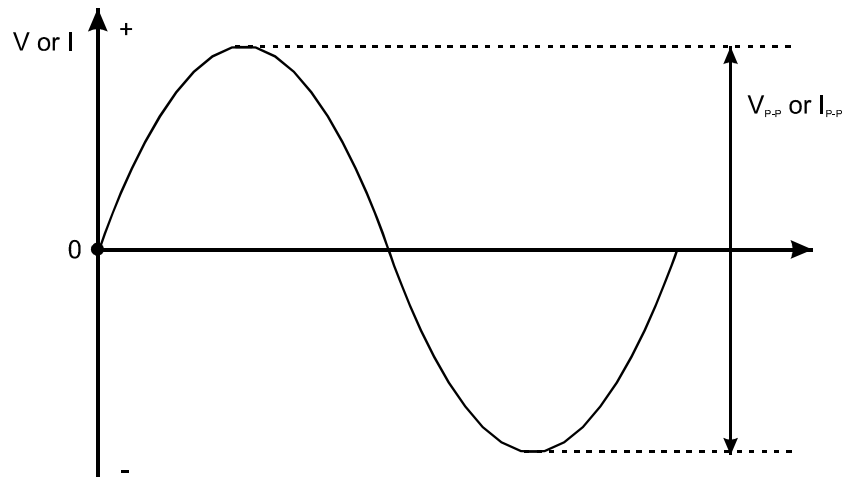


Figure 1.20
Peak-to-Peak Values

- **Average value.** Normally the average value of a sine wave is zero. During the first half-period the electrical current transfers electrical charges in one direction through the conductor. Then, during the second half-period the current transfers the same amount of charges in the opposite direction. Therefore, the net charge transferred is zero. For this reason, the average value of a sine wave is calculated for a half-period (half-cycle) only. The polarity of the half-cycle is not important and the average value could be calculated for either one of the half-cycles. Figure 1.21 shows the average value graphically.

The area enclosed from the positive half-period represents the amount of electrical charges transferred through the conductor for this half-period. Let us assume that the area enclosed from the positive half-period is equal to the area of the rectangular ABCD (see Figure 1.21). The area enclosed from the rectangular ABCD represents the same amount of electrical charges, transferred for the same time, through the same conductor, but as if a direct current with a magnitude, equal to AB flows through it. The energy dissipated, and the work done from the alternating current will be equivalent to the energy dissipated and the work done from the direct current.

Therefore, the average value of a sine wave is equal to the value of a direct current, which for the time of a *half-period* transfers through the conductor the same amount of charge as the alternating current. Through mathematical calculations that go beyond the scope of this manual, it is possible to prove that the relationship between the peak value and the average value is as follows:

$$I_{av} = \frac{2I_p}{\pi} = 0.637I_p$$

$$V_{av} = \frac{2V_p}{\pi} = 0.637V_p$$

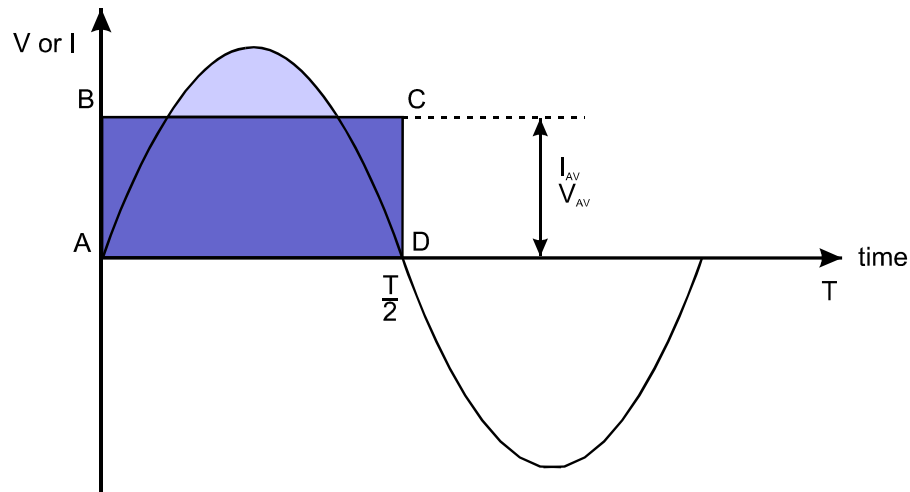


Figure 1.21
Average Value

- **Root-mean-square value (rms).** An alternating current flowing through the load (for example a heater or a stove) has a heating effect. The produced amount of heat does not depend on the current's direction. Let's assume that the amount of heat, produced from an alternating current for 1 second is equal to Q . It is possible to find out a source of a direct current, producing the same amount of heat, equal to Q for the same time. Therefore, the root-mean-square value can be defined as equal to the value of the direct current, producing the same amount of heat, in the same load, for the same time, as an alternating current. The formula describing the relationship between the rms and the peak value is given below:

$$I = \frac{I_p}{\sqrt{2}} = 0.707I_p$$

$$V = \frac{V_p}{\sqrt{2}} = 0.707V_p$$

These formulae can be proved mathematically, but this will not be developed here. The rms is the value read from the voltmeters or the ammeters, as you will see in Chapter 6. In electronics, all specified values are given in root-mean-square values, unless specifically written otherwise.

1.4.5 Summary

An electric current can be either direct (dc) or alternating (ac). The direct current flows all the time in one direction only. The alternating current constantly reverses its direction. A battery is the most common source of direct current. Batteries are made from either primary or secondary cells. Primary cells are discarded, once their chemicals are used up. Secondary cells can be charged and used repeatedly. The basic waveform of an alternating current is the sine wave. A cycle of a sine wave is a sequence of values in a certain order, after which the sine wave begins to repeat them in the same order. The period of a sine wave is the smallest interval of time, taken for the wave to complete one full cycle. The frequency of a sine wave is the

number of full cycles that occur for one second. The sine wave values are described with peak, peak-to-peak, average and root-mean-square (rms) values. In electronics, all specified values are rms, unless specifically written otherwise.

1.5 Units and Abbreviations

Every science has its own unique language. Electronics is no exception. In this manual, some of the most common definitions and abbreviations are given in the text, where they appear first. To learn more definitions and abbreviations you should consider buying an electronic dictionary. They are available in most specialized bookshops and in most cases are excellent value for money.

Most countries in the world, including Australia, use the international system of units, called *Systeme International d'Unites*, known simply as SI. This system is discussed in detail in Section 6.1.1. The common electric terms, according to SI have the symbols and units shown in Table 1.1.

Table 1.1
Common Electronic Units

Item	Symbol	Unit	Symbol
<i>Current</i>	<i>I</i>	<i>Ampere</i>	<i>A</i>
<i>Voltage</i>	<i>V</i>	<i>Volts</i>	<i>V</i>
<i>Charge</i>	<i>Q</i>	<i>Coulomb</i>	<i>C</i>
<i>Energy</i>	<i>W</i>	<i>Joule</i>	<i>J</i>
<i>Power</i>	<i>P</i>	<i>Watt</i>	<i>W</i>
<i>Resistance</i>	<i>R</i>	<i>Ohm</i>	Ω
<i>Resistivity</i>	ρ	<i>Ohm.Meter</i>	Ωm
<i>Reactance</i>	<i>X</i>	<i>Ohm</i>	Ω
<i>Inductance</i>	<i>L</i>	<i>Henry</i>	<i>H</i>
<i>Capacitance</i>	<i>C</i>	<i>Farad</i>	<i>F</i>
<i>Frequency</i>	<i>f</i>	<i>Hertz</i>	<i>Hz</i>
<i>Time</i>	<i>t</i>	<i>Second</i>	<i>s</i>

1.6 Quiz

1. Describe what a proton, neutron and electron is.
2. What is the physical relationship between them?
3. Is it true that the electrons in shell L have more energy, compared to those in shell M?
4. An atom has 77 orbiting electrons. Is it true that its net charge is zero?
5. Describe the process of ionization.
6. What is the basic difference between insulators and conductors?
7. How many valence electrons does a semiconductor have?
8. List the most common conductors and insulators.
9. Which is the most widely used semiconductor material?
10. Is it true that at 0°K semiconductors behave as ideal conductors?
11. What is the name of the process that dramatically increases the conductivity of the semiconductors?
12. What happens with the electrons when an emf is applied to a conductor?
13. What is the difference between resistance and resistivity?
14. A charge of 2 C is moved through a circuit with an emf of 12 V. How much work is done?
15. Is it true that semiconductors increase their conductivity at higher temperatures?
16. It takes 10 s for an emf of 6 V to drive a charge of 6 C. How much power was expended?
17. What is the main difference between a direct and an alternating current?
18. What is the name of the process that creates a potential difference in the battery?
19. What type of dry battery produces the most stable voltage over time?
20. What is the memory effect and how can it be overcome?
21. Describe three methods for a full constant-current charge.
22. What formula describes the sine wave?
23. What has to be done in order to generate a sine wave with a higher frequency?
24. What is the peak value of the voltage for the domestic power supply in Australia?

Discrete Components

2.1 Resistors, Inductors and Capacitors

When you have completed study of this section you should be able to:

- Recognize resistor, inductor and capacitor circuit symbols
- Describe the factors that affect resistance, inductance and capacitance
- List types of resistors, inductors and capacitors
- Specify component values, using color codes
- Determine effective resistance, inductance and capacitance in series and parallel circuits

2.1.1 Resistors

2.1.1.1 Overview

As it was explained already in Section 1.3.3, the quality that describes how much the material restricts the electrical current flowing through it, is called **electrical resistance** and it is measured in Ohms (Ω). A good analogy of resistance is water flowing through a pipe. If the pipe is of sufficient diameter it will impose little or no restriction to the flow. In this case the speed of water will depend only on the pressure (which can be compared to the voltage). Any blockage in the pipe will impose a restriction to the flow. An electrical resistor is equivalent to a restriction in the pipe.

Also from Chapter 1.2 we already know that depending on their conductive properties all materials can be classified as conductors, insulators and semiconductors. The conductors offer minimal resistance to the electrical current, while insulators restrict it from flowing. The reason for this is that materials have

different number of free electrons in the valence shell that determines their resistivity. Materials with a higher resistivity impose a higher restriction to the flow of electrical current.

The resistivity is not the only factor that affects the resistance. The length and the cross-sectional area of the material are also important. This is the formula that describes the electrical resistance:

$$R = \rho \frac{L}{A}$$

where:

R is the electrical resistance in Ohms (Ω),

ρ is the resistivity of the material in Ohms.meter(Ω .m),

L is the length in meters (m),

A is the cross sectional area of the conductor in square meters (m^2).

The resistivity remains the primarily factor, which determines the resistance and can vary greatly for different materials. The resistivity of the most common materials, used in electronics is given in Table 2.1.

Table 2.1
Resistivity of Common Materials, used in Electronics

<i>Material</i>	ρ , (Ω .m), t = 20°C
<i>Silver</i>	1.65×10^{-8}
<i>Copper</i>	1.75×10^{-8}
<i>Aluminum</i>	2.9×10^{-8}
<i>Nickel</i>	1.25×10^{-7}
<i>Tin</i>	1.2×10^{-7}
<i>Lead</i>	2.08×10^{-7}
<i>Nichrome</i>	4.2×10^{-7}
<i>Constantan</i>	5×10^{-7}
<i>Chrome/Nickel</i>	1.1×10^{-6}
<i>Hard rubber</i>	10^{16} - 10^{17}
<i>Polyethylene</i>	10^{17} - 10^{18}
<i>Ceramics</i>	10^{15} - 10^{16}

A **resistor** is a device, made of material that conducts an electrical current, but possess a larger resistance, compared to the normal wires. The resistance, offered by a resistor does not depend on the current frequency. It is designated with the letter **R** and it is measured in **Ohms**. The resistance, provided from a resistor is called **ohmic, real** and sometimes **dc resistance**.

Resistors are mostly used as voltage and current dividers and as current-limiting devices, as it will be discussed later. A resistor circuit symbol is shown in Figure 2.1.

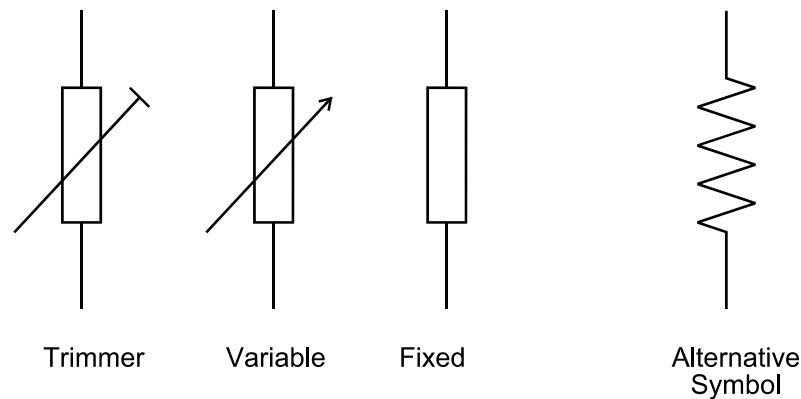


Figure 2.1
Resistor Circuit Symbols

The resistors are manufactured with resistance values that vary widely. For example a resistor may have a value of 0.1Ω , or a value of $100,000,000 \Omega$. According to the international system of units SI, the value of a resistor may be given in Ohms or in units, which are multipliers of the Ohm. Here are some of them:

$$\begin{aligned} 1 \text{ k}\Omega &= 1 \times 10^3 \Omega = 1,000 \Omega \\ 1 \text{ M}\Omega &= 1 \times 10^6 \Omega = 1,000,000 \Omega \\ 1 \text{ G}\Omega &= 1 \times 10^{12} \Omega = 1,000,000,000 \Omega \end{aligned}$$

The resistor tolerance is the maximum allowable deviation from the specified value. It also can vary for different types of resistors. For example the real value of a resistor with a specified value of 100Ω , and tolerance of 1%, has to be in the range from 99 to 101Ω , when measured.

It is obvious that no single resistive material can satisfy all required ranges and tolerances. For this reason, many different material types and designs are employed. These are described in short in the following section.

2.1.1.2 Types of Resistors

- **Wirewound Resistors**

The simplest way to make a resistor is to wind some resistive material around an insulating rod. The resistive material can be made from an alloy with a reasonably high resistance, such as a *nichrome*. A nichrome is the material often used as an element in electrical heaters and toasters. The thinner and the longer the material, the higher the resistance. These resistors are available in a range

from $1\ \Omega$ to $100\ \Omega$. They are primarily used for three applications: high precision, low resistance and high power dissipation. Typical tolerances range from 0.01% to 1%. There are wirewound resistors that can dissipate up to 200 W of power. Figure 2.2 shows a wirewound resistor.

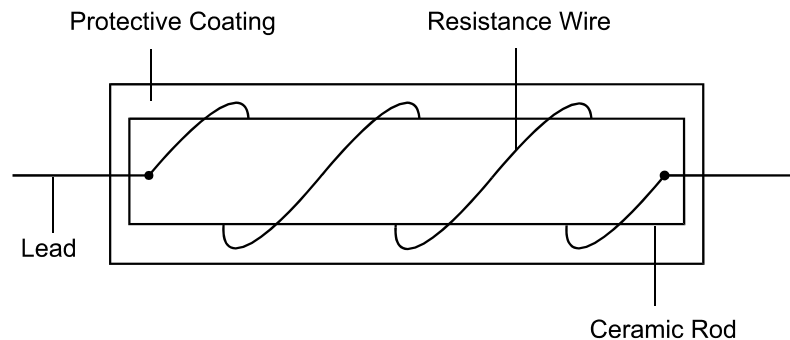


Figure 2.2
Wirewound Resistor

- **Carbon-film Resistors**

These are the most common resistors, used in electronics today. A special carbon compound is used as a resistive material. The compound is deposited as a film on the outside of a small ceramic rod. In order to keep the size of the resistor fairly small, a spiral is cut on the rod, thus increasing the effective electrical length of the film. The resistor leads are welded to end caps, which are fitted to the resistor's body. The end caps provide an electrical connection between the leads and the carbon compound. The carbon-film resistors are easily recognized by these caps, which are not used with other type of resistors. They are produced in a range from $10\ \Omega$ to $10\ \text{M}\Omega$ and can dissipate a maximum of 2 W of power. A diagram of a carbon-film resistor is given in Figure 2.3.

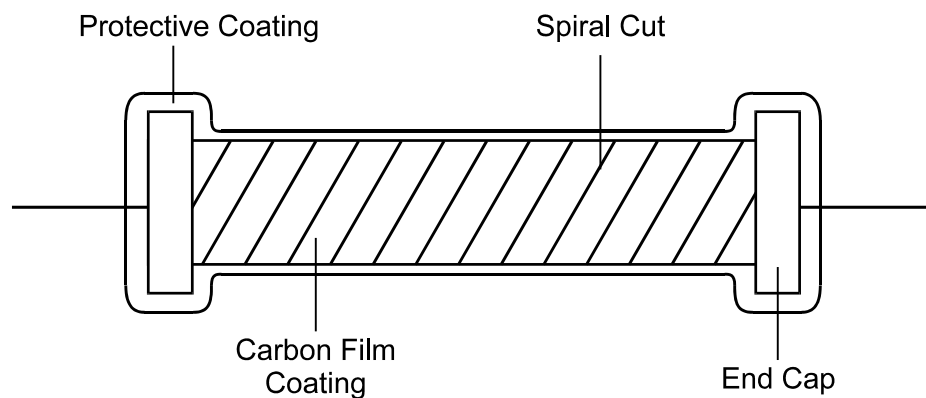


Figure 2.3
Carbon-film Resistor

- **Metal-film Resistors**

These are made in much the same way as the carbon-film resistors, but use a metal oxide as the resistive material. Their value remains fairly constant over time, which means that they are more stable. The price of metal-film resistors is usually higher than the carbon-film resistors.

2.1.1.3 Variable Resistors

So far we have talked about resistors that have a fixed specified value. It is determined at the time of their manufacture and cannot be changed. There is also another type of resistors which value can be changed. They are called **variable**. Variable resistors can be classified into two groups:

- **Potentiometers**

The word potentiometer comes from “variable potential”. A potentiometer construction is shown in Figure 2.4. Potentiometers have three leads - two fixed and one movable. The basic parameters of a potentiometer are the overall resistance and the power rating. These parameters are usually specified on the their cases. The resistive element of a potentiometer can be either carbon, deposited onto an insulating surface, or a resistive wire, wound on an insulating material. In either case the insulating material is usually constructed in a circular shape. A moving wiper rotates with respect to the center of the circular shape. The wiper slides on the resistive element surface and provides an electrical contact with it. Depending on the wiper's position, the resistance values between the points A and C on one side, and the points B and C on the other are changing.

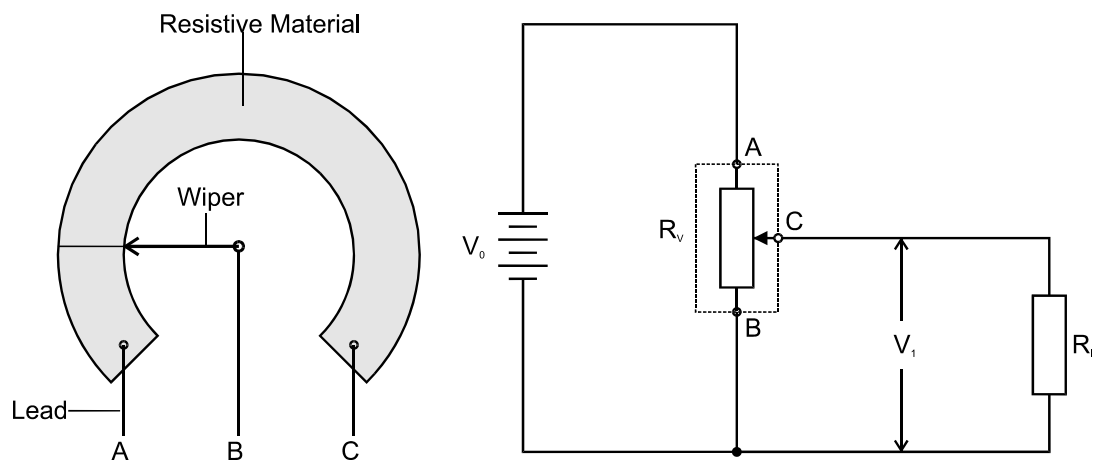


Figure 2.4
Potentiometer

It is not difficult to see that:

$$R_{AB} = R_{AC} + R_{CB}$$

where:

R_{AB} is the overall resistance of the potentiometer,

R_{AC} is the resistance between the *left* lead of the potentiometer and the wiper,

R_{CB} is the resistance between the *right* lead of the potentiometer and the wiper.

Their main application of potentiometers is as voltage dividers in circuits. A potentiometer method of connection is shown in Figure 2.4.

The resistive element often follows a certain mathematical law, which refers to the way the resistance changes, as the wiper moves along the resistive material. If the resistance between the points A and C (or B and C) is directly proportional to the distance at which the wiper is moved, the potentiometer is called **linear**. Such potentiometers are often marked with the letter "A". The other type of potentiometers follows a **logarithmic** law. They are used for example as a volume control in amplifiers. They are often designated with the letter "C" on their cases. If in doubt, the wiper has to be placed in the middle position and the resistance R_{AC} and R_{CB} have to be measured. They both should be approximately equal to the half of the overall potentiometer resistance. The curves of the linear and the logarithmic potentiometer are plotted in Figure 2.5.

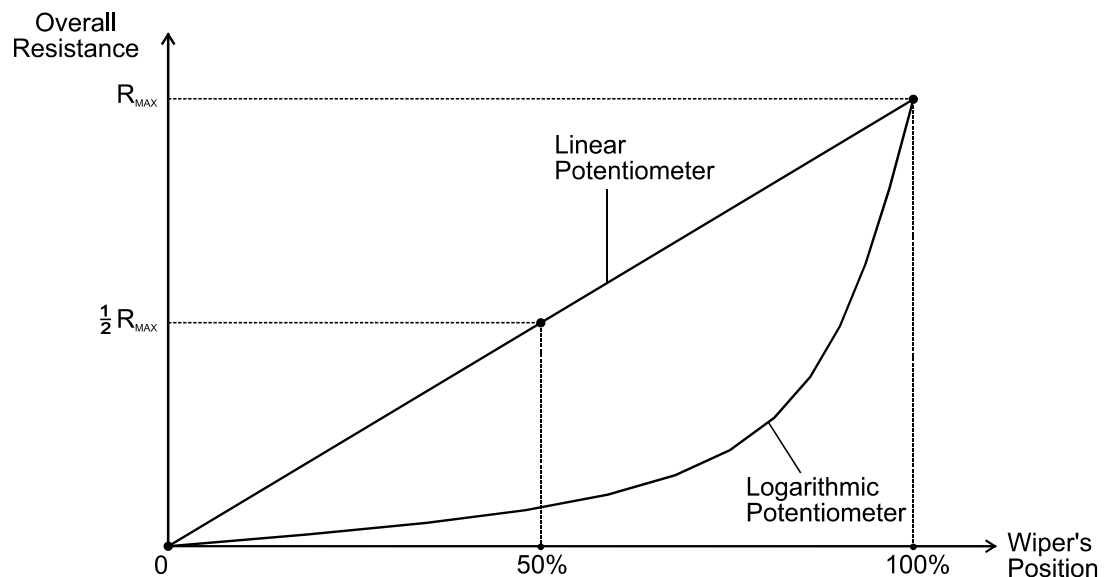


Figure 2.5
Plots of Linear and Logarithmic Potentiometers

- **Rheostats**

A variable resistor that has two leads only is called a **rheostat**. The rheostats are primarily used to vary a current value. It is done by varying the resistance between the two leads of the rheostat. They are manufactured in a similar way to that of potentiometers. Figure 2.6 shows how a potentiometer can be employed as a rheostat.

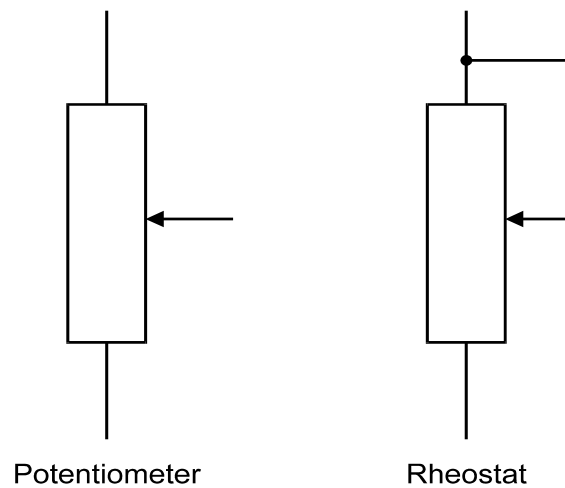


Figure 2.6
A potentiometer, connected as a rheostat

2.1.1.4 Power Rating and Nominal Values

An important consideration, when choosing the right resistor is its **power rating**. An electrical current flowing through a resistor generates heat. This heat is proportional to the applied voltage and the magnitude of the current. To enable the resistors to withstand the dissipated heat they are manufactured in a range of wattage ratings. The wattage ratings are standard, with the most common being $\frac{1}{4}$, $\frac{1}{2}$, 1, 5 and 10 watts. One should avoid touching resistors that dissipate more than 5W, as they may be too hot. Usually such resistors are mounted on radiators or are installed away from other components for better air-cooling.

Resistors are available in **standard nominal values** as described in Table 2.2. Resistors are also standardized by their tolerance. The most common tolerance values are 10% and 5%. The standard nominal values are calculated on the basis that the real values of the resistor (i.e. their nominal value plus, minus the tolerance) overlap each other. Thus, it is always possible to find a resistor, with a standard nominal value that is fairly close to the value required for a circuit.

Table 2.2
Nominal Values and Tolerance Ranges for Resistors

<i>Nominal Resistor Values</i>												
20%	10	15	22	33	47	68						
5%and10%	10	12	15	18	22	27	33	39	47	56	68	82
5%	11	13	16	20	24	30	36	43	51	62	75	91
<i>Decades: 0.01; 0.1; 1; 10; 100; 1k; 10k; 100k; 1M; 10M</i>												

The table also specifies other available decades. A **decade** is the same range of 12 values, as listed in the table for every tolerance range, but all of those values are multiplied by one of the multipliers shown below. Note that the tolerance range of 5% uses two decades of values. The table lists the decade for 10% tolerance range resistors as follows: 10 Ω ; 12 Ω ; 15 Ω ; 18 Ω ; 22 Ω ; 27 Ω ; 33 Ω ; 39 Ω ; 47 Ω ; 56 Ω ; 68 Ω and 82 Ω . For example, using the multiplier of 10k we can see that another decade of resistors is available and as follows: 100k Ω ; 120k Ω ; 150k Ω ; 180k Ω ; 220k Ω ; 270k Ω ; 330k Ω ; 390k Ω ; 470k Ω ; 560k Ω ; 680k Ω and 820k Ω . The list of available resistors can be continued using the other multipliers.

2.1.1.5 Color Coding

The physical size of low power resistors is very small. Thus it is not possible to mark their value numerically due to space constraints. For this reason, a very simple system of color codes is established. The resistor is marked with 4 or 5 color bands and each one of them represents a certain figure. A 4-band resistor is shown in Figure 2.7.

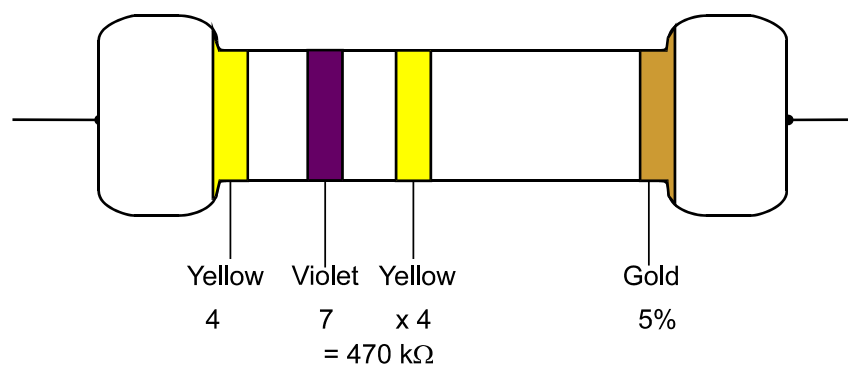


Figure 2.7
4-band Resistor

The first two bands determine the first two digits of the nominal value of the resistor, the third one is the multiplier and the fourth band is used to specify the tolerance value. In cases, where three digits are needed to express the nominal value, 5-band resistors are used. The colours are allocated in a similar manner, as with the 4-band resistors. In this case the first *three* colours give the nominal value of the resistor, the fourth band determines the multiplier and the last one specifies the tolerance value. Figure 2.8 shows a 5-band resistor.

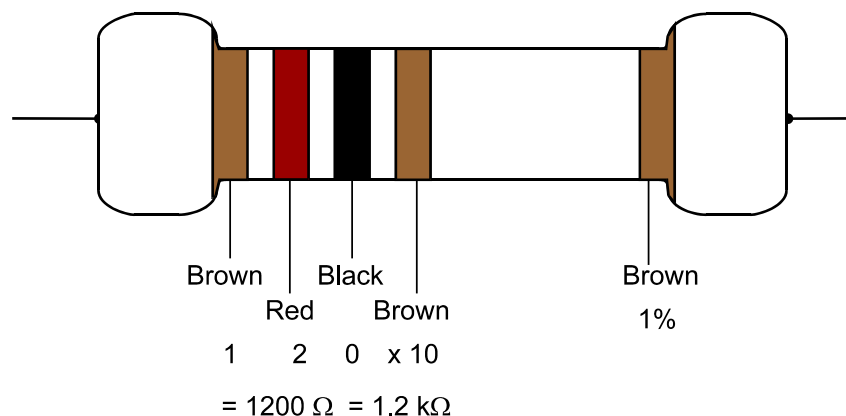


Figure 2.8
5-band Resistor

The resistor color code is given in Table 2.3. Note that gold and silver are used only as multipliers or as tolerance indicators.

Table 2.3
Resistor Color Code

Color	Digits	Multiplier	Tolerance
Black	0	x 1	
Brown	1	x 10	1%
Red	2	x100	2%
Orange	3	x 1,000	
Yellow	4	x 10,000	
Green	5	x100,000	
Blue	6	x 1,000,000	
Violet	7		
Grey	8		
White	9		
Gold		x 0.1(divide by 10)	5%
Silver		x 0.01(divide by 100)	10%

2.1.1.6 Combining Resistors

Sometimes it becomes necessary to combine two or more resistors together. This happens when the required resistance value or power rating is not available. Resistors can be combined together in two fundamental ways – in **series** or in **parallel**.

Figure 2.9 shows three resistors, connected in series. In this case the same *current* flows in all resistors in the circuit. The total resistance in the circuit is equal to the sum of all resistors and can be calculated with the formula:

$$R_T = R_1 + R_2 + R_3$$

where:

R_T is the total resistance in the circuit.

The golden rule to remember when calculating resistors, connected in series is that the total resistance of the circuit is always larger than the largest single value.

$$R_{\text{TOTAL}} = R_1 + R_2 + R_3$$

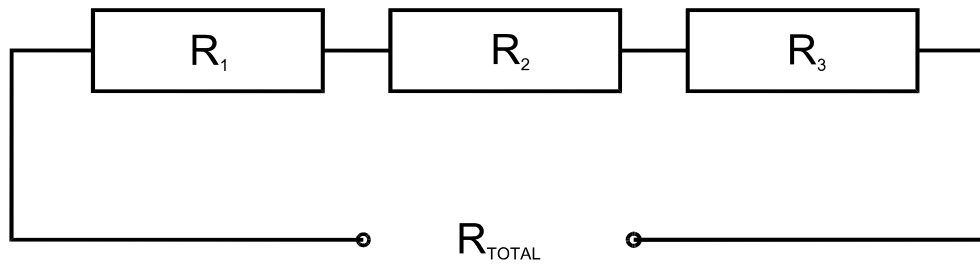


Figure 2.9
Resistors in Series

Figure 2.10 shows three resistors connected in parallel. In this case the *voltage* applied to each one of them is equal to the voltage of the source. The calculations here are a little more difficult, but still can be done with a simple pocket calculator, using the formula:

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

The golden rule for resistors connected in parallel is that the total resistance is always less than the smallest value.

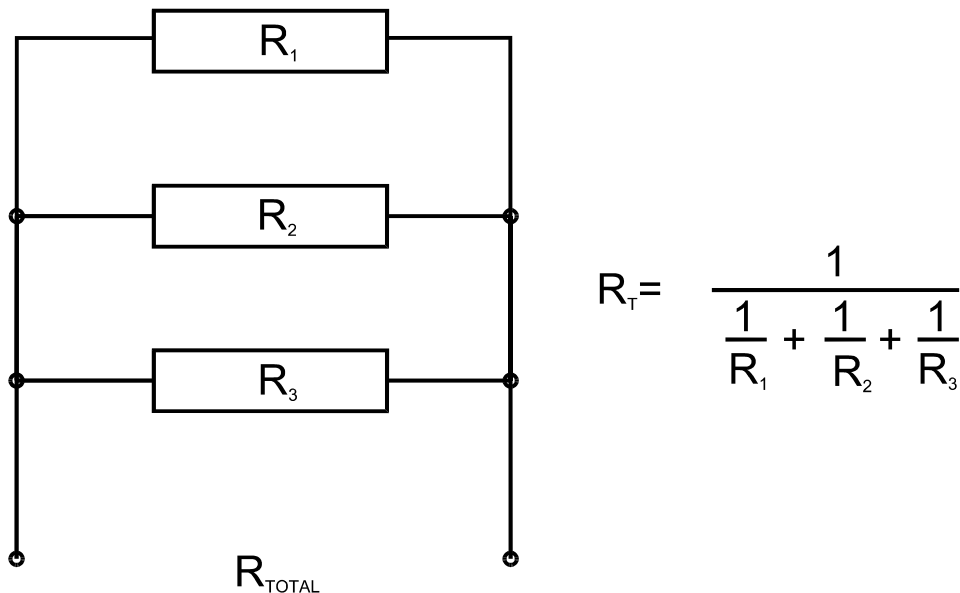


Figure 2.10
Resistors in Parallel

2.1.2 Inductors

We have already observed in Section 1.4.2 that moving a conductor in a magnetic field causes an electrical current to flow through it. Conversely an electrical current

flowing through a conductor creates a magnetic field around it. This is illustrated in Figure 2.11.

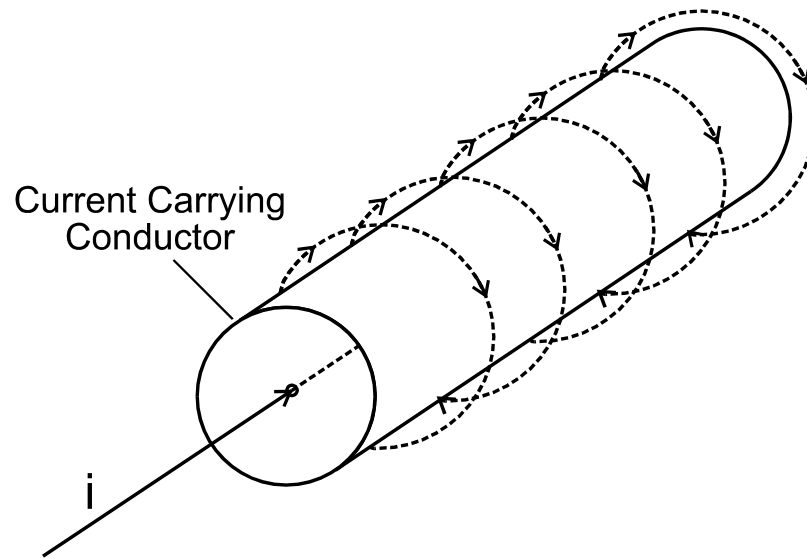


Figure 2.11
Current-carrying Conductor

An inductor can simply be a coil of wire. The current in every loop produces a magnetic field that is passed through the next loop. In this way the coil has the ability to concentrate the magnetic field in its core. The core can be made of materials that can conduct the magnetic flux more effectively than the air, such as iron or ferrite. In this case the magnetic flux density in the core becomes vastly greater. The larger flux density increases the overall inductance of the structure. This is shown in Figure 2.12.

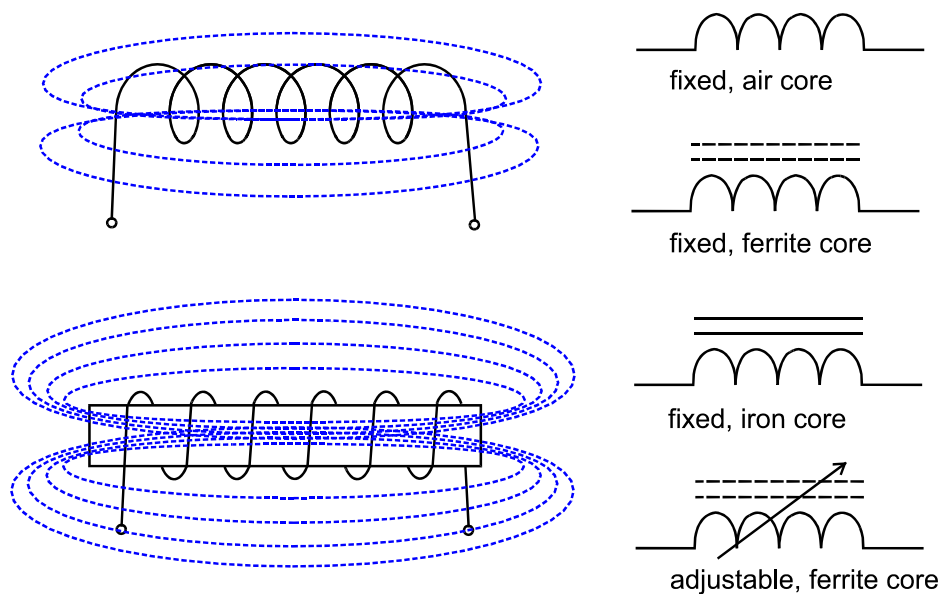


Figure 2.12
Inductors and their Symbols

When current first starts flowing, it creates a magnetic field around the conductor. This magnetic field itself induces a voltage in the conductor, which opposes the

current. Due to this effect any change in the current will produce a change in the magnetic field that will then oppose the current.

In other words, **inductance** is the total effect of current, creating a magnetic field and the magnetic field then creating current, and it is that property of a device that reacts against the *change* in current through the device, or the *change* in magnetic field around the device. **Inductors** are similar to resistors as they also provide resistance to a current, but only when the current *changes* its value or direction. A resistor will always offer the same resistance to the current, regardless of the current's parameters, such as value, frequency and so on. The inductor's resistance to a dc is zero, as the current does not change with time. However an alternating current changes its direction and value over time. These changes occur faster if the frequency of the current is higher. The higher frequency, the larger resistance of the inductor to the current.

The resistance of the inductor is called **inductive reactance**. It is denoted with X_L and has the units of Ohms. In the abbreviation X_L , X stands for reactance and L for inductance. The inductive reactance differs from the ohmic resistance in that X_L is frequency dependent, and it is given by the formula:

$$X_L = 2 \pi f L$$

where:

X_L is the inductive reactance in (Ω),

f is the frequency in (Hz),

L is the inductance in (H).

This formula shows that the inductive reactance is not only directly proportional to the current frequency, but also to the coil inductance. For this reason coils are often wound to have a specific value of inductance. Inductance is denoted with the letter L, and the measurement unit is Henry (see Table 1.1).

Some applications, such as tuning circuits in amplifiers or radios need variable, rather than fixed inductors. Varying the inductance can be achieved in a number of ways. The factors that affect inductance value are the number of loops in the coil, the area of the loops, the length of the coil and the properties of the coil. Changing any of these parameters can vary inductance values. The most common method to vary the inductance in a coil is to employ a movable core. The principle used here is that the inductance increases as the core is inserted further into the coil. Normally, a screw type ferrite core is used and its position can be adjusted with a screwdriver. The symbol of an adjustable coil with ferrite core is shown in Figure 2.12.

The calculation of total inductance for a combination of individual inductors is done in a similar way to that for resistance. Figure 2.13 shows three inductors, connected in series. The current through the circuit is common to all of them. The total inductance is derived with the formula:

$$L_T = L_1 + L_2 + L_3$$

where:

L_T is the total inductance in the circuit.

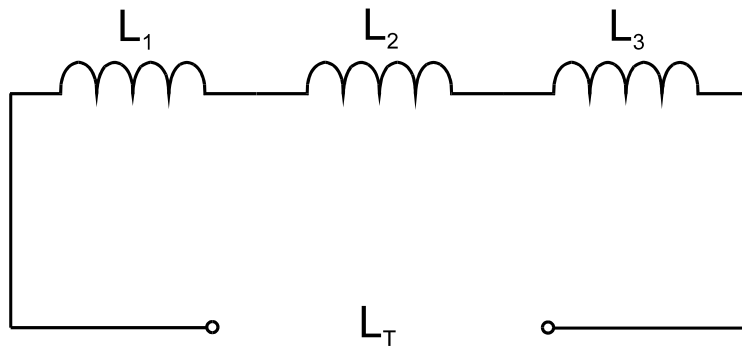


Figure 2.13
Inductors in Series

Similarly the total inductance of inductors in a parallel is according to the formula:

$$L_T = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3}}$$

A circuit of parallel inductors is shown in Figure 1.14.

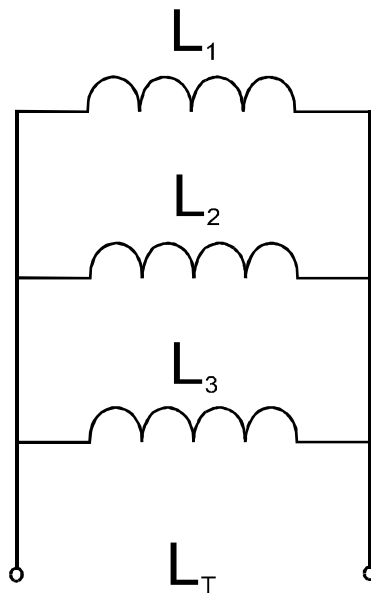


Figure 2.14
Inductors in Parallel

2.1.3 Capacitors

2.1.3.1 Overview

Resistors, inductors and capacitors are the fundamental components used in electronics. Two conductors, usually referred to as **plates** and an insulator, placed between them, called **dielectric** form the simplest **capacitor**. This is shown in Figure 2.15.

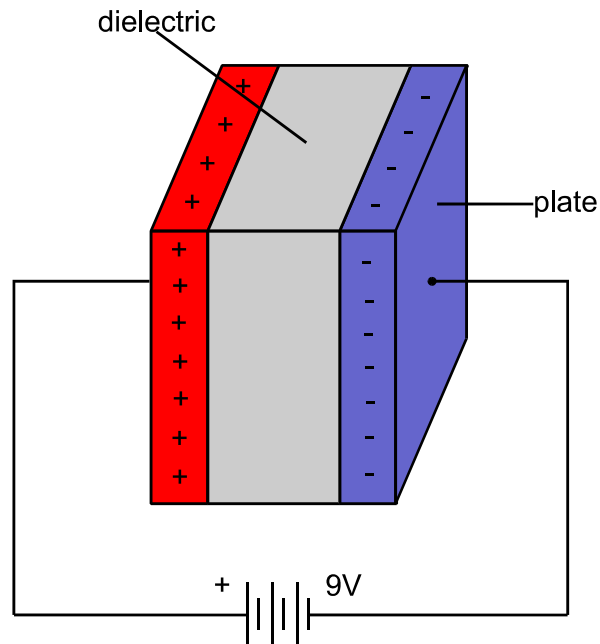


Figure 2.15
Capacitor in Charge

The word capacitor comes from the term “capacity”, which implies the capability of storing something. It is not difficult to guess that a component, such as a capacitor is used in electronics to store electrical charges. Now we will discuss how it happens with minimal use of theory.

When a voltage from a battery is initially applied to the capacitor an electric field is created between the two plates. The electrons of the dielectric are tightly bonded to their atoms. The valence electrons cannot be dislodged, but they acquire additional energy and their orbits change from a circle to an ellipse. This is illustrated in Figure 2.16. This process is called **polarization**. Atoms of the dielectric in electric field can be regarded as particles with two poles – negative and positive. Due to polarization, an electric field is created inside the dielectric, which opposes the electric field from the external source.

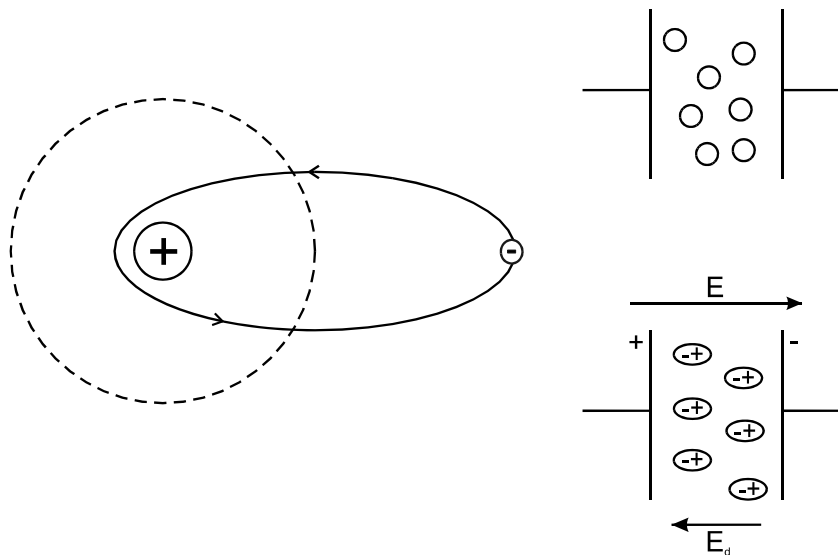


Figure 2.16
Dielectric Polarization

Electrons in the battery are attracted from the positive charge created in the dielectric and start flowing from the negative terminal of the battery to the plate connected to it. Electrons from the other plate are repelled from the negative charges in the dielectric and will flow back to the battery. The plate with more electrons will become negatively charged and the plate with less electrons - positively charged. After a certain time no more electrons can flow and the capacitor becomes fully charged. If the voltage source is then disconnected, a potential difference will exist between the two plates, as the electrical charges accumulated in the plates have nowhere to go. Effectively, a capacitor resembles a rechargeable battery. A charged capacitor is shown in Figure 2.17.

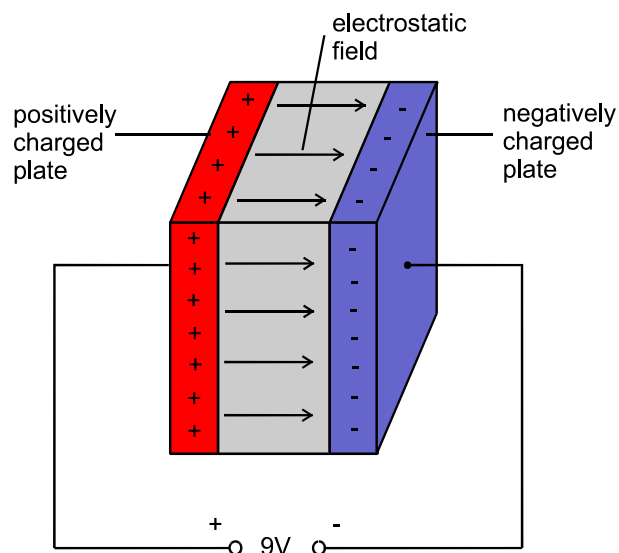


Figure 2.17
Charged Capacitor

In other words **capacitance** refers to the amount of charge the device can store, for each volt of potential difference between two bodies. A **capacitor** is a component, which is deliberately built to possess a specific capacitance value. The unit of capacitance is the **Farad (F)** and is expressed as:

$$1 \text{ Farad} = \frac{1 \text{ Coulomb of charge}}{1 \text{ Volt}}$$

A Coulomb is a very large amount of charge and it makes it inconvenient to work with Farads. For this reason in electronics it is more common to work with the following units:

$$1 \mu\text{F (microfarad)} = 1 \times 10^{-6} \text{ F}$$

$$1 \text{ nF (nanofarad)} = 1 \times 10^{-9} \text{ F}$$

$$1 \text{ pF (picofarad)} = 1 \times 10^{-12} \text{ F}$$

Electrical symbols for different types of capacitors are shown in Figure 2.18.

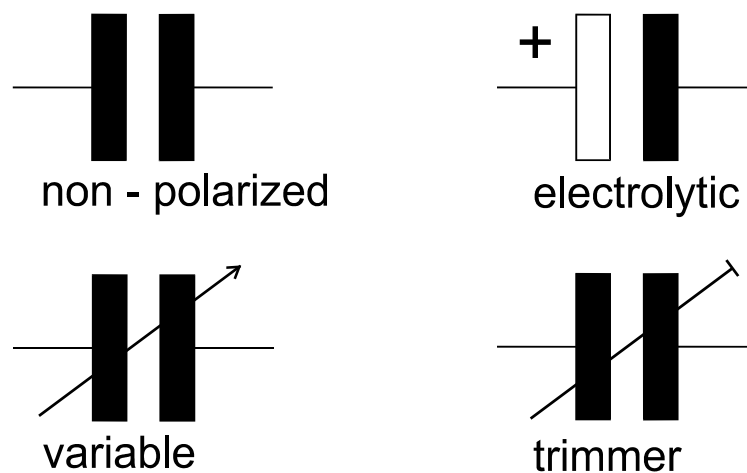


Figure 2.18
Symbols for Different Types of Capacitors

2.1.3.2 Types of Capacitors

The most common capacitor types are:

- **Paper Capacitors**

These are the most widely used type of capacitors. Their popularity is due to their low cost and wide range of values (from 500pF to 50μF). Their tolerances are relatively poor (10%-20%) and leakage currents can be a problem. The paper is usually soaked in substances such as oil, wax or plastic.

- **Plastic-film Capacitors**

A dielectric is coated from both sides with two layers of metal and wound into a roll. The dielectric can be polyester, mylar, or teflon and the coating is usually made from silver. The dielectric minimizes the leakage currents, but increase

the cost of the capacitor. The working voltages are 100 V or more. The values of plastic-film capacitors range from 500pF to 10 μ F. They are widely used in many areas of electronics, such as computers, radios and TV sets.

- **Ceramic Capacitors**

These are built on a manufacturing process where silver is vaporized to a ceramic plate. Usually the plate has the shape of a disk. The type of ceramic used will determine the properties of the capacitor. Some ceramic capacitors are made to withstand high voltages. Others provide a high capacitance. Ceramic capacitors have poor tolerances and are only suitable for applications, where approximate capacitance values are required.

- **Electrolytic Capacitors**

They are usually made from aluminum or tantalum. These capacitors provide the largest capacitance value per unit of volume. The basic structure of an aluminum electrolytic capacitor consists of two pieces of aluminum foil, one of which is coated with an extremely thin oxide. A paper, soaked in an electrolytic solution is placed between the two sections of foil. The electrolyte acts as a conductor. Only an extremely thin oxide film effectively separates the two oppositely charged plates. The two pieces of foil are rolled into a tube. It is necessary to emphasize that the electrolytic capacitor should only be connected in a circuit with its proper polarity. Failure to do so can destroy the capacitor or even cause an explosion.

- **Variable Capacitors**

The capacitance of a two-plate capacitor depends on the plate area, the distance between the plates, and the properties of the dielectric. By varying any of these parameters the value of capacitance will vary. This is used to build variable capacitors. One such common type is the **air-variable** capacitor. It is constructed by mounting a set of metal plates on a shaft and meshing them with another set of fixed metal plates. The air between them is the dielectric. This type of variable capacitor is often used for tuning in radios. The **trimmer** capacitor is also a variable capacitor, but it is adjustable in a narrower range. It is used primarily in circuits that need only one-off or infrequent tuning adjustments. Usually the value of their capacitance is adjusted with a screwdriver. Variable capacitors have a relatively small value of maximum capacitance, which is typically around 10 to 500 pF.

2.1.3.3 Color Coding for Capacitors

Most capacitor's values are marked on their cases and only the smallest of them use color codes. The nominal values of capacitors follow the same set of values, used for resistors, multiplied with the appropriate multipliers. Therefore Tables 2.2 and 2.3 can be used for both resistors and capacitors. Some differences occur only with the tolerance bands for capacitors. For capacitor tolerances Table 2.4 can be used. Figure 2.19 shows color coding for capacitors.

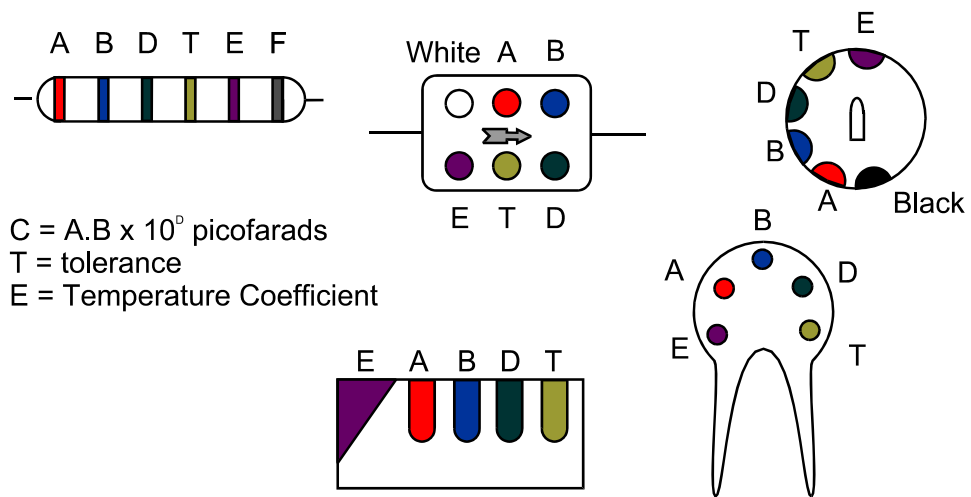


Figure 2.19
Color Coding for Capacitors

Table 2.4
Capacitors Tolerance Color Coding

Color	Tolerance
Black	20%
Brown	1%
Red	2%
Orange	3%
Yellow	4%
Green	5%
Blue	
Violet	
Grey	
White	
Gold	
Silver	

2.1.3.4 Capacitance Reactance

Similarly to inductors, the resistance of capacitors is frequency dependent. A capacitor opposes any changes in the *voltage* across it, and exhibits a **capacitance reactance**. A capacitance reactance has units of Ohms and is calculated from the formula:

$$X_C = \frac{1}{2\pi fC}$$

where:

X_C is the capacitance reactance (Ω),

f is the frequency (Hz),

C is the capacitance (F).

It is easy to see that a capacitor appears as an open circuit to a dc, while offering a minimal resistance to very high frequency alternating current.

2.1.3.5 Combining Capacitors

Figure 2.20 shows three capacitors, connected in series. The total capacitance is formed between the plates situated at the two extreme ends of the chain of capacitors connected together. For example the total capacitance between C_1 and C_2 in Figure 2.20 is formed on plates A and D. This effectively increases the distance between the plates of a capacitor, and as it was already said earlier in this section, the total capacitance will decrease. The equation for capacitors connected in series is as follows:

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

where:

C_T is the total capacitance in the circuit.

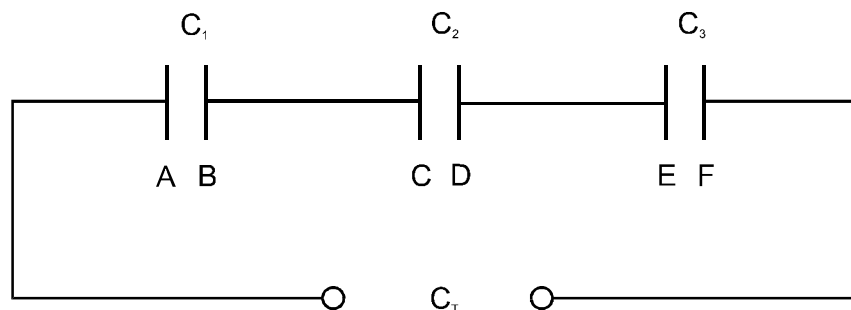


Figure 2.20
Capacitors in Series

Figure 2.21 shows three capacitors, connected in parallel. This effectively increases the size of the plates. As it was mentioned earlier in this chapter this is one of the ways to increase the total capacitance. The three capacitors in Figure 2.21 are equal to one capacitor, with two plates formed from A+C+E from one side and B+D+F from another. The total capacitance is given by the formula:

$$C_T = C_1 + C_2 + C_3$$

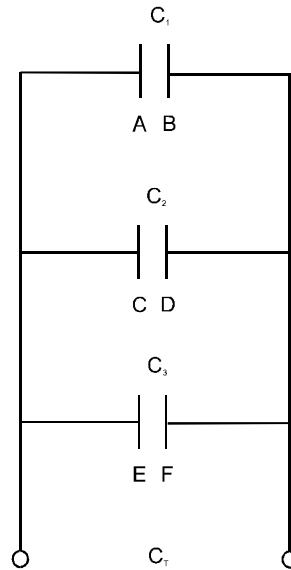


Figure 2.21
Capacitors in Parallel

2.1.4 Impedance

As was shown in the sections above, resistors, inductors and capacitors respond differently to variations in the frequency. Resistors are not frequency dependent. At high frequencies inductors exhibit higher resistance, while at the same time capacitors exhibit lower resistance. This is very important, because there are many circuits in electronics, where resistors, inductors and capacitors are connected together. Even single components (such as resistors, for example) still can be represented as equivalent circuits of resistors, inductors and capacitors connected together (**RLC circuits**). This can be explained with the fact that every single piece of wire possesses some inductance, and every two isolated wires that are at a close distance from each other form a capacitor. An RLC circuit, where all components are connected in series is shown in Figure 2.22.

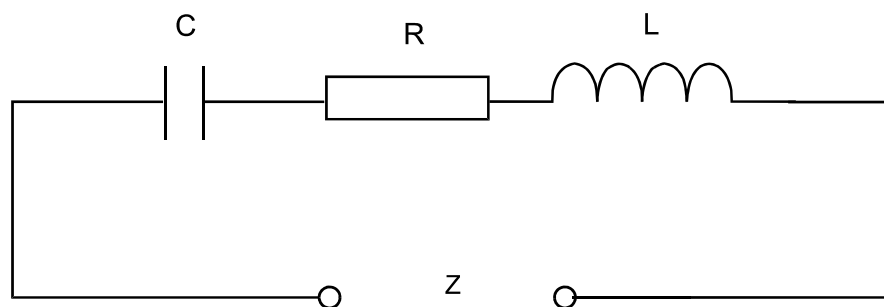


Figure 2.22
A Resistor, a Capacitor and an Inductor, Connected in Series

Ideal components do not exist. For example, a capacitor can be represented as an equivalent resistor and a capacitor connected in parallel. The reason for this is that the dielectric of the capacitor does not possess an infinite resistance. This causes some leak of current between the plates of the capacitor. Therefore, the dielectric acts as an equivalent resistor with a very high value. In fact, *everything* in electronics can be represented with such combination circuits. Since the frequency response of resistors, inductors and capacitors differs, the total frequency response of such combination circuits is a complex quantity.

The total resistance of a circuit to the alternating current is called **impedance**. The impedance is the combination of the ohmic resistance, the inductive reactance and the capacitive reactance of the circuit. It is measured in Ohms and it is denoted with the letter Z. The impedance of a circuit, comprising a resistor, an inductor and a capacitor, connected in series is given with the formula:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

where:

- Z is the impedance of the circuit,
- R is the resistance of the resistor,
- X_L is the inductive reactance of the inductor,
- X_C is the capacitive reactance of the capacitor.

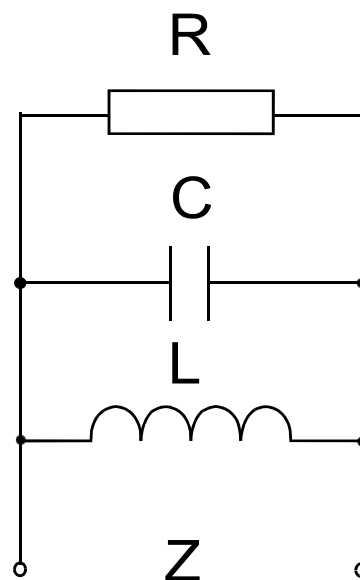


Figure 2.23
A Resistor, a Capacitor and an Inductor, Connected in Parallel

The impedance of the circuit, where the resistor, the inductor and the capacitor are connected in parallel (Figure 2.23) can be derived from the formula:

$$Z = \frac{R(X_L - X_C)}{\sqrt{R^2 + (X_L - X_C)^2}}$$

It is assumed in electronics that the inductive reactance is a positive quantity, while the capacitance reactance is a negative quantity. If there are several resistors, capacitors or inductors in the circuit, initially the equivalent impedance has to be found for each branch, and then the total impedance of the circuit. The calculations are performed, using the formulae above for impedance and the formulae given in the previous sections for resistors, capacitors and inductors, connected in series and in parallel.

Further consideration of impedance involves calculations with complex numbers. Such discussion goes beyond the scope of this manual and it is not provided here. However, what is important to remember is that if the circuit is supplied with an *alternating current*, the term impedance should be used, instead of resistance. Using the term impedance indicates that not only the ohmic resistance of the circuit is considered, but also its inductive and capacitance reactance for the frequency range of operation.

2.1.5 Summary

A resistor is a device made of material that conducts an electrical current, but possess a larger resistance, compared to the normal wires. The primary factor, for determining the resistance, is the material's resistivity. A resistor provides a resistance that is called ohmic, real or dc resistance. It is measured in Ohms and it is denoted with the letter R. Resistors are manufactured with standard nominal values, power ratings and tolerances. These parameters are often marked on resistors, using color codes. Color codes and standard values are obtainable from tables. The total resistance of several resistors, connected in series is always larger than the largest single value. The total resistance of several resistors, connected in parallel is always less than the smallest single value. The simplest inductor is made from a coil of wire. The resistance of an inductor is frequency dependent and it is called inductive reactance. It is measured in Ohms and it is denoted with the letters X_L . An inductor appears as an open circuit to a current with a very high frequency and offers no resistance to a direct current. Inductors or resistors connected in series or parallel are calculated with the same formula. A capacitor is a component, which is deliberately built to store a specific amount of electrical charge. The resistance of a capacitor is also frequency dependent and it is called capacitance reactance. It is measured in Ohms and it is denoted with the letters X_C . A capacitor appears as an open circuit to direct current and offers little or no resistance to a current with a very high frequency. In a converse manner to that of resistors and inductors, capacitors add their values, when connected in parallel. The total resistance of a circuit to the alternating current is called impedance. Using the term impedance indicates that not only the ohmic resistance of the circuit is considered, but also its inductive and capacitance reactance for the frequency range of operation.

2.2 Transformers, Diodes and Bridge Rectifiers

When you have completed study of this section you should be able to:

- Identify the transformer and the diode circuit symbols
- Describe the operation of transformers
- Describe the operation and bias of a PN junction
- Describe the operation of diodes
- List common types of diodes
- Describe the operation of bridge rectifiers

2.2.1 Transformers

In Section 2.1.2 we have discussed what inductance is, and we noted there that a wire wound around a material, such as iron or ferrite will greatly increase the magnetic flux density in this material. But what will happen if not one, but two coils are wound on the same former, as is shown in Figure 2.24?

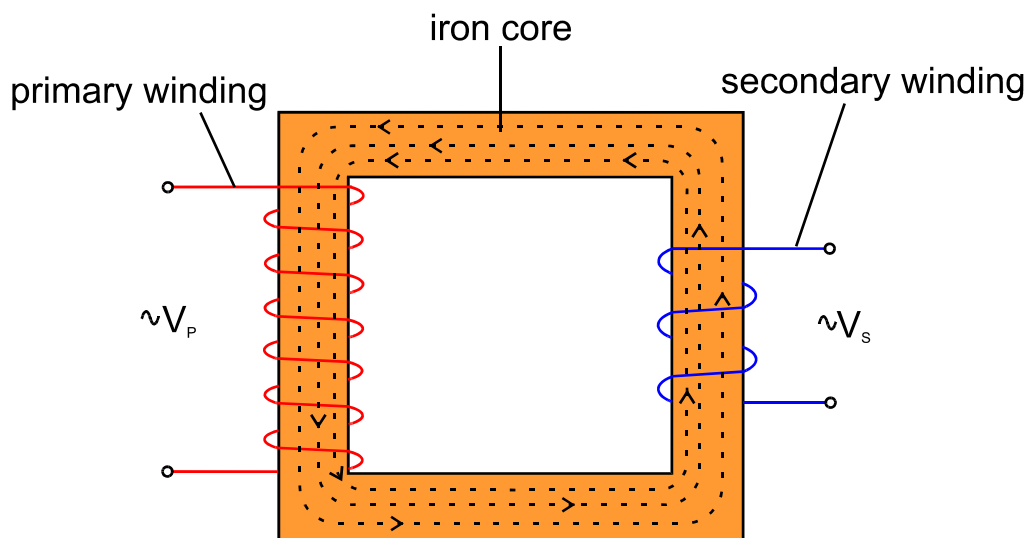


Figure 2.24
Transformer

If an alternating current flows in one of the windings, referred to as the **primary winding**, it creates a magnetic flux, which is concentrated in the iron core. The reason why this happens is that the resistance of the air is much greater for the magnetic flux, compared to that of an iron. The magnetic flux in the iron core induces an emf into the other winding, called the **secondary**.

If the current in the primary winding is dc it will not affect the secondary winding, as there are no changes in the magnetic field. If the current in the primary is ac, the

magnetic field will change. A changing magnetic field will appear across the secondary winding. Since there are a lot of free electrons in the conductor they will experience the force of the changing magnetic field and a current in the secondary winding will start flowing. As a result a current that flows in the primary winding will induce an emf into the secondary winding.

The electrical phenomenon, when a magnetic field from one element influences other elements in its vicinity is called **mutual inductance**. A **transformer** operates on this principle. Obviously the value of the induced voltage in the secondary winding depends on the rate of change of current in the primary (i.e. on its frequency). Another factor contributing to the value of the induced voltage is the turns ratio. The **turns ratio** expresses the number of turns on the primary winding to the number of turns on the secondary winding, and it is written as:

$$\text{turns ratio} = \frac{N_p}{N_s}$$

where:

N_p is the number of turns on the primary winding

N_s is the number of turns on the secondary winding

The turns ratio is a very important quantity and determines how much a transformer can step up or step down a voltage. It is good to know that when the voltage is stepped down, the current is stepped up and vice versa. This will be discussed further in the next chapter.

The turns ratio is also equal to the voltage across the primary and the voltage across the secondary:

$$\text{turns ratio} = \frac{N_p}{N_s} = \frac{V_p}{V_s}$$

where:

V_p is the voltage across the primary winding,

V_s is the voltage across the secondary winding.

Transformers are not solely used for stepping the voltage up or down. They can also connect two parts of a circuit, without the necessity to connect them *electrically*. Since only the magnetic field actually links the two parts, they can remain *electrically* isolated from each other. This is a very useful property, especially in applications, where a dc voltage has to be restricted to one part of the system.

The two most common transformers are the high-frequency pulse transformers and the power transformers. High-frequency pulse transformers are used for coupling

ac signals while isolating the dc voltages. Power transformers are used to step up or step down voltages to the appropriate level for different parts of the equipment.

The electrical symbol of a transformer is given in Figure 2.25. The winding can be divided into several sections. The lead dividing a winding into sections is called a **tap**, and can be connected anywhere along the length of the winding.

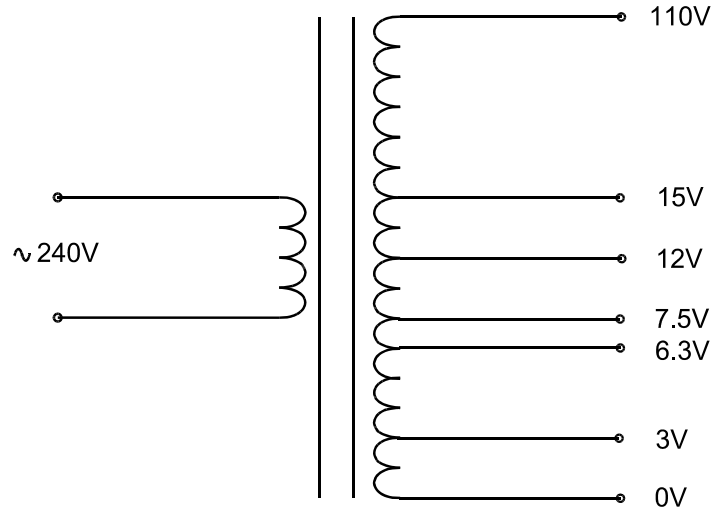


Figure 2.25
Transformer Circuit Symbol

2.2.2 Diodes

2.2.2.1 PN Junction

We already know from Section 1.2.3 that a semiconductor's conductivity can be substantially increased through doping. Adding pentavalent impurity atoms produce an n-type semiconductor, rich of free electrons. Adding trivalent impurity atoms produces a p-type semiconductor, containing many holes. Let us see now what will happen if one half of an intrinsic semiconductor is doped with a trivalent material and the other half with a pentavalent material. This is shown in Figure 2.26.

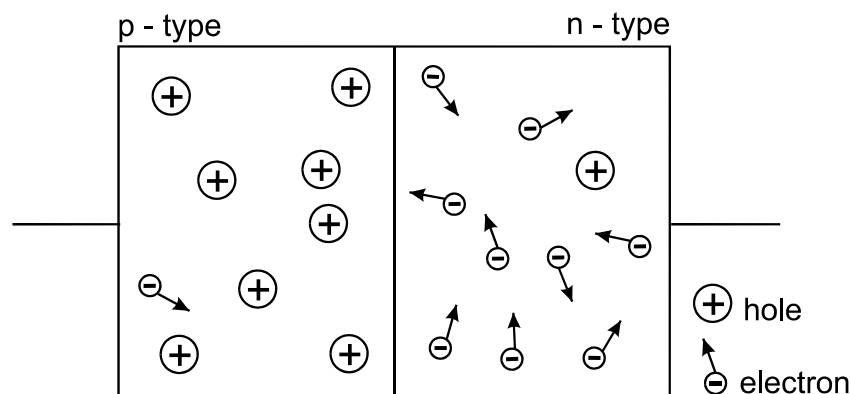


Figure 2.26
PN Structure

When a bar of semiconductor material is doped in such manner that one of its halves is a n-type and the other half is a p-type, a **pn junction** is formed between the two halves. The n-region contains many free electrons and they are attracted towards the holes in the p-region. Then for an instant, the electrons diffuse across the junction and combine with the holes from the p-region. When they do this, they leave behind more and more positively charged ions. The same happens with the holes. They leave behind more and more negatively charged ions. As a result very quickly a potential difference is formed on the border of the pn junction. The value of this potential is approximately 0.7 V for a silicon pn junction and around 0.2 V for a germanium one. This potential is called a **barrier potential** and is sufficient to prevent any more electrons or holes from crossing the border. The two layers of positive and negative charges near the junction are called **depletion region**. A depletion region is formed very quickly and it is very thin. Figure 2.27 shows a pn junction with an exaggerated depletion region for illustration purposes:

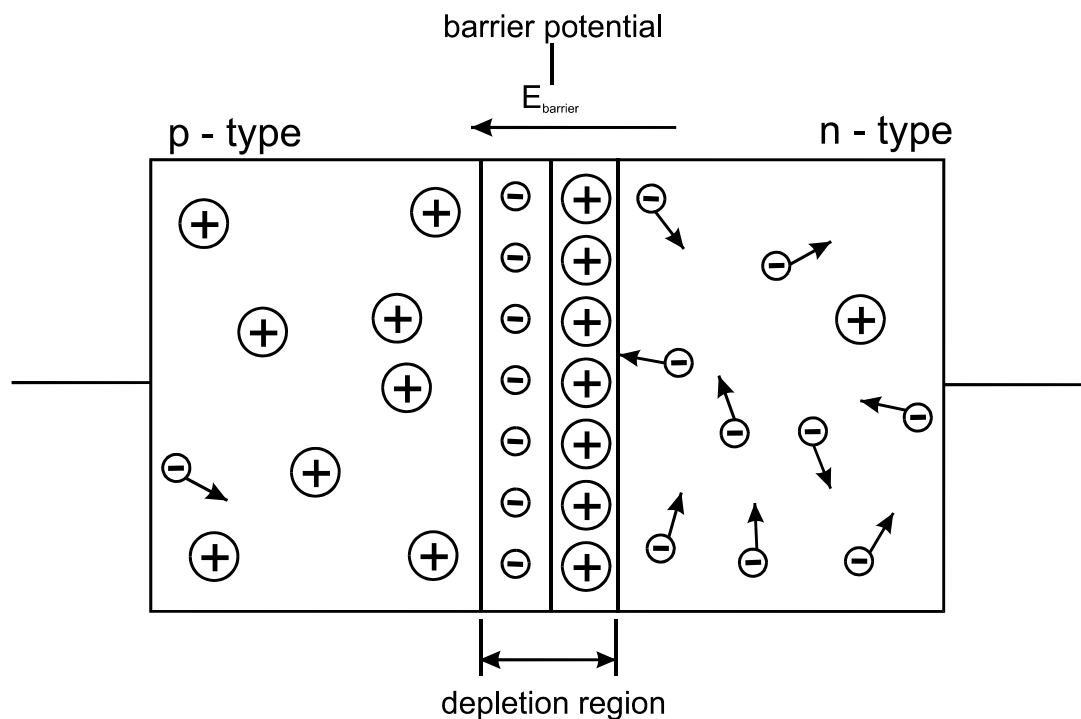


Figure 2.27
PN Junction

2.2.2.2 Biasing the pn Junction

In Figure 2.28 an external voltage V_S is applied to the pn junction. The negative terminal of V_S is connected to the n region of the pn junction and the positive terminal to the p region. In this case the external potential opposes the barrier potential. If V_S is greater than V_{barrier} the effect of the latter will be overcome and the electrons will flow through the junction. The fixed voltage that *allows* current through the pn junction is called the **forward bias**. Two conditions have to be satisfied for forward biasing:

- The negative terminal of the voltage supply has to be connected to the n region of the pn junction, and the positive terminal of the voltage supply has to be connected to the p region of the pn junction.
- The value of the voltage supply has to be greater than the barrier potential, which is approximately 0.7V for a silicon pn junction and around 0.2V for a germanium pn junction.

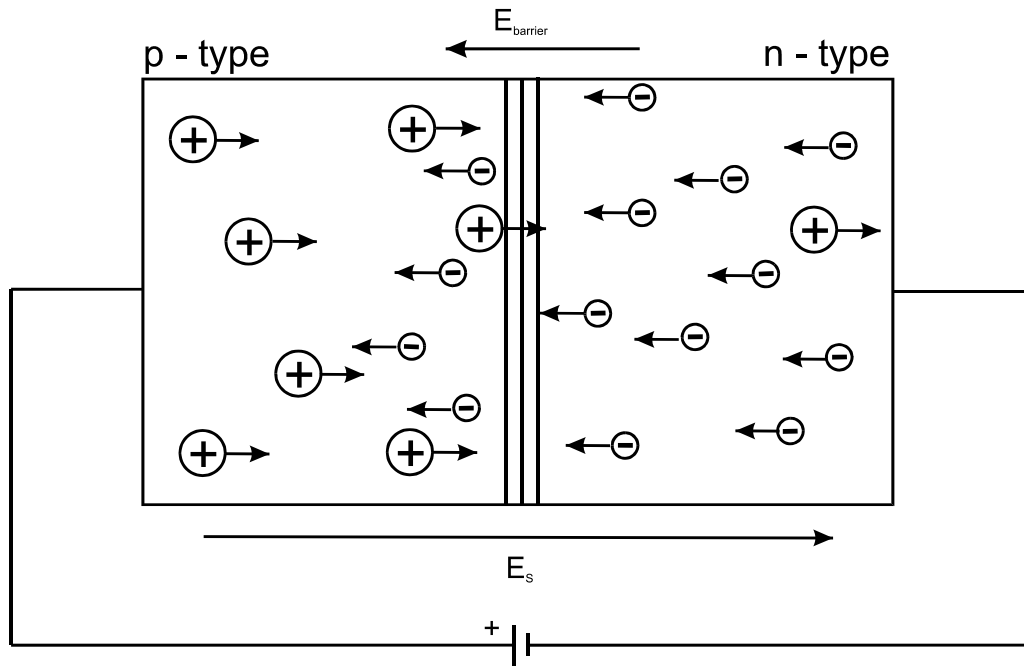


Figure 2.28
A Forward-biased pn junction

In Figure 2.29 the external voltage is applied to the same pn junction, but in the opposite direction. In this case the supply voltage enhances the effect of the barrier voltage. The depletion region becomes wider and makes the electron flow even more difficult. The fixed voltage that *prevents* current flow through the pn junction is called the **reversed bias**. If the value of the reversed bias is increased greatly, a point will be reached where the flow of electrons will be forced in a reverse direction. This condition is dangerous for the pn junction and usually destroys it, through extreme heat dissipation. The point in which the reverse voltage becomes so great that the current increases dramatically is called a **breakdown voltage**. Manufacturers specify this voltage as a **peak inverse voltage** or abbreviated **PIV**. Two conditions have to be satisfied for reversed biasing:

- The negative terminal of the voltage supply has to be connected to the p region of the pn junction, and the positive terminal of the voltage supply has to be connected to the n region of the pn junction.
- The reverse bias has to be less than the breakdown voltage.

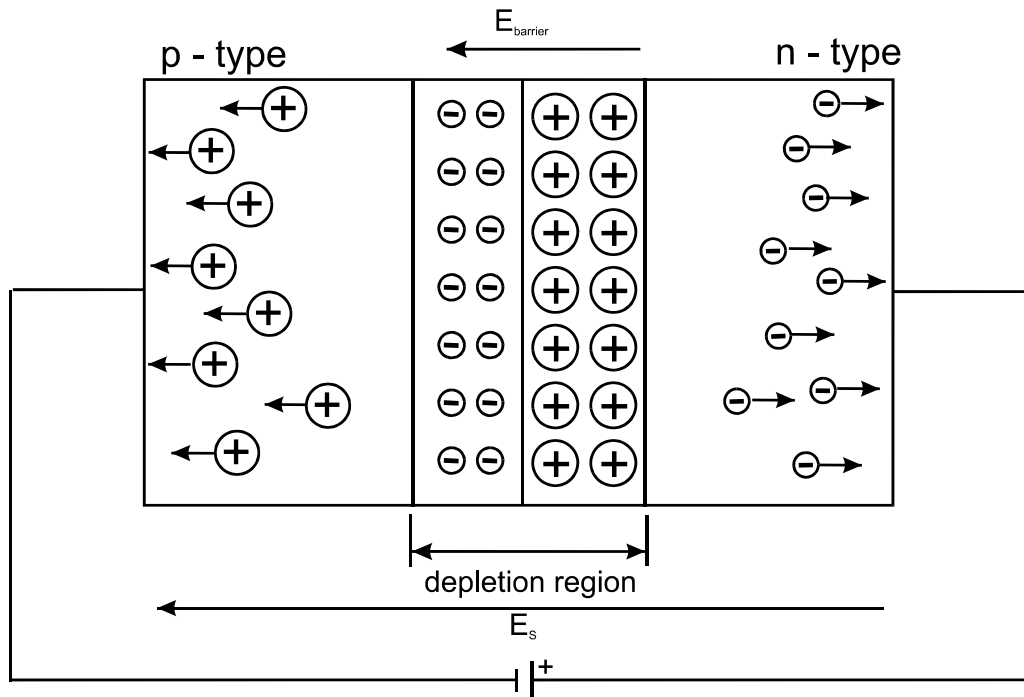


Figure 2.29
A Reverse-biased pn junction

2.2.2.3 A Diode

A **diode** is a single pn junction device, with two terminals connected to each region. The p region of a diode is called the **anode**, abbreviated as **A** and the n region is called the **cathode**, abbreviated as **K**. The forward and reverse bias of a diode is identical to that of a pn junction. When the positive terminal of a voltage source is connected to the anode of the diode and the negative terminal is connected to the cathode through a current-limiting resistor, the diode is forward-biased and there is a current flow. When the polarity of the voltage source is reversed the diode is reverse-biased and there is no current flow in the circuit. The resistor in the second case is not necessary. The symbol of the diode and the two methods of connection are shown in Figure 2.30.

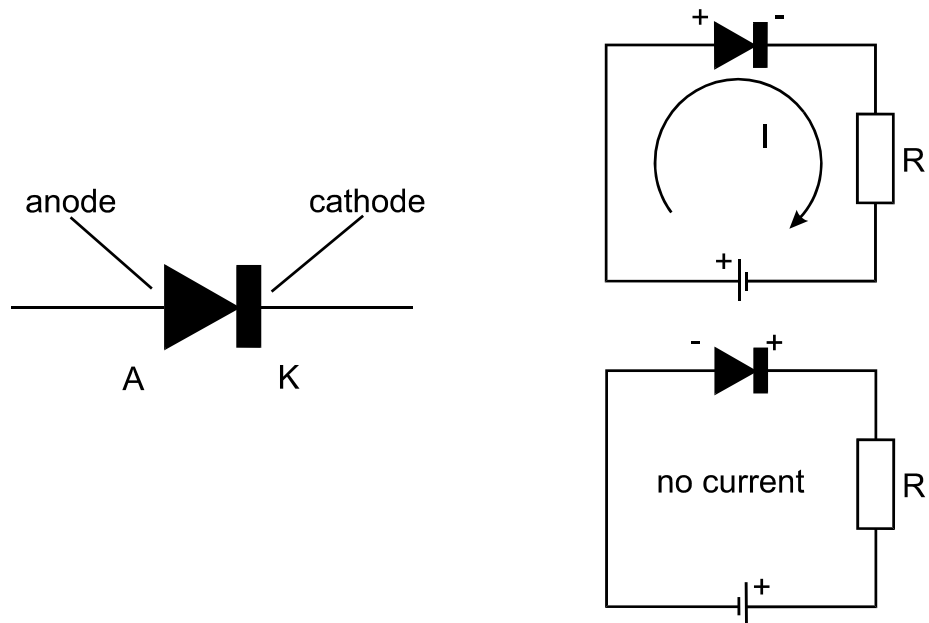


Figure 2.30
Forward and Reverse Bias of a Diode

One may regard the diode as a simple switch. When forward-biased it acts as a closed switch and when reverse-biased it acts as an open switch. This is true, but for the so-called ideal diode. The real diode has a characteristic that differs slightly from that of the ideal diode. Figure 2.31 shows the current-versus-voltage characteristic of a real diode.

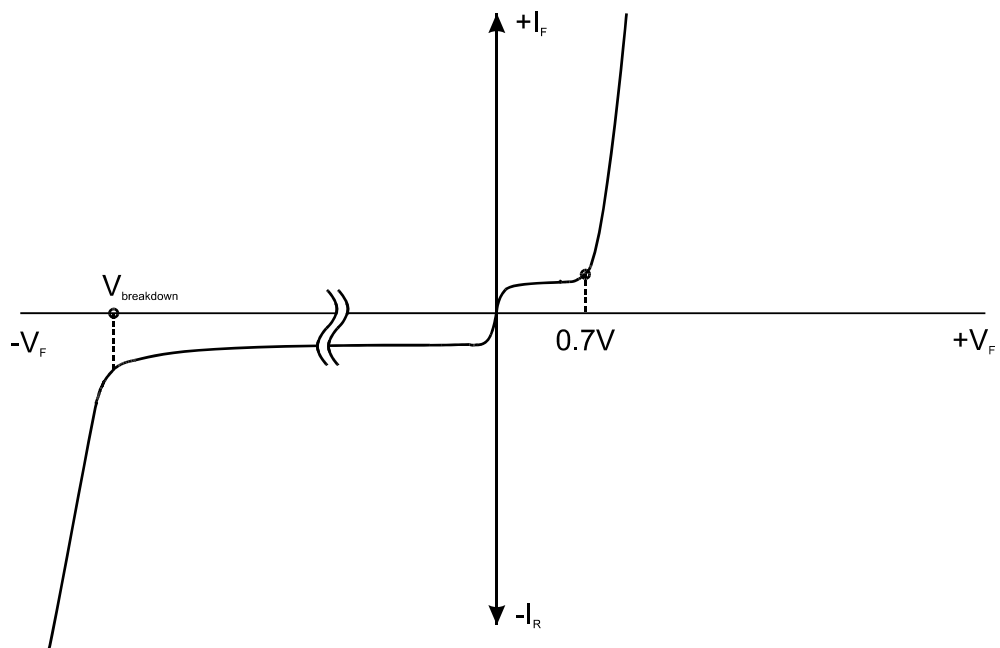


Figure 2.31
Current-versus-Voltage Characteristic of a Diode

When the diode is forward-biased it will act as a closed switch, allowing the current flow. The current does not start increasing at 0 V, as an ideal switch would do, because the barrier potential has to be overcome first. For this reason, the current starts increasing after 0.2V or 0.7 V, depending on the type of junction. When the diode is reverse-biased, the curve does not follow the horizontal axis precisely, as there is a small reverse current, flowing through the junction. As the reverse-bias is gradually increased, it reaches the breakdown value, at which point the reverse current begins to increase rapidly. Breakdown (with a few exceptions) is not a diode's normal mode of operation and has to be avoided.

2.2.2.4 Characteristics of Diodes

Diodes are manufactured in various shapes and sizes, but usually they are divided into two groups of signal diodes and power diodes. Regardless of the group, to which they belong, there are several important characteristics of diodes that determine their applications. These are:

- **The forward current I_F** is the value of the maximum current that the diode can handle, when forward-biased. The power diodes usually can handle currents of 1A and more, while the signal diodes usually cannot pass more than 100 mA. It is important not to exceed the forward current limit, as the diode may overheat and eventually fail.
- **The reversed current I_R** is the value of the current that flows, when the diode is reverse-biased. It flows in the opposite direction and has a very small value, which is thousands of times less than the forward current. The value of the reversed current depends on the ambient temperature and on the properties of the semiconductor. The reversed current of the signal diodes is usually measured in μA , while those of a power diodes is several nA.
- **The peak-inverse-voltage (PIV)** is the value of the maximum voltage that the diode can withstand, when reverse-biased, before a breakdown occurs. This rating depends on the diode's construction and can vary greatly for the different types of diodes. Usually the PIV of the power diodes is rated at around 400V to 1kV, while the PIV of the signal diodes rarely exceeds 100V.

2.2.3 Bridge Rectifiers

Most items of electronic equipment that operate from the domestic ac power supply require dc at a specified voltage. The process of converting the ac voltage into dc is called **rectification**. In this section we will discuss how this is done.

2.2.3.1 Half-wave Rectifier

Figure 2.32 shows a transformer connected to the mains power supply. A diode and a load resistor are connected in series to the secondary winding of the transformer. V_s is the secondary voltage from the transformer, which is 12 V in our example. The

mains power supply with a rms value of 240V will induce an inverted waveform in the secondary winding with a rms value of 12V.

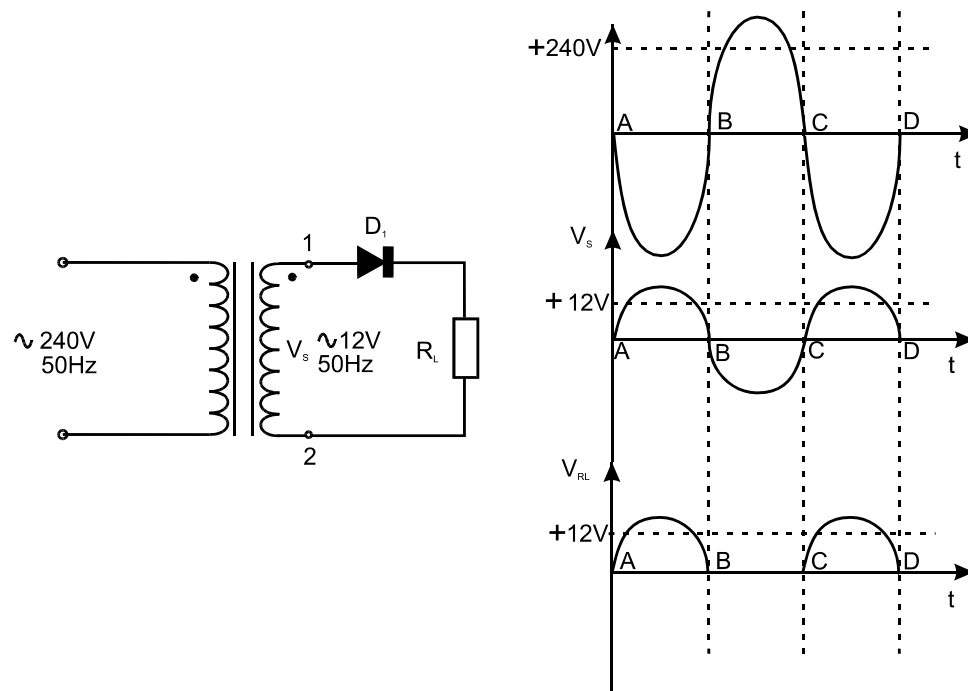


Figure 2.32
Half-wave Rectifier

As shown in the graphics, during the first half-period of the secondary voltage, point 1 is positive with respect to point 2. This means that the diode is forward-biased and the current flows through the resistive load. The voltage on the resistor will follow the voltage in the secondary winding.

During the second half period point 1 is negative with respect to point 2. The diode is reversed-biased and no current flows through the resistor. Therefore the current through the load follows the waveform of the secondary winding during the positive half-cycles only. If the diode connection is reversed, the current through the load will follow the waveform of the secondary winding during the *negative* half-cycles only. In either case, the voltage across the resistor has a half-sine waveform, but it is a direct current, as it does not change its polarity.

The term **half-wave rectification** comes from the fact that a voltage on the load is produced only during half of the time, i.e. only during the positive or negative half-cycles.

2.2.3.2 Full-wave Bridge Rectifier

Obviously, the half-wave rectifier is only half-efficient as it uses the energy just from the positive or the negative half-cycles of the sine wave. A circuit that overcomes this disadvantage is shown in Figure 2.33. It is called a full-wave bridge rectifier.

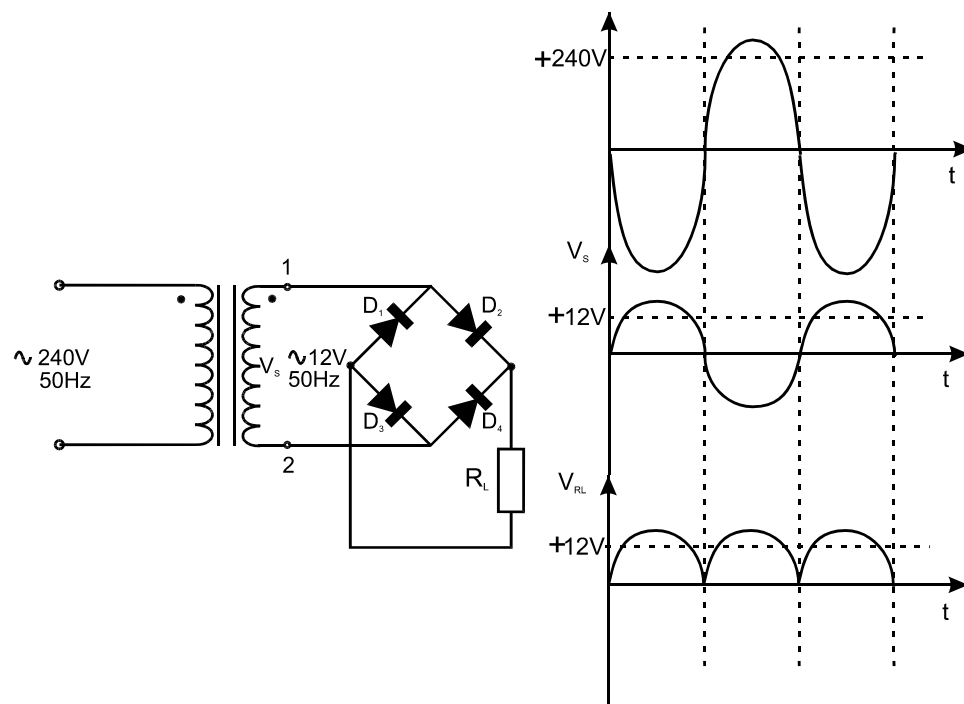


Figure 2.33
Full-wave Bridge Rectifier

It requires four diodes, but it is much more efficient than the half-wave rectifier. When point 1 is positive with respect to point 2 diodes D_1 and D_4 are biased off and they will appear as an open circuit to the current. Therefore the current passes from point 1 through forward-biased D_2 , flows through the load resistor, passes through the forward-biased D_3 , passes point 2 and flows through the secondary winding back to point 1. Thus the circuit is closed and the positive half-wave of the secondary voltage is applied to the resistor.

When point 1 is negative with respect to point 2 diodes D_2 and D_3 are biased off and they appear as an open circuit to the current. In this case the current starts flowing from point 2, passes through the forward-biased D_4 , flows through the load resistor, passes through the forward-biased D_1 , passes point 1 and through the secondary winding flows back to point 2. The circuit is closed. The negative half-wave of the secondary voltage is applied as a positive half-wave on the resistor. This is so, because in both cases the current through the load resistor flows in the same direction.

Reversing the points 1 and 2 in the circuit does not change the polarity on the load resistor. The only way to reverse the output polarity is to disconnect all four diodes and connect them again, all with reversed polarity. You can check this yourself.

Bridge rectifiers are common for many applications. For this reason most component manufacturers supply them in a package, containing all four diodes in one device. They are available in many different voltage and power ratings.

2.2.4 Summary

A transformer is a device consisting of two windings, wound on the same former. They are called a primary winding and a secondary winding. The transformer operates on the principle of mutual inductance, where current flowing in the primary winding induces voltage in the secondary winding. The ratio that expresses the number of turns on the primary winding to the number of turns on the secondary winding is called the turns ratio. This ratio also determines how much a transformer can step up or step down a voltage. A pn junction is formed in the middle of a semiconductor bar, one half of which is a p type and the other half is a n type. A diode is a single pn junction device, with a lead connected to each half. When the positive terminal of a voltage source is connected to the anode of the diode (p type region) and the negative terminal connected to the cathode (n type region) the diode is forward-biased and there is a current flow. When the polarity of the voltage source is reversed the diode is reversed-biased and there is no current flow. The process of converting the ac voltage into dc is called rectification. A full wave bridge rectifier employs four diodes and is the most efficient rectifier circuit.

2.3 Bipolar Junction Transistors, JFETs and MOSFETs

When you have completed study of this section you should be able to:

- Identify symbols of bipolar junction transistors, JFETs and MOSFETs
- Describe their construction
- Describe their operation
- Identify and describe transistor bias
- List transistor parameters

2.3.1 Bipolar Junction Transistor

2.3.1.1 Transistors Construction and Symbols

The bipolar junction transistor (abbreviated **BJT**) consists of three regions of doped semiconductor, forming two pn junctions. There are two possible combinations, shown in Figure 2.34. When an n-type region separates two p-type regions the transistor is called a **pn_p transistor**. Conversely, when a p-type region separates two n-type regions the transistor is called a **np_n transistor**. Wire terminals are connected to each region. The semiconductor and lead connections are covered with an insulation material.

Transistor terminals are referred to as **emitter**, **base** and **collector**. They are designated with the letters **E**, **B** and **C** respectively. The emitter and collector are not identical and cannot be interchanged randomly, as there are physical differences between them. Usually the collector is much larger physically than the emitter. Also it is only moderately doped with impurities, while the emitter is very heavily

doped. The base is always thinner than the emitter and collector, and it is only lightly doped. This is essential for the good operation of the transistor. Therefore, simply connecting two diodes together cannot make a transistor. The reasons for this will be discussed later in this section.

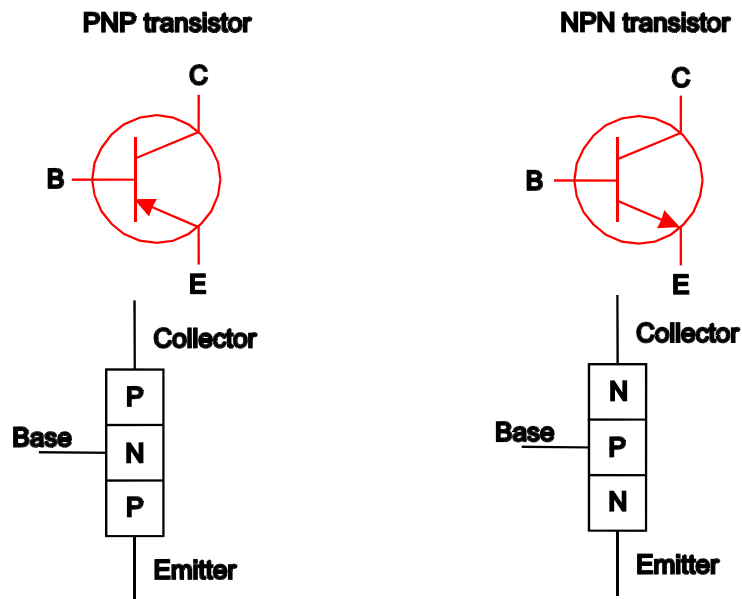


Figure 2.34
NPN and PNP Type Transistors

A transistor can operate in two basic modes - as a switch and as an amplifier. When operating as a switch, the current through the transistor is either flowing at its maximum value or it is not flowing at all. A small input current controls the output current. When the transistor operates as an amplifier, small changes in the input current result in large changes in the output current.

2.3.1.2 A Transistor as an Amplifier

The theory describing the amplifying effect of the transistor is a rather complex topic, though the effect of it is very simple. Here we will provide a basic explanation, concentrating on *how* things happen, instead of delving too much into the theory of *why* things happen.

A npn transistor connected as an amplifier is shown in Figure 2.35. A forward-biased voltage called V_{BE} , which is greater than 0.7 V is applied to the base-emitter junction. We already know from Section 2.2.2.2 that this is the threshold, where the barrier potential is overcome, and therefore the base-emitter junction is forward-biased. A current, called **base current (I_B)** flows through the junction. The base is only lightly doped and the number of vacant holes in the base is limited. For this reason, I_B remains very small. Another, much higher voltage is applied between the emitter and the collector. It is called V_{CE} . The voltage on the collector is more positive, than the base. The base appears as negative compared to the collector, and the base-collector junction is therefore reverse-biased.

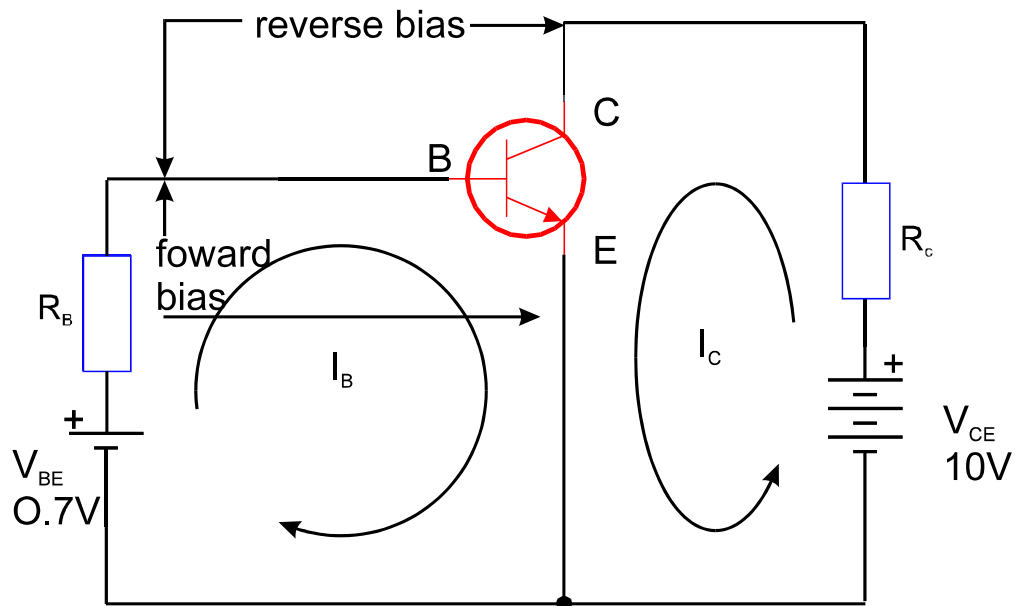


Figure 2.35
Collector Characteristic Curves

When electrons flow from the emitter to the base, they experience the large positive potential V_{CE} . They are attracted to it and cross the base-collector junction. Thus, they form a current called the **collector current (I_C)**. Almost all electrons are attracted from the large positive potential V_{CE} , and only a small amount of them are diverted to the base and form a base current. If the base current varies, the collector current also varies, following exactly the changes in the base current. The base current is small, while the collector current is large. Therefore, small changes in the small base current lead to large changes in the large collector current. In this way, an **amplifying** device is created.

As you can see from Figure.2.35 the current of the npn transistor flows *out* of the emitter. The emitter of the transistor symbol is an arrow showing the direction of conventional current flow. For this reason a npn transistor is depicted with an emitter arrow pointing outwards, while the emitter arrow of a pnp transistor points inwards. The principle of operation of the pnp transistor is absolutely identical to the principle of operation of the npn transistor, with the only difference that it operates with reversed polarities and directions of current flow.

The ratio of the dc collector current I_C to the dc base current I_B is called **dc beta**. This characteristic determines the **current gain** of the transistor. It is denoted with β_{DC} . It is quite common to see this parameter, given in the transistor data sheets as an equivalent hybrid parameter h_{FE} . What is important to remember is that:

$$h_{FE} = \beta_{DC} = \frac{I_C}{I_B}$$

where:

h_{FE} is an equivalent hybrid parameter = β_{DC} ,

β_{DC} is the dc current gain of the transistor,
 I_C is the dc collector current,
 I_B is the dc base current.

The ratio of dc collector current I_C to the dc emitter current I_E is called **dc alpha** and it is denoted with α .

$$\alpha_{DC} = \frac{I_C}{I_E}$$

The values of α_{DC} are typically in the range of 0.95 to 0.99 or greater, but they are always less than 1. The relationship between β_{DC} and α_{DC} is given with the formula:

$$\beta_{DC} = \frac{\alpha_{DC}}{1 - \alpha_{DC}}$$

2.3.1.3 A Transistor as a Switch

A simple circuit, as in Figure 2.36 can be used to generate a set of collector characteristic curves of transistor. The base current I_B is kept constant at specified values, and the variations of the collector current I_C are plotted against the collector-to-emitter voltage V_{CE} .

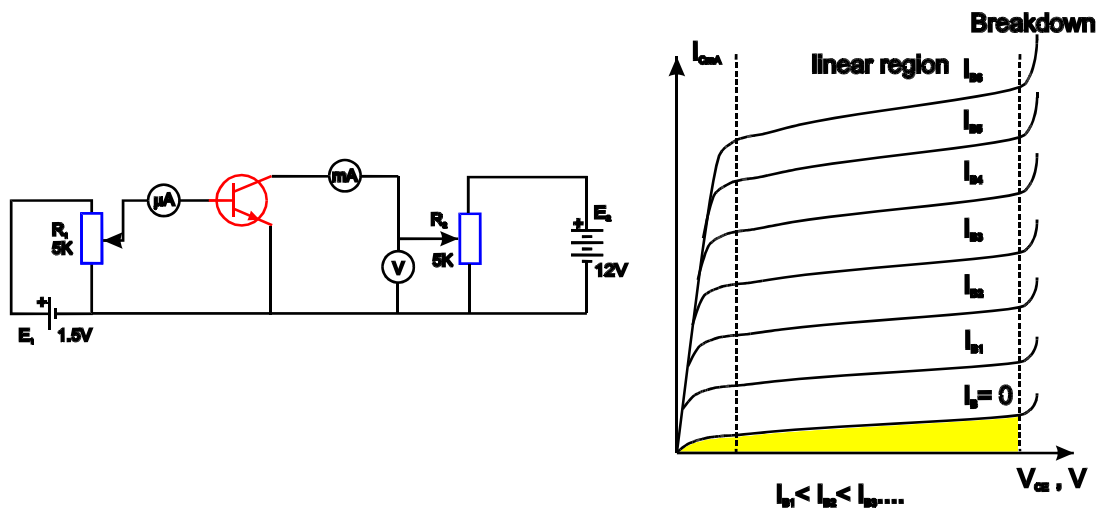


Figure 2.36
Collector Characteristic Curves

If $I_B = 0$, there is practically no collector current and the transistor is in the cutoff region. If V_{CE} exceeds 0.7 V the base-collector junction becomes forward-biased and the transistor goes into the linear (or active) region of operation. In this mode, for a given constant value of I_B the level of I_C remains essentially constant with the increase of V_{CE} . In fact, I_C increases slightly. This can be explained with the fact that when V_{CE} becomes higher, the base becomes narrower. Narrower base means fewer holes for recombination and greater I_C . Thus β_{DC} is also increased.

If the value of V_{CE} reaches a certain point, the collector current increases rapidly and the transistor goes into breakdown. In this case, it can be damaged permanently due to overheating. A transistor should never be operated in the breakdown region.

The collector characteristic curves are generated with different values of I_B . According to the formula $I_C = \beta I_B$, the greater the value of I_B , the greater the value of I_C . This is true up to a certain point, called a point of **saturation**, where further increases in I_B does not produce any further increases in I_C , and the formula above is no longer valid. The value of the collector-emitter voltage at which I_C stops to increase further, and remains essentially constant is called the **voltage of saturation**, and is denoted with $V_{CE(sat)}$. For silicon transistors $V_{CE(sat)}$ is usually only a few tenths of a volt. On the other side, when the I_B is zero, the I_C is also approximately zero, even at the maximum value of V_{CE} . This condition is known as **cutoff**.

The curve that connects the saturation point with a cutoff point is called the **load curve**. In other words, this is the straight line, which connects the maximum value of the output current with the maximum value of the output voltage. A transistor is always operated between these two points of the load curve. The load curve is drawn in Figure 2.37.

In Figure 2.36 the V_{BE} can be regulated in such manner that only two values of I_B exist, causing the transistor to operate only in two extreme modes - cutoff and saturation. When in cutoff mode, there is no collector current and the transistor behaves as an open switch between its emitter and collector terminals. When in saturation mode, the base-emitter junction and the base-collector junction are forward biased and the collector current is saturated. In this condition, the transistor appears as a short cut between its emitter and collector terminals. Therefore, the availability of a small base current in the input determines the availability of a much larger collector current at the output. In other words, the transistor acts as a switch.

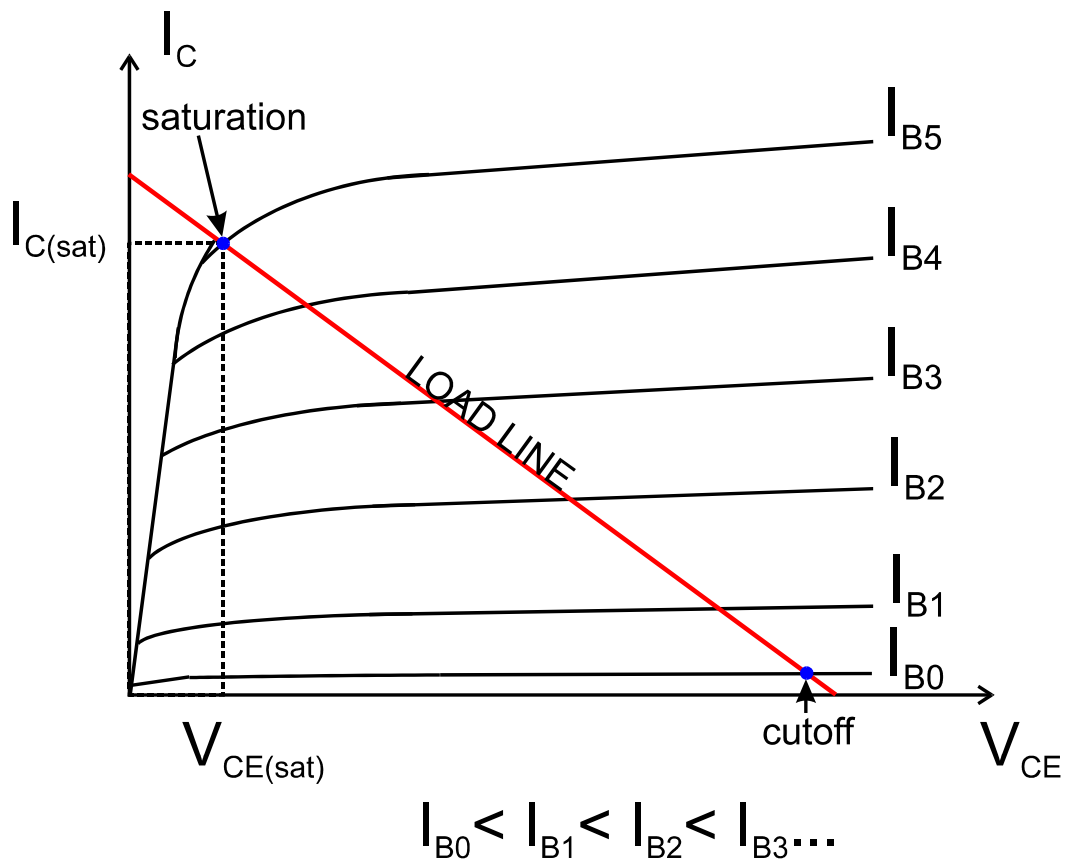


Figure 2.37
Cutoff and Saturation Conditions

2.3.1.4 Transistor Characteristics

The most obvious difference between transistors is their physical size and type of case. It is logical to assume that bigger transistors are more powerful. It is true, but the best way to get more information about a specific transistor is to refer to the transistor data sheets. The most common specifications, obtainable from data sheets are specified below:

- **Type of semiconductor used.**

Most transistors manufactured nowadays are made from silicon, though some applications still use germanium-type devices. The material of the semiconductor can be easily checked by looking at V_{EBO} (emitter-base voltage, open collector), which should be approximately 0.2 V for germanium transistors, and around 0.7 V for silicon transistors. Also the base-emitter saturation voltage (denoted as $V_{BE(sat)}$) of the silicon transistors is higher, than the germanium one and is typically several tenths of a volt. In any case, it is always safer to assume that the transistor is a silicon type, if this is not specified in the data sheets.

- **Type of junctions.**

A transistor can be either a pnp or npn type. Most transistors nowadays are manufactured as an npn type. This parameter is important to know in order to

supply the transistor with the correct polarity. Usually it is always specified in the data sheets, but if this is not the case, it can be checked with a simple ohmmeter. How this is done is discussed in Chapter 8.

- **Current gain β_{DC} .**

This is one of the most important transistor parameters, as it tells us how many times greater the output current is, compared to the input current. Usually it is specified in the data sheets with two figures, stating the lowest and the highest limit of the value. A typical range can be somewhere between 50 and 800. Transistor manufacturing technologies today are still imperfect, and cannot guarantee that two transistors even of the same type will have the same current gain. However, the actual value of h_{FE} is not so important, providing that it exceeds 100. We will discuss this further in Chapter 4.

- **Maximum collector current I_{Cmax} .**

This is the maximum collector current that the transistor can handle. It flows in the emitter-collector circuit and its value is typically between 50mA and 1A. I_C can reach even 15 A for large transistors. It is important not to exceed the specified maximum collector current, as the transistor goes in a breakdown region and can be damaged permanently.

- **Maximum collector-emitter voltage V_{CEOmax} .**

The letter “o” in the abbreviation means that the maximum voltage between the emitter and the collector is measured with the base open circuit. It is typically between 10 and 100 V, but can reach several thousand volts for some of the more powerful transistors.

- **Maximum frequency of operation.**

As the frequency becomes higher, the capacitance of the pn junction becomes increasingly significant, as it tends to shortcut the signal (Recall that a capacitor has a very low resistance at high frequencies). Therefore, with the increase of frequency, the current gain decreases. Data sheets often specify a parameter called f_T , which is the frequency at which the current gain becomes equal to 1. If for example f_T is specified as 100 MHz, it does not mean that the transistor can actually operate without any problems up to this frequency. Most of the time the transistor can only be used if the current gain does not drop below 100. Therefore, the frequency at which the transistor will still have an acceptable current gain is probably somewhere around 1MHz.

2.3.2 JFETs and MOSFETs

Field-effect transistors (or abbreviated **FET**) look physically very similar to the traditional bipolar junction transistors. For this reason, it is very difficult to distinguish between them at the first sight, and the only certain way to determine which one is which is to check the part number on the package in the data book. FETs are divided into two major categories. The first category is called **junction**

field-effect transistors (JFET) and the second one is known as **metal oxide semiconductor field-effect transistors (MOSFET)**. The second category is also sometimes called **IGFET**, where **IG** stands for "**insulated gate**". Each category can be divided into subcategories and this is discussed further in this section.

2.3.2.1 Construction and Operation of JFETs

Though similar at first sight, JFETs are totally different in their operation from the bipolar junction transistors. The main difference is that the BJTs are *current operated*, while the JFETs are *voltage operated*. In a BJT an input current controls the output current, while in a JFET an input voltage controls the output current, which makes it a totally different device. Similarly, to the npn and the pnp type BJTs, JFETs are also manufactured in two types: as a **n-channel** and as a **p channel**. The construction of an n-channel JFET transistor is shown in Figure 2.38.

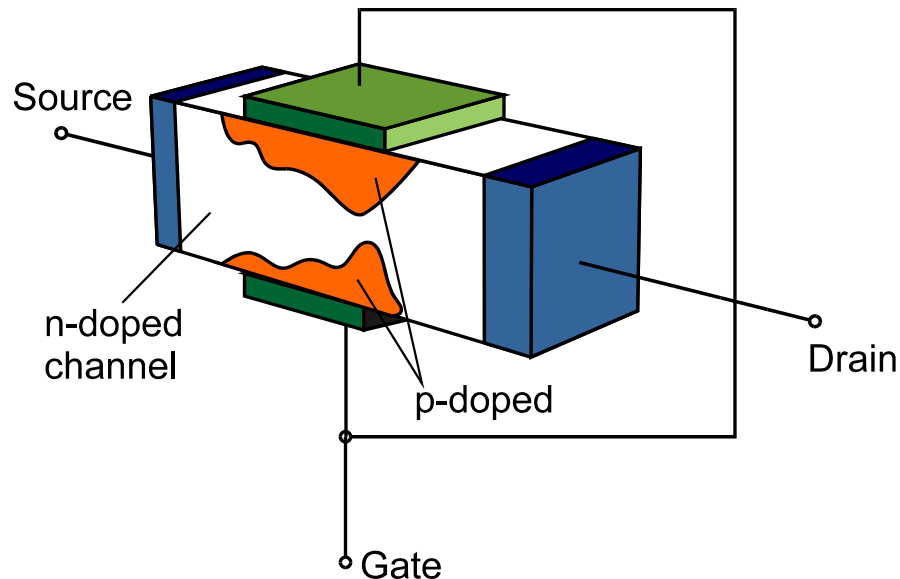


Figure 2.38
Construction of a JFETs Transistor

We will discuss the construction and principle of operation of a n-channel JFET transistor, bearing in mind that the construction and the principle of operation of a p-channel JFET transistor is absolutely identical. The only differences between the two types are the opposite type of semiconductor materials used, and the reversed polarities of the voltages around the devices.

The JFET transistor consists of a channel of doped semiconductor material. The channel can be n-doped or p-doped and determines the type of the transistor. The transistor, shown in Figure 2.38 is an n-channel JFET transistor. Wire terminals are connected at both ends of the channel. One of these leads is called the **drain**, and the other is called the **source**. A second piece of semiconductor material, which has an opposite polarity, surrounds the middle of the channel, forming essentially a pn junction. The terminal connected to this second piece of semiconductor is called **gate**.

The symbols of the n-channel and p-channel JFET transistors are shown in Figure 2.39. The symbols marked with "preferred Australian standard" are recommended for use in Australia, but often both standards are used. Similarly to the BJTs, the gate of the JFETs is depicted as an arrow, showing the direction of the current flow.

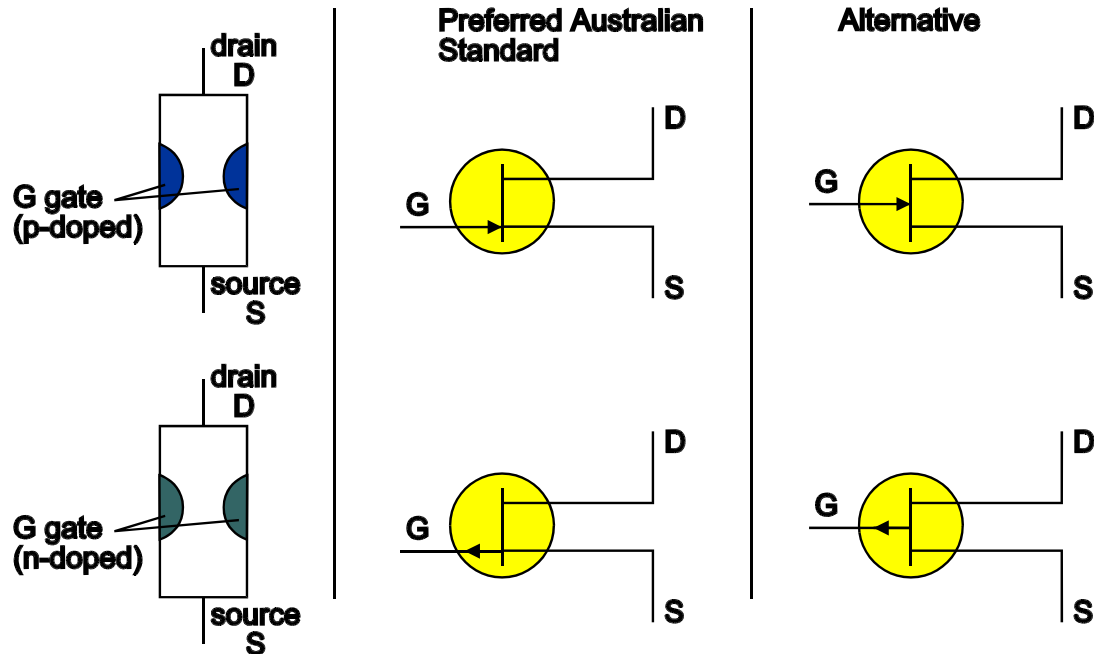


Figure 2.39
JFET Symbols

The junction field effect transistor is always operated with the gate-source pn junction reversed-biased. Reverse-biasing of this junction produces a depletion region, along the pn junction. The width of the depletion region varies according to the voltage applied to the junction. If the voltage on the gate-source pn junction V_{GS} is equal to zero, the current flowing from the drain to the source is not limited in any way and it depends only on the properties of the channel. This current is known as I_{DSS} . It is practically independent of the voltage between the drain and the source, and it is specified for each transistor type. I_{DSS} can only be limited by applying a reversed biased voltage V_{GS} on the gate-source junction. The greater the value of V_{GS} , the smaller the current I_D . In other words, the value of the gate-source voltage controls the value of the drain current.

2.3.2.2 JFET Characteristics and Parameters

A simple circuit, as with the one shown in Figure 2.40 can be used to create a set of drain characteristics for a JFET transistor. If the wiper of the potentiometer R_1 is set to its lowest position, the external voltage supply E_1 is eliminated. The gate and the source of the transistor are short-circuited and $V_{GS} = 0$. When the V_{DS} is then gradually increased from zero to a certain point, the I_D also increases

proportionally. In Figure 2.40, the region from point A to point B is called the **Ohmic region**, as I_D and V_{DS} are related by Ohm's law. (Ohm's law is discussed in Chapter 3).

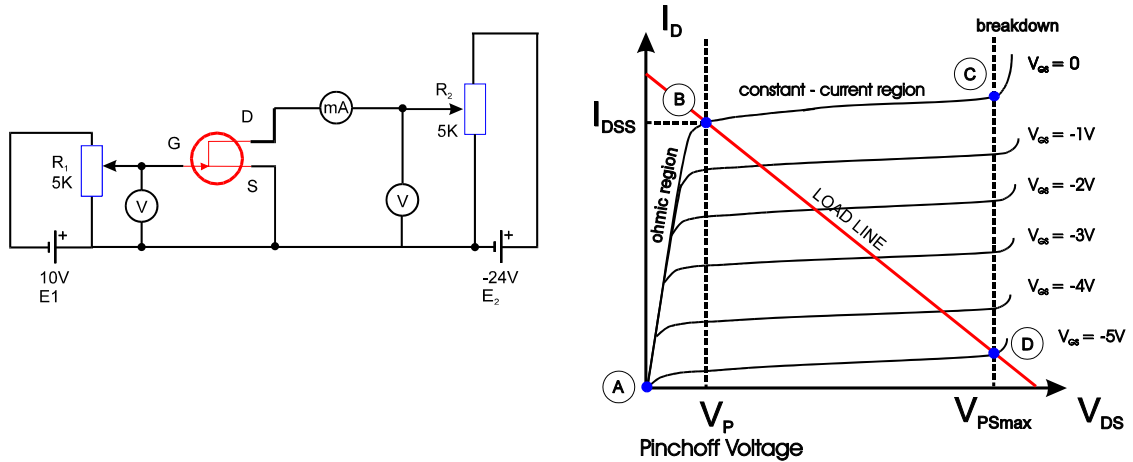


Figure 2.40
Drain Characteristic Curves

A further increase in V_{DS} does not produce any further increases of the I_D . The I_D characteristic levels off, and becomes essentially constant. In Figure 2.40, this is the region from point B to point C. This occurs, because the reverse-bias voltage from gate to drain V_{GD} produces a depletion region, large enough to offset any further increase of V_{DS} . For this reason, I_D remains relatively constant. Point B in the curve is called the **pinch-off voltage**, and it is denoted with V_P . It is the point at which I_D becomes essentially constant with further increase in V_{DS} , when $V_{GS} = 0$. At this point the $I_D = I_{DSS}$. I_{DSS} is an abbreviation for “**drain to source current, with gate shorted**”, and it is given for each specific transistor type in the data sheets.

If the V_{DS} is increased further after the point C, the drain current increases rapidly and a breakdown occurs. The voltage at the point C is called **maximum operating voltage** and it is denoted with V_{DSmax} . Exceeding this voltage can lead to a permanent damage of the transistor. For this reason JFET transistors are always operated at voltages well below V_{DSmax} in the region between the points B and C. This region is called the **constant-current region**.

So far we discussed how the transistor behaves at zero input current. By regulating R_1 the transistor can be biased. Different values of gate-to-source voltages can be applied to it (for example -1V, -2V, -3V etc.), and a set of drain characteristic curves can be produced.

Larger values of V_{GS} narrow the channel and the drain current decreases. The value of V_{GS} , at which I_D becomes practically zero, is called the **cutoff voltage**, and it is denoted with $V_{GS(off)}$. This voltage corresponds to the letter D in the Figure 2.40. Similarly, to the bipolar junction transistor, the straight line connecting the maximum output current with the maximum output voltage, gives us the load

curve. The JFET transistor operates at gate-to-source voltages, ranging from V_P to $V_{DS(max)}$. For this range of voltages, the drain current I_D varies from I_{DSS} to nearly zero.

Another important characteristic of the JFET transistors is the **forward transfer conductance (transconductance)**, which is abbreviated to g_m . g_m expresses the change in drain current, for a given change in gate-to-source voltage, and it is measured in A/V, or **Siemens (S)**. 1 Siemens is the unit for conductance and it is the reciprocal of 1 Ohm. In other words:

$$1S = \frac{1}{1\text{Ohm}}$$

If for example 1 Volt change in V_{GS} , produces a 5 mA change in the drain current, then this transistor has a g_m of 5 mA/V, or 5 millisiemens (mS). In other words, the transconductance is given with the ratio:

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

You can see this parameter sometimes abbreviated as y_{fs} (**forward transfer admittance**), or as G_{fs} (**forward transconductance**). In any case, this is an important parameter that determines the voltage gain, and it is always specified in the data sheets.

Because a JFET transistor operates with its gate-source junction reverse-biased it has a very high input resistance. This is the definite advantage of the MOS transistors over the bipolar junction transistors. Why it is important for a transistor to have a high input resistance is discussed in Chapter 4.

2.3.2.3 Construction and Operation of MOSFETs

MOSFETs form the second category of field effect transistors. They differ from JFETs in that, they have no pn junction. The gate junction and the channel are separated with a very thin layer of metal oxide insulation. Usually SiO_2 is used as an insulator. As it was remarked earlier in this section, because of the insulated gate these devices are sometimes called IGFETs, where IG means "insulated gate". Consequently the gate-source leakage current is reduced and most importantly of all, the input resistance is drastically increased. In fact, it is increased so much, that the transistor becomes vulnerable to damage even from small static voltages.

This imposes some difficulties when testing and installing MOSFETs. Usually they come packed in conducting material that short-circuits all transistor terminals and prevents static charges from building up. They have to be handled with an extreme care, using an earthed soldering iron, and earthing the operator when soldering them to the circuit.

MOSFETs can be divided into two categories - **depletion MOSFETs (D-MOSFETs)** and **enhancement MOSFETs (E-MOSFETs)**. The structure and symbols of D-MOSFET are shown in Figure 2.41.

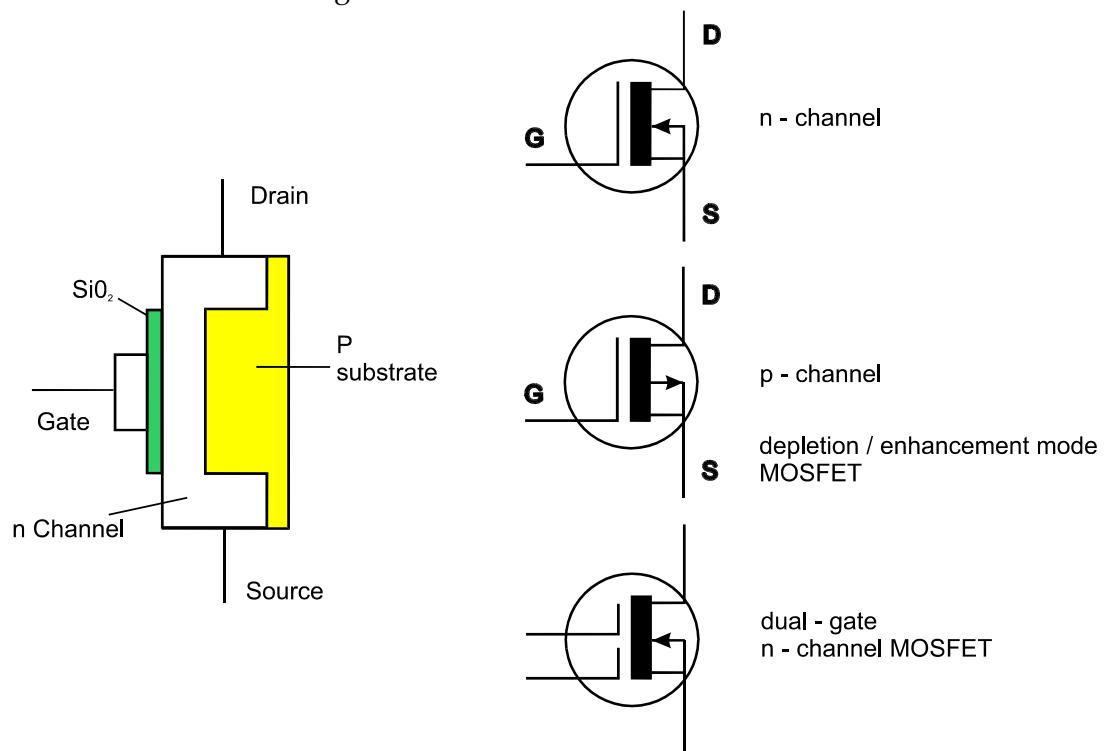


Figure 2.41
D-MOSFET

The D-MOSFET consists of a **substrate** and a **channel** from opposite semiconductor materials. The drain and the source are connected to both ends of the channel. The gate surrounds the channel, but it is electrically isolated from it by a thin layer of SiO₂. Note that though the D-MOSFET is operated in most cases in depletion mode, it can also be operated in an enhancement mode. The gate is insulated from the channel and can be supplied either with a negative or with a positive voltage. When a negative gate-to-source voltage is applied, the transistor operates in a depletion mode. Conversely when a positive gate-to-source voltage is applied it will operate in an enhancement mode.

When in depletion mode, the negative voltage on the gate repels some free electrons from the channel, leaving in their place positive ions. Because the channel is depleted of some of its free electrons, it loses some of its conductivity. The higher the negative voltage on the gate, the less the conductivity of the channel. When the negative gate-to-source voltage reaches the value of $V_{GS(off)}$, all electrons from the channel are totally depleted and there is no drain current at all. In such manner, the voltage on the gate can control the drain current. When the transistor works in an enhancement mode, the positive gate-to-source voltage increases (enhance) the channel conductivity, thus controlling the drain current, in a similar way.

The symbol of the D-MOSFET depicts the substrate as an arrow, which in the case of the n-channel transistor points inward, and in the case of a p-channel transistor points outward. Most D-MOSFETs are manufactured with an internally connected

substrate and source, but some of them may have the substrate, connected to a separate pin.

Because the channel and the gate are insulated with a dielectric, they actually form a capacitor. The value of this capacitor can reach several pF, and is capable to short-circuit the gate and the drain at high frequencies. This is obviously a drawback of MOSFET transistors. To overcome it, **dual-gate MOSFETs** are manufactured. The only difference is that they have two, instead of one gate. This feature reduces the high input capacitance and is useful for some applications, such as for automatic gain control (AGC) input in certain radio-frequency amplifiers. The symbol of a dual-gate MOSFET is shown in Figure 2.41.

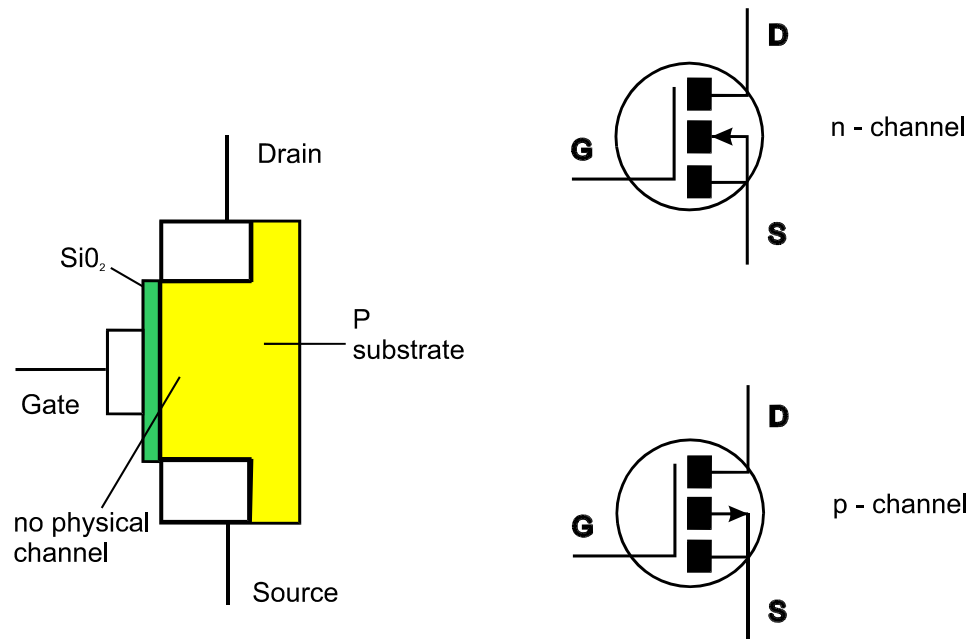


Figure 2.42
E-MOSFET

The structure and symbols of an enhancement MOSFET (E-MOSFET) transistor is shown in Figure 2.42. It differs from a D-MOSFET in that there is no structural channel. The channel can only be induced from a positive gate voltage, above the threshold value. Thus, the E-MOSFET transistor can operate only in enhancement mode. The broken lines in its symbol mean that a physical channel does not exist.

2.3.3 Summary

The bipolar junction transistor (BJT) is a device, made of three pieces of doped semiconductor, forming two pn junctions. There are two types of BJTs - pnp type and npn type. Transistor leads are called emitter, base and collector. A transistor can operate in two basic modes - as a switch and as an amplifier. When operating as a switch, the current through the transistor is either flowing at its maximum value or it is not flowing at all. When operating as an amplifier, small changes in the input current result in large changes in the output current. The most important transistor characteristics are: current gain β_{DC} , maximum collector current I_{Cmax} , maximum

collector-emitter voltage V_{CEOmax} , and maximum frequency of operation. A field effect transistor (FET) is a device, that is *voltage* (and not current) operated. Its three terminals are called source, drain, and gate. In a BJT an input current controls the output current, while in a FET an input voltage controls the output current. FETs can be junction FETs (JFETs), or metal oxide semiconductor FETs (MOSFETs or IGFETs). In a JFET the gate forms a pn junction with the channel, while in a MOSFET the gate is insulated from the channel with SiO_2 . JFETs are divided into p-channel or n-channel type. MOSFETs can be depletion (D-MOSFETs) or enhancement (E-MOSFETs). Each one of them can be p-channel or n-channel type too. MOSFETs can be also manufactured with two gates and are then called dual-channel MOSFETs.

2.4 SCRs, DIACs, TRIACs and LEDs

When you have completed study of this section, you should be able to:

- Identify the symbols for a SCR, DIAC, TRIAC and LED
- Describe their construction
- Describe their operation
- Define some of their parameters

2.4.1 The SCR

SCR is an abbreviation for a **silicon-controlled rectifier**. Its construction and schematic symbol are shown in Figure 2.43. It is a device made of four pieces of doped semiconductor, forming three pn junctions. In other words, it is a four-layer pnpn device. In a similar way to the classic diode, the SCR has two terminals, called an **anode (A)** and a **cathode (K)**. In addition, it has one more terminal called a **gate**, which is denoted with the letter **G**. Very often a SCR is called a *thyristor*. This name originates from the gas filled thyatron valves. They were used for many years in similar applications, as the SCRs are used nowadays.

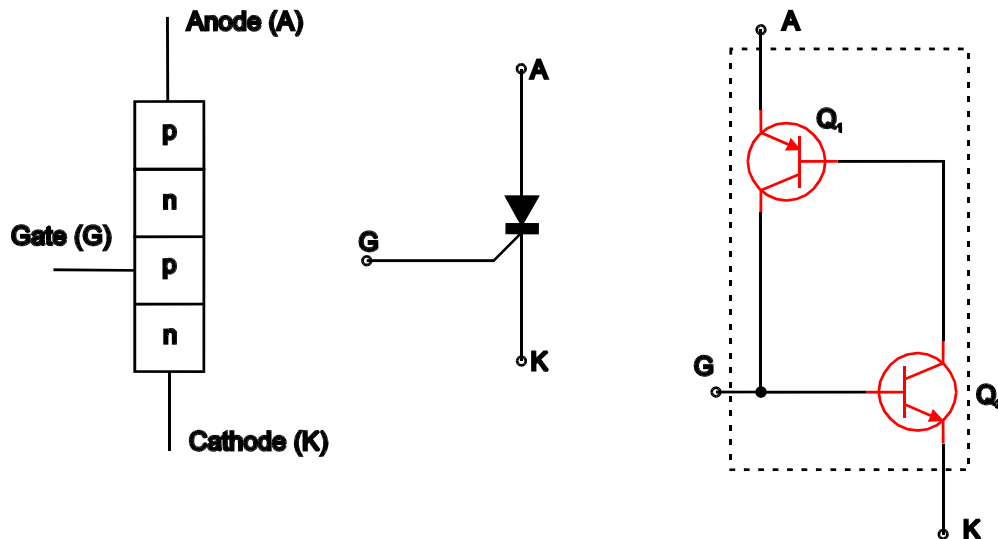


Figure 2.43
SCR

The operation of an SCR can be understood better if it is regarded as a two-transistor arrangement, as it is shown in Figure 2.44. The inner n and p layers are shared between both transistors. The positive terminal of an external power supply is connected to the anode. When the gate current I_G is zero, both transistors are turned off. The resistance between the anode and the cathode is very high and there is no anode current. In this case, the SCR acts as an open switch.

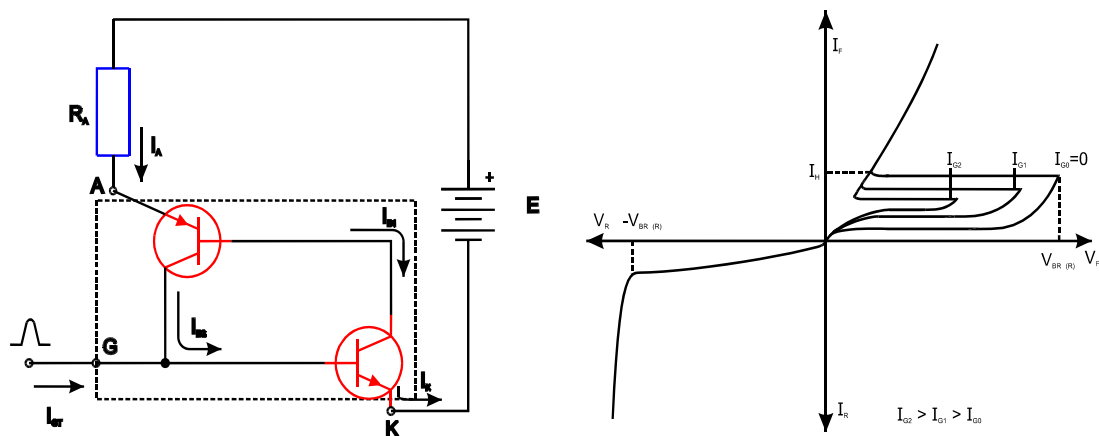


Figure 2.44
SCR Operation and Characteristic Curves

If a positive pulse of current, called a gate trigger current I_{GT} is applied to the gate, the current I_{B2} starts flowing, which turns on Q_2 . This allows the current I_{B1} to start flowing, which turns Q_1 on. The collector current of Q_1 provides a further increase of I_{B1} that keeps Q_2 on, even after the trigger pulse is removed from the gate. On the other hand, when Q_2 stays in conduction, it provides a path for the current I_{B1} , that keeps Q_1 on too. In other words, after a short trigger pulse, both transistors are turned on, and maintain a conductive path across themselves, even after the trigger pulse is removed. The resistance between the anode and the cathode becomes minimal, and the current through the thyristor reaches a maximum. In this case, the SCR acts as a closed switch.

The pulse required to trigger the SCR can be as short as 1ms (or less), with a voltage as low as 0.6 V. As it can be seen from the characteristic curves, the SCR can be switched on even without gate triggering. In order to do that the positive voltage between the anode and the cathode has to be increased above the forward-breakover voltage $V_{BR(F)}$. If I_G is increased, the $V_{BR(F)}$ decreases. Eventually a point can be reached, where the SCR can be turned on at a very low anode-to-cathode voltage. Once triggered, the SCR can only be turned off, if the anode current drops below a certain value. This value is called **holding current** I_H .

SCRs are used in industrial electronics to control large amounts of power. They have only two states of operation - on and off. A very small and short pulse can trigger a large current (for example 20 A and more) to flow through the SCR. For this reason they are used in various control applications, such as motor controls, time-delay circuits, heater controls and so on.

2.4.2 The TRIAC

The triac is very similar in its operation to the SCR. Physically, they are also very much alike. As it was discussed in the previous section, the SCR has two modes of operation - on and off. The anode current is either flowing or not flowing at all. Effectively this means that the SCR behaves as a diode, rectifying the current flowing through it. The triac can be regarded as consisting of two SCRs connected in parallel but with inverted polarity and with a common gate terminal. This means that the triac behaves exactly as an SCR, but it conducts current in both directions. The construction and the symbol of the triac is shown in Figure 2.45.

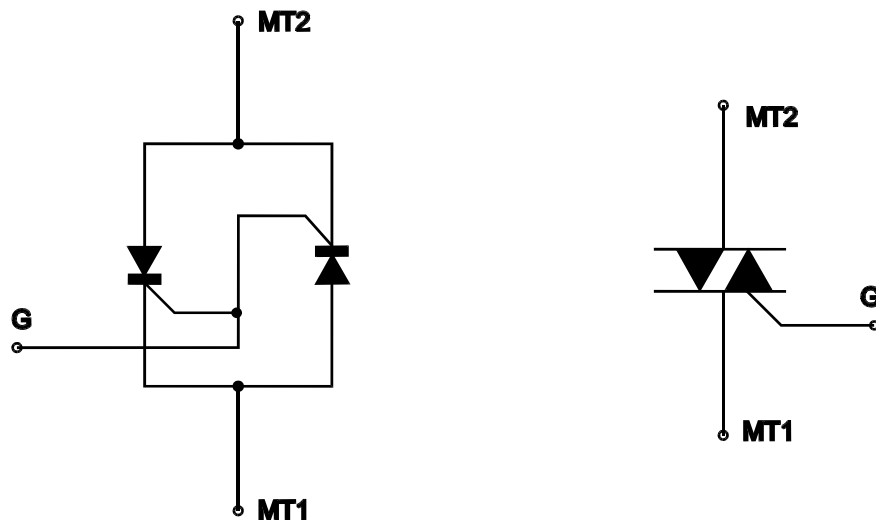


Figure 2.45
TRIAC

The two main triac terminals are called the **main terminal 1 (MT1)** and the **main terminal 2 (MT2)**. Sometimes they are referred to as anode 1 and anode 2. The third terminal is called gate, and is equivalent to that for an SCR. A small complication arises from the fact that when the MT2 is positively biased with respect to MT1, a positive trigger pulse is required at the gate terminal, but when MT2 is negatively

biased with respect to MT1, a negative trigger pulse is needed. This can be seen from the triac characteristic curves in Figure 2.46.

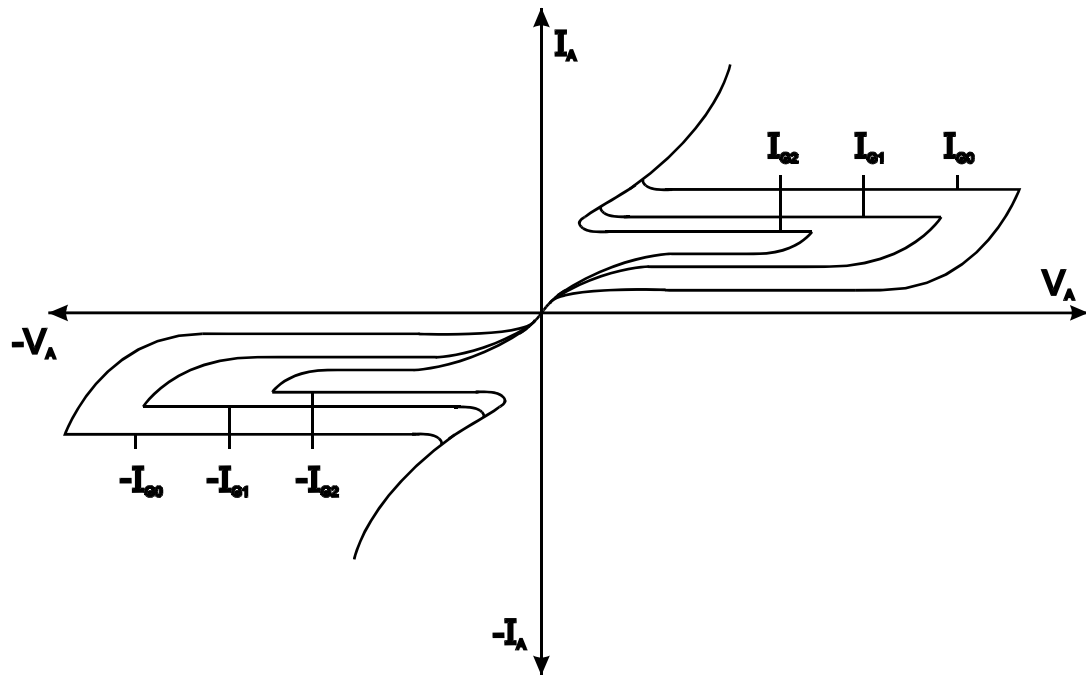


Figure 2.46
Triac Characteristic Curves

They are very similar to the SCR characteristic curves. The only difference is that the triac exhibits the same curve at forward and reverse bias, and therefore can conduct in both directions. Triacs are used in similar applications, to these of SCRs, such as control applications for a large amount of load current. The difference is that the triacs exercise an effective control during the positive and negative half-periods of the sine wave, while the SCRs can control the application during the one half-period only.

2.4.3 The DIAC

The construction and the symbol of a diac is shown in Figure 2.47. This device is very similar to the triac, with the only difference that there is no gate terminal. The symbol of the diac is also identical to the symbol of the triac, with the gate terminal omitted. As the diac can conduct in both directions, and it does not need a trigger pulse to initiate conduction, both of its terminals can be inverted. In other words, the diac terminals do not need to be identified and in most cases, they are simply called A1 and A2.

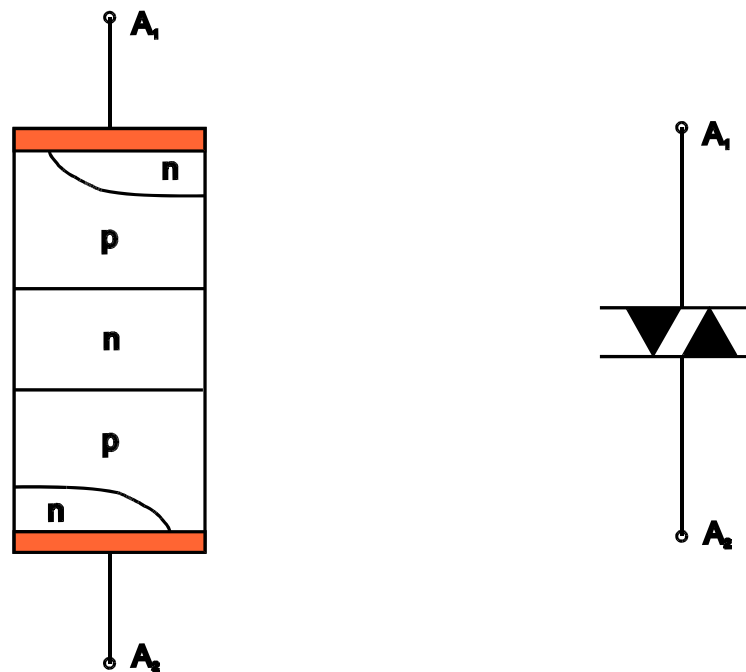


Figure 2.47
DIAC

A triac has a set of characteristics, but for a diac only one characteristic can be drawn. The reason for this is that there is no gate, and no gate current in the diac. Its characteristic is shown in Figure 2.48.

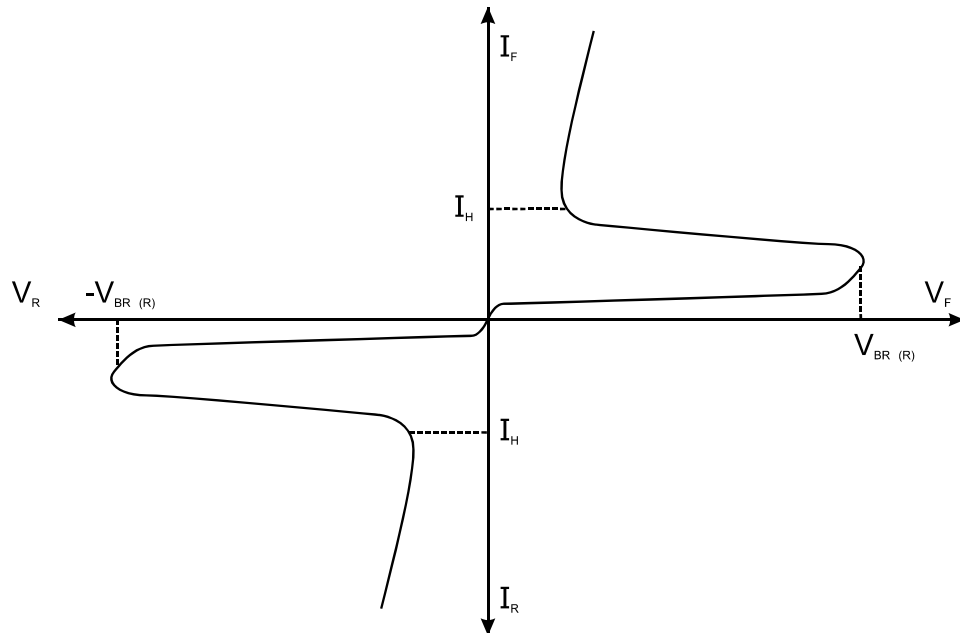


Figure 2.48
Diac Characteristic Curve

A typical breakover voltage $V_{BR(F)}$ of a diac could be around 30 V, with a forward current of around 10mA and a holding current of 100 μ A or so. In Figure 2.48, the forward and the reversed characteristics are the same. A diac with such characteristics is called a symmetrical diac. There is another group of diacs, where breakover voltage in both directions may differ. These are called asymmetrical

diacs. They are polarity sensitive and their terminals have to be connected accordingly.

2.4.4 LEDs

LED is an abbreviation for **light-emitting diode**. As indicated by its name, this device emits light, when an electrical current passes through it in a forward direction. To better understand its principle of operation, review Section 1.2.3 (especially Figure 1.5). The structure and symbol of the LED is shown in Figure 2.49.

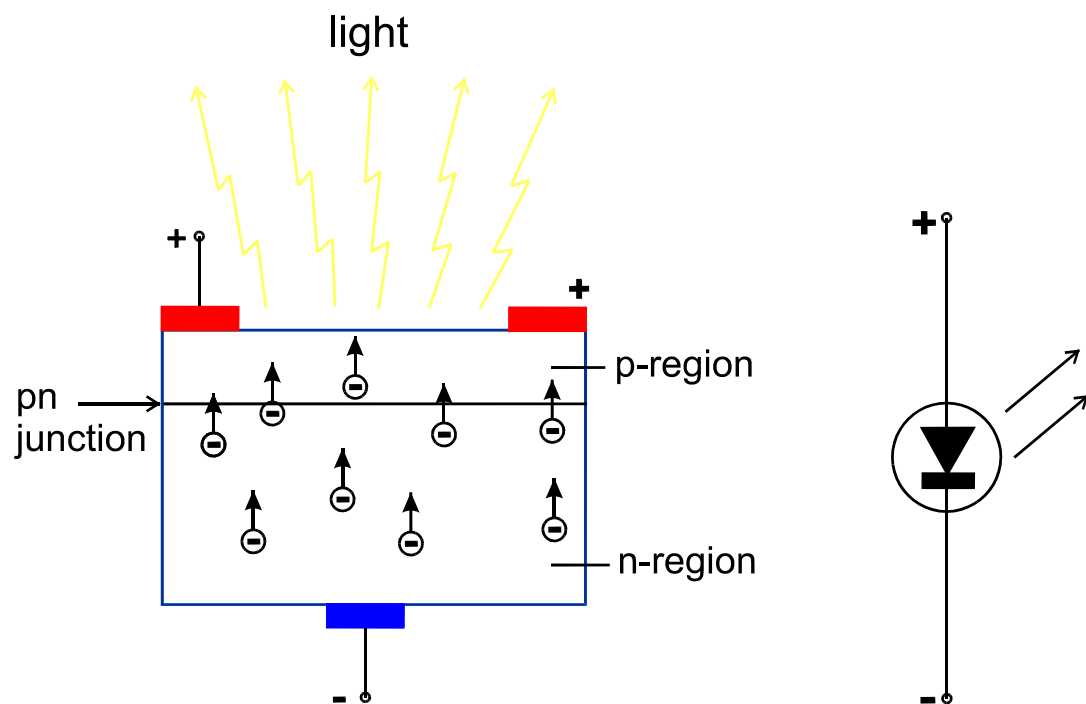


Figure 2.49
LED

When a diode is forward biased, the electrons from the conduction band of the n-type material cross the pn junction and recombine with the holes in the p-type material. There is a difference in the energy levels of the electrons in the conduction band and the electrons in the valence band. Recall that this difference is equal to the energy gap. When electrons are in the conduction band, they have high energy. In the process of recombination, they actually have to fill the existing holes in the valence band. In order to leave the conduction band and to descend to the valence band, electrons have to release the extra energy (equal to the energy gap), that they possess. In the classical diode, this extra energy is dissipated in the form of heat. However, the diode may be specifically designed to have a large exposed surface area of one layer of semiconductor material. In this case the extra energy is released in the form of heat and light. The process of emitting photons, when a pn junction is forward-biased is called **electroluminescence**.

The spectrum of the emitted light can be changed by using different semiconductor materials, or by adding different impurities in them, during the doping process. For example, LEDs made from gallium arsenide phosphide (GaAsP) emit red or yellow light, those made from gallium phosphide (GaP) produce red or green light and LEDs manufactured with gallium arsenide (GaAs) emit infrared light. Though red LEDs are the most commonly used, there are many other colours available, such as amber, blue and even white. Apart from the different colours, sizes and shapes in which they are manufactured, LEDs can also be high-brightness LEDs, dual or three colour LEDs, flashing LEDs and so on.

The forward voltage of an LED is about 2 or 3 times greater than for a silicon diode and can reach the value of 3.2 V. Also the reverse breakdown of an LED is much less than for a silicon rectifier and it is typically between 3 to 10 V. It means that the LED will only emit light, when sufficient forward voltage is applied to it, and that the reverse biasing of the LED should be avoided. The characteristic in Figure 2.50 shows the relationship between the light output and the forward current.

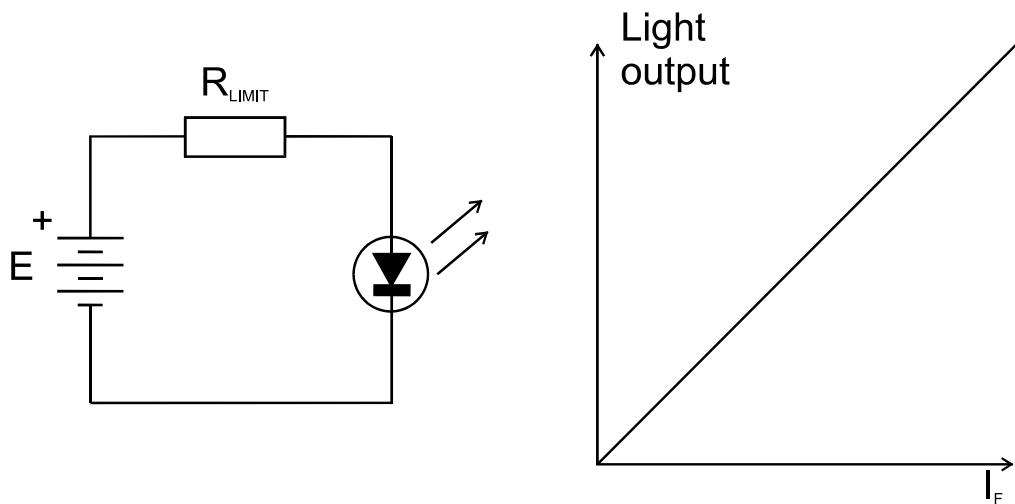


Figure 2.50
Operation and Characteristic of a Forward-biased LED

It shows that the light output is directly proportional to the forward current. The greater I_F , the greater the light output. Three basic parameters of LEDs should be remembered. These are:

- The **maximum forward current** that the LED can handle,
- The **forward voltage** at which the LED starts conducting, and therefore light emitting,
- The **maximum reverse voltage** that the LED can withstand.

Since their invention, LEDs have become extremely popular in many areas of industrial electronics. It is mainly because they are quite inexpensive, very reliable, and easy to implement. Typical LED applications include pilot lights, readout displays, indicators in domestic appliances, just to mention a few. A very common device, using LEDs is the seven-segment display. Each segment is a LED. By forward-biasing selected combination of segments, all decimal figures and some symbols can be displayed.

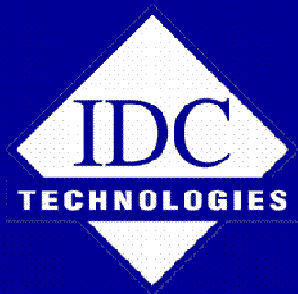
2.4.5 Summary

Various types of discrete semiconductor components exist. The most common ones are SCRs, DIACs, TRIACs and LEDs. A silicon-controlled rectifier (SCR) is a four-layer pnpn device, with three terminals called an anode, a cathode, and a gate. If a short trigger pulse is applied to the gate, the SCR turns on and conducts current from the anode to the cathode. The current flows until its value drops below the holding current. The triac is another multilayer semiconductor device, similar in operation to the SCR, with the only difference being that it can conduct current in both directions. It has also three terminals, one of which is the gate, but the other two are reversible and are called main terminal 1 (MT1) and main terminal 2 (MT2). It is important that the trigger pulse is with the same polarity as the applied voltage to the main terminals, in order to turn the triac on. The diac is actually a triac, without the gate terminal. It can be switched on by sufficient forward or reversed voltage. Its terminals are called A1 and A2. The light-emitting diode (LED) has the capability of emitting light, when its pn junction is forward biased. LEDs are available in many different shapes, sizes, and colours. The LED color depends on the type of the semiconductor used and its doping. Reverse biasing of LEDs should be avoided and they should only be operated at sufficient forward voltages.

2.5 Quiz questions

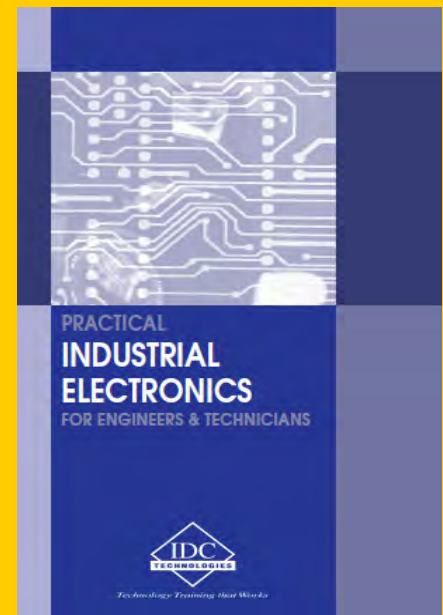
1. Which factors determine the electrical resistance?
2. A resistor has a nominal value of 4.7 k Ω and a tolerance of 5%. What should be the value of that resistor, when measured?
3. Is it true that metal-film resistors are usually made with very high power ratings?
4. What is the difference between a potentiometer and a rheostat?
5. A 4-band resistor is marked as follows: red, red, yellow, and silver band. What does this mean?
6. Two resistors with nominal values of 470k Ω and 270 Ω are connected in parallel. What is the total resistance of the circuit?
7. An inductor has an inductance of 50 mH. What resistance does it offer to a current with a frequency of 50Hz and a current with a frequency of 50 MHz?
8. Which factors affect inductance?
9. What does polarization mean?
10. Which capacitor type provides the largest capacitance value per unit of volume?
11. What will happen if a paper capacitor is connected with a particular polarity and then the connection is swapped around?
12. Calculate the capacitance reactance for a current with frequency of 50Hz and current with a frequency of 50MHz, if the capacitor has a nominal value of 47 μ F.
13. Two capacitors, with nominal values of 2700pF and 12nF are connected in parallel. What is the total capacitance of the circuit?
14. A transformer steps down the voltage from 240 V to 6 V and has 2200 turns in its primary winding. What is the number of turns in the secondary winding?
15. What are the typical applications of a transformer?
16. Explain how a barrier potential is formed in the pn junction.
17. List conditions to be satisfied for reverse biasing.
18. What is a breakdown voltage?

19. Explain how the full-wave bridge rectifier operates.
20. What are the advantages of the full-wave bridge rectifier, compared to the half-wave rectifier?
21. Draw the symbol of a npn-type BJT transistor and show the direction of the emitter and the collector current.
22. What does h_{FE} mean and what is it relevant for?
23. Draw the collector characteristic curves and specify which two points are used for a transistor in a switch mode.
24. What is $V_{CEO(max)}$? Explain the meaning of each letter.
25. Describe the construction and the principle of operation of a p-type JFET transistor.
26. Which parameter specifies the voltage gain of a FET transistor?
27. What are the advantages and disadvantages of the FETs, compared to the BJTs?
28. What are the advantages of the MOSFET, compared to the JFETs?
29. Draw symbols of a SCR, DIAC, TRIAC and LED.
30. Describe the principle of operation of an SCR.
31. What is the difference between a diac and a triac?
32. Is it possible to connect an SCR with a reversed polarity between its anode and cathode, without this having any impact on the circuit?
33. Is it possible to do the same with a triac?
34. Why does the LED emit light, while the silicon transistor does not?
35. Name the three most important parameters of an LED.



**YOU MAY ALSO BE INTERESTED
IN THE FOLLOWING TITLE:**

Practical Industrial Electronics for Engineers and Technicians



**This book is available for purchase in
printed form or as an ebook**

**For more details about this title
please visit the link below:**

<http://bit.ly/1BYBWAZ>

**If you would like a quote or any further information
please do not hesitate to contact us:**

books@idc-online.com