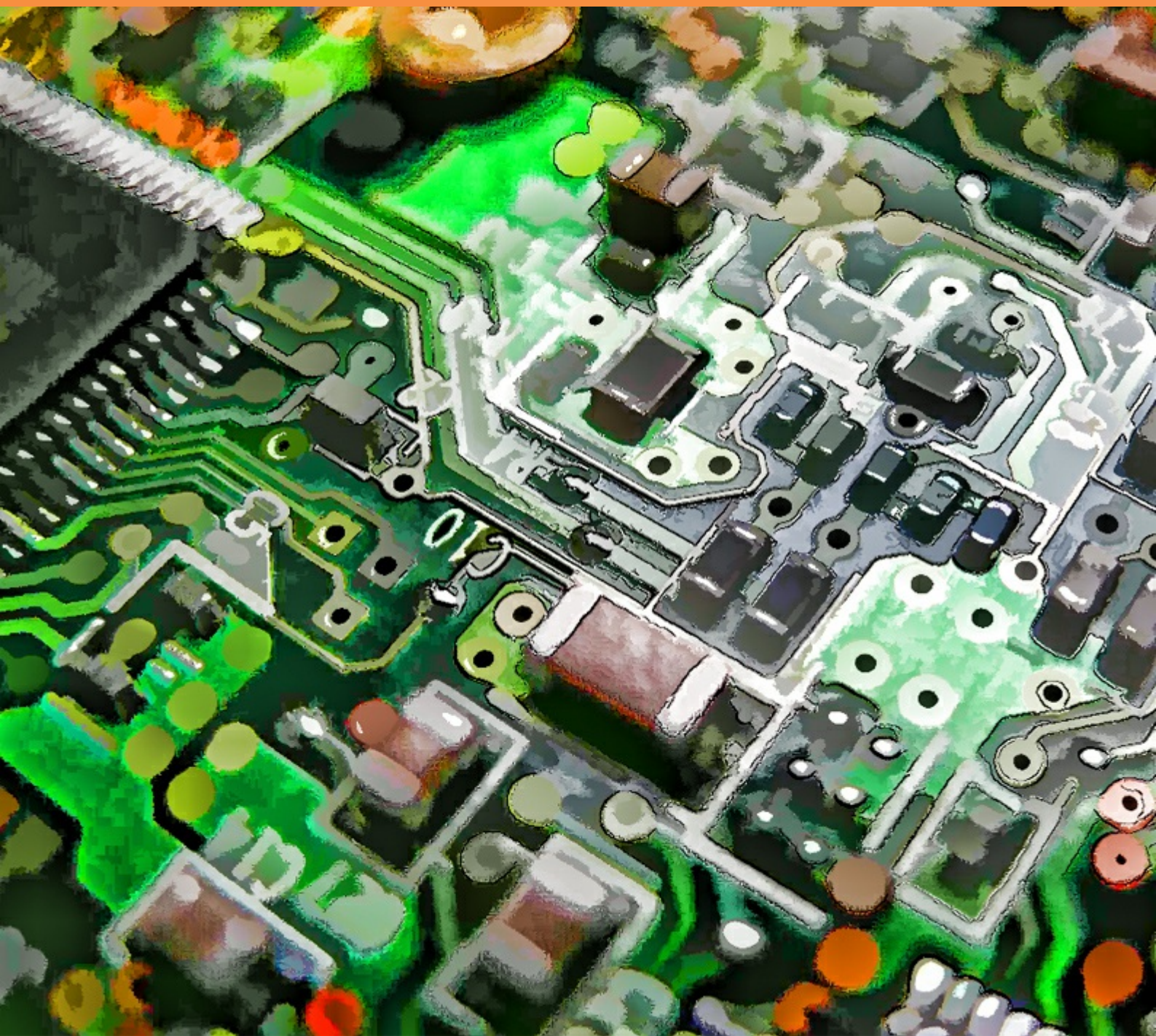
bookboon
The eBook company

Introduction to Electronic Engineering

Valery Vodovozov



Valery Vodovozov

Introduction to Electronic Engineering

Introduction to Electronic Engineering
1st edition

ISBN 978-87-7681-539-4

Contents

Designations

C	capacitor	K	amplification, gain	W	energy
D	diode, thyristor	L	inductance	X	reactance
L	inductor, choke	P	power	Z	impedance
R	resistor	q	duty cycle	α	dc alpha, firing angle
T	transistor	Q	multiplication, selectivity	β	current gain
w	number of turns			δ	error, loss
C	capacitance	r	ripple factor	η	efficiency
$\cos \varphi$	power factor	R	resistance	φ	phase angle
f	frequency	t	time	ω	angular frequency
G	conductivity	T	period		
I	current	U	voltage		

Abbreviations

A	Ampere
ac	alternating current
ADC	analog-to-digital converter
AM	amplitude modulation
BiFET	bipolar FET
BiMOS	bipolar MOS
BJT	bipolar junction transistor
CB	common base
	complementary bipolar technology
CC	common collector
CD	coder
CE	common emitter
CMOS	complementary MOS
DAC	digital-to-analog converter
dc	direct current
DC	decoder
DMOS	double-diffused transistor
F	Farad
FET	field-effect transistor
FM	frequency modulation
FS	full scale
G	Giga = 10^9 (prefix)
GaAsFET	gallium arsenide FET
GTO	gate turn-off thyristor
H	Henry
Hz	Hertz
IC	integrated circuit
IGBT	insulated gate bipolar transistor
JFET	junction FET
k	kilo = 10^3 (prefix)
LDR	light-dependent resistor
LED	light-emitting diode
LSI	large-scale integration circuit
LSB	least significant bit
M	Mega = 10^6 (prefix)
m	milli = 10^{-3} (prefix)
MOS	metal-oxide semiconductor
MCT	MOS-controlled thyristor
MPP	maximum peak-to-peak
MSB	most significant bit
MSI	medium-scale integration circuit
MUX	multiplexer
n	nano = 10^{-9} (prefix)
n	negative
p	pico = 10^{-12} (prefix)
p	positive
PWM	pulse-width modulation
PZT	piezoelectric transducer
RDC	resolver-to-digital converter
rms	root mean square
RMS	rms volts
S	Siemens
s	second
SADC	sub-ADC
SAR	successive approximation register
SCR	silicon-controlled rectifier
SDAC	sub-DAC
S/H	sample-and-hold
SSI	small-scale integration circuit
T	flip-flop
TTL	transistor-transistor logic
V	Volt
VDC	dc volts
VCO	voltage-controlled oscillator
VFC	voltage-to-frequency converter
W	Watt
WA	Volt-Ampere
XFCB	extra fast CB technology
μ	micro = 10^{-6} (prefix)
Ω	Ohm

Preface

Electronics is a science about the devices and processes that use electromagnetic energy conversion to transfer, process, and store energy, signals and data in energy, control, and computer systems. This science plays an important role in the world progress. Implementation of electronic devices in various spheres of human activity largely contributes to the successful development of complex scientific and technical problems, productivity increase of physical and mental labour, and production improvement in various forms of communications, automation, television, radiolocation, computer engineering, control systems, instrument engineering, as well as lighting equipment, wireless technology, and others. Contemporary electronics is under intense development, which is characterized by emergence of the new areas and creation the new directions in existing fields.

The goal of this work is to introduce a reader to the basics of electronic engineering. The book is recommended for those who study electronics. Here, students may get their first knowledge of electronic concepts and basic components. Emphasis is on the devices used in day-to-day consumer electronic products. Therefore, semiconductor components diodes, transistors, and thyristors are discussed in the first step. Next, the most common electronic circuits, such as analogue, differential and operation amplifiers, suppliers and references, filters, math converters, pulsers, logical gates, etc. are covered.

After this course, students can proceed to advanced topics in electronics. It is necessary to offer an insight into the general operation of loading as well as into the network distortions caused by variables, and possibilities for reducing these disturbances, partly in power electronics with different kinds of load. Such problems, as the design and methods for implementing digital equipment, Boolean algebra, digital arithmetic and codes, combinatorial and sequential circuits, network instruments, and computers are to be covered later. Modeling circuits and analysis tools should be a subject of interest for future engineers as well. Further, electronics concerns the theory of generalized energy transfer; control and protection of electronic converters; problems of electromagnetic compatibility; selection of electronic components; control algorithms, programs, and microprocessor control devices of electronic converters; cooling of devices; design of electronic converters.

Clearly, in a wide coverage such, as presented in this book, deficiencies may be encountered. Thus, your commentary and criticisms are appreciated: **valery.vodovozov@ttu.ee**.

Author

Introduction

Electronic system. Any technical *system* is an assembly of components that are connected together to form a functioning machine or an operational procedure. An *electronic system* includes some common used electrical devices, such as resistors, capacitors, transformers, inductors (choke coils), frames, etc., and a few classes of semiconductor devices (diodes, thyristors, and transistors). They are joined to control the load operation.

Historical facts. An English physicist W. Hilbert proposed the term "*electricity*" as far back as 1700. In 1744, H. Rihman founded the first electrotechnical laboratory in the Russian Academy of Science. Here, M. Lomonosov stated the relation of electricity on the "nature of things".

A major electronic development occurred in about 1819 when H. Oersted, a Danish physicist, found the correlation between an electric and a magnetic field. In 1831, M. Faraday opened the electromagnetic induction phenomenon. The first to develop an electromechanical rotational converter (1834) was M.H. Jacobi, an Estonian architect and Russian electrician. Also, he arranged the arrow *telegraph* receiver in 1843 and the letter-printing machine in 1850. In 1853, an American painter S. Morse built a telegraph with the original coding system and W. Kelvin, a Scottish physicist and mathematician, implemented a digital-to-analog converter using resistors and relays.

In 1866, D. Kaselly, an Italian physicist, invented a pantelegraph for the long-line transmission of drawings that became a prototype of the *fax*. A.G. Bell was experimenting with a telegraph when he recognized a possibility of voice transmission. His invention of the *telephone* in 1875 was the most significant event in the entire history of communications. A. Popov and G. Marcony demonstrated their first *radio* transmitting and receiving systems in 1895–1897.

In 1882, a French physicist J. Jasmin discovered a phenomenon of semiconductance and proposed this effect to be used for rectifying alternating current instead of mechanical switches. In 1892, a German researcher L. Arons invented the first mercury arc vacuum valve. P.C. Hewitt developed the first arc valve in 1901 in the USA and a year later, he patented the mercury rectifier. In 1906, J.A. Fleming has invented the first vacuum diode, an American electrician G.W. Pickard invented the silicon valve, and L. Forest patented the vacuum tube and a vacuum triode in 1907. The development of electronic amplifiers started with this invention. Later, based on the same principles, many types of electronic devices were worked out. A key technology was the invention of the feedback amplifier by H. Black in 1927. In 1921, F. Meyer from Germany first formulated the main principles and trends of power electronics.

In the first half of the 20th century, electronic equipment was mainly based on *vacuum tubes*, such as gas-discharge valves, thyratrons, mercury arc rectifiers, and ignitrons. In the 1930s, they were replaced by more efficient mercury equipment. The majority of valves were arranged as coaxial closed cylinders round the cathode. Valves that are more complex contained several gridded electrodes between the cathode and anode. In this way, triode, tetrode, and pentode valves were designed.

The vacuum tube has a number of disadvantages: it has an internal power filament; its life is limited before its filament burns out; it takes up much space, and gives off heat that rises the internal temperature of equipment. Because of vacuum tube technology, the first electronic devices were very expansive, bulky, and dissipated much power.

In the middle of the 1920s, H. Nyquist studied telegraph to find the maximum signaling rate. His conclusion was that the pulse rate could not be increased beyond double channel bandwidth. His ideas were used in the first *television* translation provided by J. Baird in Scotland, 1920, and V. Zworykin in Russia, 1931. In 1948, C. Shannon solidified the signal transmitting theory based on the Nyquist theorem.

The digital *computer* was a significant early driving force behind digital electronics development. The first computer project was started in 1942, revealed to the public in 1946. The ENIAC led to the development of the first commercially available computer UNIAC by Eckert and Mauchly in 1951. Later, the IBM-360 mainframe computer and DEC PDP-series minicomputers, industrial, and military computer systems were developed.

The era of semiconductor devices began in 1947, when American scientists J. Bardeen, W. Brattain, and W. Shockley from the Bell Labs invented a germanium transistor. Later they were awarded the Nobel Prize for this invention. The advantages of a transistor overcome the disadvantages of the vacuum tube. From 1952, General Electric manufactured the first germanium diodes. In 1954, G. Teal at Texas Instruments produced the silicon transistor, which gained a wide commercial acceptance because of the increased temperature performance and reliability. During the middle of the 1950s through to the early 1960s, electronic circuit designs began to migrate from vacuum tubes to transistors, thereby opening up many new possibilities in research and development projects.

The invention of the integrated circuit by J. Kilby from Texas Instruments in 1958 was followed by the planar process in 1959 of Fairchild Semiconductor that became the key of solid-state electronics.

Before the 1960s, semiconductor engineering was regarded as part of low-current and low-voltage electronic engineering. The currents used in solid-state devices were below one ampere and voltages only a few tens of volts. The year 1970 began one of the most exciting decades in the history of low-current electronics. A number of companies entered the field, including Analog Devices, Computer Labs, and National Semiconductor. The 1980s represented high growth years for integrated circuits, hybrid, and modular data converters. The 1990s major applications were industrial process control, measurement, instrumentation, medicine, audio, video, and computers. In addition, communications became an even bigger driving force for low-cost, low-power, high-performance converters in modems, cell-phone handsets, wireless infrastructure, and other portable applications. The trends of more highly integrated functions and power dissipation drop have continued into the 2000s.

The period of power semiconductors began in 1956, when the silicon-based thyristors were invented by an American research team led by J. Moll. Based on these inventions, several generations of semiconductor devices have been worked out. The time of 1956–1975 can be considered as the era of the first generation power devices. During of second-generation from 1975 until 1990, the metal-oxide semiconductor field-effect transistors, bipolar *npn* and *pnp* transistors, junction transistors, and gate turn-off thyristors were developed. Later the microprocessors, specified integral circuits, and power integral circuits were produced. In the 1990s, the insulated gate bipolar transistor was established as the power switch of the third generation. A new trend in electronics arrived with the use of intelligent power devices and intelligent power modules.

Now, electronics is a rapidly expanding field in electrical engineering and a scope of the technology covers a wide spectrum.

Basic quantities. The main laws that describe the operation of electronic systems are *Ohm's law* and *Kirchhoff's laws*. The main quantities that describe the operation of electronic systems are *resistance R*, *capacitance C*, and *inductance L*. The derivative quantities are *reactance X, impedance Z*, and *admittance*, or *full conductivity G*.

Inductive reactance (*reluctance*) is presented by

$$X_L = \omega L,$$

and capacitive reactance is equal to

$$X_C = 1 / (\omega C),$$

where $\omega = 2\pi f$ is the *angular frequency* and f is the supply *frequency*. The impedance depends on the type of the circuit. In a series-connected *RLC* circuit, reactance is as follows:

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

In the case of a parallel *RLC* connection

$$G = 1 / X_L - 1 / X_C, Z = \sqrt{(G^2 + 1 / R^2)}.$$

Resonance. Any connection of an inductor and a capacitor is called a *tank circuit, tuned circuit*, or *resonant circuit.* In these circuits, *resonance* may occur. At the *resonance frequency*, the reluctance and the capacitive reactance are equal to

$$X_L = X_C = \sqrt{(L / C)},$$

therefore the characteristic impedance is

$$Z_r = R.$$

The resonance frequencies are as follows:

$$\omega_r = 1 / \sqrt{(LC)}, f_r = 1 / (2\pi\sqrt{(LC)}).$$

In series connections, the low impedance occurs, whereas in parallel connections, high impedance is the case because the series circuit behaves as a low-value resistor and a parallel circuit as a large-value resistor. Below the resonance frequency, the series circuit behaves like a resistive-capacitive circuit and the parallel circuit behaves like a resistive-inductive circuit. Above the frequency of resonance, the series circuit behaves like a resistive-inductive circuit and the parallel circuit behaves like a resistive-capacitive circuit.

Signals. Any circuit passes signals. The main signal magnitudes are current I, voltage U, and powers – P (true power or active power) and P_S (apparent power). The power is an instant quantity of energy that inputs in or outputs from an electronic element. The ratio of the active power P to apparent power P_S is defined as a *power factor*. It is often called cos φ, where

$$\varphi = \text{arctg}\ (X / R).$$

The displacement between the voltage and the current is called the *phase displacement angle* and is designated with the Greek letter φ. Thus, the power is defined as

$$P = UI \cos \varphi = P_S \cos \varphi.$$

The load value should be agreed with the electronic circuit.

In the case of *direct current* (*dc*), the main laws describe the level of changing the mentioned quantities. In terms of electrical engineering, dc is a unipolar current flow that may contain considerable ac components. These ac components result in fluctuations, called *a ripple*, at the dc output level. The average voltage level is symbolized as U_d, measured in dc volts, VDC. The average current level is I_d, measured in dc amperes.

In the case of *alternating current* (*ac*), one should take into account primarily the sign of signals, as well as their shape and repetition. The wave of a repetitive signal has a *cycle*, which *period* T is the amount of time between the beginning of the positive half-cycle and the start of the next positive half-cycle. Frequency is the number of cycles per period. For the repetitive signal, it is equal to

$$f = 1 / T.$$

European power companies usually supply a sinusoidal voltage 230 V of frequency $f = 50$ Hz with period $T = 20$ ms.

Usually, an instantaneous value of an ac signal varies during the time of operation. Once a signal is a continuous wave of sinusoidal shape, the peak-to-peak value consists of two amplitude values. The on-state ac value, which is equal to the dc value with the same power, is called a root mean square value, rms, or effective value:

$$U_{rms} = \sqrt{(1 / (2\pi) \times \int (U^2 \times d\omega t))} = U_{\max} / \sqrt{2} = 0{,}707\ U_{\max},$$

where U is the instantaneous value, U_{max} is the *amplitude* value of a sinusoidal wave. This level is measured in ac volts, rms.

The ac value, which is equal to the area enveloped by a signal during its positive alternation of period T, is called an *average value*. The average value of the sinusoidal wave that a voltmeter reads is equal to

$$U_d = 1 / \pi \times \int (U \times d\omega t) = 2U_{\max} / \pi = 0{,}637\ U_{\max}.$$

Passive and active devices. The devices that can only reduce signal amplitude or bring it down to a smaller value are generally called *passive devices* or *attenuators, pads*. Examples are as follows: a resistor, a capacitor, and an inductor.

When the magnitude of a signal is increased during the operation, it is said to have *amplification.* Components of this type are known as *active devices*. Transistors and circuits built on their base are examples of active components. The amount of amplification accomplished by an active device is called *a gain*. Electronically, a gain is a ratio of the output signal to the input signal. An equation for a *voltage gain* or amplification is

$$K_U = U_{out} / U_{in}.$$

Formula

$$K_I = I_{out} / I_{in}$$

expresses a current amplification and

$$K_P = P_{out} / P_{in} = K_U K_I$$

is a power amplification. Here, index "*in*" denotes the input signal and index "*out*" is the output signal of a device.

The resonant circuit can provide voltage amplification without power amplification. This quantity is termed a *voltage multiplication Q*

$$Q = U_{out} / U_{in} = \omega_r L / R,$$
$$Q = 1 / (\omega_r CR),$$
$$Q = \sqrt{(L / C)} / R.$$

Efficiency. To evaluate the power quality of an electronic system, *efficiency* is used. Efficiency is given by

$$\eta = P_L / P_S \times 100\%.$$

This means that the efficiency is the ratio of the load power P_L to the supply power P_S. Here

$$P_S = U_S I_S, P_L = UI,$$

where U_S is the supply voltage, I_S is the total supply current or *current drain*, U is the load voltage amplitude, and I is the load current amplitude. System efficiency is a value between 0 and 100 percent. It is a way of measuring how well a circuit uses the power from the supply to produce useful load power. One can calculate the power of *losses* as

$$P_{loss} = P_S - P_L = P_L (100 / \eta - 1).$$

Features and standards. In today's electronic engineering, two branches are distinguished – *low-signal electronics* that belongs to the field of *signal processing* or *radio-electronics*, and *power electronics* that belongs to the field of *power supplies* and *energy conversion*. Modern electronic technologies include the manufacture of low-signal electronic chips, printed circuits, and logic arrays, as well as power electronic devices, and their modules. The important features of electronic devices and circuits are as follows:

- breakdown and cutoff voltages and currents;
- instantaneous and on-state voltages, currents, and powers;
- turn-on and turn-off speeds;
- power losses and power dissipation;
- frequency response;
- efficiency.

Another two fields include analog and digital (pulse or switching) electronics. Note that there is no pure analog or digital devices and all the systems include both components. However, traditionally these two modes of device operation are discussed independently because of their different features and characteristics.

The following standards have been used in the book to present electronic elements, circuits, and devices and to measure their quality:

- ISO 3.1-11. Quantities and units. Mathematical signs and symbols for use in physical sciences and technology;
- ISO 129. Technical drawings. – Dimensioning. – General principles, definitions, methods of execution and special indications;
- EN 60617 / IEC 617. Graphical symbols for diagrams.

1 Semiconductor Devices

1.1 Semiconductors

1.1.1 Current in Conductors and Insulators

To understand how electronic devices operate, one has first to learn about the atomic structure of matter.

Structure of matter. The matter consists of *atoms*, which contain *electrons* and *a nucleus* with *protons* and *neutrons* in a particularly intimate association. The electron has a negative *charge*. The proton has a positive charge equal to the negative charge carried by the electron. The neutron, as its name implies, has no charge; it is electrically neutral. Each element possesses a certain number of protons and an equal number of electrons to keep the atom electrically neutral. Each element is characterized by its number of electrons, or as it is called, its *atomic number*. The electrons are spread out in space around the nucleus in shells, which have been compared to the orbits of the planets round the sun. The electrons can be often stripped off the atom rather easily, leaving it positively charged, naturally, but it is much more difficult to break up the nucleus.

Current. Electric *current* flows in a material being a result of the interaction of charged pieces called *carriers*. A review of the mechanism for conducting electricity through various kinds of matter shows that in *electrolytes* and in gases, conduction occurs through the motion of *ions*. In metallic *conductors*, conduction takes place via the motion of electrons, and there is no conduction in *insulators,* but only a slight displacement of the charges within the atoms themselves. The number of free carriers in different materials varies in an extremely wide range. In metals, the density of free electrons is in order of 10^{23} 1/cm^3. In insulators, the free electron density is less than 10^3 1/cm^3. For this reason, the electrical conductivity of various materials is very different, more than 10^6 S/cm for metals and less than 10^{-15} S/cm for insulators.

Energy levels. The negatively charged electrons possess energy in discrete amounts, and therefore they are placed only in certain energy levels without gaps between them. In the normal state, the electrons tend to fill the lowest energy levels, leaving only the highest energy level unfilled. Electrons in this outer shell are loosely bound to the nucleus and can be freed or tied to neighboring atoms. In solids, atoms are situated very closely to each other. Neighboring atoms can derange their energy levels and combine to form energy bonds. Only the outer orbit is of interest to understanding the conductivity properties in a solid, also called the *valence bond* where electrons can move and participate in an electric current. Between the valence and other bonds, there is a forbidden gap, which the electrons can cross but where they cannot remain.

Conductivity. The key to electrical conductivity of chemical elements is the number of electrons in the valence orbit. Insulators have up to eight valence electrons. Some of the atoms of the conductor have only one *valence electron* in their outer orbit. Since this single electron can be easily dislodged from its atom, it is called a *free electron* or a *conduction-bond electron* because it travels in a large orbit, equivalent to a high energy level. The slightest voltage causes free electrons to flow from one atom to another.

The density of free carriers of metals and insulators is approximately constant and cannot be changed in a marked range. The electrical resistance of a metal changes slightly with temperature. The variation of resistance with temperature is accounted for as follows. In a metal only very few electrons are free to move upon application of a potential difference. The temperature of the conductor being lowered, the thermal vibration of its atoms' lattice is decreased. As a result, the atoms interfere less with the motion of electrons, and consequently, the resistance is lowered. Such kind of resistance is known as an *ohmic resistance* or *positive resistance.* Only near the absolute zero does an abrupt change occur.

Summary. Electric current is a flow and interaction of charged carriers. In conductors, conduction takes place via the motion of negatively charged electrons. The electrical conductivity depends on the number of electrons in the valence orbit of chemical elements. Voltage causes free electrons to flow from one atom to another. The density of electrons in metal and therefore its resistance is approximately constant. Nevertheless, due to thermal vibration, the metal resistance slightly lowers when the temperature drops. Consequently, it is referred to as positive ohmic resistance of metals.

1.1.2 Current in Semiconductors

Semiconductors are neither conductors nor insulators. The commonly used semiconductor elements are silicon, germanium, and gallium arsenide. *Silicon* is the most widely used semiconductor material. It has 14 protons and 14 electrons in orbits. An isolated silicon atom has four electrons in the valence bond. *Germanium* has 32 protons, 32 electrons, and 4 valence electrons like silicon.

Crystal. Each atom that is normally bonded with the nearest neighbor atoms results in a special shape called a *crystal* (Fig 1.1). A silicon atom that is a part of a crystal has eight electrons in the valence orbit and four neighbor atoms. Each of the four neighbors shares one electron. Since each shared electron in Fig. 1.1 is being pulled in opposite directions, it is a kind of a bond between the opposite cores. This type of a bond is known as a *covalent bond*. The covalent bonds hold the tetravalent crystal together, ensuring its stability.

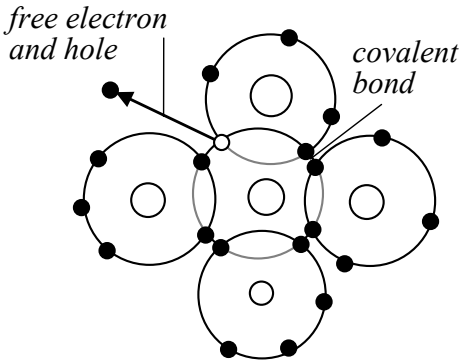


Fig. 1.1

Intrinsic semiconductors. The density of free carriers defines the conductivity of semiconductors as an intermediate between that of insulators and conductors. As mentioned above, the density of free carriers of metals and insulators is approximately constant. This is exact opposite for semiconductors, where the free carrier density can be changed by many orders. This feature of semiconductors, their ability to manipulate by free carrier density, is very significant in many electronic applications. The reason of this phenomenon is next.

Conduction of semiconductors takes place by electrons just as in metals, but, contrary to the behavior of metals, a substance of this kind exhibits a growing of resistance as the temperature falls. The resistance of the semiconductor material is called a *bulk resistance*. Since the resistance decreases as the temperature increases, it is a *negative resistance*, and semiconductor is called a negative *temperature coefficient* device. Such a substance is referred to as a semiconductor because at the absolute zero of temperature, it would be an insulator and at a very high temperature, it is a conductor. At room temperature, a pure silicon crystal has only a few thermally produced free electrons. Any temperature rise will result in thermal motion of atoms. This process is called *thermal ionization*.

The higher the ambient temperature, the stronger is the mechanical vibration of atoms and the lattice. These vibrations can dislodge an electron from the valence orbit. For example, if the temperature changes some ten degrees centigrade, the electrical resistance of pure germanium changes several hundred times. The materials the conductivity of which is found to increase very strongly with increasing temperature are called *intrinsic semiconductors*. The name "intrinsic" implies that the property is a characteristic of pure material that has nothing but silicon or germanium atoms. They are not only characterized by the resistive factor but also by the great influence that various factors, such as heat and light, have upon conductivity.

Recombination. The departure of the electron leaves a vacancy in the valence orbit. Such a vacant spot in the valence bond is called a *hole*. This hole acts in many respects as a positive charge because it will attract and capture any electron in the immediate vicinity, as presented in Fig. 1.1. Occasionally, a free electron will approach a hole, fill its attraction, and fall into it. This merging of a free electron and a hole is called *recombination*. In this way, valence electrons travel along the material. As far as both electrons and holes contribute to the conductivity, the holes in each case contribute about half as much as electrons. The average amount of time between the creation and recombination of a free electron and a hole is called the *lifetime*.

Voltage influence. The applied voltage will force the free electrons and holes to flow between the positive and negative terminals in the crystal. If the external voltage is applied to the semiconductor, the free electrons flow toward the positive terminal, and the holes flow toward the negative source terminal. In Fig. 1.2, the free electrons and holes move in opposite directions. From now on, we will visualize the current in a semiconductor as the combined effect of the two types of flow – the flow of free electrons through larger orbits in one direction and the flow of holes through the large and smaller orbits in other direction. Thus, free electrons and holes carry a charge from one place to another. They both are carriers in semiconductors in contrast to electrons in metals.

Doping. One way to raise conductivity is by *doping*. This means adding *impurity* atoms to a pure tetravalent crystal (*intrinsic crystal*). A doped material is called an *extrinsic semiconductor*. Impurity atoms added to the semiconductor change the thermal equilibrium density of electrons and holes. In the case of silicon, the appropriate impurities are elements from the 5^{th} and 3^{rd} columns of the periodic table, e.g. such as phosphorus and boron. By doping, two types of semiconductors may be produced.

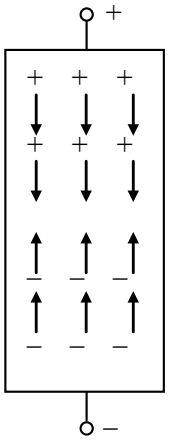


Fig. 1.2

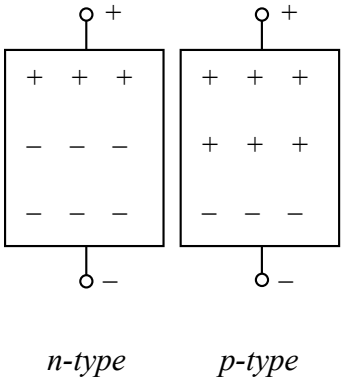


Fig. 1.3

First of them are *n-type semiconductors* with a pentavalent (phosphorus) impurity where the *n* stands for negative (Fig. 1.3) because their conduction is due to a transfer of excess electrons. A pentavalent atom, the one that has five valence electrons is called *a donor*. Each donor produces one free electron in a silicon crystal. In an *n*-type semiconductor, the free electrons are the *majority carriers*, while the holes are the *minority carriers* because the free electrons outnumber the holes.

Another type of semiconductors with a trivalent (boron) impurity has the hole type of conduction or deficit conduction by transfer from atom to atom of electrons into available holes. A semiconductor in which the conduction is due to holes referred to as a *p-type semiconductor*. Here, *p* stands for positive because of the carriers acting like positive charges, for the hole travels in a direction opposite to that of the electrons filling it. A trivalent atom, the one that has three valence electrons is called an *acceptor* or *recipient*. Each acceptor produces one hole in a silicon crystal. In a *p*-type semiconductor, the holes are the majority carriers, while the free electrons are the minority carriers because of the holes outnumber the free electrons.

Summary. Semiconductor crystals are very stable thanks to the covalent bond. However, unlike the metals their free carriers' density can be changed by many orders. Moreover, semiconductors exhibit a growth of resistance as the temperature falls, that is a bulk or a negative resistance. Because of thermal ionization, any temperature or light rise will result in significant motion of atoms that dislodges electrons from their valence orbits. The departure of the electron leaves the holes that carry the current together with electrons by the join recombination. This process speeds up when the voltage is applied. Doping additionally increases the conductivity of semiconductors. By doping, two types of semiconductors are produced – *p*-type with extra holes and *n*-type with excess electrons.

1.1.3 *pn* Junction

When a manufacturer dopes a crystal so that one half of it is *p*-type and the other half is *n*-type, something new occurs. The area between *p*-type and *n*-type is called *a pn junction*. To form the *pn* junction of semiconductor, an *n*-type region of the silicon crystal must be adjacent to or abuts a *p*-type region in the same crystal. The *pn* junction is characterized by the changing of doping from *p*-type to *n*-type.

Depletion layer. When the two substances are placed in contact, the free electrons of both come into equilibrium, both their number and the forces that bind them being unequal. Therefore, a transfer of electrons occurs, which continues until the charge accumulated is large enough to repel a further transfer of electrons. The accumulation of the charge at the interface acts as a barrier layer, called so due to its interfering with the passage of current.

As shown in Fig. 1.4, the *pn* junction is the border where the *p*-type and the *n*-type regions meet. Each circled plus sign represents a pentavalent atom, and each minus sign is the free electron. Similarly, each circled minus sign is the trivalent atom and each plus sign is the hole. Each piece of a semiconductor is electrically neutral, i.e., the number of pluses and minuses is equal.

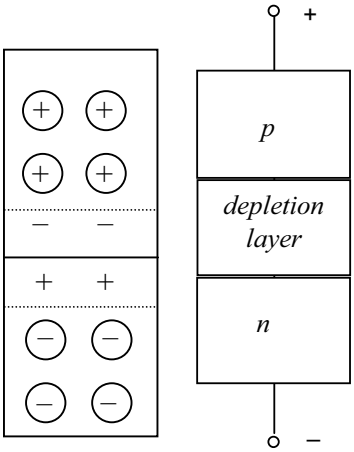


Fig. 1.4 **Fig.1.5**

The pair of positive and negative ions of the junction is called *a dipole*. In the dipole, the ions are fixed in the crystal structure and they cannot move around like free electrons and holes. Thus, the region near the junction is emptied of carriers. This charge-empty region is called the *depletion layer* also because it is depleted of free electrons and holes.

The ions in the depletion layer produce a voltage across the depletion layer known as the *barrier potential*. This voltage is built into the *pn* junction because it is the difference of potentials between the ions on both sides of the junction. At room temperature, this barrier potential is equal approximately to 0,7 V for a silicon dipole.

Biasing. Fig. 1.5 shows a dc source (*battery*) across a *pn* junction. The negative source terminal is connected to the *n*-type material, and the positive terminal is connected to the *p*-type material. Applying an external voltage to overcome the barrier potential is called the *forward bias*. If the applied voltage is greater than the barrier potential, the current flows easily across the junction. After leaving the negative source terminal, an electron enters the lower end of the crystal. It travels through the *n* region as a free electron. At the junction, it recombines with a hole, becomes a valence electron, and travels through the *p* region. After leaving the upper end of the crystal, it flows into the positive source terminal.

Application of an external voltage across a dipole to aid the barrier potential by turning the dc source around is called the *reverse bias*. The negative source terminal attracts the holes and the positive terminal attracts the free electrons. Because of this, holes and free electrons flow away from the junction. Therefore, the depletion layer is widened. The greater the reverse bias, the wider the depletion layer will be. Therefore, the current will be almost zero.

Avalanche effect. The only exception is exceeding the applied voltage. Any *pn* junction has maximum voltage ratings. The increase of the reverse-biased voltage over the specified value will cause a rapid strengthening of current. There is a limit to maximum reverse voltage, a *pn* junction can withstand without destroying. That is called a *breakdown voltage*. Once the breakdown voltage is reached, a large number of the carriers appear in the depletion layer causing the junction to conduct heavily. Such carriers are produced by geometric sequence. Each free electron liberates one valence electron to get two free electrons. These two free electrons then free two more electrons to get four free electrons and so on until the reverse current becomes huge. A phenomenon that occurs for large (at least 6…8 V) reverse voltages across a *pn* junction is known as an *avalanche effect*. The process when the free electrons are accelerated to such high speed that they can dislodge valence electrons is called an *avalanche breakdown* and the current is called a *reverse breakdown current*. When this happens, the valence electrons become free electrons that dislodge other valence electrons.

Operation of a *pn* junction in the *breakdown region* must be avoided. A simultaneous high current and voltage lead to a high power dissipation in a semiconductor and will quickly destroy the device. In general, *pn* junctions are never operated in the breakdown region except for some special-purpose devices, such as the Zener diode.

Zener effect. Another phenomenon occurs when the *intensity of the electric field* (voltage divided by distance known as a *field strength*) becomes high enough to pull valence electrons out of their valence orbits. This is known as a *Zener effect* or *high-field emission*. The breakdown voltage of the Zener effect (approximately 4 to 5 V) is called the *Zener voltage*. This effect is distinctly different from the avalanche effect, which depends on high-speed minority carriers dislodging valence electrons. When the breakdown voltage is between the Zener voltage and the avalanche voltage, both effects may occur.

Summary. When *p*-type to *n*-type substances are placed in contact, a depletion layer appears, which is emptied of free electrons and holes. A barrier potential of the silicon depletion layer is approximately 0,7 V and this value of germanium is about 0,3 V. In the case of forward bias, the voltage of which is greater than the barrier potential, the current flows easily across the junction. In the case of reverse bias there is almost no current. The exception is the avalanche effect of exceeding the applied reverse voltage 6…8 V across a *pn* junction. A simultaneous high current and voltage leads to a high power dissipation in a semiconductor and will quickly destroy the device. The similar phenomenon occurs when the intensity of electric field becomes very high. This Zener voltage of 4 to 5 V may destroy the device also.

1.2 Diodes

1.2.1 Rectifier Diode

A *diode* is a device that conducts easily being the forward biased and conducts poorly being the reverse biased.

Term and symbol. The word "diode" originates from Greek "di", that is "double". One of its main applications is in rectifiers, circuits that convert the alternating voltage or alternating current into direct voltage or direct current. It is also applied in detectors, which find the signals in the noisy operation conditions. The third application is in switching circuits because an ideal rectifier acts like a perfect conductor when forward biased and acts like a perfect insulator when reverse biased. A schematic symbol for a diode is given in Fig. 1.6.

The *p* side is called the *anode* from Greek "anodos" that is "moving up". An anode has positive potential and therefore collects electrons in the device. The *n* side is the *cathode*; it has negative potential and therefore emits electrons to anode. The diode symbol looks like an arrow that points from the anode (*A*) to the cathode (*C*) and reminds that *conventional current* flows easily from the *p* side to the *n* side. Note that the real direction of electron flow is opposite that is against the diode arrow.

Output characteristic. A diode is a nonlinear device meaning that its output current is not proportional to the voltage. Because of the barrier potential, a plot of current versus voltage for a diode produces a nonlinear trace. Fig. 1.7 illustrates the graph of diode current versus voltage named an *output characteristic* or a *volt-ampere characteristic*. Here, the current is small for the first few tenths of a volt. After approaching some voltage, free electrons start crossing the junction in large numbers. Above this voltage border, the slightest increase in diode voltage produces a large growth in current. A small rise in the diode voltage causes a large increase in the diode current because all that impedes the current is the bulk resistance of the *p* and *n* regions. Typically, the bulk resistance is less than 1 Ω depending upon the doping level and the size of the *p* and *n* regions. The point on a graph where the forward current suddenly increases is called the *knee voltage*. It is approximately equal to the barrier potential of the dipole. A silicon diode has a knee voltage of about 0,7 V. In a germanium diode it is about 0,3 V.

Forward biasing. If the current in a diode is too large, excessive heat will destroy the device. Even approaching the *burnout current* value without reaching it can shorten the diode life and degrade other properties. For this reason, a manufacturer's *data sheet* specifies the *maximum forward current* I_F that a diode can withstand before being degraded. This average current is the rate a diode can handle up to the forward direction when used as a rectifier. Another entry of interest in the data sheet is the *forward voltage drop* $U_{F\max}$ when the maximum forward current occurs. A usual rectifier diode has this value between 0,7 and 2 V.

Closely related to the maximum forward current and forward voltage drop is the *maximum power dissipation* that indicates how much power the diode can safely dissipate without shortening its life. When the diode current is a direct current, the product of the diode voltage and the current equals the power dissipated by the diode.

When an ambient temperature rises, the power rises also therefore the output characteristic is distorted, as shown in Fig. 1.7 by the dotted line. Fig. 1.8 shows the simple forward biased diode circuit. A current-limiting resistor **R** has to keep the diode current lower than the maximum rating. The diode current is given by: $I_A = (U_S - U_{AC}) / R$, where U_S is the source voltage and U_{AC} is the voltage drop across the diode.

Reverse biasing. Usually, the reverse resistance of a diode is some megohms under the room temperature and decreases by tens times as the temperature rises. The *reverse current* is a *leakage current* at the source rated voltage. Typically, silicon diodes have 1 to 10 μA and germanium 200 to 700 μA of leakage current. This value includes thermally produced current and surface-leakage current. When a diode is reverse biased, only these currents take place. The diode current is very small for all reverse voltages lower than the breakdown voltage. Nevertheless, it is much more dependent on temperature.

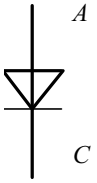


Fig. 1.6

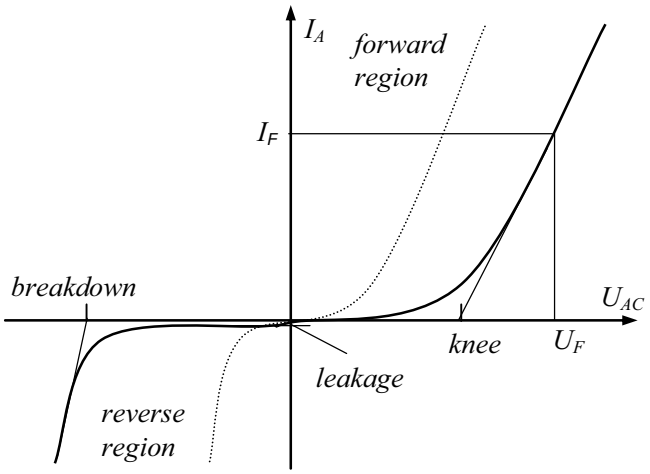


Fig. 1.7

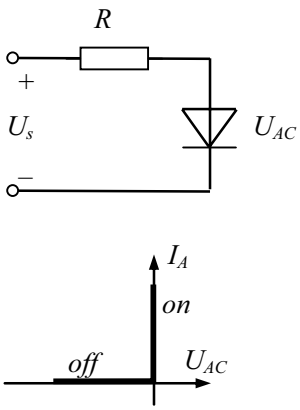


Fig. 1.8

At breakdown, the diode goes into avalanche where many carriers appear suddenly in the depletion layer. With a rectifier diode, breakdown is usually destructive. To avoid the destructive level under all operating conditions, a designer includes a *derating* (*safety factor*), usually of two.

Idealized characteristic. In view of a very small leakage current in the reverse-bias state and a small voltage drop in the forward-bias state as compared to the operating voltages and currents of a circuit in which the diode is used, the output characteristic of the diode can be idealized as shown in Fig. 1.8. This idealized corner can be used for analyzing the circuit topology but should not be used for actual circuit design. At turn on, the diode can be considered as an ideal switch because it turns on rapidly as compared to transients in the circuit. In a number of circuits, the leakage current does not affect significantly the circuit and thus the diode can be considered as an ideal switch.

Summary. The forward biased diode conducts easily whereas the reverse biased diode conducts poorly. The diode is the simplest non-controlled semiconductor device that acts like a switch for switching on the current flow in one direction and switching it off in the other direction. Unlike the ideal switch, a diode is a nonlinear device meaning that its output current is not proportional to the voltage. Its typical bulk resistance is near 1 Ω and forward voltage drop between 0,7 and 2 V. When an ambient temperature rises, the diode′s characteristic is slightly distorted. Due to high reverse resistance, a diode has a low leakage current, typically 1 to 700 μA for all reverse voltages lower than the breakdown. At breakdown, the diode goes into avalanche that may destroy it. This destructive level should be avoided.

1.2.2 Power Diode

A power diode is more complicated in structure and operational characteristics than the small-signal diode. It is a two-terminal semiconductor device with a relatively large single *pn* junction, which consists of a two-layer silicon wafer attached to a substantial copper base. The base acts as a heat sink, a support for the enclosure and one of electrical leads of the device. The extra complexity arises from the modifications made to the small-signal device to be adapted for power applications. These features are common for all types of power semiconductor devices.

Characteristics. In a diode, large currents cause a significant voltage drop. Instead of the conventional exponential output relationship for small-signal diodes, the forward bias characteristic of the power diode is approximately linear. This means the voltage drop is proportional both to the current and to ohmic resistance. The maximum current in the forward bias is a function of the area of the *pn* junction. Today, the rated currents of power diodes are thousands of amperes and the area of the *pn* junction may be tens of square centimeters.

The structure and the method of biasing of a power diode are displayed in Fig. 1.9. The anode is connected to the *p* layer and the cathode to the *substrate layer n*. In the case of power diode, an additional n^- layer exists between these two layers. This layer termed as a *drift region* can be quite wide for the diode. The wide lightly doped region adds significant ohmic resistance to the forward-biased diode and causes larger power dissipation in the diode when it is conducting current.

Forward biasing. Most power is dissipated in a diode in the forward-biased *on-state operation*. For small-signal diodes, power dissipation is approximately proportional to the forward current of the diode. For power diodes, this formula is true only with small currents. For large currents, the effect of ohmic resistance must be added. In a high frequency *switching operation*, significant switching losses will appear when the diode goes from the off-state to the on-state, or vice versa. Real operation currents and voltages of power diodes are essentially restricted due to power losses and the thermal effect of power dissipation. Therefore, in power devices cooling is very important.

Reverse biasing. In the case of reverse-biased voltage, only the small leakage current flows through the diode. This current is independent of the reverse voltage until the breakdown voltage is reached. After that, the diode voltage remains essentially constant while the current increases dramatically. Only the resistance of the external circuit limits the maximum value of current. Large current at the breakdown voltage operation leads to excessive power dissipation that should quickly destroy the diode. Therefore, the breakdown operation of the diode must be avoided.

To obtain a higher value of breakdown voltage, the three measures could be taken. First, to grow the breakdown voltage, lightly doped junctions are required because the breakdown voltage is inversely proportional to the doping density. Second, the drift layer of high voltage diodes must be sufficiently wide. It is possible to have a shorter drift region (at the same breakdown voltage) if the depletion layer is elongated. In this case, the diode is called *a punch-through diode*. The third way to obtain higher breakdown voltage is the boundary control of the depletion layer. All of these technological measures will result in the more complex design of power diodes.

Switching. For power devices, switching process is the most common operation mode. A power diode requires a finite time interval to switch over from the off state to the on state and backwards. During there transitions, current and voltage in a circuit vary in a wide range. This process is accompanied with energy conversion in the circuit components. A power circuit contains many components that can store energy (reactors, capacitors, electric motors, etc.). Their energy level cannot vary instantaneously because the power used is restricted. Therefore, switching properties of power devices are analyzed at a given rate of current change, as shown *transients* in Fig. 1.10.

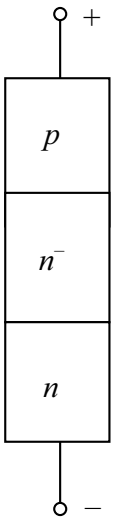


Fig. 1.9

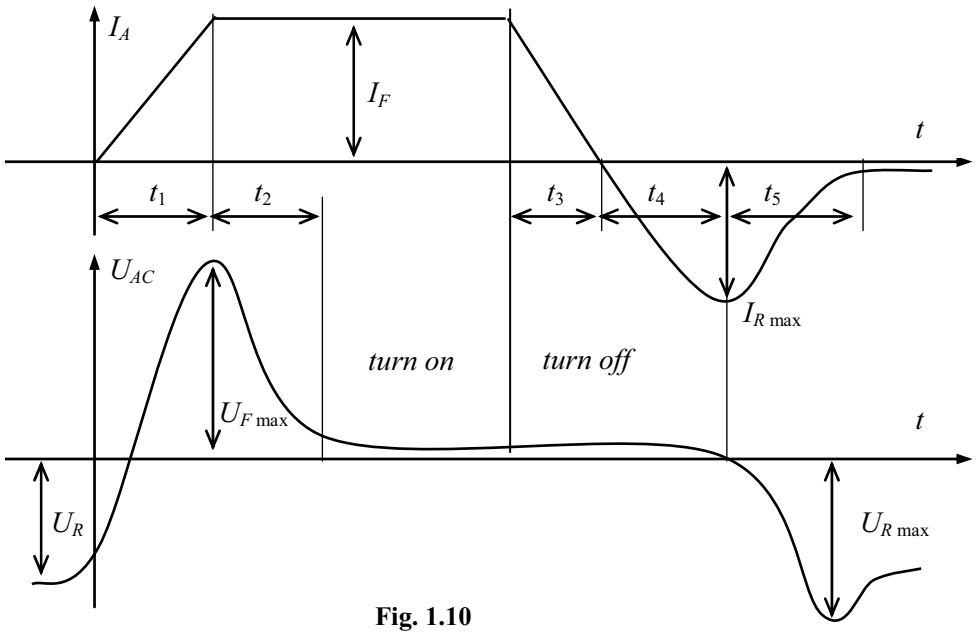


Fig. 1.10

The most essential data of power switching are the *forward voltage overshoot* $U_{F\,max}$ when a diode turns on and the *reverse current peak value* $I_{R\,max}$ when a diode turns off.

During the process, when the space charge is removed from the depletion region, the ohmic and inductive resistances cause a forward voltage overshoot of tens volts. The duration of the turn-on process of the power diode is the sum of two time intervals – the current growing time t_1 up to the steady state value I_F of the diode and the time t_2 up to stabilizing the forward on-state voltage. With high-voltage diodes (some kilovolts), the first time interval is approximately some hundreds of nanoseconds and the second about one microsecond, whereas usual diodes have these values tenfold less. Commonly, a shorter turn-on transients and lower on-state losses cannot be achieved simultaneously. The turn-off current and voltage transient process duration is the sum of three time intervals – the decreasing time t_3 of the forward current, the rise time t_4 of the reverse current, and the stabilizing time t_5 of the reverse voltage. The maximum value of the reverse current $I_{R\,\max}$ is fixed at the end of the second time interval and then the current value drops quickly. After the diode turns off, the current drops almost to zero with only small leakage current flows. A decrease in the diode reverse current raises the reverse voltage U_R, the maximum value of which reaches $U_{R\,\max}$. The sum of t_4 and t_5 is called a *reverse recovery time*.

Summary. Power diode is adapted for switching power applications. In addition to bulk resistance, it has high ohmic resistance. To withstand the essential losses that appear when the diode goes from the off state to the on state and backward, cooling is very important. To obtain a higher value of breakdown voltage, some measures are usually taken, such as lightly doped junctions, sufficiently wide drift layer, and the boundary control of the depletion layer. These measures result in a more complex design of power diodes but shorten the reverse recovery time and increase their lifetime.

1.2.3 Special-Purpose Diodes

Rectifier diodes are used in the circuits of 50 Hz to 50 kHz frequencies. They are never intentionally operated in the breakdown region because this may damage them. They cannot operate properly under abnormal conditions and high frequency. Devices of other types have been developed for such kind of operations.

Varactor. All the junction diodes have a measurable capacitance between anode and cathode when the junction is reverse biased, and this capacitance varies with the value of the reverse voltage, being least when the reverse voltage is high. In *a varactor* (Fig. 1.11) also called *voltage-variable capacitance*, *varicap* or *tuning diode*, the width of the depletion layer increases with the reverse voltage. Since the depletion layer gets wider with more reverse voltage, the capacitance becomes smaller. This is why the reverse voltage can control the capacitance of the varactor. This phenomenon is used in remote tuning of radio and television sets.

Zener diode. A *Zener diode* sometimes called *breakdown diode* or *stabilitrone,* is designed to operate in the reverse breakdown, or Zener, region, beyond the peak inverse voltage rating of normal diodes. This reverse breakdown voltage is called the Zener, or *reference voltage*, which can range between –2,4 V and –200 V (Fig. 1.12). The Zener effect causes a *"soft" breakdown* whereas the avalanche effect causes a sharper turnover. Both effects are used in the Zener diode. The manufacturer predetermines the Zener and avalanche voltages.

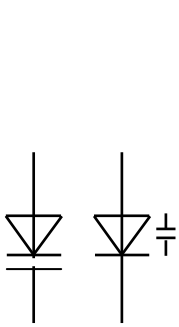


Fig. 1.11

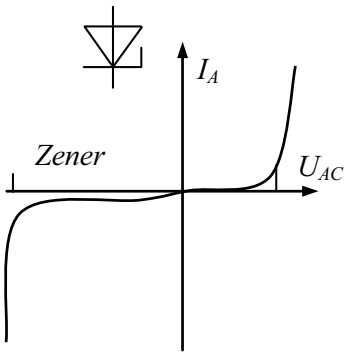


Fig. 1.12

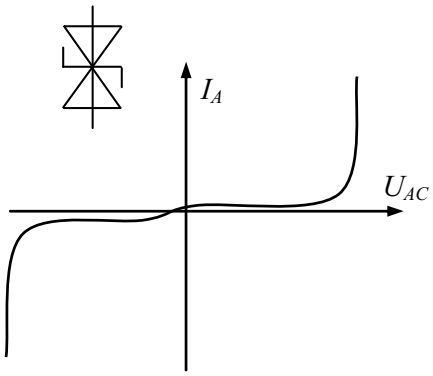


Fig. 1.13

A significant parameter of the Zener diode is the temperature coefficient that is the breakdown voltage deviation during the temperature rise or fall. The temperature coefficient of the Zener diode changes from negative to positive near –6 V. Because of this, by selecting the current value the designer may minimize the instability of the device. In all types of devices, the output levels are affected by variations in the load. Lower percentage values, approaching 0%, indicate better regulation. The Zener diode is the backbone of voltage regulators, circuits that hold the load voltage constant despite the large changes in line voltage and load resistance. When used as a voltage regulator, the Zener diode is reverse biased so that it will operate in the breakdown region with highly stable Zener voltage. In this region, changes in current through the diode have no effect on the voltage across it. The Zener diode establishes a constant voltage across the load within a range of output voltages and currents. Out of this range, the voltage drop remains constant and the current flow through the diode will vary to compensate the changes in load resistance.

A power Zener diode is called an *avalanche diode.* It can withstand kilovolts voltages and currents of some thousands of amperes.

Bi-directional breakdown diode. Lightning, power-line faults, etc. can pollute the line voltage by superimposing dips, spikes, and other transients on the normal voltage. *Dips* are severe voltage drops lasting microseconds or less. *Spikes* are short overvoltages of 500 or more than 2000 V. One of the devices used for line filtering is a set of two reverse-parallel-connected Zener diodes with a high breakdown voltage in both directions known as a *transient suppressor* or *voltage suppressor* (Fig. 1.13). It contains a pair of Zener diodes that are connected back-to-back, making the voltage suppressor bi-directional. This characteristic enables it to operate in either direction to monitor under-voltage dips and over-voltage spikes of the ac input. It is used as a filtering device to protect voltage-sensitive electronic devices from high-energy voltage transients. The voltage suppressor is connected across a primary winding of transformers to clip voltage dips and spikes and protect the equipment. The voltage suppressor must have extremely high power dissipation ratings because most of surges in ac power line contain a relatively high amount of power, in the hundreds of watts or higher. It must also be able to turn on rapidly to prevent damage to the power supply. In dc applications, a single unidirectional voltage suppressor can be used instead of a bi-directional voltage suppressor. It is shunt-connected with the dc input and reverse biased (cathode to positive dc). Often, a *varistor* (nonlinear *voltage-dependent resistor*) is used instead of the breakdown diode.

Schottky diode. As the frequency increases, the ordinary diode reaches a point where it cannot turn off fast enough to prevent noticeable current during the reverse half cycle. A special-purpose high frequency diode with no depletion layer, no *pn* junction, and extremely short reverse recovery time is called a *Schottky diode* or *reverse diode* (Fig. 1.14).

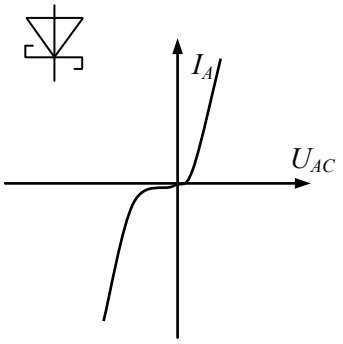


Fig. 1.14

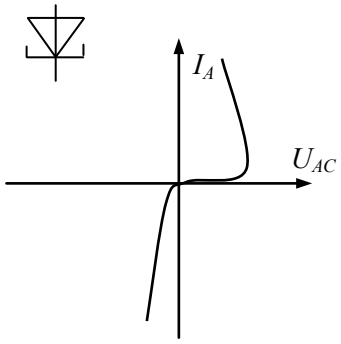


Fig. 1.15

The Schottky diodes are much faster than the rectifier diodes, but their breakdown voltage is relatively low. The operation of the Schottky diode is based on the concept that electrons in different materials have different absolute potential energies and potential energy of electrons in materials is lower than the potential energy of the free electrons. If an *n*-type semiconductor is in contact with a metal the electrons of which have a lower potential energy than the electrons in the semiconductor, the flux of electrons from the semiconductor into the metal will be much larger than the opposite flux because of the higher potential energy of electrons in the semiconductor. As a result, the metal will become negatively charged and the semiconductor will be charged positively. By that way, a *metal-semiconductor junction* is formed (*ms junction*), where the metal replaces the *p*-type side of the *pn*-junction. Compared with the *pn*-junction bipolar devices with a minority carrier current flow, in the Schottky diodes only the flow of majority carrier occurs.

The on-state voltage drop of the Schottky diode is approximately 0,3 V that is much less than the voltage drop of a rectifier diode (0,7…1 V). This will lead to smaller energy losses. The main advantage of the Schottky diodes over rectifier diodes is their very fast switching process near zero voltage with very small junction capacitance. They can operate at frequencies up to 20 GHz. These devices have a limited blocking voltage capability of 50 to 100 V (some series up to 1200 V) and sufficiently high current rating available is well below 100 A. The most important application area of the Schottky diodes belongs to computers the speed of which depends on how fast their electronic devices can turn on and turn off.

Tunnel diode. Diodes with a breakdown level equal to zero are called *tunnel diodes,* or *Shockley diodes*. The tunnel diode is a heavily doped diode that is used in high-frequency communication circuits for such applications as amplifiers, oscillators, modulators, and demodulators. The unique operating curve of the tunnel diode is a result of the heavy doping used in the manufacturing of the diode. The tunnel diode is doped about one thousand times as heavily as a standard *pn*-junction diode. This type of a diode exhibits a negative resistance. This means that a decrease in voltage produces an increase in current (Fig. 1.15). The negative resistance is useful in high-frequency circuits called oscillators, which create the sinusoidal signals.

Optoelectronics. Fig. 1.16,a displays a *light-emitting diode* (LED). This diode emits visible and invisible light rays when forward current through it exceeds the turn-on current. In the forward-biased LED, free electrons cross the junction and fall into holes. As these electrons fall from the higher to a lower energy level, they radiate energy. In rectifier diodes, this energy goes off in the form of heat. However, in a LED the energy is radiated as light. LEDs have replaced incandescent lamps in many applications because of their low voltage, long life, and fast on-off switching. LEDs are constructed of gallium arsenide or gallium arsenide phosphide. While their efficiency can be obtained when conducting as little as 2 mA of current, the usual design goal is in the vicinity of 10 mA. During conduction, a voltage drop on the diode is about 2 to 3 V that is twice more than the rectified diode.

Until the low-power liquid-crystal displays were developed, LED displays were common, despite high current demands in battery-powered instruments, calculators and watches. They are still commonly used as on-board enunciators, displays, and solid-state indicator lamps. Manufacturers produce LEDs that radiate green, yellow, blue, orange, or infrared (invisible) rays.

The same principle is used in *photoelectric cells*. When light energy bombards a *pn* junction, it can dislodge valence electrons. The more light striking the junction, the larger is the reverse current in a diode. Among the photoelectric cells that use this phenomenon, the most popular optoelectronic device is a *photodiode*. A photodiode is the one that has been optimized for its sensitivity to light. In this diode, a window lets light pass through the package to the junction. The incoming light produces free electrons and holes. The stronger the light, the greater the number of minority carriers and the larger the reverse current. Fig. 1.16,b shows of reverse biasing of the photodiode, where light becomes brighter and the reverse current increases.

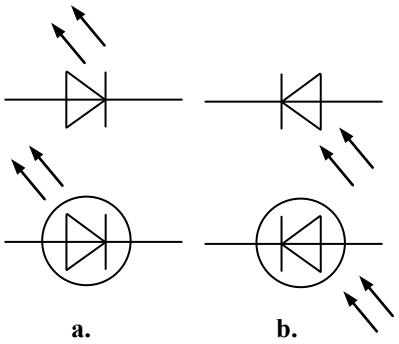


Fig. 1.16

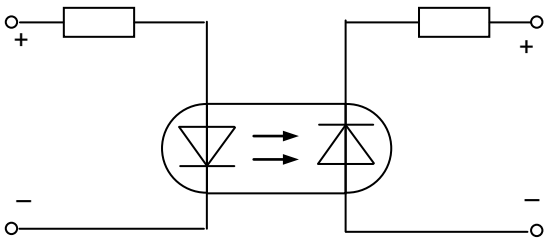


Fig. 1.17

The sensitivity zone of a photodiode spectrum is between 0,45 and 0,95 μm, which corresponds to the interval from blue to infrared light. A human eye perceives waves in the range of 0,45 to 0,65 μm therefore the photodiode can operate in the invisible rays.

In a sense, the photodiode is similar to a photoresistor also known as a light-dependent resistor (LDR) or a photovoltaic cell (FVC).

Another optoelectronic device is an *optocoupler* also called *optoisolator* that combines a LED and a photodiode in a single package. Fig. 1.17 illustrates the optocoupler that has a LED on the input side and a photodiode on the output side. The left source voltage and the series resistor set up a current through the LED. Then the light from the LED hits the photodiode, and this sets up a reverse current in the output circuit. This reverse current produces a voltage across the output resistor. The output voltage then equals the output supply voltage minus the voltage across the resistors. When the input voltage is varying, the amount of light is fluctuating and the output voltage is varying in step with the input voltage. In this way, the device can couple an input signal to the output circuit.

The key benefit of the optocoupler is electrical isolation between the input and output circuits as the only contact between the input and the output is a beam of light. Because of this, it is possible to have an insulation resistance between the two circuits in the thousands of megohms. Power optoelectronic modules can operate on 2 kV and 0,5 kA.

More diodes. Besides the special-purpose diodes discussed so far, there are a few more. A *constant-current diode* works in a way exactly opposite to the Zener diodes. Instead of holding the voltage constant, this diode holds the current constant when the voltage changes.

A *step-recovery diode* has an unusual doping profile because the density of carriers decreases near the junction. This phenomenon is called a *reverse snap-off*. During the positive half cycle, the diode conducts like any rectifier diode. Nevertheless, during the negative half cycle, the reverse current exists for a while because of the stored charges, and then suddenly drops to zero. This phenomenon is useful in frequency multipliers.

Zener diodes normally have breakdown voltages greater than –2 V. By increasing the doping level, a manufacturer achieves the Zener effect to occur near zero (approximately –0,1 V). A diode like this is called a *back diode* because it conducts better in the reverse than in the forward direction. Back diodes are occasionally used to rectify weak signals.

Summary. Special-purpose diodes successfully operate in the breakdown region, high-frequency applications, and other ad hoc conditions. The most widespread of them are Zener and Schottky diodes used in low-signal and middle-power applications, as well as optoelectronic devices for signal circuits and control systems.

1.3 Transistors

1.3.1 Common Features of Transistors

The word "*transistor*" was coined to describe the operation of a "transfer resistor". First, a *point-contact transistor* was produced. It included two diodes placed very closely together such that the current in either diode had an important effect upon the current in the other diode. By the proper biasing the diodes, it was possible to obtain power amplification of electric signals between the diode common layer, which lead was called *a base*, and other layers. One of the leads of this device was designated as *an emitter*, the corresponding diode was biased in the forward direction, the other was *a collector* and its diode was biased in the reverse direction. Power amplification was obtained by virtue of the fact that the few variations in the base current caused a large variation in the emitter-collector current. The point-contact transistor had certain drawbacks:

- high sensitivity to temperature, either ambient or self-generated;
- production problems, i.e., a difficulty to reproduce the same electrical qualities in close tolerance for mass production;
- low amplification, especially at high frequencies.

Intensive research has been done to diminish or remove these drawbacks. As a result, developers have produced semiconductor materials that are not so sensitive to temperature, inexpensive, operate at high frequencies, have low power dissipation, and internal noise of the transistor. A device, which is more stable both mechanically and electrically, has been constructed by forming junctions rather than point contacts. General classes of transistors that are used in electronics today are as follows:

- bipolar junction transistors (BJT);
- junction field-effect transistors (JFET);
- metal-oxide semiconductor field-effect transistors (MOSFET) up to some kilowatts, hundreds amperes, and tenths gigahertz;
- insulated-gate bipolar transistors (IGBT) up to thousands of kilowatts, some kiloamperes, and hundreds kilohertz.

More powerful devices have been built on the thyristors though IGBTs have the potential to replace them.

1.3.2 Bipolar Junction Transistors (BJT)

A *junction transistor* has three doped regions as shown in Fig. 1.18. The bottom region is the emitter, the middle region is the base, and the top one is the collector. This particular device is an *npn transistor*. Transistors are also manufactured as *pnp transistors*, which have all currents and voltages reversed from their *npn* counterparts. They may be used with negative power supplies and with positive once in an upside-down configuration.

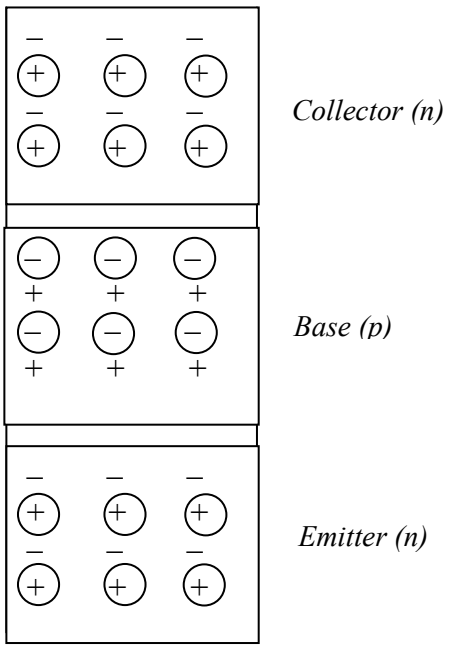


Fig. 1.18

Structure. A transistor has two junctions on opposite sides of a thin slab of semiconductor crystal – one between the emitter and the base, and another between the base and the collector. Because of this, a transistor is similar to two back-to-back connected diodes. The emitter and the base form one of the diodes, while the collector and the base form the other diode. From now on, we refer to these diodes as the *emitter diode* (the top one) and the *collector diode* (the bottom one). Accordingly, a *bipolar transistor* has three terminals: a collector, an emitter, and a base. Before diffusion has occurred, the depletion layers with the barrier potentials are at both junctions. The most common low-frequency transistor is the alloy type. The collector junction is made larger than the emitter one to improve the collector action.

After connecting of external voltage sources to the transistor, some new phenomena will occur. For normal operation, the emitter diode is forward biased and the collector diode is reverse biased (Fig. 1.19). Under these conditions, the emitter sends free electrons into the base. Since the base is lightly doped and thin, most of these free electrons pass through the base to the collector, which collects, or gathers, electrons from the base.

Basic topologies. Fig. 1.20 presents schematic symbols of *npn* and *pnp* transistors. There are three different currents in a transistor: emitter current I_E, base current I_B, and collector current I_C. Accordingly, the three *basic schemes* of the transistor connection in electronic circuits are usually discussed: *common emitter* (CE) connection, *common base* (CB) connection, and *common collector* (CC) connection.

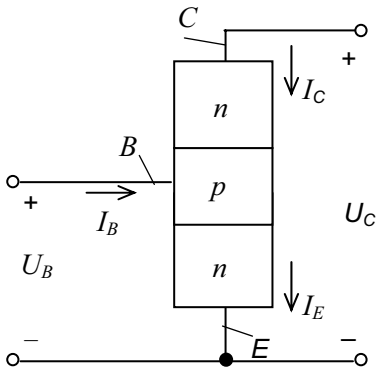


Fig. 1.19

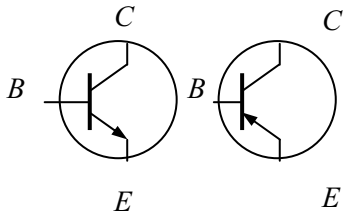


Fig. 1.20

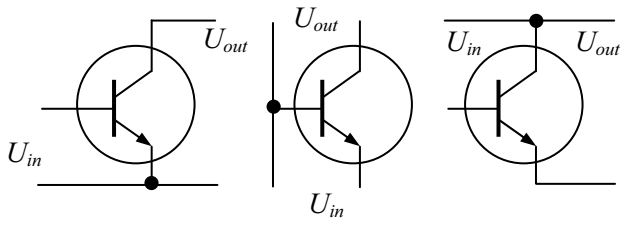


Fig. 1.21

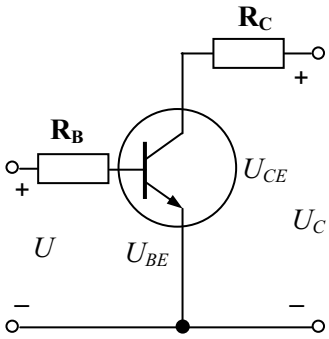


Fig. 1.22

In the first, shown in Fig. 1.21, the common node is an emitter and it is known as a *grounded emitter circuit*. Here, the input signal drives the base whereas the output signal occurs between the collector and the emitter. It is the most popular circuit because of its high flexibility and gain.

The second variant is a *grounded base circuit* because it has a common base node. Here, the input signal drives the emitter whereas the output signal occurs between the collector and the base. This connection is known as a low-gain circuit with high frequency selectivity Q. The common node of the third circuit is a collector. That is why this is a *grounded collector circuit*. Usually, this circuit is called also an *emitter follower*. Its input signal drives the base, and the output signal comes from the emitter. When connected between the CE transistor device and the small load resistance, the emitter follower can drive the small load under the stable voltage gain with no *overloading* and little distortion.

Beta and alpha gains. In Fig. 1.22, the common side, or groundside of each voltage source is connected to the emitter. Because of this, the circuit is an example of a CE connection with the *base circuit* to the left and the *collector circuit* to the right. Current from the energy supply enters the collector, flows through the base, and exits via the emitter. The collector current approximately equals to the emitter current. The base current is much smaller, typically less than 5 percent of the emitter current. The ratio of the collector current I_C to the base current I_B is called a *current gain* or *static gain* or *dc beta* of the transistor, expressed as

$$\beta = I_C / I_B.$$

This parameter is also called a *forward-current transfer ratio.* It is the main property of the transistor in the CE connection. For small-signal transistors, this is typically 100 to 300. The current gain of a transistor is an unpredictable quantity and may vary as much as a 3:1 range when changing in the temperature, the load, and from one transistor to another.

The *dc alpha* of a transistor indicates how close in value the collector current and the emitter current are; it is defined as

$$\alpha = I_C / I_E.$$

Alpha gain is the main property of the transistor in the CB connection. Consequently, a formula of alpha in terms of beta is

$$\alpha = \beta / (\beta + 1)$$

and vice versa

$$\beta = \alpha / (1 - \alpha).$$

Alpha gain is always less than unity and is near unity. Both gains depend on the signal frequency. In the data sheets, the limit frequency is shown, which reduces dc beta to unity.

Input characteristic. Fig. 1.23 displays an *input characteristic* or *transconductance* (*base*) *curve* of the CE connection. This graph of I_B versus U_{BE} looks like the graph of an ordinary rectifier diode. The maximum value of U_{BE} is limited in the transistor data sheets.

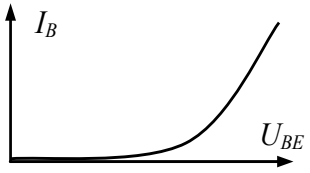


Fig. 1.23

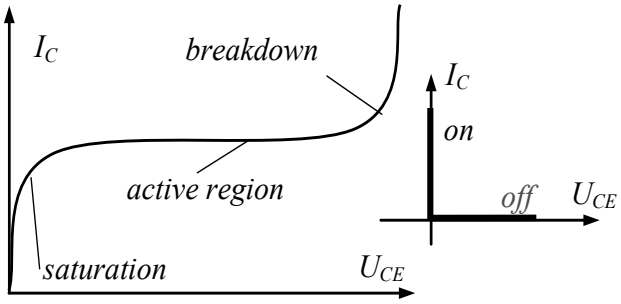


Fig. 1.24

Output characteristics. Fig 1.24 shows the output characteristic known here as a *collector curve* that is the collector current I_C as a function of the collector-emitter voltage U_{CE}. The collector curve has three distinct operating regions. First, there is the most important region in the middle called an *active region*. When the transistor is used as an amplifier, it operates in the active region. Another region is a *breakdown region*. The transistor should never operate in this region because it is very likely to be destroyed. The rising part of the curve, where U_{CE} is between 0 and approximately 1 V is called a *saturation region* or *ohmic region*. Here, the resistance of the device is very low and it is fully open. When it is used in digital circuits, the transistor usually operates in this region in a long time.

The idealized output characteristic of BJT operating as a switch is given in Fig. 1.24 as well.

Fig. 1.25 illustrates the set of collector characteristic curves under the different values of I_B. The bottom curve when there is no base current limits a *cutoff region* of the transistor where resistance is high, and the small collector current is called a *collector cutoff current*. As usual, a designer never allows voltage to get close to the maximum breakdown voltage U_{CE}, which is given in the data sheets for the transistor with an open base ($I_B = 0$).

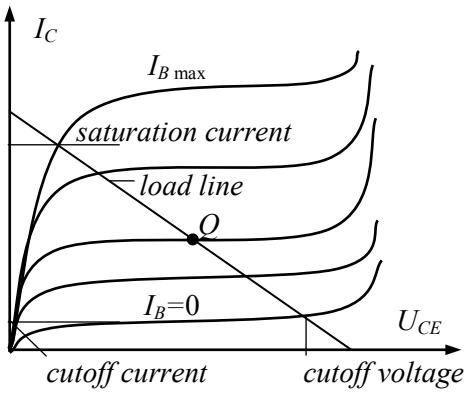


Fig. 1.25

A safety factor of two is common to keep U_{CE} well below the rating value. In digital circuits, the transistor may operate in the cutoff region. The upper curve in Fig. 1.25 limits the maximum collector rating. At this maximum, the transistor is in saturation and there is no sense to raise the base current more than $I_{B\,max}$.

Load line. A line in Fig. 1.25 drawn over the collector curves to show every possible operating point of a transistor is called a *load line*. Every transistor circuit has a load line. The top end of the load line is called *saturation*, and the bottom end is called *cutoff*. The first expresses the maximum possible collector current for the circuit, and the last gives the maximum possible collector-emitter voltage. The key step in finding the saturation current is to visualize a short circuit between the collector and the emitter. The key step to finding the cutoff voltage is to visualize an open between the collector and emitter.

The load line is expressed by the following equation:

$$I_C = (U_C - U_{CE}) / R_C.$$

Here U_C and U_{CE} are shown in Fig. 1.22. An *operating point* or *quiescent point Q* of the transistor lies on the load line. The collector current, collector-emitter voltage, and current gain determine the location of this point.

To calculate the maximum power dissipation of the transistor, we should write

$$P = I_C U_{CE} = (U_C U_{CE} - U_{CE}^2) / R_C$$

and solve the equation

$$dP / dU_{CE} = 0.$$

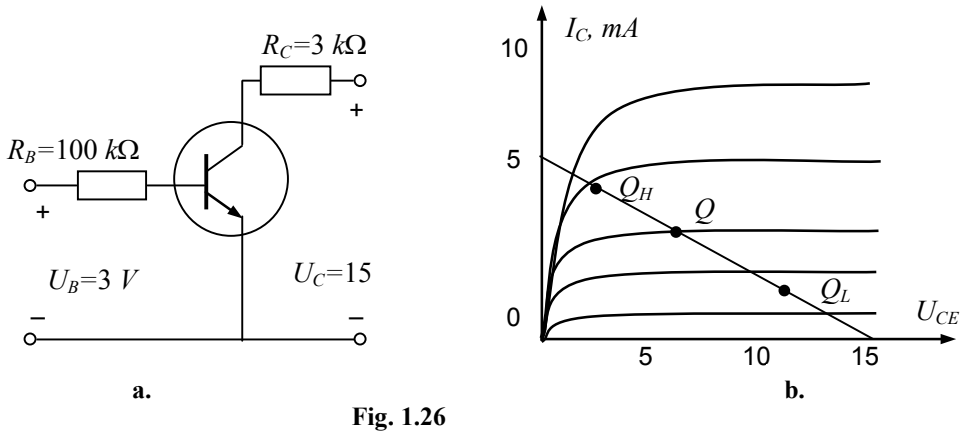
Thus, it seems by such a way that the maximum power dissipation occurs in the case of

$$U_{CE} = U_C / 2.$$

This power is equal

$$P_{\max} = U_C{}^2 / (4R_C).$$

Example. Fig. 1.26 is an example of a base-biased circuit. In the case of a short circuit across the collector-emitter terminals, the saturation current is 15 V / 3 kΩ = 5 mA. In the case when collector-emitter terminals are open, the cutoff voltage is 15 V. The load line shows the saturation current and cutoff voltage. The base current is approximately equal $I_B \approx 3$ V / 100 kΩ = 30 μA. Let the current gain of the transistor is β = 100. Then the collector current is $I_C = \beta \times I_B = 100 \times 30$ μA = 3 mA. This current flowing through 3 kΩ produces a voltage of 9 V across the collector resistor. Here, voltage across the transistor is calculated as follows: $U_{CE} = U_C - U_{RC} = 15 - 9 = 6$ V. Plotting 3 mA and 6 V gives the operating point *Q* shown on the load line of Fig. 1.26. If the current gain varies from 50 to 150, for example, the base current remains the same because the current gain has no effect on it. Plotting the new values gives the low point Q_L and the high point Q_H shown in Fig. 1.26.



a. **b.**

Fig. 1.26

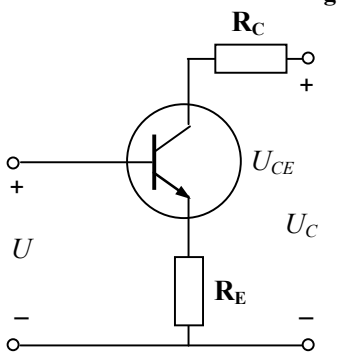


Fig. 1.27

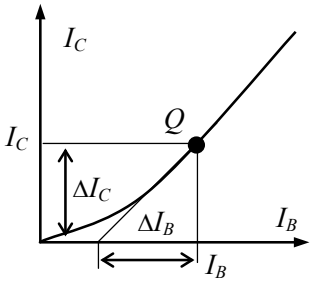


Fig. 1.28

In the emitter bias presented in Fig. 1.27, the resistor has been moved from the base circuit to the emitter circuit. Thanks to that one change, the Q point is now rock-solid and when the current gain changes, it shows no movement along the load line. The reason may be found by analyzing the circuit currents

$$I_E = I_C + I_B = I_C + I_C / \beta.$$

Solving this to the collector current gives

$$I_C = I_E \beta / (\beta + 1).$$

The quantity that multiplies I_E is called a *correction factor*. When the current gain is high, the correction factor may be ignored. Because of this, the emitter-biased circuits are usually designed to operate in the active region.

Transfer characteristic. Another important feature of the transistor is its *transfer characteristic* that sets the relation of the collector current versus the base current (Fig. 1.28). An *ac current gain* β_{ac} (*ac beta*) may be calculated from this curve in operating point Q as

$$\beta_{ac} = \Delta I_C / \Delta I_B.$$

Summary. The major benefits of BJT are as follows:

- stable output characteristics due to easy saturation;
- enough power handling capabilities, power dissipation is proportional to the current;
- low (less than 1 V) forward conduction voltage drop.

The main disadvantages of BJT are:

- relatively slow switching times, thus the operation frequencies are lower than 10 kHz;
- high control power by virtue of the current control;
- complex requirements to build the current controller.

1.3.3 Power Bipolar Transistors

Small-signal transistors usually dissipate half a watt or less. To dissipate more values, *power transistors* are needed. This rating is the limit of the transistor currents, voltages, and other quantities, which are much higher than those of the small-signal devices.

Structure. In most applications, *power bipolar transistors* are used in a CE circuit with the base as an input terminal and the collector output. In power electronic circuits, the bipolar *npn* transistors are more common than *pnp* transistors.

To obtain high current and high voltage capabilities, the structure of a power bipolar junction transistor shown in Fig. 1.29 is substantially different from that of the small-signal bipolar transistor. It has a low-doped drift region n– between the high-doped emitter and base layers. The drift region of power transistors is relatively large (up to 200 micrometers) and their breakdown voltage is hundreds of volts. To reduce the effect of current crowding in a small area (unequal current density), the base and emitter of power transistors are composed of many parts interleaved between each other. This multiple-emitter layout reduces the ohmic resistance and power dissipation in the transistor. The base thickness of a transistor must be made as small as possible in order to have a high amplification effect, but too small base thickness will reduce the breakdown voltage capability of the transistor. Thus, a compromise between these two considerations has been found. Therefore, as a rule, the current gain of high voltage power transistors is essentially lower than that of low-voltage transistors, typically 5 to 20.

The allowed maximum voltage U_{CE} between the collector and the emitter depends slightly on the base current. In power circuits, commutation losses should be diminished and the switching time of transistors must be sufficiently short. The turn-off process can be made much faster when the negative base pulses with abrupt fronts are applied. To adjust the switching processes and protect the transistor, special protection circuits (snubbers) are used.

Darlington transistors. Since the current gain of power bipolar transistors is small, two transistors are usually connected as a pair (Fig. 1.30,a). Such connection consists of the cascaded emitter followers. The emitter of the first transistor is connected to the base of the second one. A connected pair of bipolar transistors could raise the current gain of a power device. Commonly, this connection is designed monolithic because manufacturers put two transistors inside a single housing. This three-terminal device is known as a *Darlington transistor*. The summary current gain of such connection of two transistors $\mathbf{T}_1$, $\mathbf{T}_2$ is expressed as

$$\beta = \beta_1 + \beta_2 + \beta_1\beta_2,$$

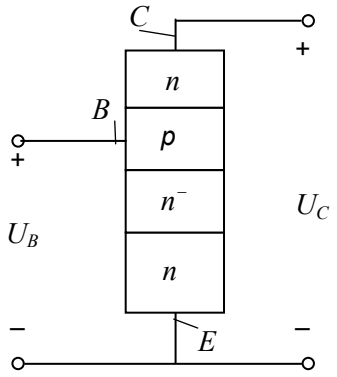


Fig. 1. 29

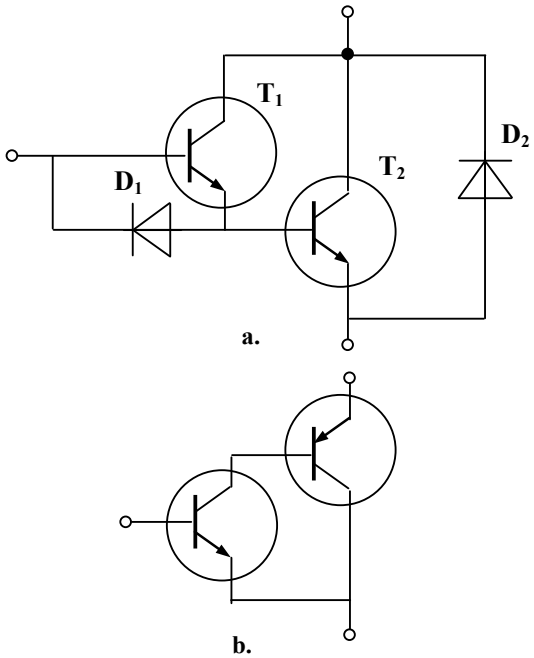


Fig. 1.30

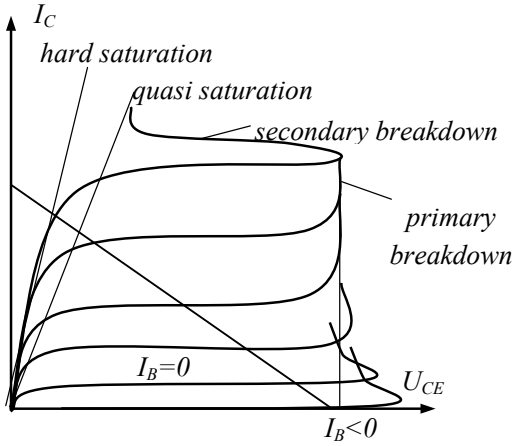


Fig. 1.31

i.e., the pair of transistors has a total current gain that is more than the product of the individual current gains. To speed up the turn-off time of the Darlington transistor, diodes $\mathbf{D}_1$ and $\mathbf{D}_2$ are added.

The complementary Darlington circuit shown in Fig. 1.30,b is a combination of the pair of bipolar transistors of different structures. Its current gain is equal to

$$\beta = (\beta_1 + 1)\beta_2 \approx \beta_1\beta_2,$$

i.e., the two transistors have a total current gain equal to the product of the individual current gains. In practice, the gain is somewhat less due to the difference of emitter currents. To equalize them, a resistor is added across the emitter junction of the right transistor. As a result, β approaches 100 to 5000.

Output characteristics. The output characteristics of a typical *npn* power transistor are shown in Fig. 1.31. The curves are given for the different base currents. The differences between power transistors and low-current transistors, shown in Fig. 1.31, are the regions labeled as a *primary breakdown* and a *secondary breakdown* as well as a *quasi saturation* on the power transistor characteristics. The small-signal transistors have no such regions. The operation of a power bipolar transistor in the primary and secondary breakdown regions should be avoided because of simultaneous high voltage and current and large power dissipation within the semiconductor. The difference of these breakdowns is that after the primary breakdown, the transistor can operate but the secondary breakdown destroys the transistor. As a result, a narrow *safe operating area* is the remarkable disadvantage of the transistor.

The forward voltage drop and power dissipation of a transistor in the quasi-saturation region are more significant than in the hard saturation region. The effect of the quasi-saturation operation appears in the switching processes when the transistor commutates from the off state to the on state or backward. An additional time interval is needed to move across the quasi-saturation operation region and the resultant switching time of power transistors will be higher than that of the small-signal transistors.

Summary. The main advantages of the power BJT are as follows:

- high power handling capabilities, up to 100 kVA, 1500 V, 500 A;
- sufficiently low forward conduction voltage drop.

The major drawbacks of the power BJT are:

- relatively slow switching;
- inferior safe operating area, thus the overvoltage protection is needed;
- complex requirements to build the current controller.

1.3.4 Junction Field-Effect Transistors (JFET)

In some applications, a *unipolar transistor* suits better than a bipolar one. The operation of the unipolar transistor depends only on one type of charge, either electrons or holes. A *field-effect transistor* (FET) is an example of the unipolar device. It is a special type of a transistor, which is particularly suitable for high-speed switching application. Its main advantage is that the control signal is voltage rather than current. Thus, it behaves like a voltage-controlled resistance with the capacity of high frequency performance. A *junction field-effect transistor* (JFET) is the first kind of FET.

Structure. Fig. 1.32 illustrates the normal way to bias a JFET. The bottom lead is called a *source*, and the top lead is a *drain*. The source and the drain of a JFET are analogous to the emitter and collector of the bipolar transistor. In the case of a *p*-channel JFET, a *p*-type material with different islands of *n*-type material is used. The action of a *p*-channel JFET is complementary, which means that all voltages and currents are reversed.

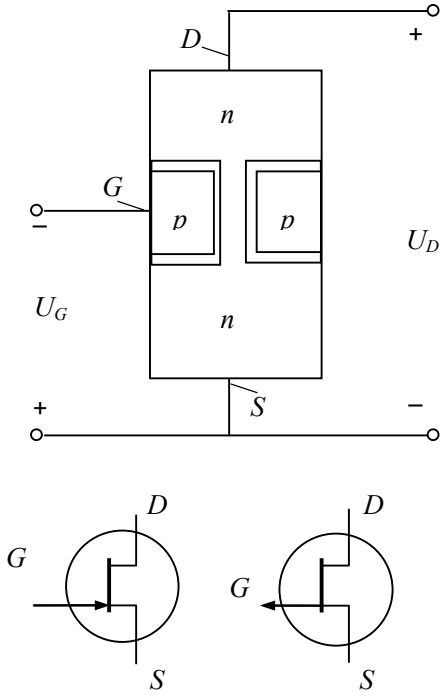


Fig. 1.32

To produce a JFET, two areas of a *p*-type semiconductor have been diffused into the *n*-type semiconductor. Each of these *p* regions is called a *gate*. When a manufacturer connects a separate lead to each gate, the device is called a *dual-gate JFET*. A dual-gate JFET is mostly used with a mixer, a special circuit applied in communications equipment. Most JFETs have two gates joined internally to achieve a single external gate lead, thus the device acts as though it has only a single gate. Incidentally, the gate of the JFET is analogous to the base of the bipolar transistor. Instead of the emitter current, a JFET has a source current I_S, rather than the base current it has a gate current I_G, and instead of the collector current it has a drain current I_D.

Biasing of the JFET is distinctly different from that of the bipolar transistor. In the bipolar transistor, the base-emitter diode is forward biased, but in the JFET, the gate-source diode is always reverse biased. Because of the reverse bias, only a very small reverse current can exist in the gate lead. As an approximation, the gate current is zero. This means that the input impedance of the device is close to infinity.

The supply voltage U_D forces free electrons to flow from the source to the drain. When electrons flow from the source to the drain, they pass through the channel between the two depletion layers. Unlike the current-controlled bipolar transistor, the JFET acts as a voltage-controlled device and the more negative the gate voltage U_G is, the narrower the channel and the smaller the drain current. The popular circuits built on the JFETS are as follows: a *common-source biasing*, a *common gate* topology, and a *source follower,* similar to those of a bipolar transistor.

Fig. 1.32 shows schematic symbols of *n*-channel and *p*-channel JFETs also. A schematic symbol of the *p*-channel JFET is similar to that of the *n*-channel JFET, except that the gate arrow points from the channel to the gate.

Output characteristics. Fig. 1.33 illustrates a set of *drain curves* of a JFET. The drain current I_D versus drain-source voltage U_{DS} increases rapidly at the first ohmic region, then levels off and becomes almost horizontal at the second active region. If the drain voltage is too high, the JFET breaks down. The minimum voltage of the second active region is called a *pinchoff*, and the maximum voltage is called the breakdown. Between the pinchoff and breakdown, the JFET acts approximately like a stable current device with a shorted gate. The gate voltage $U_{G\,off}$ of the bottom curve is called a *gate-source cutoff voltage*. This voltage closes the transistor. As shown in Fig. 1.33, in the ohmic region, the drain resistance depends on U_G. Unlike the bipolar transistors, one can change this quantity by altering the gate voltage. Typically, the on resistance of a FET device is on the order of 10 Ω to 100 Ω.

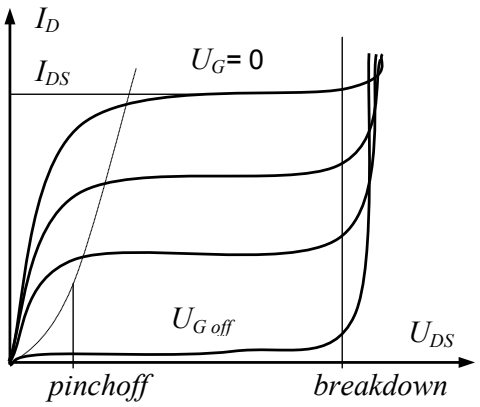


Fig. 1.33

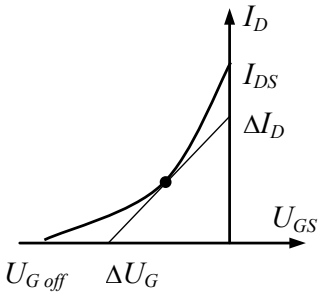


Fig. 1.34

Input characteristic. The input curve of a JFET, presented in Fig. 1.34, is a trace of the drain current I_D versus gate voltage U_G. It is the graphical solution of the following equation:

$$I_D = I_{DS}(1 - U_G / U_{G\,off})^2.$$

The quantity defined as

$$K = 1 - U_G / U_{G\,off}$$

is called a *K factor*. Because of the parabolic *K* factor, JFET is called *a square-low device*. This property gives the JFET some advantages over a bipolar transistor. Since instead of the current, the input voltage controls JFET, there is no current gain. The input conductivity (transconductance) is calculated as

$$G = \Delta I_D / \Delta U_G.$$

The unit of conductivity is Siemens ($1\ S = 1\ \Omega^{-1}$).

Summary. The main benefits of the JFET device are as follows:

- due to the voltage adjustment, the control circuit is simple, with a low control power;
- because a JFET is an electron majority carrier device, the switching transient speed grows essentially;
- for the same reason, its on-state resistance has a positive temperature coefficient, that is the resistance rise with the temperature rise;
- accordingly, the current falls with the load and the parallel connection of such devices is not the problem;
- due to the absence of the second breakdown, the safe operating area is large, therefore the overvoltage protection is not needed.

The drawbacks of the JFET are as follows:

- due to the high transistor resistance of the current flow, efficiency of FET is not high when a number of transistors are connected in parallel;
- additional losses between the source and the drain (*Miller's effect*) complicate the control processes.

1.3.5 Metal-Oxide Semiconductor Field-Effect Transistors (MOSFET)

MOSFET is an *n*-channel voltage-controlled *metal-oxide semiconductor field-effect transistor* that has a source, a drain, and a gate. Unlike a JFET, however, its metallic gate is electrically insulated from the channel by a thin layer of silicon dioxide. Because of this, the input resistance is even higher than that of a JFET.

Depletion-mode MOSFET. Fig. 1.35 shows a structure and a way to bias an *n*-channel *depletion-mode MOSFET* with a *p*-region called a *substrate*. Usually, the manufacturer internally connects the substrate to the source that results in a three-terminal device. Schematic symbols of *n*-channel and *p*-channel depletion-mode MOSFETs are shown in Fig. 1.36. As with a JFET, the gate voltage controls the width of the channel between the gate and substrate where electrons pass from the source to the drain. The more negative the gate voltage, the smaller the drain current. When the gate voltage is negative enough, the drain current is cut off. However, because the gate is electrically insulated from the channel, one can apply a positive voltage to increase the number of free electrons flowing through the channel. Being able to use a positive gate voltage is what distinguishes the MOSFET from the JFET again. There exists also a *p*-channel MOSFET.

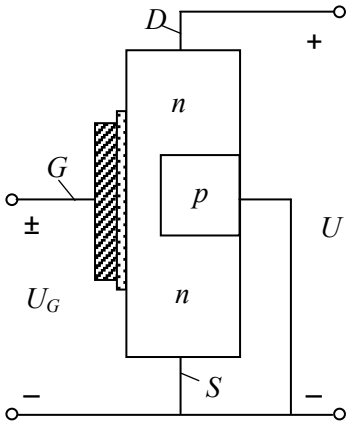


Fig. 1.35

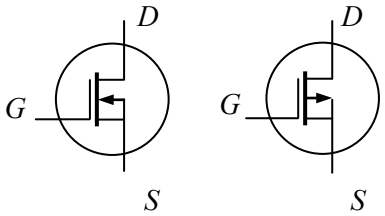


Fig. 1.36

The drain curves (Fig. 1.37) and the transconductance curve (Fig. 1.38) of the depletion-mode MOSFET are similar to the characteristics of JFET. The region where U_{GS} is between $U_{G\,off}$ and zero is called a *depletion-mode operation* area. If U_{GS} is greater than zero, *an enhanced-mode operation* occurs. The drain curves again display the ohmic region, the current-source region, and the cutoff region. I_{DS} is the drain current with a shorted gate, which is no longer the maximum possible drain current.

In its most basic form, the MOSFET looks like a voltage-controlled resistor, the resistance of which varies nonlinearly with the input voltage. In the on state, resistance can be less than 1 Ω, while in the off state, resistance increases to several hundreds of megohms, with picoampere leakage currents. Its fast switching characteristics are well controlled with the minimal parasitic circuit. MOSFET is bilateral that is it can switch positive and negative voltages and conduct positive and negative currents with an equal ease.

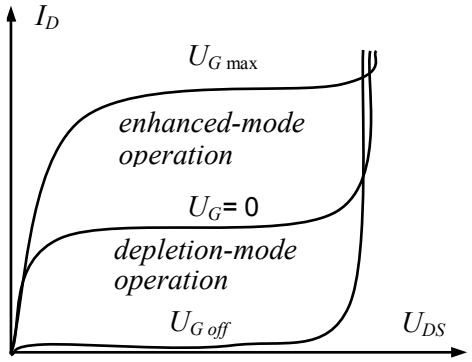


Fig. 1.37

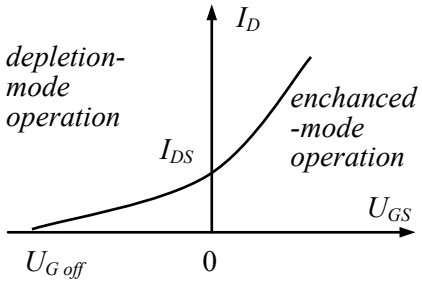


Fig. 1.38

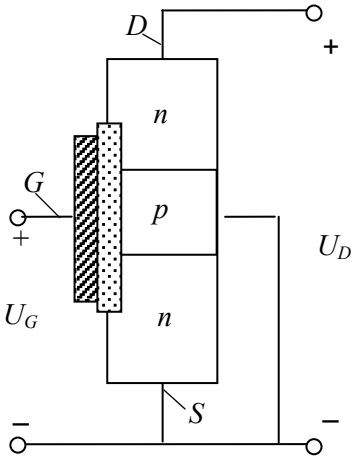


Fig. 1.39

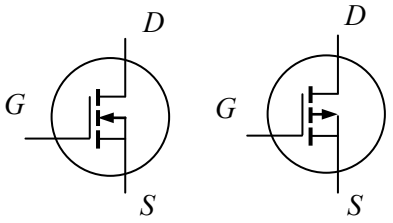


Fig. 1.40

Enhancement-mode MOSFET. Figs. 1.39 and 1.40 display an *n*-channel *enhancement-mode MOSFET*. These devices have revolutionized the electronics industry. Because there is no longer an *n* channel between the source and the drain, an enhancement-mode MOSFET is normally off when the gate voltage is zero, whereas a depletion-mode device is normally on. When the gate voltage is positive enough, electrons fill all the holes touching the silicon dioxide. The effect is the same as creating of a thin layer in *n*-type material next to silicon dioxide. This conducting layer is called an *n*-type *inversion layer*. The normally off device suddenly turns on and free electrons begin to flow easily from the source to the drain. The minimum U_{GS} that creates the inversion layer is called a *threshold voltage*, $U_{G\,th}$. Figs. 1.41 and 1.42 reflect a set of drain curves and an input curve for an enhancement-mode MOSFET, where the vertex (starting point) of the transconductance parabola and lowest drain curve are at $U_{G\,th}$.

Because of its threshold voltage, the enhanced-mode MOSFET is ideal for use as a switching device and its on-off action is the key to building personal computers and power applications.

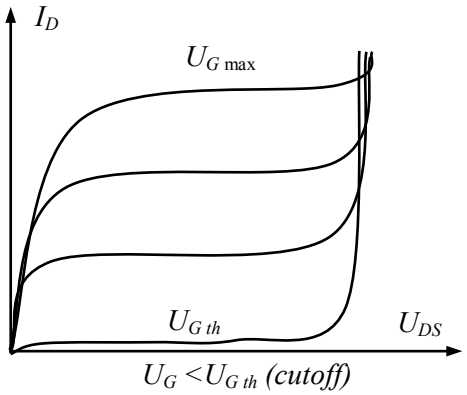


Fig. 1.41

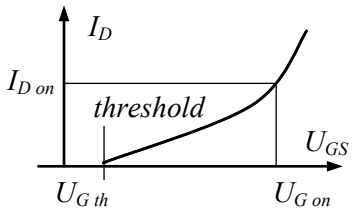


Fig. 1.42

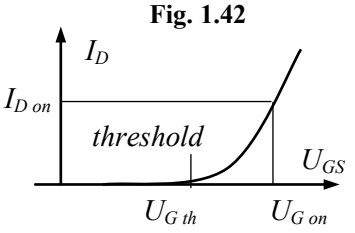


Fig. 1.43

Power enhancement-mode MOSFET. In the *power enhancement-mode MOSFET*, the structure of a semiconductor is composed of many thousands of cells connected in parallel to achieve a large gain and low on-state resistance. Overall, the input curve of the power MOSFET is almost linear compared with the parabolic transfer curve of the small-signal MOSFET (Fig. 1.43). When the MOSFET is driven by high gate-source voltage, it is operating in the ohmic region and its gain value is approximately constant. The threshold voltage is in the range of 2 to 4 volts. To keep the transistor in the off state when the gate voltage is zero, the drain-source voltage must not be higher rather than the maximum allowed value. In comparison with BJT, power MOSFET does not suffer the second breakdown and does not require the large base currents. The switching processes of the power MOSFET are much faster than those of bipolar transistors because they have no excess minority carriers that must be moved into or out of the device as it turns on or off. The switching intervals of the power MOSFET are in the range of 10 to 300 ns. These advantages make the power MOSFET suited to switching applications.

Double-diffused transistor (DMOS). This is a kind of a power MOSFET fabricated on a lightly doped *n*-type substrate. The DMOS device has a heavily doped region at the bottom for the drain contact. Two diffusions are used, one to create the *p*-type body region and another to create the *n*-type source region. It is operated by applying a positive gate voltage greater than the threshold voltage, which induces a lateral *n* channel in the *p*-type body region underneath the gate oxide. Current is conducted through the resulting short channel to the substrate and then vertically down the substrate to the drain. The DMOS transistor can have a breakdown voltage as high as 600 V and the current capability as high as 50 A is possible.

GaAsFET. This component is a high-speed field-effect transistor that uses gallium arsenide (GaAs) as the semiconductor material rather than silicon. It is generally used as a very high frequency amplifier (into the gigahertz range). The channel of the GaAsFET consists of a set of *n*-type or *p*-type doped germanium layers. The ends of the channel are the source and the drain. The terminal with the arrowhead represents the gate. GaAsFET is commonly used in microwave applications.

Data sheets. In data sheets, such parameters of FETs are usually shown as: signature, drain-source voltage U_{DS}, drain-source resistance R_{DS}, maximum drain current $I_{D\,max}$, maximum gate voltage $U_{G\,max}$, threshold gate voltage $U_{G\,th}$, and maximum power P_{max}. Examples are in the following data sheet of MOSFETs:

MOSFET	U_{DS}, V	R_{DS}, Ω	$I_{D\,max}$, A	$U_{G\,max}$, V	$U_{G\,th}$, V	P_{max}, W
IRFZ44	60	0,028	50	±20	4	150
IRF710	400	3,600	2	±20	4	36
IRFP710	100	0,055	41	±20	4	230
IRF820	500	3,000	2,5	±20	4	50

Summary. The advantages of the MOSFET are as follows:

- high speed switching capability, that is the operational frequencies up to 10 GHz with the transient speed 10…100 ns because of almost no saturation;
- switching of positive and negative voltages and conducting of positive and negative currents with equal ease;
- simple protection circuits;
- simple voltage control;
- normally off device if the enhancement-mode MOSFET is used;
- positive temperature coefficient makes it easy to be applied for parallel devices to increase their current-handling capability.

The drawbacks of the MOSFET are as follows:

- relatively low power handling capabilities (less than 10 kVA, 1000 V and 200 A); power losses are proportional to the square of current value;
- relatively high (more than 2 V) forward voltage drop, which results in higher losses than BJT.

1.3.6 Insulated Gate Bipolar Transistors (IGBT)

Bipolar transistors and MOSFETs have technical parameters and characteristics that complement each other. Bipolar junction transistors have lower conduction losses in the on state, especially at larger blocking voltages, but they have longer switching times. MOSFETs are much faster, but their on-state conduction losses are higher. Therefore, attempts were made to combine these two types of transistors on the same silicon wafer to achieve better technical features. These investigations resulted in the development of an *insulated gate bipolar transistor* (IGBT), which is becoming the device of choice in most new power applications.

Structure. Fig. 1.44 displays the structure of IGBT that is quite similar to that of enhanced-mode MOSFET. The principal difference is the presence of *p* layer that forms the collector of the IGBT. This layer arranges the *pn* junction, which injects minority carriers. The circuit symbols of *n*-channel and *p*-channel IGBTs are also presented. Their leads are the collector, the emitter, and the gate.

An equivalent circuit to simulate the IGBT operation is given in Fig. 1.45. This circuit presents the IGBT as a Darlington circuit with the bipolar transistors as the main transistors and the MOSFET as the driver device in the single housing. The current of $\mathbf{T}_1$ drives the base current of $\mathbf{T}_2$ and backward. By adjusting $\mathbf{R}_1$ and $\mathbf{R}_2$, the manufacturer sets a very high gain of IGBT.

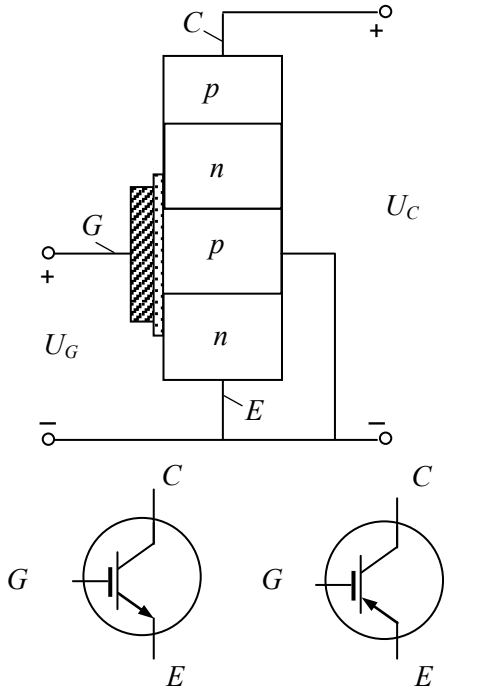


Fig. 1.44

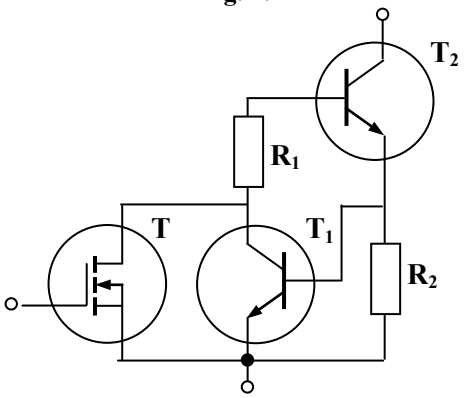


Fig. 1.45

Characteristics. The output curves of the *n*-channel IGBT and the input characteristic are shown in Fig. 1.46. The output curves are very similar to those of the small-signal *npn*-type bipolar transistor. The difference is that the gate signal of the IGBT is the gate voltage rather than the base current as for the bipolar transistors. Accordingly, the current of the input signal of IGBT does not flow through the gate. The transfer curve is identical to that of the power MOSFET. The curve is reasonably linear over most of the collector current range, becoming nonlinear only at low collector currents where the gate voltage is approaching the threshold.

The typical graphs of collector current versus frequency are shown in Fig. 1.47. In accordance with the frequency response, the higher is the switching frequency the lower is the maximum current. Nevertheless, overloading capacity of IGBT is 7 to 10 that is the pulse maximum current is 7 to 10 times greater than the rated collector current is.

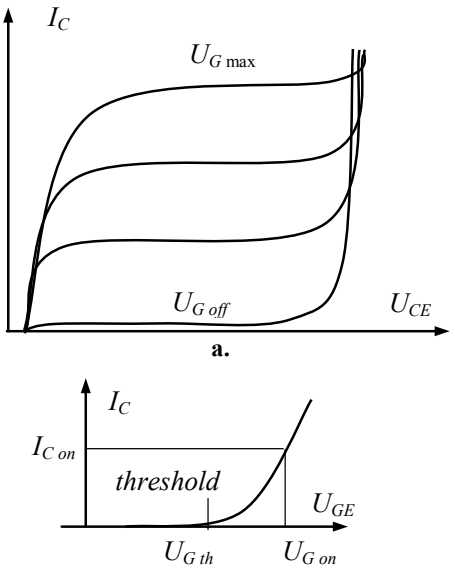


Fig. 1.46

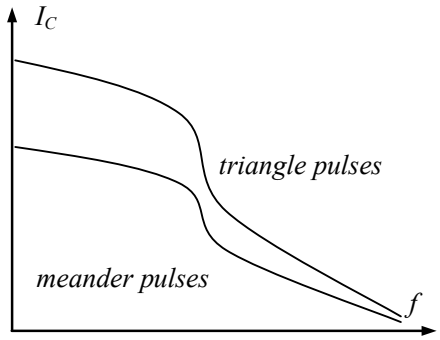


Fig. 1.47

Data sheets. In data sheets, such parameters of IGBTs are usually shown as: signature, collector-emitter voltage U_{CE}, maximum collector current $I_{C\,max}$, and maximum power P_{max}. Examples are in the following data sheet of IGBTs:

IGBT	U_{CE}, V	$I_{C\,max}$, A	P_{max}, W
IRGPH40U	1200	30	160
IRGPH50F	120	45	200
IRGDDN200M12	1200	200	1800
IRGDDN600K06	600	600	2600

Summary. The main features of the IGBT are as follows:

- the highest power capabilities up to 1700 kVA, 2000 V, 800 A;
- thanks to the lower resistance than that of the MOSFET, the heating losses of the IGBT are low too;
- highest switching capabilities;
- forward voltage drop is 2 to 3 V, that is higher than that of a bipolar transistor but lower than that of the MOSFET;

- due to the negative temperature coefficient, when the temperature rises, the power and heating is lowed, therefore, the device withstands the overloading and the operation in parallel well;
- the reliability is higher than with the FET thanks to the absence of a secondary breakdown;
- relatively simple voltage controlled gate driver and low gate current.

IGBTs are not prospective for the high frequency supply sources. The switching times of power IGBT modules are within the range of units to hundreds nanoseconds. For this reason, the leading manufacturer of IGBTs, International Rectifier, classifies the production by the four categories: "W" – warp speed devices for 17 to 150 kHz; "U" – ultra fast speed devices for 10 to 75 kHz; "F" – fast speed devices for 3 to 10 kHz; "S" – standard speed devices for 1 to 3 kHz.

1.4 Thyristors

1.4.1 Rectifier Thyristor (SCR)

A *thyristor* was invented in 1956 in General Electric. Its name is derived from the Greek "thyra" and means "door", that is allowing something to pass through. The main group of thyristors is composed by SCR, and others are the special-purpose devices.

Structure. A *silicon-controlled rectifier* (SCR) consists of a four-layer silicon wafer with three *pn* junctions. It has four doped regions, the anode (A), the cathode (C), and the gate (G). The gate is the control lead. The SCR is triggered into conduction by applying a gate-cathode voltage, which causes a specific level of gate current. The device is returned to its non-conducting state by either anode current interruption or *forced commutation*. When the SCR is turned off, it stays in a non-conducting state until it receives another trigger. Therefore, the SCR can be termed as *one-operation thyristor* or *rectifier thyristor*.

The structure, biasing circuit, and possible symbols of thyristors are shown in Fig. 1.48. First of them displays the anode-side SCR with an *n*-gate lead, the second is the cathode-side thyristor with a *p*-gate lead, and the last is the most common device. High-voltage high-power thyristors sometimes also have a fourth terminal, called an *auxiliary cathode*, used for connection to the triggering circuit. This prevents the main circuit from interfering with the gate circuit.

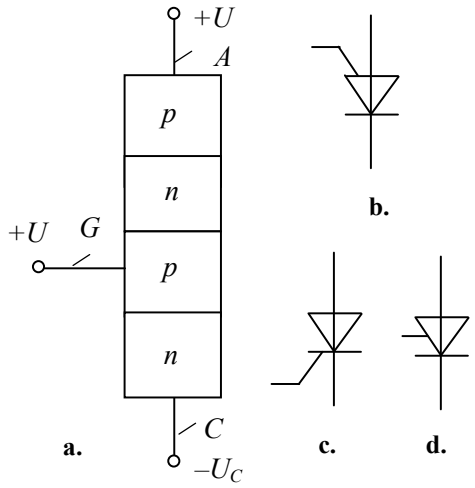


Fig. 1.48

Thyristors are commonly used in adjustable ac rectifier circuits, especially in power units up to 100 MVA. Their frequency capabilities are not high, in fact lower than 10 kHz.

Output characteristics. Fig. 1.49 illustrates the output curves and idealized output characteristics of a thyristor. The device has two operating regions: non-conducting and conducting. The current–voltage output characteristics for different gate currents show the forward bias. The output characteristic of a thyristor in the reverse bias is very similar to the same curve of the diode with a small leakage current. Using the same arguments as for diodes, the thyristor can be represented by the idealized characteristic in analyzing the circuit-desired topologies.

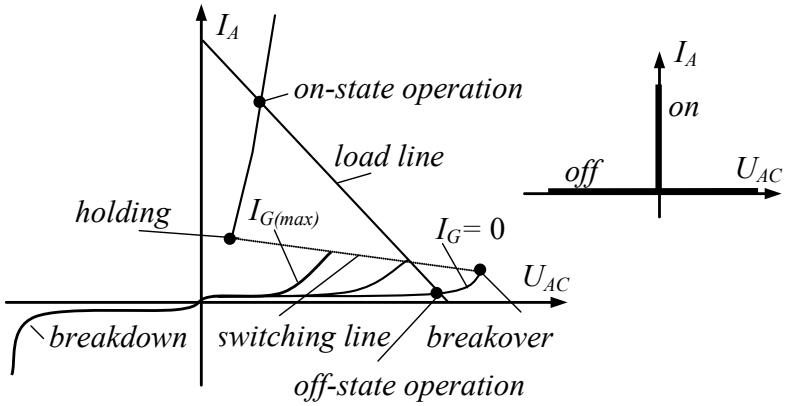


Fig. 1.49

When it is non-conducting, the thyristor operates on the lower line in the forward blocking state (off state) with a small leakage current. The thyristor is in off state until no current flows in the gate. The short firing pulse below the *breakover voltage* from the *gate driver* triggers the thyristor. This current pulse may be of triangle, rectangle, saw-tooth, or trapezoidal shape.

When a thyristor is supplied by ac, the moment of a thyristor firing should be adjusted by shifting the control pulse relative to the starting point of the positive alternation of anode voltage. This delay is called a *control angle* or *firing angle* α. In dc circuits, the use of thyristors is complicated due to their turning on/off.

After the pulse of the gate driver is given, the thyristor breaks over and switches along the dashed line to the conducting region. The dashed line in this graph indicates an unstable or temporary condition. The device can have current and voltage values on this line only briefly as it switches between the two stable operating regions. Once turned to the on state and the current higher than the *holding current*, the thyristor remains in this state after the end of the gate pulse.

When the thyristor is conducting, it is operating on the upper line. The current (up to thousands of amperes) flows from the anode to the cathode and a small voltage drop (1 to 2 V) exists between them. If the current tries to decrease to less than the holding border, the device switches back to the non-conducting region.

Turning off by gate pulse is impossible. Thyristor turns off when the anode current drops under the value of the holding current.

Input characteristics. Fig. 1.50 illustrates the input characteristics of the thyristor. The curves show the relation between the gate current and the gate voltage. This relation has a broad coherence area with a width that depends on the temperature and design properties of the device.

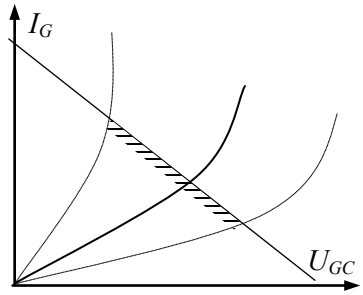


Fig. 1.50

The gate current has an effect upon the form of the characteristic. The value of the breakover voltage is the function of the gate current. The more is the gate current the lower is the voltage level required to switch on the thyristor. Maximum breakover voltage of a thyristor reaches up to thousands of volts. If the applied voltage exceeds the breakover level, SCR triggers without the gate pulse. This prohibited mode should be avoided.

Transients. Fig. 1.51 reflects the current and voltage transients of a thyristor when it turns on after the gate pulse appears and turns off after the current direction changes. During the thyristor opening process, the anode current will be distributed through the full crystal surface at the speed near 100 μm/μs. The current distribution is not homogeneous. The local overloading is possible; therefore, the growing rate of forward current I_F should be limited by hundreds of amperes per microseconds. For the best control of the thyristor firing process, the gate electrode has a specific spreading shape. The turn-on process includes three time intervals – the turn-on delay t_0, the current rise time t_1, and the current spreading time t_2.

The turn-off process of the thyristor is similar to that of a diode. For that, the anode current must be kept well below the hold current. The decreasing rate of the current depends on the circuit inductance. The density of excess carriers will diminish by the recombination. Although the current direction changes, the thyristor remains opened until the current attains its peak negative value $I_{R(max)}$. The voltage of the device remains small and positive. During the next time intervals of the reverse recovery time (t_4, t_5), the SCR will switch off and the reverse voltage U_R is stabilized. At the end of the turn-off process, the excess carriers remain in the medium layers and recombine until the forward voltage appears.

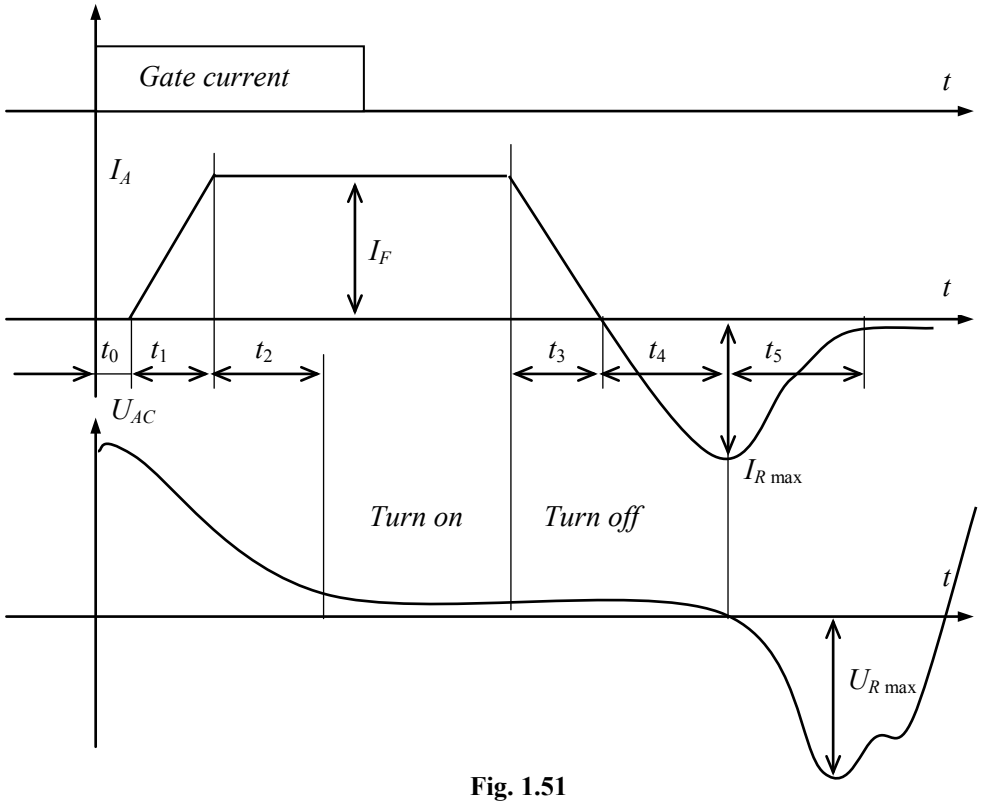


Fig. 1.51

Summary. The highest benefit of SCR is the ability to control its firing instant. The device withstands short circuit currents and has low on-state losses. Nevertheless, the semi-controlling is the drawback of the SRC devices.

1.4.2 Special-Purpose Thyristors

Besides the SCR, other thyristors have been developed for a multitude of application fields, capacities, and frequency ranges.

Diac. A *diac* is a bi-directional diode that can be triggered into conduction by reaching a specific voltage value. General Electric introduced this term as a "diode ac semiconductor device". It functions as two parallel Shockley diodes aligned back-to-back. The diac can pass current in either direction. Its equivalent circuit is a pair of non-controlled reverse-parallel-connected thyristors. The crystal structure of this device is the same as a *pnp* transistor with no base connection. The current-voltage characteristic and a symbol of a diac are shown in Fig. 1.52. A diac has neither an anode, nor a cathode. Its terminals are marked as MT_1 (main terminal 1) and MT_2 (main terminal 2). Like a rectified diode, every diode of a diac conducts the current in one direction only after the knee voltage exceeding. Once the diac is conducting, the only way to turn it off is by the current drop out. With the voltage values lower than the breakover level, the device cannot start conduction.

Triac. A *triac* (*bi-directional thyristor, simistor, tetrode thyristor*) is a three-terminal five-layer device capable of conducting current in both directions. It is identified as a three-electrode ac semiconductor switch that switches conduction on and off during each alternation. Fig. 1.53 gives a typical current-voltage curve and a schematic symbol of the triac. The triac is the equivalent of the two reverse-parallel-connected thyristors with one common gate. Its terminals are marked as MT_1 (main terminal 1), MT_2 (main terminal 2), and G (gate).

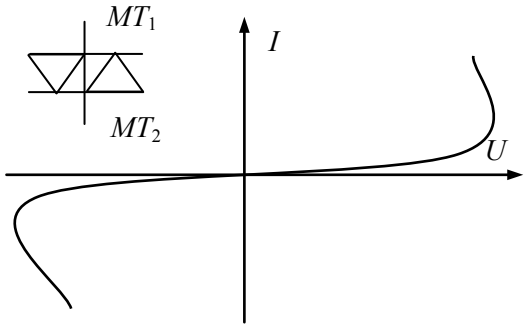


Fig. 1.52

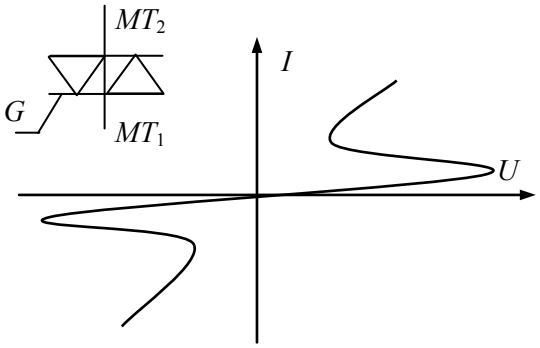


Fig. 1.53

Just as the rectifier thyristor, the device will conduct when triggered by a gate signal. The breakover voltage is usually high, so that the normal way to turn on the triac is by applying the forward bias trigger. The gate pulse is started in regard to MT_1. Conduction can be achieved in either direction with an appropriate gate current. Selection depends on the polarity of the source. During one alternation, conduction is through a *pnpn* combination. Conduction for the next alternation is by *npnp* combination. Triacs can operate in power modes of 1,5 kV and 100 A.

Gate turn-off thyristor. Besides the power rectifier thyristors, a *gate turn-off thyristor* (GTO) is produced. This device has two adjustable operations, thus it is known as a *two-operation thyristor* switch. The GTO can be turned on by the positive current gate pulse, and turned off by the negative current gate pulse. The cross terminals in Fig. 1.54 show that the symbol belongs to the GTO thyristors.

The turn-on control pulse of the GTO should be more powerful rather than that of the SCR, because the GTO has no regenerative effect on the gate electrode. The firing pulse has a very short front and long duration. This guarantees full and fast switching and minimum switching losses of the GTO. In danger, the anode current rapidly decreases and the thyristor can be closed. Since the temperature rises, the gate current should be diminished.

Commonly, the turn-on process of the GTO thyristor is the same as for the rectifier thyristor. The process includes the turn-on delay, the current rise and the stabilizing interval, similarly to those shown in Fig. 1.51. The switching speeds are in the range of a few microseconds to 25 μs. It is a sufficiently fast switching time. A switching frequency range is a few hundred hertz to 10 kHz. The on-state voltage drop (2 to 3 V) of the GTO thyristor is higher rather than that of the SCR.

For turning off, a powerful negative current control pulse must be applied to the gate electrode. The magnitude of the off-pulse depends on the value of the current in the power circuit, typically 20% of the anode current. Consequently, the triggering power is high and this results in additional commutation loss. The turn-off process consists of the three steps. The first one is a *storage time* when the negative current grows. The next is an *avalanche breakdown time*. During the last interval, the *tail current* flows between the anode and the gate. The gate terminal in the closed state of the GTO device should be on the negative voltage to achieve the best blocking and to minimize the influence of spikes and noise.

Because of their capability to handle large voltages (up to 5 kV) and large currents (up to a few kiloamperes and 10 MVA), the GTO thyristors are more convenient to use than the SCR in applications where high price and high power are acceptable.

MOS-controlled thyristor. A *MOS-controlled thyristor* (MCT) is a voltage-controlled device like the IGBT and the MOSFET, and approximately the same energy is required to switch an MCT as for a MOSFET or an IGBT. There may be a *p*-MCT and an *n*-MCT, as given in Fig. 1.55. The difference between the two arises from different locations of the gate.

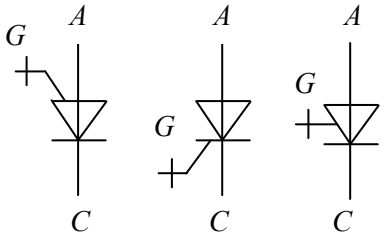


Fig. 1.54

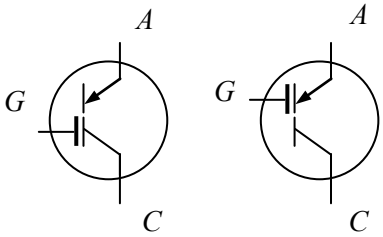


Fig. 1.55

The MCT has many of the properties of the GTO thyristors, including a low voltage drop at high currents. Here, turn on is controlled by applying a positive voltage signal to the gate, and turn off by a negative voltage. Therefore, the MCT has two principal advantages over the GTO thyristors, including much simpler drive requirements (voltage rather than current) and faster switching speeds (a few microseconds). Its available voltage rating is 1500 to 3000 V and currents of hundreds amperes. The last is less than those of the GTO thyristors. However, the MCT technology is in a state of rapid expansion, and significant improvements in the device capabilities are possible.

2 Electronic Circuits

2.1 Circuit Composition

2.1.1 Electronic Components

The primary components of electronics are the electronic devices:

- elementary components – resistors, capacitors, and inductors;
- diodes, including Zener, optoelectronic, diacs, and Schottky diodes;
- transistors, such as bipolar junction (BJT), field-effect (FET), and insulated gate bipolar (IGBT) transistors;
- thyristors, particularly silicon-controlled rectifiers (SCR), triacs, gate turn-off thyristors (GTO), and MOS-controlled thyristors (MCT).

The comparative diagram of power rating and switching frequencies of active devices is given in Fig. 2.1. The power range of some devices is shown in Fig. 2.2.

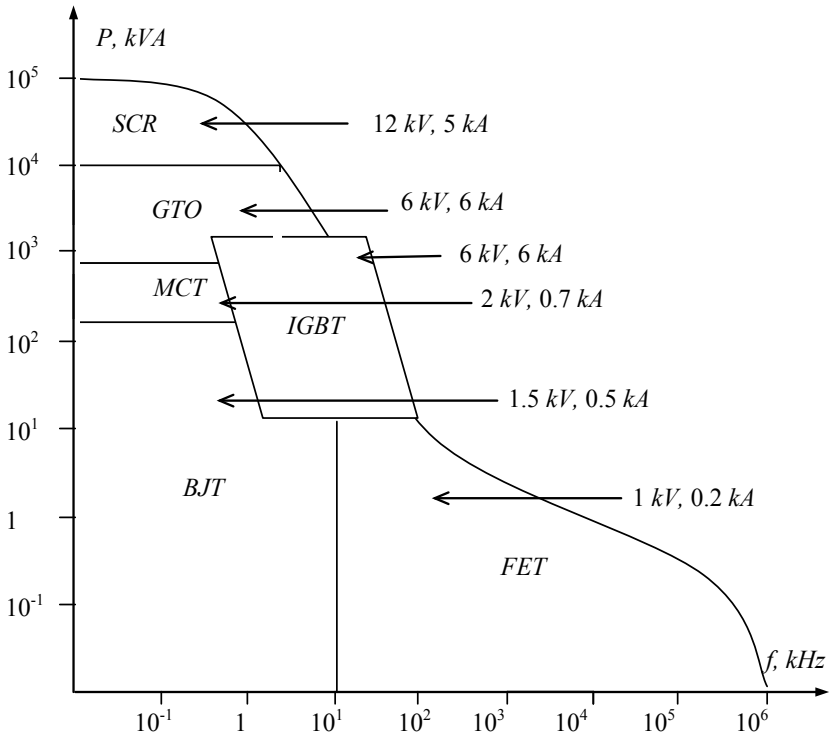


Fig. 2.1

The widespread classes of electronic circuits that are built on the primary components are as follows:

- ac amplifiers that change and control voltage and current magnitude;
- dc amplifiers that change and control current, voltage, and power magnitude with some forms of smoothing;
- analog circuits, such as filters and math converters;
- switching circuits, such as pulsers and digital gates;
- digital-to-analog and analog-to-digital data converters.

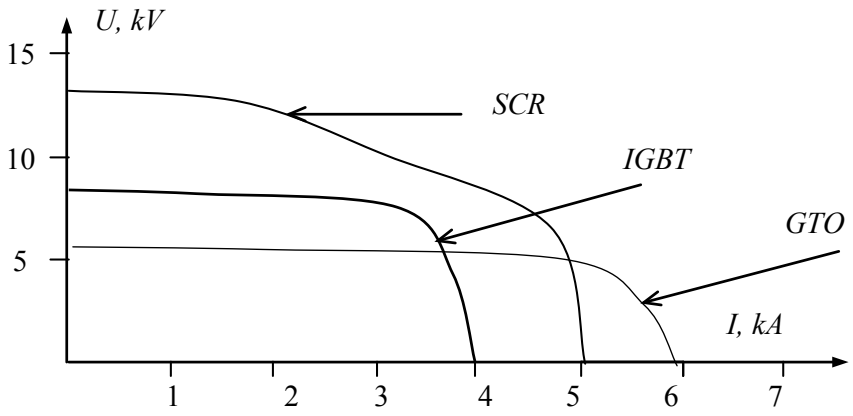


Fig. 2.2

Linear and nonlinear devices. Some electronic devices are linear, meaning that their current is directly proportional to their voltage. The reason they are called linear is that a graph of current plotted against voltage is a straight line. Resistors are commonly described as having linear characteristics, whereas capacitors and inductors, which store energy in magnetic fields, are nonlinear electronic elements. Diodes, transistors, and thyristors are normally classified as nonlinear devices and their behavior is represented on a graph by curved lines or lines which do not pass through the zero-voltage, zero-current point. Such behavior can be caused by temperature changes, by voltage-generating effects, and by conductivity being affected by voltage.

Resistors. *Resistors* come in a variety of sizes, related to the power they can safely dissipate. Color-coded stripes on a real-world resistor specify its *resistance R* and tolerance. Larger resistors have these specifications printed on them. Any electrical wire has resistance, depending on its material, diameter and length. The wires that must conduct very heavy currents (e.g. ground wires on lightning rods) have large diameters to reduce resistance. The power dissipated by a resistive circuit carrying electric current is in the form of heat. Circuits dissipating excessive energy will literally burn up. Practical circuits must consider power capacity. The power coupled by a resistor R with a current I flowing through it is as follows:

$$P = I^2R.$$

Inductors. An *inductor* is a *coil* of wire with *turns*. An inductance L specifies the inductor ability to oppose a change in the current flow. It reacts to being placed in a changing magnetic field by developing an induced voltage across the turns of the inductor, and will provide current to a load across the inductor. The inductors store energy in magnetic fields. Their charge and discharge times make them useful in time-delay circuits. The power of an inductor passing the current I upon the frequency f is expressed as follows:

$$P = LI^2f / 2.$$

Transformers. A *transformer* is one of the most common and useful applications of the inductors. It can step up or step down an input primary voltage U_1 to the secondary voltage U_2. The supply voltage is commonly too high for most of the devices used in electronics equipment; therefore, the transformer is used in almost all applications to step the supply voltage down to lower levels that are more suitable for use. The supply coil is called a *primary winding* and the load coil is called a *secondary winding*. The number of turns on the primary winding is w_1, and the number of turns on the secondary winding is w_2.

The turns are wrapped on a common *core*. For the low frequency applications, the massive core made of the transformer steel alloy must be used. The transformers intended only for higher audio frequencies can make use of considerably smaller cores. At radio frequencies, the losses caused by the transformer steels make such materials unacceptable and the ferrite materials are used as the cores. For the highest frequencies, no form of the core material is suitable and only the self-supporting, air-cored coils, usually of thick silver-plated wire, can be used. In the higher ultra high frequency bands, inductors consist of the straight wire or metal strips because the high frequency signals flow mainly along the outer surfaces of conductors.

Since the coefficient of coupling of the transformer approaches one, almost all the flux produced by the primary winding cuts through the secondary winding. Thus, the transformer is usually represented as a linear device. The voltage induced in the secondary winding is given by

$$U_2 = U_1 w_2 / w_1,$$

therefore the current is defined as

$$I_2 = I_1 w_1 / w_2.$$

In a *step-down transformer*, the *turns ratio* w_2 / w_1 is less than unity. Consequently, for a step-down transformer, the voltage is stepped down but the current is stepped up. The output apparent power of a transformer P_{S2} almost equals the input power P_{S1} or

$$U_2 I_2 = U_1 I_1.$$

The rated power of the transformer P_S is the arithmetic mean of the secondary and primary power.

The transformer can also be used in a center-tapped configuration. The voltage across the center-tap usually is half of the total secondary voltage.

Capacitors. A *capacitor* stores electrical energy in the form of an electrostatic field. Capacitors are widely used to filter or remove unnecessary ac components from a variety of circuits – ac ripple in dc supplies, ac noise from computer circuits, etc. They prevent the flow of direct current in a number of ac circuits while allowing ac signals to pass. Using capacitors to couple one circuit to another is a common practice. Capacitors h a predictable time to charge and discharge, and can be used in a variety of time-delay circuits. They are similar to inductors and are often used with them for this purpose.

The basic construction of all capacitors involves two metal plates separated by an insulator. Electric current cannot flow through the insulator, so more electrons pile up on one plate than on the other. The result is a difference in voltage level from one plate to the other. The power of a capacitive element operated under the voltage U on the frequency f is

$$P = CU^2 f / 2.$$

Loads. Every electronic circuit drives a *load* connected to the output. There are three kinds of loading. The load can be entirely ohmic (*resistive load*). There is no displacement between the current and the voltage of the load in this case, as shown in Fig. 2.3,a. When the load is ohmic-inductive (*resistive-inductive load*), the current is delayed in time compared to the voltage (Fig. 2.3,b). When the load is ohmic-capacitive (*resistive-capacitive load*), the current is time-wised in advance of the voltage (Fig. 2.3,c).

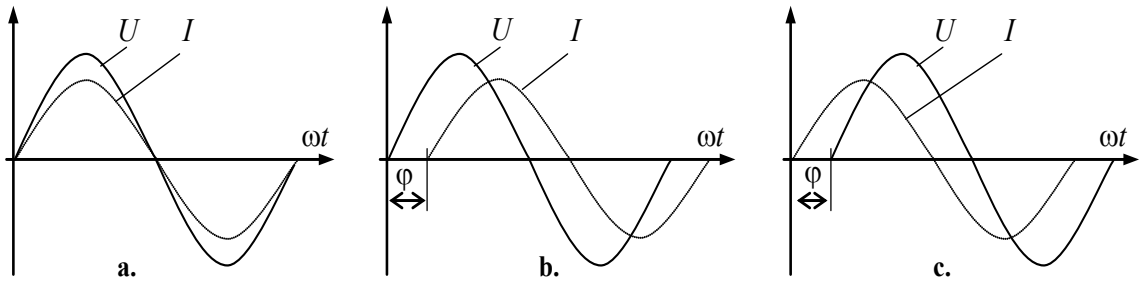


Fig. 2.3

Circuit efficiency depends on the load value, as shown in Fig. 2.4.

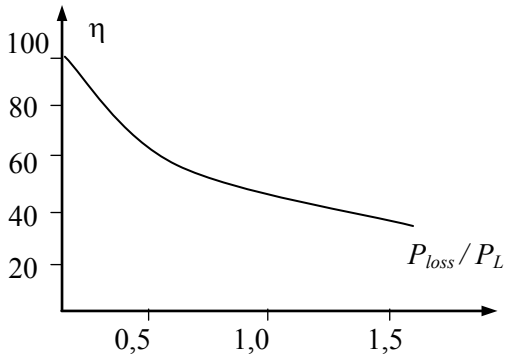


Fig. 2.4

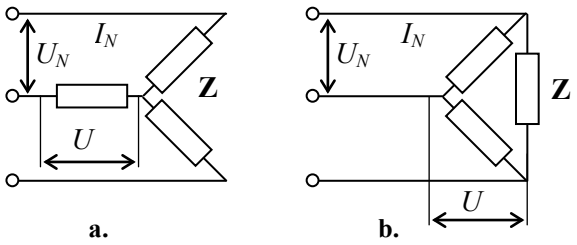


Fig. 2.5

In a three-phase system, the voltages are displaced 120 electrical degrees according to each other. The voltage between a phase wire and a neural wire is called a *phase voltage U*. The voltage between the two-phase wires is called a *mains voltage* U_N. Accordingly, the three-phase systems can have a *star-connected load* (*wye-connected load*) (Fig. 2.5,a) or a *delta-connected load* (Fig. 2.5,b).

In the star connection, one output of electronic circuit is connected to one of the load ends, whereas the other ends are short-circuited in the *star point*. Here,

$$U = U_N / \sqrt{3}.$$

For the currents, the following applies:

$$I_1 = I_2 = I_3 = I_N.$$

In the delta connection, the three branches are connected in series and each link is connected to the system output. The voltages above the various load branches are

$$U = U_N.$$

For the currents, the following applies:

$$I_1 = I_2 = I_3 = I_N / \sqrt{3}.$$

Summary. Linear and non-linear analog, switching, and mixed circuits present a multitude of analog and switching electronic equipment. Resistors, inductors, transformers, and capacitors participate in signal generation and conversion. Their operation depends on the load and has major effect upon the load behavior. The resistive load is the simplest one; it is easily described and controlled. In practice, the resistive-inductive load is the most widespread kind of an energy consumer and a signal provider. Sometimes, the electronic circuits include a resistive-capacitive load. Low-power single-phase and high-power three-phase loads meet the different requirements of domestic and industrial applications.

2.1.2 Circuit Properties

The power range and efficiency, the frequency response and step response of open loop and closed loop chains are the main properties of analog electronic circuits.

Frequency response. Fig. 2.6 shows the typical *frequency response* of an electronic system. This is a graph of the gain or output voltage versus the frequency of a sinusoidal signal. At low and high frequencies, the gain and the output voltage decrease because of the input and output capacitances of the system. In the middle range of the frequencies, the electronic system produces a maximum output signal. The frequencies above and below this middle range are avoided in most applications because of *amplitude distortion* and *frequency distortion*.

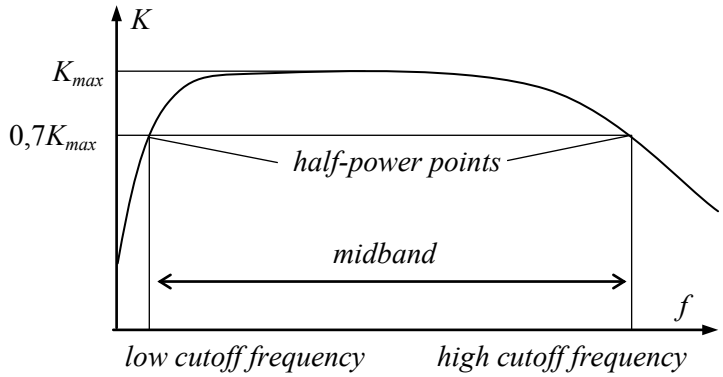


Fig. 2.6

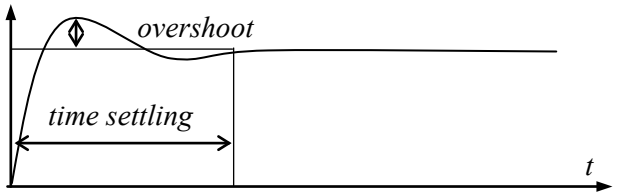


Fig. 2.7

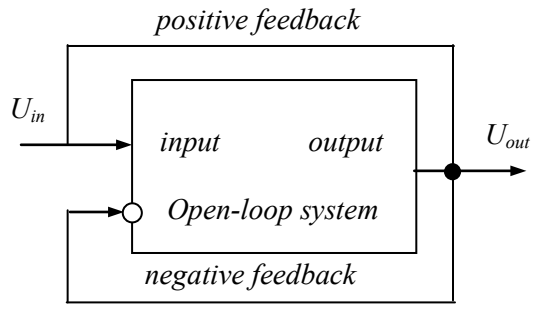
The *critical frequencies* of a system are the frequencies where the output signal is 0,7 (-3 decibel) of its maximum value. The alternating names for the critical frequencies are as follows: *cutoff frequencies*, *half-power points, break frequencies*, *corner frequencies*, 3-*dB frequencies,* etc. The range of frequencies between the cutoff frequencies where the output signal has its maximum value is called a *midband* or *bandwidth*. It is the area where the system is supported to operation. Other parts of the frequency curve are known as *sidebands*.

The midband of *audio signals* lies between 16 Hz and 20 kHz. When the frequency is higher than 10 kHz, the following applies to the *radio frequencies*:

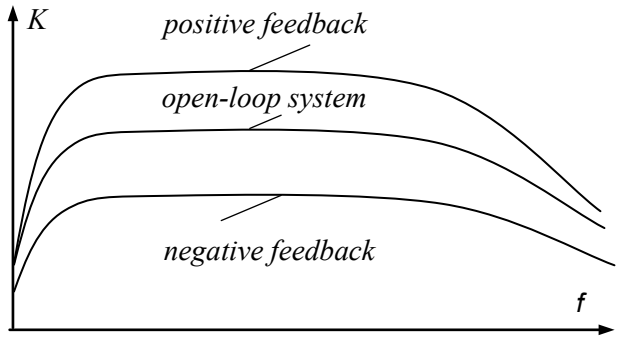
- 10 to 100 kHz – very low radio frequencies (VLF),
- 100 kHz to 2 MHz – long (LF) and medium (AM-radio, MF) radio frequencies, often called a broadcast band,
- 2 to 30 MHz – short radio waves of high frequency (HF) and video band,
- 30 to 300 MHz – meter television band (FM-radio, VHF),
- 300 MHz to 2 GHz – decimeter television and cell phone band,
- more than 2 GHz – ultra-high frequencies (UHF).

Step response Another typical characteristic of electronic circuit is a transient that is called a *step response*. An example given in Fig. 2.7 describes the system output when the step change in the input occurs. Ideally, when a device input changes, the output should change instantly. In practice, the output is likely to *overshoot*, *undershoot*, or both during the *settling time*. This uncontrolled movement of output during a transition is known as a *glitch*. The settling time of an electronic system is the time from a change of input to when the output comes within and remains within some error band. The shorter is the settling time and glitch the better is the system.

Feedbacks. The main toop to improve the frequency response or the step response is a feedback. The circuit, the output of which changes the input, at least partly, is called a *closed loop circuit* or a circuit with *feedback*. We refer to the *negative feedback* when the output signal enters the input with the negative polarity (Fig 2.8,a). The other term of this circuit is an *inverse feedback*. In this case, the voltage across the feedback input opposes the reference input voltage. The negative feedback reduces the gain, but improves the gain stability, decreases distortion, and enlarges the midband, as is seen in Fig 2.8,b.



a.



b.

Fig. 2.8

The loop is called a *positive feedback* when the output signal enters the input with the same polarity. In this case, the voltage across the feedback input corresponds to the reference input voltage. The positive feedback enlarges the gain, but deteriorates the gain stability and distortion and narrows the midband, as illustrated in Fig 2.8,b.

Summary. Electronic systems have some typical characteristics. First, it is the frequency response. In accordance with the frequency possibilities, different classes of circuits are in use, from zero to hundreds of gigahertz. Another circuit property is the step response. This feature determines the speed of operations, their starting and ending processes, and deals with the frequency response in detail. The feedbacks help to improve and correct both features.

2.2 Amplifiers

2.2.1 AC Amplifiers

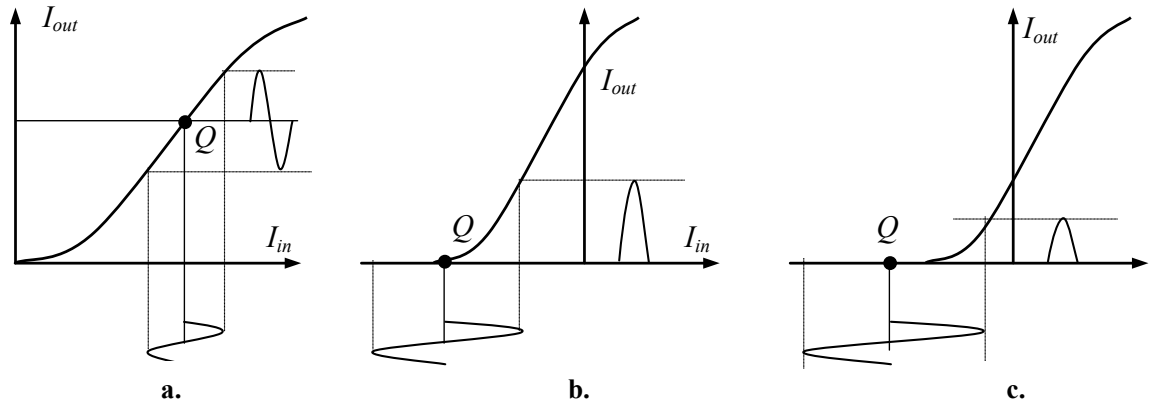
Altering of a voltage or current signal size as it is passed through a system is called an *amplitude control*. An *amplifier* is a circuit for the amplitude control provision. Except for early relatively inefficient electromechanical amplifiers, electronic amplifier development started with the invention of the vacuum tube.

Classes of amplifiers. Amplifiers are classified according to the polarity and properties of the output current or voltage. Their characteristics cover one, two, or four quadrants on the axes plane. The *ac amplifiers* and *dc amplifiers* are distinguished.

The fundamental specifications of ac amplifiers are listed on their data sheets; usually they include

- small-signal and large-signal bandwidths,
- voltage and current band noise,
- harmonic distortion level,
- input and output impedances,
- current and voltage gains.

It is common knowledge that amplifiers are divided into some general classes – A, B, C, etc., depending on the type of service in which they are to be used.



a. **b.** **c.**

Fig. 2.9

A *class A amplifier* is one which operates in the transistor's active region so that the output wave shapes of current are practically the same as those of the existing input signal at all times. Fig. 2.9 illustrates the typical transfer graphs of the collector current versus the base current. For the class A amplifier (Fig. 2.9,a), if the input signal is sinusoidal, the output signal is also sinusoidal. Consequently, the low *clipping* is the main advantage of this mode of operation. For this reason, the amplifiers of such kind are known as *linear amplifiers*. Low efficiency (30 to 45%) is the main drawback of the class A amplifier. For this reason, it is commonly used in low-power applications and *preamplifiers*.

A *class B amplifier* operates with a negative bias approximately equal to cutoff. Its base voltage is more negative than in the class A amplifier. Therefore, the output current is almost zero when the alternating input signal is removed or negative (Fig. 2.9,b). With a sinusoidal signal applied, the output consists of a series of half-sine waves. A bottom part of this half-wave is distorted, and the border of this distortion is called a *cutoff zone*. The amplifiers of such kind are known as *pulse amplifiers* with high clipping. Efficiency of the class B amplifier is higher (45 to 70%) than in the class A amplifier. For this reason, they are used as the balanced output stages. More often, the intermediate class AB is selected, the clipping of which is much less.

A *class C amplifier* operates with a negative bias essentially less than cutoff. It passes the current during the part of the positive alternation only. The output current has narrow width and its shape distortion is maximal (Fig. 2.9,c). Its high efficiency (70 to 90%) is the primary considerations at radio frequencies higher than 20 kHz. The class C amplifiers are preferable in power amplifiers with resonance load, for example, transmitters.

A *class D amplifier* uses transistors as switches where the only modes are switch on and switch off. It is used in different switching circuits.

The dc and ac load lines. The maximum unclipped peak-to-peak output of an amplifier is called an *output voltage swing* or *MPP*. Earlier, the dc load line was used to analyze biasing circuits. On the typical transfer characteristic shown in Fig. 2.10, which is the graph of the collector current versus the base current, *Q* point corresponds to the current gain β that is the slope of the curve at the point *Q*.

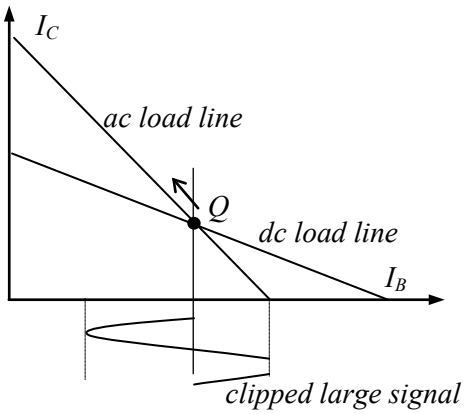


Fig. 2.10

It is normal to distinguish the small-signal operations of unclipped signals and the large-scale operations with the signal clipping. Under the small-signal operation, the emitter current has the same frequency and phase as the ac base voltage and approximately similar shape, usually with some distortion. The same is not true for the large-signal operation.

Because of this, most of amplifiers have two load lines: a *dc load line* and an *ac load line*. Commonly, the ac signal is considered small when a peak-to-peak ac emitter current is significantly less than the dc emitter current is. The dc current gain was defined earlier as β. The ac current gain β_{ac} equals the ratio of the change in the collector current to the change in the base current.

The ac load line helps to analyze large-signal operations. As is seen in Fig. 2.10, the saturation and cutoff points on the ac load line are different from those on the dc load line. In addition, the ac load line is steeper (has a higher slope) because the ac collector resistance is smaller than the dc one. The maximum load power occurs when an amplifier produces the MPP unclipped output as discussed earlier. Efficiency of an amplifier is equal to the ac load power P divided by the dc power from the supply P_S times 100 percent. The class A amplifiers have poor efficiency, typically well under 45 percent. This is because of power losses in the biasing resistors, the collector resistor, the emitter resistor, and the transistor.

The key to building the more efficient amplifiers is to reduce the unwanted power losses. One way to reduce the power losses is to derate (reduce) the power rating when the ambient temperature increases in accordance with the specified derating factor. Another way is to get rid of the heat faster. That is why heat sinks are used. Large power transistors have a collector connected directly to the case to allow heat escape as easily as possible.

CE current amplifiers. Fig. 2.11,a displays a simple transistor linear amplifier. Because the emitter is at the ac ground, it is the CE amplifier. In this circuit, the ac input signal U_{in} is added to the dc biasing voltage U_B. They produce a voltage drop in the base resistor $\mathbf{R}_B$. As a result, the total base voltage changes in accordance with the input signal. Since the base voltage changes, the collector current changes also, as well as the voltage in the resistor $\mathbf{R}_C$ and in the load supplied by U_{out}. The amplified ac collector voltage is equivalent to being 180 degrees out of phase with the input voltage.

A transistor *current amplifier* used in practice is presented in Fig. 2.11,b. Here, variations in both resistor voltages (base and collector) and transistor currents take place similarly to as in the previous circuit. The only quantity that does not change is the emitter voltage because the emitter is at ac ground. U_C is the dc supply voltage that sets the Q point and U_{in} is the ac voltage that should be amplified. Except for external ac source, the dc biasing current enters the base circuit through the divider $\mathbf{R}_1\mathbf{R}_2$. Its value has to be higher than the maximum amplitude of U_{in}. In such a way, U_{in} will be amplified without clipping.

$\mathbf{C}_B$ and $\mathbf{C}$ are called *coupling capacitors*. *Coupling* is the method of circuit connection without an air gap. $\mathbf{C}_B$ couples the reference signal into the base, while $\mathbf{C}$ couples the amplified signal into the load. $\mathbf{C}_E$ is a *bypass capacitor* that shunts the emitter to the ground. Thanks to capacitor coupling (instead of direct coupling), only the alternating part of the signal passes through the circuit. For proper operation, the reactance of the capacitor should be at least ten times smaller than the load resistance of the external load $\mathbf{R}_L$ or the emitter resistor $\mathbf{R}_E$,

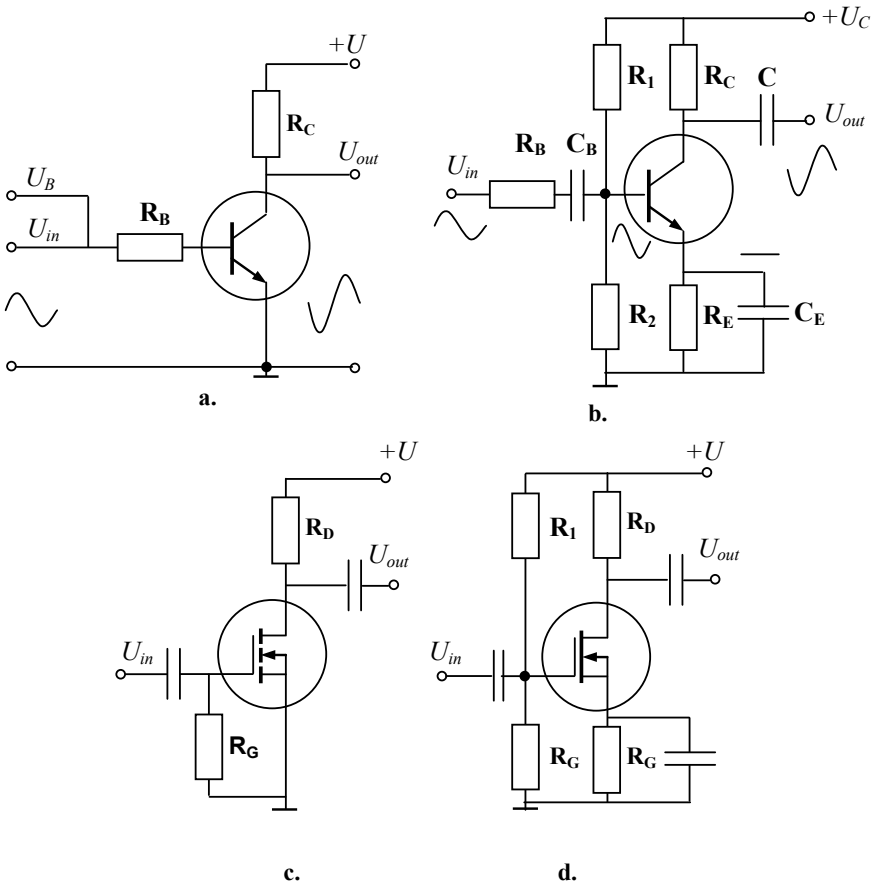


Fig. 2.11

$$C >> 1 / (2\pi f R_L),\ C_E >> 1 / (2\pi f R_E),$$

where f is the minimum reference signal frequency. This condition is equivalent to the high-frequency border

$$f << f_C,$$

where $f_C = 1 / (2\pi RC)$ is the critical frequency of the circuit.

Because an ac signal is coupled into the base, it produces ac variations in the base current. The ac base voltage is smaller than the reference voltage because there is some loss across the internal resistance of the ac voltage source. The ac variations are multiplied by the current gain β to produce the ac variations in the collector current. Since the current gain β is high while the voltage gain K_U is low and unpredictable, we call this circuit a current amplifier.

The ac collector current is approximately equal to the ac emitter current. Because the collector current flows through the collector resistor $\mathbf{R}_C$, the collector voltage has large ac ripples. On the positive half cycle of the input voltage, the total collector current increases, which means there is more voltage across the collector resistor and less total voltage at the collector. In other words, the amplified ac collector voltage is inverted, equivalent to being 180 degrees out of phase with the input voltage. The total collector voltage is the superposition of the dc ac voltages. Because the capacitor **C** is open to dc and shorted to ac, it will block the dc voltage but pass the ac voltage. For this reason, the final load voltage is a pure ac voltage.

The figures below illustrate the current amplifiers with *n*-channel enhancement-mode MOSFET (Fig. 2.11,c) and depletion-mode MOSFET (Fig. 2.11,d). Each circuit has the common source, and the input voltage applied to the gate changes the output voltage signal. MOSFETs have very high input impedance at low frequencies (hundreds teraohms) whereas the BJTs input impedance is tens of megohms. These impedances drop down with the frequency growing.

CE voltage amplifiers. Since β has large variations by virtue of the quiescent current, temperature change, and transistor replacement, the performance of the amplifier is beta-sensitive.

Historically, the first attempt to stabilize the *Q* point was to introduce an emitter resistor $\mathbf{R_E}$. There are two symmetrical supply sources in the circuit of Fig. 2.12,a – the positive source $+U_C$ and the negative source $-U_E$. It is known as a *balanced supply* with two equal *rails*, positive and negative. While $U_{in} = 0$, the output $U_{out} = 0$ too. Any quantity ΔU_{in} leads to the ΔU_E appearance, consequently

$$\Delta I_E = \Delta U_E / R_E,$$
$$\Delta I_C = \Delta I_E \beta / (\beta + 1),$$
$$\Delta U_{out} = \Delta U_C = -R_C \Delta I_C.$$

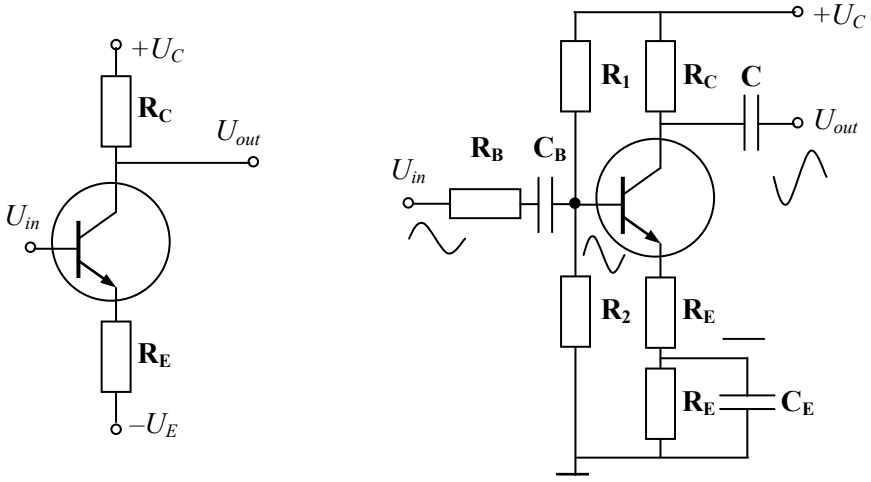


Fig. 2.12

Here, R_E includes the differential impedance of the emitter junction (approximately 25 mV / I_E) and the external resistor $\mathbf{R_E}$. Accordingly,

$$K_U = \Delta U_{out} / \Delta U_{in} = -\beta / (\beta + 1) \times R_C / R_E \approx R_C / R_E.$$

Therefore, the voltage gain does not depend on the transistor parameters while beta is high. In this case, we have a *voltage amplifier*.

A feedback *voltage divider* $\mathbf{R_E}$ shown in Fig. 2.12,b is usually called a *bleeder*. Such a feedback amplifier was invented in 1927 by H.S. Black. When the gain increases, so does the output quantity too. This output quantity flows through the emitter resistor, which diminishes an input quantity. In other words, the output influences the input. It is called an *emitter current feedback*, and refers to the output controlling of the input, at least partly. This *staircase divider* is a part of the loop that stabilizes the voltage gain. The voltage across the feedback resistor opposes the input voltage. This negative feedback reduces the voltage gain, but improves the gain stability and distortion. The resistor **R1** is another attempt to stabilize the *Q* point using a *negative collector feedback*. When the current gain increases, the collector current reduces the collector voltage, which means a lower base current and, therefore, a lower collector current.

Emitter followers. In the emitter followers, the load is connected to the emitter as shown in Fig. 2.13,a. Typically, the voltage gain of the emitter follower is ultra-stable and close to unity, also the current gain is much higher. Both of them are defined as

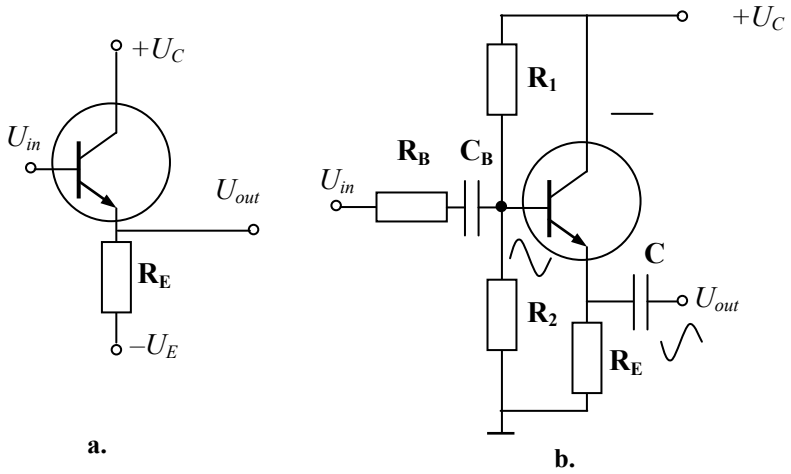


Fig. 2.13

$$K_U = \Delta U_{out} / \Delta U_{in} \approx 1,$$
$$K_I = \Delta I_E / \Delta I_B = (\beta + 1)\ (R_L + R_E) / R_E,$$

where R_L is the load resistance. The output impedance of this circuit is significantly lower than the input impedance. That is, the circuit is especially useful to decrease the output resistance of the electronic device. Another benefit of the circuit is that almost no distortion of the signal occurs. That is why the emitter follower is often used as an intermediate stage of a power amplifier for current amplification.

Fig. 2.13,b shows another design of the emitter follower. There, the base ac voltage produces an emitter ac current. Thanks to the limiting resistor $\mathbf{R_B}$ and the coupling capacitor $\mathbf{C_B}$, an ac voltage appears at the emitter. The biasing is arranged with the help of $\mathbf{R_1}$ and $\mathbf{R_2}$. Because of the output capacitor **C**, this voltage is coupled to the load. Since the emitter is no longer at ac ground, the ac voltage across the emitter is approximately equal to the input voltage at the base. The reason the circuit is called an emitter follower is that the output voltage follows the input voltage.

Two-stage ac amplifiers. To obtain higher voltage gain of an amplifier, one can connect two stages, as shown in Fig. 2.14,a. This is called a stage *cascading* and means the amplified voltage out of the first transistor is coupled into the base of the second transistor. The second transistor then amplifies the signal, so that the final signal is much higher than the input signal. The capacitor **C** insulates the collector of the first transistor from the base of the second transistor.

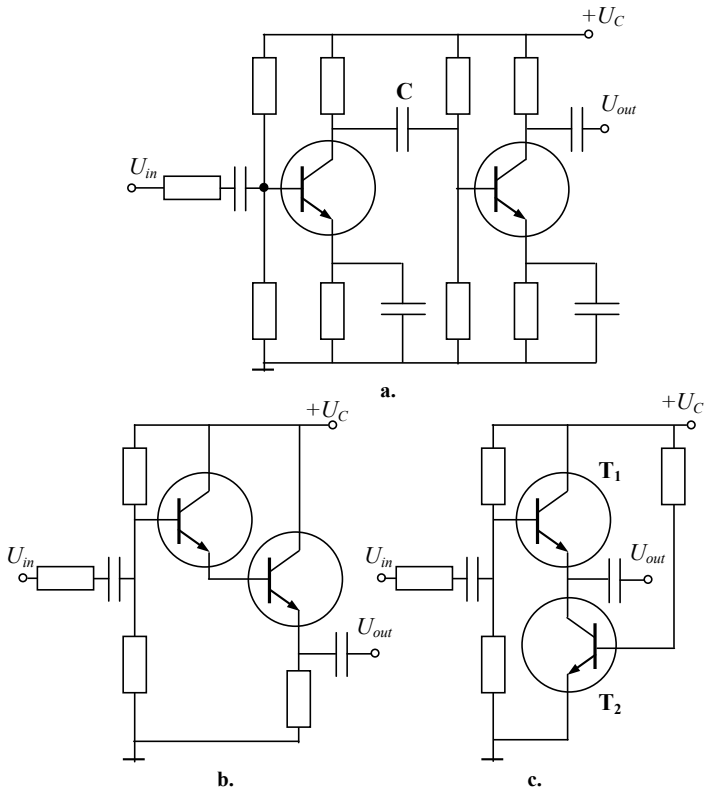


Fig. 2.14

After the signal value has been amplified, it can be used to control larger amounts of power. Large-signal amplifiers are more commonly called as *power amplifiers*. An expression of power amplification was given earlier as $K_U K_I$.

To raise the current gain and the input resistance, the emitter follower is built by cascading of two transistors. Fig. 2.14,b shows a method of emitter follower cascading where the current is amplified twice and $\beta = \beta_1\beta_2$.

Cascode amplifier. The circuit in Fig. 2.14,c is called a *cascode amplifier* that is an amplifier with the same dc current flowing through both devices. Here, the bottom transistor $\mathbf{T_2}$ having CE connection plays a role of an active load for the top CB-connected transistor $\mathbf{T_1}$, therefore, the input impedance of the amplifier is raised. The common-base resistive divider defines the dc mode of operation, whereas the coupling capacitor determines the ac mode. Here,

$$\alpha = \alpha_1\alpha_2.$$

As a result, the cascode amplifier has no advantages in the current and voltage amplification. The main idea of this circuit is the decrease of parasitic coupling between the input and the output because the constant voltage of the base $\mathbf{T_1}$ supplies $\mathbf{T_2}$. Accordingly, the collector of $\mathbf{T_2}$ is short-circuited and its amplification is near unity. The circuit is preferable in the resonant amplifiers, particularly in the high frequency receivers.

Summary. Following the classification of amplifiers, the class A ac amplifiers were discussed in this chapter.

In the CE current amplifiers, the load signal is out of phase with the input, and current clipping is low. Nevertheless, they are beta sensitive, their voltage amplification is unpredictable, and efficiency is lower than 50%.

The negative feedback reduces the voltage gain, but improves the gain stability and decreases the voltage distortion in the voltage amplifiers.

The voltage gain of the emitter follower is very stable and close to one, though the current gain is much higher. Low clipping is another benefit of this circuit.

Cascading helps to obtain higher current, voltage, and power gains of an amplifier or improves the signal coupling.

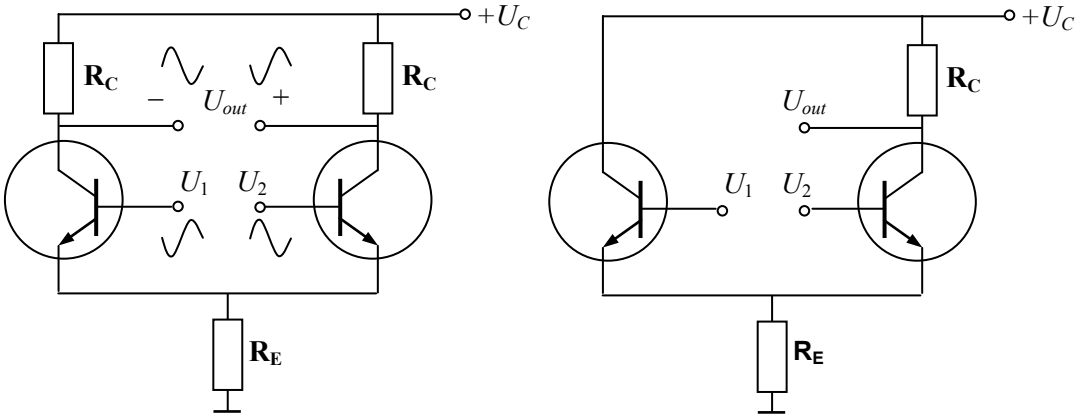
2.2.2 DC Amplifiers

The fundamental specifications of dc amplifiers are as follows:

- input/output signal range,
- offset and offset drift,
- single or balanced supply,
- input bias current,
- open loop gain,
- integral linearity,
- voltage and current noise.

To be a serious contender for the high performance applications, an amplifier should have most of them listed on the data sheet.

Differential amplifiers. A *differential amplifier*, or *diff amp* is the two-input device that amplifies the difference of both inputs. It serves as the typical input stage of many amplifiers. Fig. 2.15,a presents a general form of the diff amp that is termed as a *long-tailed pair* because $\mathbf{R_E}$ is called a *tail resistor*. The diff amp has two inputs – U_1 and U_2. Because there are no coupling or bypass capacitors, the input signals can have frequencies all the way down to zero, equivalent to dc, and the amplifier has a broad midband and high stability. The output signal is the voltage on the load connected between the collectors. Ideally, the circuit is symmetrical with identical transistors and collector resistors. The amplifier has the more linear transfer characteristic than the single bipolar transistor has.



a. **b.**

Fig. 2.15

The input difference

$$U_d = U_1 - U_2$$

is called a *differential signal*. The differential voltage gain is described by the ratio

$$K_d = \Delta U_{out} / \Delta U_d.$$

As an alternation, a *common-mode signal* is used that is a signal applied in the same phase to both inputs

$$U_c = (U_1 + U_2) / 2.$$

In the case of the common-mode input signal, the output voltage is zero, while the input voltages are equal. The *common-mode voltage gain* is

$$K_c = \Delta U_{out} / \Delta U_1 = \Delta U_{out} / \Delta U_2.$$

That is why any encumbrances and spikes of the input signals and supply voltage pulses compensate one another. On the other hand, when U_1 is greater than U_2, an output voltage with the polarity shown in Fig. 2.15,a appears. When U_1 is less than U_2, the output voltage has the opposite polarity. In any case, the output voltage is proportional to the difference of the input signals. The difference signal is amplified with a great gain.

The quality of a diff amp is evaluated by attenuation

$$K_a = K_d / K_c$$

that shows the ratio of the differential signal amplification to the common-mode one.

One may use this topology with the signal on one of the inputs, whereas another input remains grounded. For instance, the positive half-wave enters the base of the left transistor. Therefore, the emitter voltage and the current of the transistor are growing up. The voltage drop in the left $\mathbf{R}_C$ is raised and the phase shift of 180 degrees occurs between the input and output signals. This leads to the voltage rise in the joint collectors. As a result, the voltage drop and the current of the right transistor decrease, therefore the voltage drop in the right $\mathbf{R}_C$ is lowered also. The collector signal of the right transistor occurs in counter-phase to the left branch. Here, we refer to a *paraphase amplifier*.

Fig. 2.15,b illustrates the modified topology of the diff amp. Here, a growing of U_1 produces an increase in the output voltage. The U_1 input voltage is called a *non-inverting voltage* because the output voltage is in phase with U_1. On the other hand, the U_2 input voltage is called an *inverting input* because the output voltage is 180 degrees out of phase with U_1.

Two-stage dc amplifier. The capacitor between the stages shown before in Fig. 2.12 decreases the signal and shifts its phase. It is the main reason of the frequency limiting in the ac amplifiers. Moreover, the capacitor needs an additional place in the amplifier design. As there are many applications without ac signals, the dc amplifier manages without the capacitor.

A direct-coupled *two-stage dc amplifier* is shown in Fig. 2.16. As has been calculated earlier, when there is no input voltage U_{in}, the preferable output voltage should be equal to half of the supply voltage

$$U_{out} = U_C / 2.$$

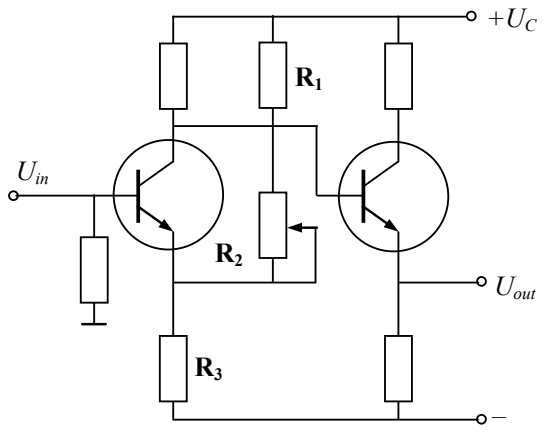


Fig. 2.16

As a result, one can obtain the maximum power and amplitude of the signal.

Concerning the dc amplifiers, this problem is solved by applying the balanced supply with equal rails. Because of the voltage divider $\mathbf{R_1}$, $\mathbf{R_2}$, $\mathbf{R_3}$, the emitter potential of the left transistor is supported slightly negative regarding to the ground. Thus, the left transistor is opened. The right transistor shifts the output voltage to the zero and amplifies the signal simultaneously. Therefore, because of the *split supply* (equal positive and negative voltages) the quiescent output is ideally zero when the input voltage is zero.

Integrated circuits. The first *integrated circuit* (*IC*) was invented by J. Kilby from Texas Instruments in 1958. Kilby's work was paralleled by R. Noyce who also developed an IC, and by J. Hoerni who developed the planar process of IC manufacturing (both of Fairchild Semiconductor, 1959). Analog Devices founded in 1965 became the first company for IC production. The basic *bipolar process* was primarily worked out there to yield a good transistor IC. Then, the *complementary-metal-oxide-semiconductor* (*CMOS*) devices began to appear. The CMOS offered the potential of much higher packing density and low power than bipolar-based devices, and soon became the IC process of choice. In the early 1970s, another process technology was developed for linear circuits requiring stable precision resistors and an ability to perform calibrations. This was *thin film resistor technology*. In summary, the bipolar processes, coupled with the thin film resistors and the laser wafer trim technology led to the proliferation of IC during the 1970s…1990s. In the 1980s, the *complementary bipolar process* (CB) was introduced. The CMOS and bipolar processes were combined to achieve both the low power high-density logic and the high accuracy low noise analog circuitry on a single *chip*.

The monolithic IC usually has power dissipations under a watt thanks to the use of the FET transistors. For higher power applications, the thin-film, thick-film, and hybrid ICs may be used. Typically, an IC fabricated on the CMOS or complementary bipolar processes has fixed input ranges that are usually at least several hundred millivolts from either rail.

Small-scale integration (SSI) of IC refers to fewer than 10 integrated components, medium-scale integration (MSI) to between 10 and 100 components, and large-scale integration (LSI) to more than 100 integrated components.

Operational amplifiers. An *operational amplifier* or *op amp* is a high-performance, directly coupled dc amplifying circuit containing a set of transistors. The main features of op amp are as follows:

- high gain,
- high input resistance,
- low output resistance,
- controlled bandwidth extended to dc.

An op amp completes all circuit functions on a single chip, such as amplifiers, voltage regulators, and computer circuits.

The first op amp on *npn* transistors were proposed by R. Widlar, and Fairchild Semiconductor produced ICs μA702 and μA709 from 1964. Some time later, the complementary bipolar technology was developed and op amps on *pnp* transistors appeared. The next step was the BiFET technology on the bipolar FET devices with high input impedance and low input currents and noise. Then, the CMOS production started with the lowest input currents, highest input impedance, and minimum losses. Many linear devices are built on the BiMOS (Bipolar Metal-Oxide Semiconductor) technology now and the fastest op amps use the XFCB (eXtra Fast Complementary Bipolar) technology of Analog Devices.

An op amp can have a single input and single output, a differential input and single output, or a differential input and differential output. Fig. 2.17,a shows the typical topology of the op amp.

The *input signals range* determines the required output *voltage swing* of the op amp. There are many single-supply amplifiers, which inputs range from zero to the positive supply voltage. However, the input range can be set so that the signal only goes to within a few hundred millivolts of each rail. Often, there is a demand for the op amps with an input voltage that includes both supply rails, i.e., *rail-to-rail operation*. *Rail-to-rail op amps* are very popular in portable systems with low-voltage supply (3 V and less) where the usual op amps cannot provide a large output swing. Eventually, in many single-supply applications it is required that the input common-mode voltage range extends to one of the supply rails (usually negative rail or ground).

The input stage is a diff amp, followed by more stages of gain and an output stage. These stages must provide the required gain and offset voltage to match the signal to a dc-coupled application.

Fig. 2.17,b is a schematic diagram of the op amp. Its input stage is a diff amp using the *pnp* transistors $\mathbf{VT}_1$ and $\mathbf{VT}_2$. $\mathbf{VT}_6$ forms an active load that replaces the tail resistor. $\mathbf{R}_2$ and $\mathbf{VD}_2$ control the bias on $\mathbf{VT}_6$, which produces the tail current of the diff amp. Instead of using an ordinary resistor, the active load $\mathbf{VT}_3$ is used. Because of this, the voltage gain of the diff amp is high. The amplified signal from the diff amp drives the base of $\mathbf{VT}_4$, which serves as an emitter follower. This stage avoids the loading down of the diff amp. The signal out of $\mathbf{VT}_4$ goes to $\mathbf{VT}_5$. Diodes $\mathbf{VD}_4$ and $\mathbf{VD}_5$ provide the biasing of the final stage. $\mathbf{VT}_7$ is an active load for $\mathbf{VT}_5$. Therefore, $\mathbf{VT}_5$ and $\mathbf{VT}_7$ are like a CE stage with a very high voltage gain. The amplified signal of the CE stage goes to the final stage, which is a class B emitter follower built on $\mathbf{VT}_8$ and $\mathbf{VT}_9$. Thanks to the balanced supply, the output is zero when the input voltage is zero. Any deviation from zero is called an *output-offset voltage* of the same sign. Ideally, U_{out} can be as positive as $+U_C$ and as negative as $-U_E$ before the clipping occurs.

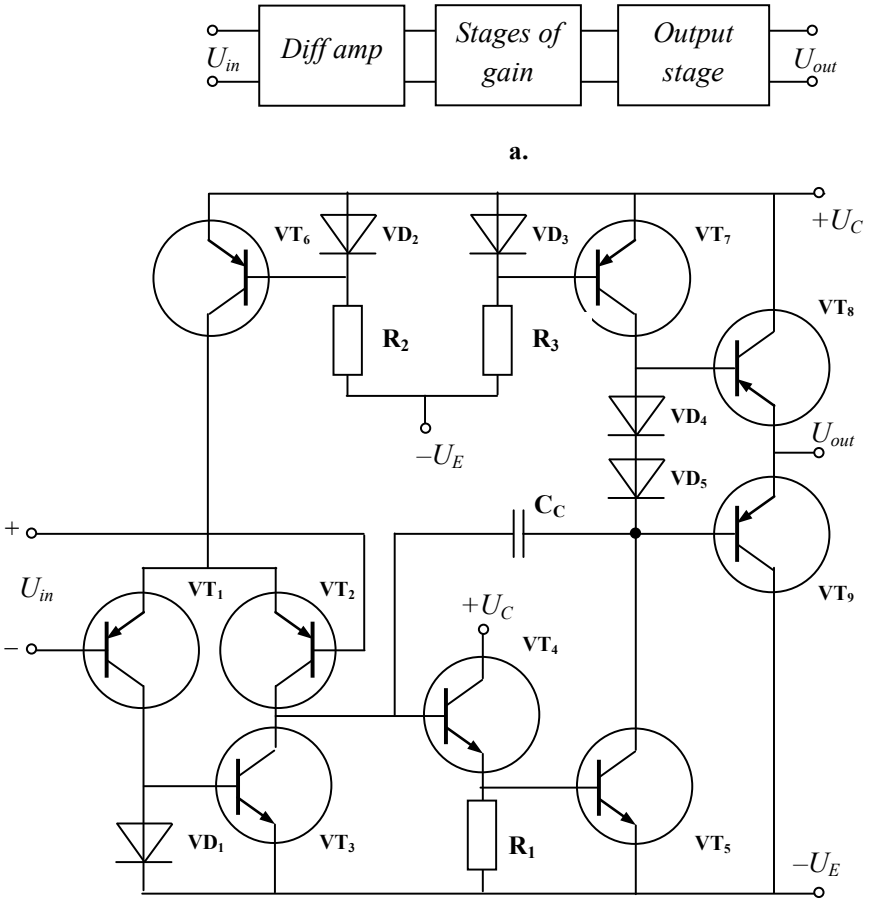


Fig. 2.17

Summary. The differential amplifier is the most popular type of amplifiers in microelectronics where the full identity of arms is provided without problems. Because there are no coupling or bypass capacitors, the input signals can have a wide range of frequencies and the amplifier has a broad midband and high stability. Other benefits of diff amp are high amplification and low clipping.

Diff amps are applied in op amps. The main features of op amp are as follows: high gain, high input resistance, low output resistance, and bandwidth extending to dc. The frequency range of op amps spreads now as far as hundreds of megahertz. It completes the circuit functions on a single IC chip, such as amplifiers, voltage regulators, and computer circuits.

2.2.3 IC Op Amps

As a rule, an op amp is a modular, multistage device with differential input and entire assembly composed on a small silicon substrate packaged as an IC.

Composition and symbols. The earliest IC op amp output stages were *npn* emitter followers with *npn* active loads or resistive pull-downs. Using a FET rather than a resistor can speed things up, but this adds complexity. With modern complementary bipolar processes, well-matched high speed *pnp* and *npn* transistors are available. The complementary output stage of the emitter follower has many advantages, and the most outstanding one is the low output impedance. The output stages constructed of CMOS FETs can provide nearly true rail-to-rail performance. Most of the modern op amps have the class B output stages of some sort.

Fig. 2.18 displays schematic symbols of an op amp. In the first of them, K_U is the voltage gain. The inverting input is U_1, and the non-inverting one is U_2. U_1, and U_2 are the node voltages measured with respect to the ground. The differential input is the difference of two node voltages, and the common-mode input is their half-sum.

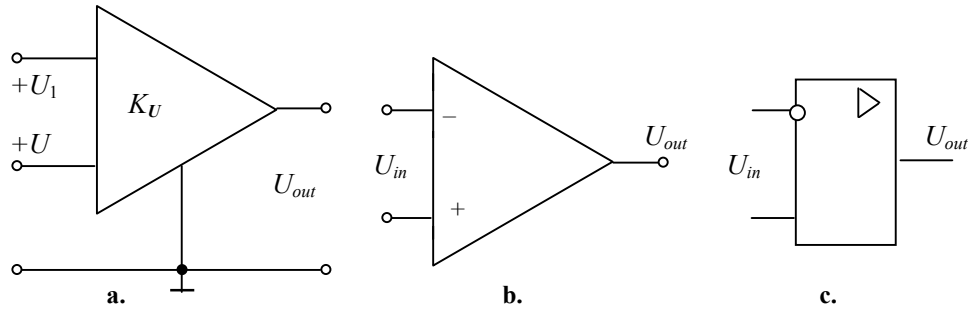


Fig. 2.18

Since the quiescent output of an op amp is ideally zero, the ac output voltage (MPP value) can swing positively or negatively. In particular, for high load resistances, the output voltage can swing almost to the supply voltages. For instance, if $U_C = +15$ V and $U_E = -15$ V, the MPP value with a load resistance of 10 kΩ or more is ideally 30 V. In reality, the output cannot swing all the way to the value of the supply voltages because there are small voltage drops in the final stages of the op amps. In other schematic symbols of an op amp, the plus sign corresponds to the non-inverting inputs and the minus or the rounded inputs are the inverting ones.

The frequency range of an op amp depends on two factors, the gain-bandwidth product (voltage gain multiplied by midband) for small signals, and the *slew rate* for a large signal. A slew rate of an amplifier is the value of the maximum rate of change of the output voltage per time. It is usually less than 10 V/μs. The slew rate limitation makes op amps unsuitable for applications, which require fast-rising pulses. Therefore, the op amps should not be used as the signal sources of the digital circuitry feed.

Non-inverting feedback voltage amplifier. As discussed earlier, one of the most valuable ideas of electronics is the negative feedback. In an amplifier with a negative voltage feedback, the output is sampled and part of it is returned to the input. The advantages of the negative feedback are as follows: more stable gain, less distortion, and higher frequency response.

In Fig. 2.19, a non-inverting op amp is presented. Here, the output voltage is sampled by the voltage divider and fed back to the inverting input of the op amp.

The differential input of the op amp is an *error voltage*, defined as

$$U_{err} = U_{in} - U_1.$$

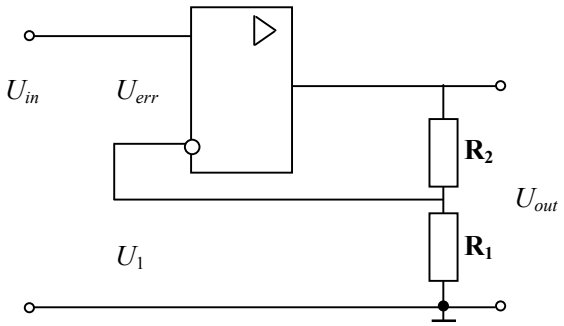


Fig. 2.19

The op amp amplifies this error voltage as

$$U_{out} = K_d U_{err},$$

where the amplification factor K_d is the differential voltage gain of the open-loop op amp. Let K_1 be the *feedback fraction* or the fraction of output voltage fed back to the input, that is

$$K_1 = R_1 / (R_1 + R_2) = U_1 / U_{out}.$$

Then, the output voltage is

$$U_{out} = K_d (U_{in} - K_1 U_{out}).$$

By rearranging,

$$K = U_{out} / U_{in} = K_d / (1 + K_d K_1).$$

This famous formula defines exactly what the effect of negative feedback is on the amplifier. Here one can see that the voltage gain K of the closed-loop amplifier with negative feedback is less than the differential voltage gain K_d of the open-loop op amp. The fraction K_1 is the key to how much effect the negative feedback has. When K_1 is very small, the negative feedback is small and the voltage gain approaches K_d. However, when K_1 is large, the negative feedback is large and the voltage gain is much smaller than K_d.

The product K_dK_1 is called a *loop gain* because it represents the voltage gain going all the way around the circuit, from the input to the output and back to the input. For the non-inverting voltage feedback to be effective, a designer must deliberately make the loop gain much greater than one. Once this condition is satisfied,

$$K = U_{out} / U_{in} \approx 1 / K_1 = (R_1 + R_2) / R_1.$$

The equation states that the voltage gain K of the closed-loop system is equal to the reciprocal of K_1, the feedback fraction, and no longer depends on the value of K_d. Since K_d does not appear in this equation, it can change with temperature or op amp replacement without affecting the voltage gain. The IC op amps approach such requirements and have extremely high differential voltage gain K_d. When the feedback path is opened, the open-loop voltage gain is approximately equal to the differential voltage gain. When $U_{in} = 0$, $U_{out} = KU_0$, where U_0 is called a *zero offset*. In the case of $R_2 = 0$, $U_{out} \approx U_{in}$. This circuit is called a *buffer*. A buffer does not amplify the voltage but it can be of high power gain and play the role of an impedance converter.

Some circuits require the positive feedbacks. The voltage gain K of the closed-loop amplifier with the positive feedback is more than the differential voltage gain of the open-loop op amp

$$K = U_{out} / U_{in} = K_d / (1 - K_dK_1).$$

The maximum value of K_dK_1 is to be less than one. In an opposite case, the output signal will grow and the system may become unstable. Typically, this effect is used in pulse generators – pulsers.

Inverting feedback voltage amplifier. An inverting amplifier given in Fig. 2.20,a also uses the negative feedback to stabilize the working conditions in the same way. Here, the output voltage drives the feedback resistor $\mathbf{R_2}$, which is connected to the inverting input. The voltage gain is given by

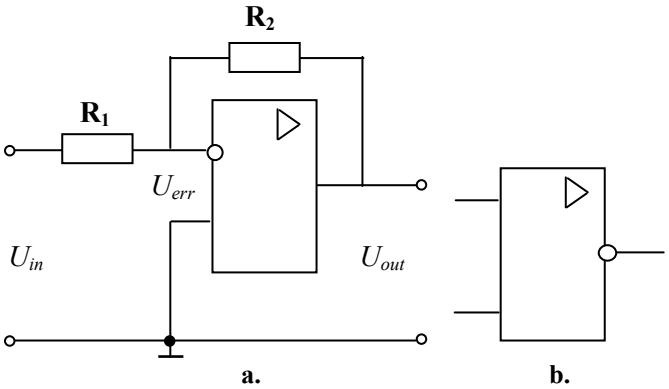


Fig. 2.20

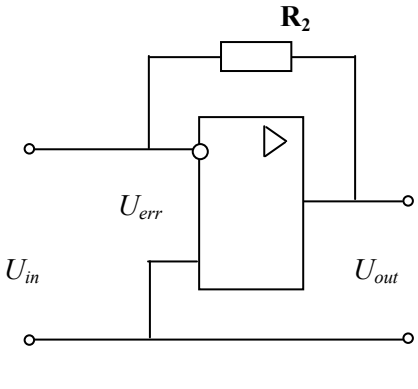


Fig. 2.21

$$K = U_{out} / U_{in} = -R_2 / R_1.$$

The circuit characteristic is linear. By virtue of the negative voltage gain, the amplifier inverts the input signal. In the case of $R_1 = R_2$ the circuit is called an *inverter* because the output signal is equal to the inverted input signal. Its circuit symbol is shown in Fig. 2.20,b.

Feedback current amplifier. Fig. 2.21 illustrates an amplifier with the inverting voltage feedback. Here, the output voltage drives the feedback resistor $\mathbf{R}_2$, which is connected to the inverting input, and the voltage gain is independent on the error. Instead of acting like a voltage amplifier, an amplifier with an inverting voltage feedback acts like an ideal *current-to-voltage converter*, a device with a constant ratio of output voltage to the input current. As K_d is much greater than unit,

$$U_{out} = K_d U_{err}, K = U_{out} / I_{in} = K_d R_2 / (K_d + 1) \approx R_2.$$

The ratio U_{out} / I_{in} is referred to as a *transresistance*. Besides stabilizing of transresistance, the inverting feedback has the same benefits as the non-inverting voltage feedback that is decreasing distortion and output offset.

When $R_2 = 0$, the current amplifier is a *voltage repeater* because the voltage gain is equal to unit.

Feedback differential amplifier. Fig. 2.22 shows an op amp connected as a diff amp with the balanced supply. It amplifies U_{in} that is the difference between U_1 and U_2. The output voltage is given by

$$U_{out} = KU_{in},$$

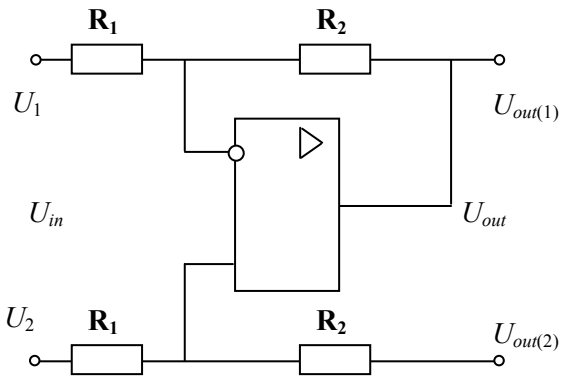


Fig. 2.22

where $K = R_2 / R_1$. When U_2 is zero, the circuit becomes an inverting amplifier with $U_{out(1)} = KU_1$. When U_1 is zero, the circuit becomes a non-inverting amplifier with

$$K = R_2 / R_1 + 1.$$

When both inputs are presented,

$$U_{out} = U_{out(1)} - U_{out(2)}.$$

Summary. To provide high efficiency and operational speed, most of the contemporary op amps have the class B output stages. The quiescent output of an op amp is zero and the MPP value can swing positively and negatively almost to the supply voltages. Op amps have a broad frequency range and limited slew rate therefore they are very popular in analog electronics and less preferable in fast-speed digital circuits.

To obtain a stable gain, low distortion, and high frequency response, the inverting or non-inverting negative feedbacks are used in the op amp circuits. The higher is the negative feedback voltage the lower is voltage gain and the higher the frequency response. The buffers, the inverters, the voltage repeaters, and the diff amps are the useful representatives of the op amps with a negative feedback.

2.3 Supplies and References

2.3.1 Sources

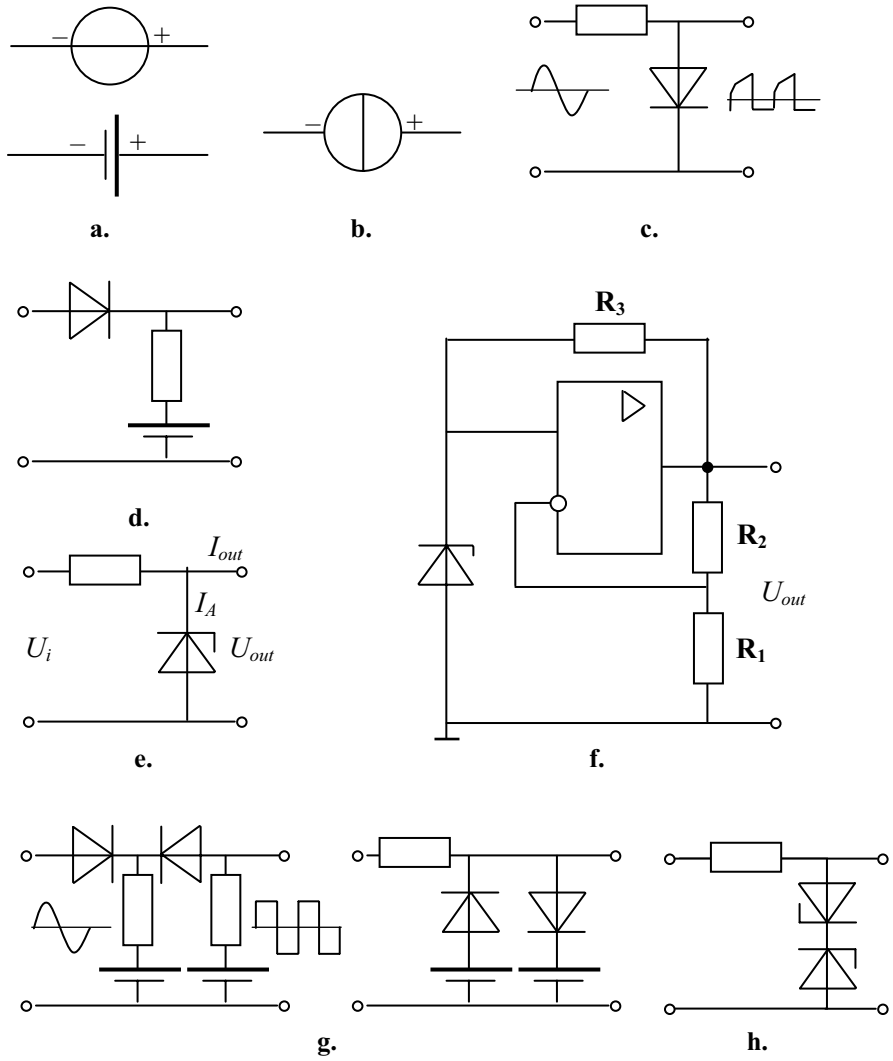


Fig. 2.23

Conventionally, energy approaches electrical end electronic systems from the power generators of different types: hydro, wind, and heat generators, atomic stations, etc. Their energy is transmitted to a consumer where the power transformation is executed. The circuits that supply electronic systems are called power supplies. They are distinguished as voltage sources, current sources, and filters. The output of a *voltage source* is the required voltage that is weakly dependent upon the load current (2.23,a). The *current source* supplies the load by the required current weakly dependent upon the load voltage (2.23,b).

Clippers and limiters. Most voltage sources are built on rectifier diodes and thyristors with different limiting and filtering circuits on the output. One-side *clippers* cut up or down the rectified voltage level, whereas double-side *limiters* provide the required voltage swing. To fix the signal level, *clampers* are also used.

A simple diode-based clipper is shown in Fig. 2.23,c. Here, a current driven forward biased diode produces a voltage. Unfortunately, while the junction drop is somewhat decoupled from the supply, it has numerous deficiencies as a clipper. These include sensitivity to loading and a rather inflexible output voltage. Therefore, such clipper is only available in some hundreds of millivolts jumps. Another limitation is that the load current is always less than the input current. More successful clipper with the additional battery is shown in Fig. 2.23,d.

Another simple clipper circuit shown in Fig. 2.23,e consists of the Zener diode and the *ballast* for the current clipping. Here, the output voltage is equal to the Zener diode voltage drop, which slightly fluctuates. The ballast resistance is calculated as follows:

$$R = (U_{in} - U_{out}) / (I_A - I_{out}),$$

where I_A is the rated Zener current and U_{out} is the Zener voltage. The ratio of the instant voltage quantities

$$K = \Delta U_{out} / \Delta U_{in}$$

is referred to as an *output stability*, which is commonly less than 100 in this circuit.

The voltage source built on an op amp is shown in Fig. 2.23,f. Here, the input signal U_{in} comes from the voltage source of Fig. 2.23,e built on the Zener diode and resistor $\mathbf{R}_3$. The output voltage is calculated as follows:

$$U_{out} = U_{in}(1 - R_2 / R_1)$$

It does not depend on the load and supply voltage of op amp. More powerful transistor output stages are often added to the voltage sources of this kind.

Sometimes, asymmetrical clipping is selected by setting the limit voltages to different values (e.g. +5 V and -2 V).

On the contrary, the voltage limiter based on the two cross-coupled Zener diodes realizes an appreciably higher output – 5 to 8 V range per one Zener pair (Fig. 2.23,h). On the positive alternation, the upper diode conducts and the lower diode breaks down. On the negative half cycle, the action is reversed. The lower diode conducts, and the upper diode breaks down. Therefore, the output is clipped as shown. The clipping level is equal to the Zener voltage. In this way, the output is almost a square wave. The larger is the input sine wave, the better the output square wave. This shunt small-scale circuit is taking only a few milliamps.

Current sources. Fig. 2.24,a shows the current source built on the BJT. Let $\mathbf{R}_L$ be the load resistor connected to the collector. While U_{in} is constant, the emitter voltage is calculated as follows:

$$U_E = U_B - U_{BE},$$

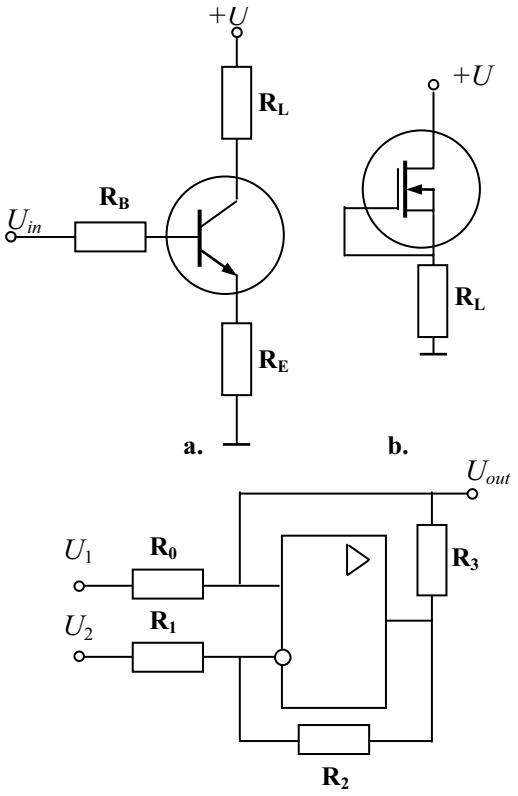


Fig. 2.24

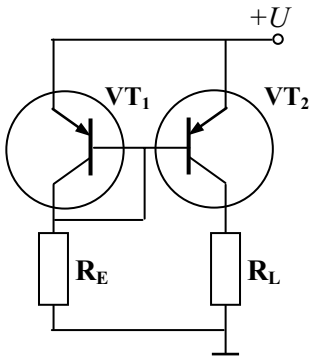


Fig. 2.25

where U_{BE} is the voltage drop of the emitter diode of the transistor. The currents are as follows:

$$I_E = U_E / R_E,$$
$$I_C = \beta I_E / (\beta + 1).$$

Since $\beta \to \infty$, the load current depends on U_B and R_E only and does not depend on the load resistance R any more, that is

$$I_C = I_E = \text{const}.$$

It is true in the case of $R_L < U_C / I_C - R_E$.

The MOSFET connected as given in Fig. 2.24,b is the current source also, because the load current of the resistor $\mathbf{R_L}$ does not depend on U_{DS} in the saturation mode.

The simple current source in Fig. 2.24,c consists of the op amp with the pair of the feedback loops. In the symmetrical circuit ($R_0R_2 = R_1R_3$) the load current is calculated as follows:

$$I_{out} = (U_1 - U_2) / R_0.$$

Current reflector. The circuit shown in Fig. 2.25 is called a *current reflector* in the case of the full identity of $\mathbf{T}_1$ and $\mathbf{T}_2$ parameters. $\mathbf{T}_1$ is connected as a diode. Thanks to the joined bases, the voltages U_{BE} are equal, therefore

$$I_{C1} = I_{C2} = \beta / (\beta + 2)\times(U_E - U_{BE}) / R_E.$$

Since $\beta \to \infty$, the load current depends on U_E and R_E only and no longer depends on the load resistance R_L, that is

$$I_C \approx (U_E - U_{BE}) / R_E.$$

Summary. A power supplier has to meet the requirements of the energy consumer, which needs the determined power, voltage, and current values and shape. Voltage sources supply fully controlled voltage, whereas the current may be unpredictable. Current sources generate adjustable current flow, whereas the voltage may change during the supply process. In practice, there is neither the pure voltage nor the exclusively current sources, but one of the features is predominant.

2.3.2 Filters

Voltage produced by most of the electronic devices is not pure dc or pure ac signal. Often, the supplier output is a *pulsating dc* voltage with ripple or ac signal with noise. For instance, the output of a SCR has a dc value and ac ripple value. The first idea is to get an almost perfect direct voltage, similar to what is obtained from a battery. Another idea is to delete noise and undesirable signals and to pass only necessary ac signals. The circuits used to remove unnecessary variations of rectified dc and amplified ac signals are called *filters*.

Terms. Filters are built on reactive components – inductors and capacitors the impedance of which depends on the frequency. Reluctance grows with the frequency, thus, a series-connected inductor has a significant resistance for the high-frequency components of a signal, whereas the parallel-connected inductor may extend them. On the contrary, capacity reactance decreases with the frequency growing, thus, a parallel-connected capacitor brings the high-frequency components of a signal down, whereas the series-connected capacitor raises them.

There are many filter designs, such as low-pass filters, high-pass filters, lead-lag filters, notch filters, Butterworth, Chebyshev, Bessel, and others. Depending upon the passive and active components, filters are classified as *passive filters* and *active filters*. The first are built on resistors, capacitors, and inductors, whereas the last include op amps and capacitors.

Passive low-pass filters. A *low-pass filter* reduces high-frequency particles of a signal and passes its low-frequency part.

Fig. 2.26,a shows a simple **RC** low-pass filter, and Fig. 2.26,b shows a simple **LC** low-pass filter. Fig. 2.26,c shows the frequency response of the filters. If the filter input is the diode rectifier, the output voltage waveform is shown in Fig. 2.26,d. The period t_1 represents diode conduction, which charges the filter capacitor to the peak voltage U_{max}. The period t_2 is the interval required for the capacitor discharging through the load. The condition of successful filtering may be written as follows:

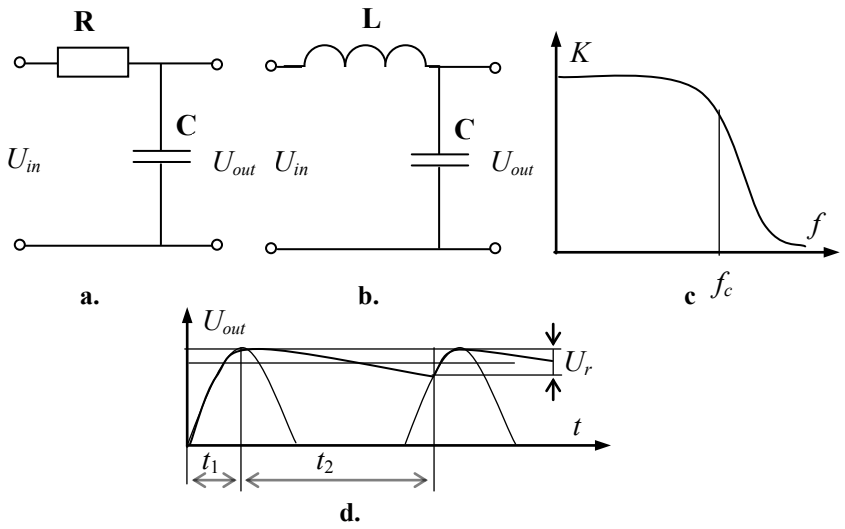


Fig. 2.26

$$T = RC >> t_1 + t_2,\ T = \sqrt{(LC)} >> t_1 + t_2,$$

where T is called a filter *time constant*. The following formula expresses the ripple (peak-to-peak output voltage) in terms of easily measured circuit values:

$$U_r = I_{out} / (fC)$$

where I_{out} is the average output current, and f is a *ripple frequency*.

Both filters are closed for high-frequency signals. For the low-frequency signals, the reactance of **L** is low. In this way, the ripple can be reduced to extremely low levels. Thus, the voltage that drops across the inductors in much smaller because only the winding resistance is involved. Simultaneously for the low-frequency signals, the reactance of **C** is high but the high-frequency signals follow across the **C**. The cutoff frequency of the low-pass filters may be calculated by the formulas:

$$f_C = 1 / (2\pi RC),\ f_C = 1 / (2\pi\sqrt{(LC)}).$$

For instance, if R = 1 kΩ and C = 1 μF, then T = 1 ms and f_c = 160 Hz. If L = 1 mH and C = 1 μF, then T = 32 μs and f_c = 5 kHz.

The circuits in Fig. 2.26 are called *single-pole filters*. Fig. 2.27,a presents a multi-stage **RC** filter. By deliberate design, the filter resistor is much greater (at least 10 times) than X_C at the ripple frequency. This means that each section attenuates the ripple by a factor at least ten times. Therefore, the ripple is dropped across the series resistors instead of across the load. The main disadvantage of the **RC** filter is the loss of voltage across each resistor. This means that the **RC** filter is suitable only for light loads.

When the load current is large, the **LC** filters of Fig. 2.27,b,c are an improvement over **RC** filters. Again, the idea is to drop the ripple across the series components; in this case, by the *filter chokes*. This idea is accomplished by making X_L much greater than X_C at the ripple frequency. Often, the **LC** filters become obsolete because of the size and cost of inductors. Nevertheless, in power circuits, they function as the protective devices for the load under the shorts.

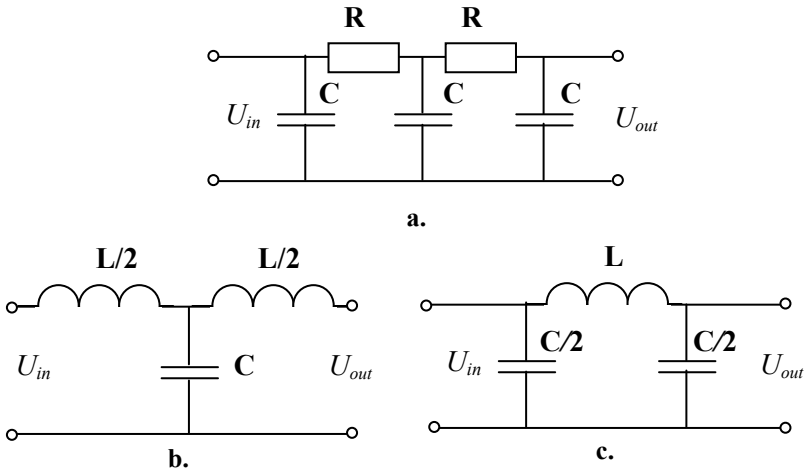


Fig. 2.27

Passive high-pass filters. Fig. 2.28 illustrates *high-pass filters* and their frequency response. The high-pass filter is open for high frequencies and attenuates the low-frequency signals. High frequencies pass through the capacitors but the low-frequency signals are attenuated by the capacitors. On the other hand, the low-frequency signals pass through the inductors, whereas the high-frequency signals cannot pass over the coils. The cutoff frequency of the high-pass filters may be calculated by the same formulas as for the low-pass filters.

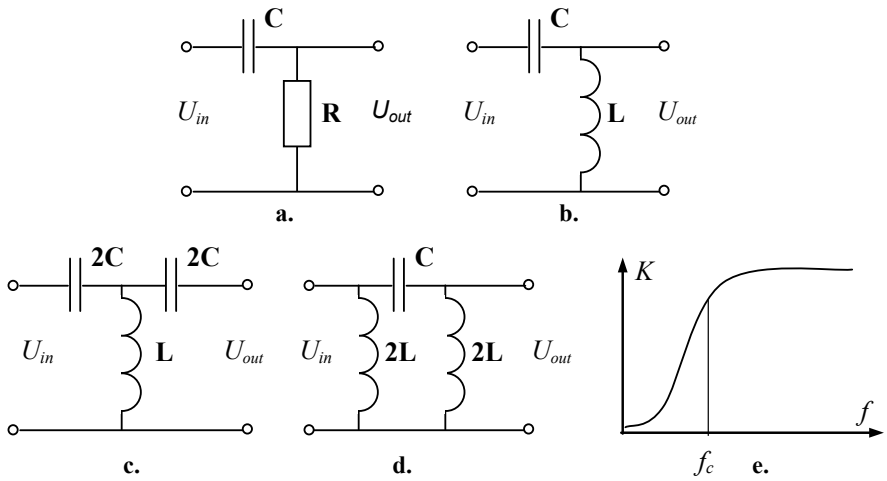


Fig. 2.28

Passive band-pass filter. Fig. 2.29 shows a *band-pass filter*, also referred to as *lead-lag filter*, and its frequency response. It is built by means of tank circuits. At very low frequencies, the series capacitor looks open to the input signal, and there is no output signal. At very high frequencies, the shunt capacitor looks short circuited, and there is no output also. In between these extremes, the output voltage reaches a maximum value at the resonant frequency

$$f_r = 1 / (2\pi\text{–}(L_1C_1)) \text{ or } f_r = 1 / (2\pi\sqrt{(L_2C_2)}).$$

For instance, if $L_1 = L_2 = 1$ mH and $C_1 = C_2 = 1$ μF, then $T_1 = T_2 = 32$ μs and $f_r = 5$ kHz.
Filter *selectivity* Q is given by

$$Q = f_r / (f_2 - f_1),$$

where f_2 and f_1 are the cutoff frequencies, which restrict the midband

$$f_2 - f_1 = R / (2\pi L_1) = 1 / (2\pi C_2 R).$$
$$(f_2 - f_1) / (f_2 \times f_1) = 2\pi L_2 / R = 2\pi C_1 R,$$

where R is the load resistance. In the case of the infinite load resistance ($R \to \infty$),

$$C_1 = (f_2 - f_1)^2 / ((f_1 f_2)^2 4\pi^2 L_2),$$
$$C_2 = 1 / (4\pi^2 L_1 (f_2 - f_1)^2).$$

For instance, if $L_1 = L_2 = 1$ mH, $f_1 = 3$ kHz, $f_2 = 7$ kHz, then $C_1 = 0{,}92$ μF and $C_2 = 1{,}6$ μF.

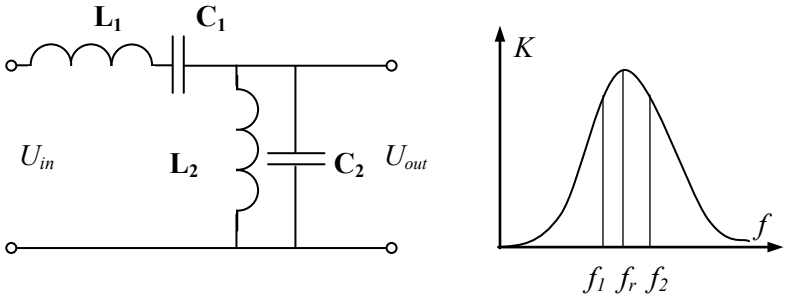


Fig. 2.29

Passive band-stop filter. A *band-stop filter*, also known as a *notch filter* is presented in Fig. 2.30,a. It is a circuit with almost zero output at the particular frequency and passing the signals, the frequencies of which are lower or higher than the cutoff frequencies (Fig. 2.30,b). The resonant frequency of the filter and selectivity Q are the same as for the band-pass filter. The cutoff frequencies are given by

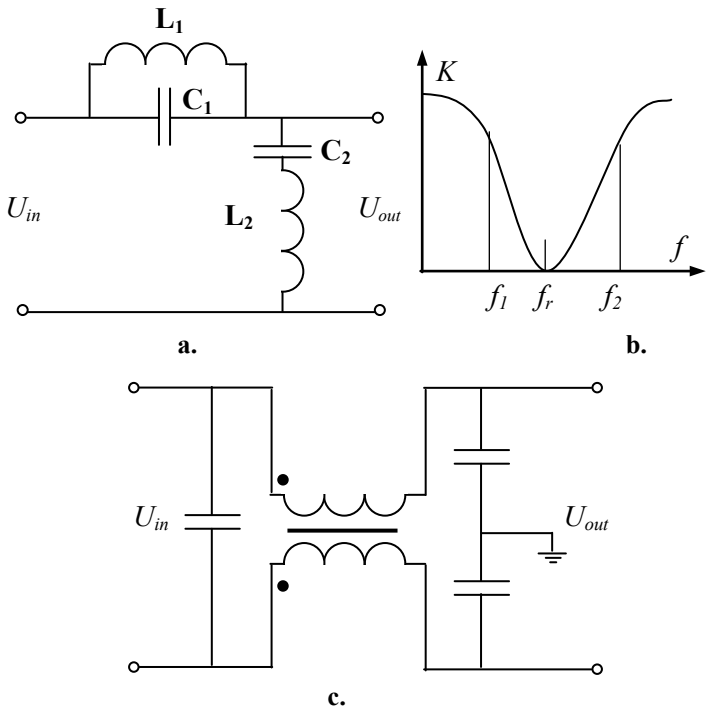


Fig. 2.30

$$f_2 - f_1 = R / (2\pi L_2) = 1 / (2\pi C_1 R).$$
$$(f_2 - f_1) / (f_2 \times f_1) = 2\pi L_1 / R = 2\pi C_2 R$$

where R is a load resistance. In the case of the infinite load resistance ($R \to \infty$),

$$C_1 = 1 / (4\pi^2 L_2 (f_2 - f_1)^2).$$
$$C_2 = (f_2 - f_1)^2 / ((f_1 f_2)^2 4\pi^2 L_1),$$

For instance, if $L_1 = L_2 = 1$ mH, $f_1 = 3$ kHz, $f_2 = 7$ kHz, then $C_1 = 1{,}6$ μF and $C_2 = 0{,}92$ μF.

A more complex band-stop filter shown in 2.30,c is used as a noise filter in low-power suppliers.

Active filters. Active filters use only resistors and capacitors together with op amps and are considerably easier to design than **LC** filters.

Active low-pass filters built on op amp are presented in Fig. 2.31. The bypass circuit on the input side passes all frequencies from zero to the cutoff frequency

$$f_c = 1 / (2\pi RC).$$

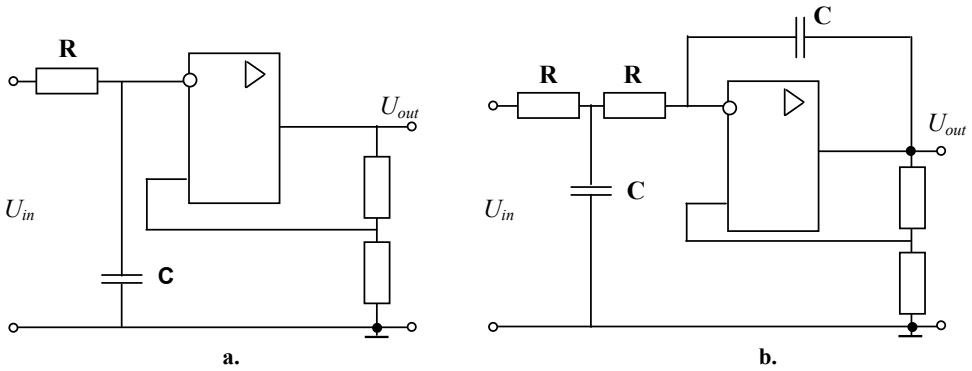


Fig. 2.31

As Fig. 2.32 displays, one can change a low-pass filter into a high-pass filter by using the coupling circuits rather than the bypass networks. The circuits like these pass the high frequencies but block the low frequencies. The cutoff frequency is still given by the same equation.

Fig. 2.33 shows a band-pass filter and Fig. 2.34 shows a notch filter. The lead-lag circuit of the notch filter is the left side of an input bridge, and the voltage divider is its right side. The *notch frequency* of the filter may be calculated as

$$f_r = 1 / (2\pi RC).$$

The gain of the amplifier determines selectivity Q of the circuit so the higher gain causes the narrower bandwidth.

Summary. Filters improve the frequency response of circuits. They are the necessary part of any electronic systems. Passive filters are often more simple and effective, but they need enough space and are the energy-consuming devices. For this reason, passive filters are preferable in power suppliers of industrial applications and are placed after the rectifiers in electronic equipment. Active filters are the low-power circuits that correct signals and couple stages by passing the signals through.

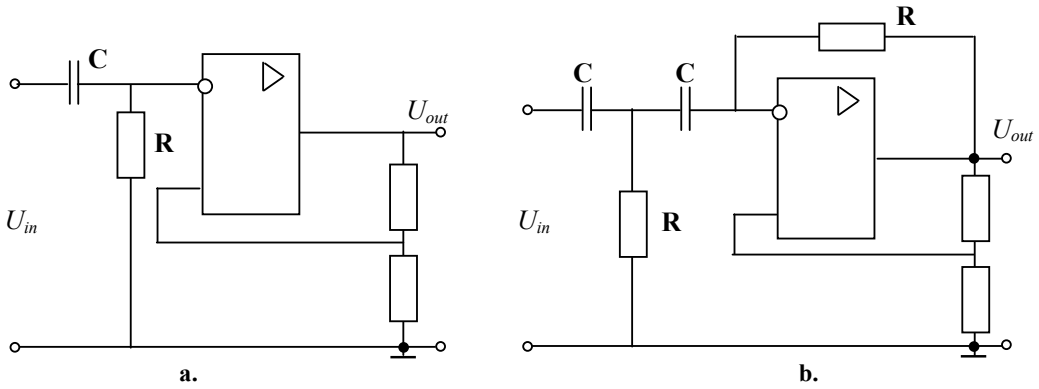


Fig. 2.32

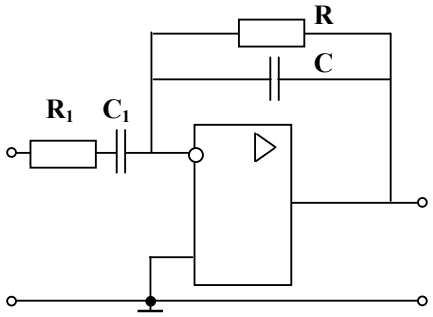


Fig. 2.33

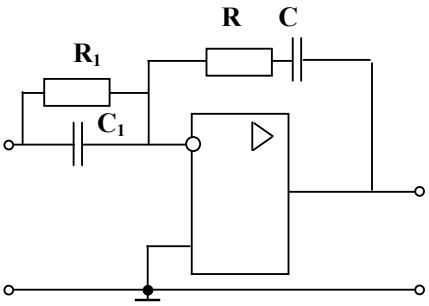


Fig. 2.34

2.3.3 Math Converters

It is the desire of all designers to achieve accurate and tight regulation of the output voltages for customer use. To accomplish this, high gain is required. However, with high gain instability comes. Therefore, the gain and the responsiveness of the feedback path must be tailored to the adjusted process.

Conventionally, an inverting differential amplifier is used to sense the difference between the ideal, or reference, voltage needed by the customer and the actual output voltage. The product of the inverse value of this difference and the amplifier gain results in an error voltage. The role the math converter is to minimize this error between the reference and the actual output by counteracting or compensating of the detrimental effects of the system. So as the demands of the load cause the output voltages to rise and fall, the converter changes the energy to maintain that specified output. If the loads and the input voltage never changed, the gain of the error amplifier would have to be considered only at 0 Hz. However, this condition never exists. Therefore, the amplifier must respond to alternating effects by having gain at higher frequencies. Such converters are called *math converters*, *regulators*, or *controllers*. The math converters serve as the cores of *reference generators*.

Summer and subtracter. Fig. 2.35 shows the simplest math converter – an op amp *summing amplifier*, named also *summer* or *adder*. The output of this circuit is the sum of the input voltages

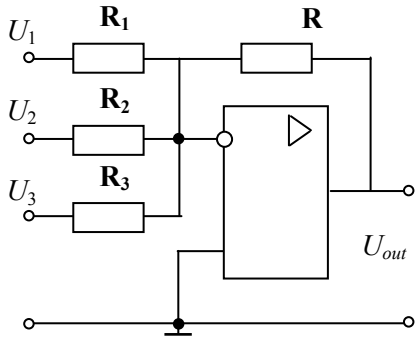


Fig. 2.35

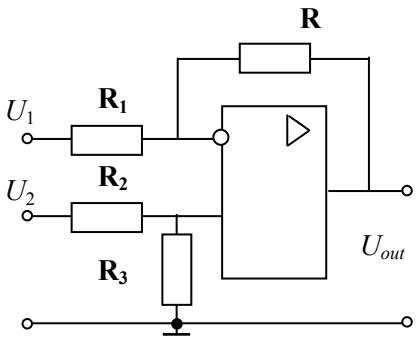


Fig. 2.36

$$U_{out} = -(U_1R / R_1 + U_2R / R_2 + U_3R / R_3).$$

In Fig. 2.36, a *subtracter* is shown, the output voltage of which is proportional to the difference of the input voltages when $R_1 = R_2$ and $R = R_3$:

$$U_{out} = (U_2 - U_1)R / R_1.$$

Integrators. Fig. 2.37 shows an op amp *integrator,* also called *I-regulator*. An integrator is a circuit that performs a mathematical operation called integration:

$$U_{out} = -1 / T \int (U_{in}\, dt),$$

where $T = RC$ is the time constant and t is time.

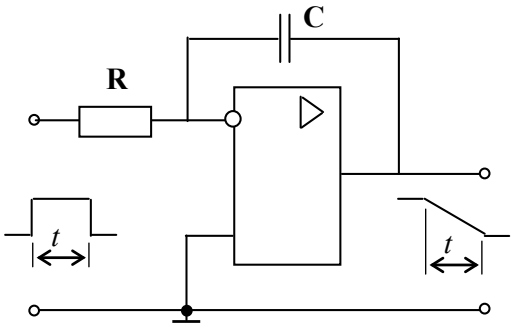


Fig. 2.37

The widespread application of the integrator is to produce a *ramp* of output voltage that is a linearly increasing or decreasing voltage value. In the integrator circuit of Fig. 2.37, the feedback component is a capacitor rather than a resistor. The usual input is a rectangular pulse of width t. As a result of the input current,

$$I_{in} = U_{in} / R,$$

the capacitor charges and its voltage increases. The virtual ground implies that the output voltage equals the voltage across the capacitor. For a positive input voltage, the output voltage will be negative and increasing in accordance with the following expression:

$$U_{out} = -I_{in}t \,/\, \mathrm{C} = -U_{in}t \,/\, T$$

while the op amp does not saturate. For the integrator to work properly, the closed-loop time constant should be higher than the width of the input pulse t. For instance, if $U_{out\ \max} = 20$ mV, $R = 1$ kΩ, $C = 10$ μF and $t = 0{,}5$ mc then $T = 10$ ms, and U_{in} should be more than 400 mV to avoid the op amp saturation.

Because a capacitor is open to dc signals, there is no negative feedback at zero frequency. Without feedback, the circuit treats any input offset voltage as a valid input signal and the output goes into saturation, where it stays indefinitely. Two ways to reduce the effect are shown in Fig. 2.38. One way (Fig. 2.38,a) is to diminish the voltage gain at zero frequency by inserting a resistor $R_2 > 10R$ across the capacitor or in series with it. Here, the rectangular wave is the input to the integrator. The ramp is decreasing during the positive half cycle and increasing during the negative half cycle. Therefore, the output is a triangle or exponential wave, the peak-to-peak value of which is given by

$$U_{out} = -U_{in} \,/\, (4fT).$$

Here, the wave of frequency f is the integrator input. This circuit is referred to as a *PI-regulator* with $K = R_2 / R$, and $T = RC$ in the case of parallel resistor and capacitor connection and $T = R_2C$ in the case of series connection. For instance, if $U_{out\ \max} = 20$ mV, $R = 1$ kΩ, $R_2 > 10$ kΩ, $C = 10$ μF and $f = 1$ kHz then $T = 10$ ms, and U_{in} should be kept more than 800 mV to avoid the op amp saturation.

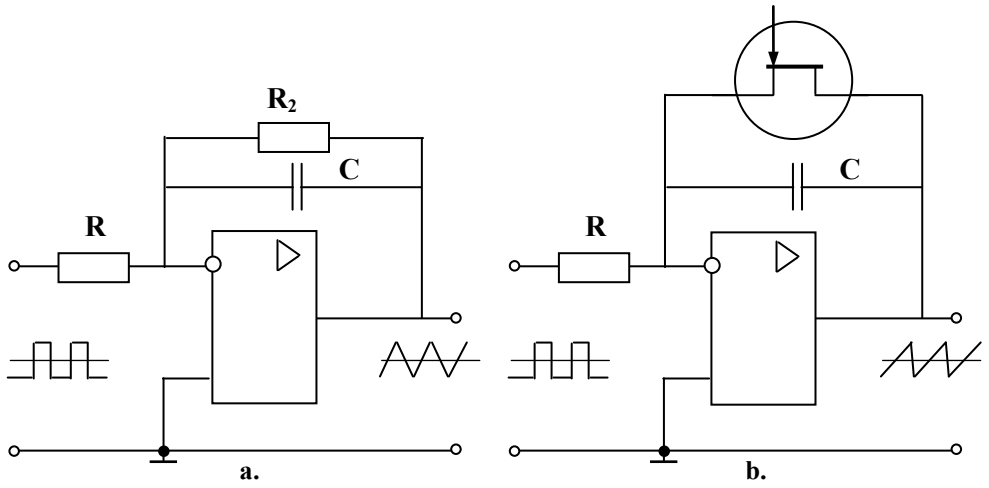


Fig. 2.38

Note that the parallel connected circuits are at the same time the low-pass and high-pass filters with the cutoff frequency $f_c = 1 / (2\pi R_2 C)$.

Another way to suppress the effect of the input offset voltage is to use a JFET switch (Fig. 2.38,b). One can set the JFET to a low resistance when the integrator is idle and to a high resistance when the integrator is active. Therefore, the output is a sawtooth wave where the JFET plays a role of the capacitor reset.

Differentiators. Fig. 2.39,a illustrates the op amp *differentiator* or *D-regulator*. A differentiator is a circuit that performs a calculus operation called differentiation

$$U_{out} = -T\, dU_{in} / dt$$

where $T = RC$ and t is time. It produces an output voltage proportional to the instantaneous rate of change of the input voltage. Common applications of a differentiator are to detect the leading and trailing edges of a rectangular pulse or to produce a rectangular output from a ramp input. Another application is to produce very narrow spikes.

One drawback of this circuit is its tendency to oscillate with a *flywheel effect*. To avoid this, a differentiator usually includes some resistance in series with the capacitor, as shown in Fig. 2.39,b or across the capacitor. A typical value of this added resistance is between 0,01 R and 0,1 R. With the resistor, the closed-loop voltage gain is between 10 and 100. The effect is to limit the gain at higher frequencies, where the oscillation problem arises. Such a circuit is called a *PD-regulator* and has $K = R / R_1$ and two time constants: $T_1 = RC$ and $T_2 = R_1 C$.

Note that these circuits are also the high-pass filters and low-pass filters with the cutoff frequency $f_c = 1 / (2\pi RC)$.

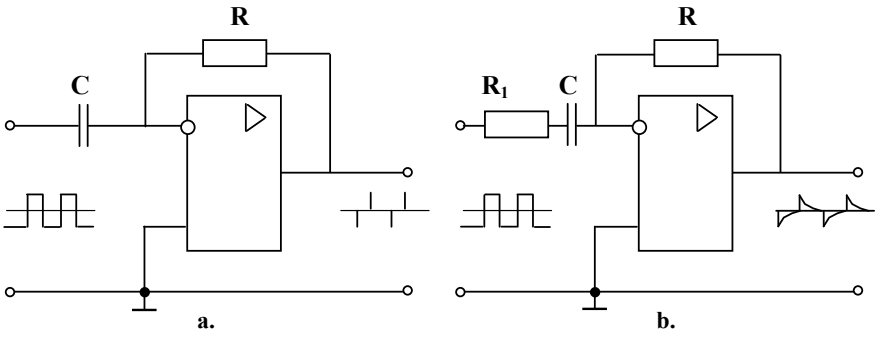


Fig. 2.39

PID-circuits. Two variants of proportional-integrated-differential circuit (*PID-regulator, PID-controller*) are shown in Fig. 2.40. They amplify the beginning and the end of the pulse signal. Circuit parameters are as follows: $K = R_2 / R_1$, $T_1 = R_1C_1$, $T_2 = R_2C_2$.

Logarithmic and exponential amplifiers. A *logarithmic amplifier* is the inverting amplifier with a feedback diode rather than feedback resistor, as given in Fig. 2.41,a. The nonlinear diode characteristic gives

$$U_{out} = U_0 \ln (U_{in} / (I_0 R)),$$

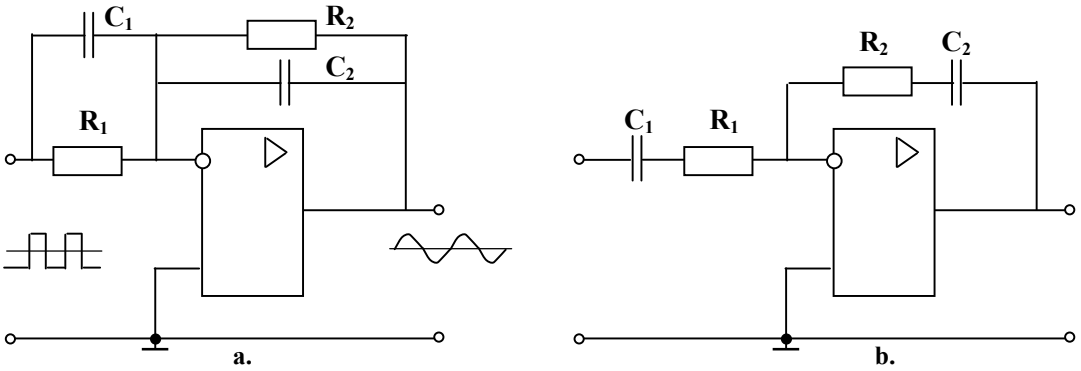


Fig. 2.40

where U_0 is near 0,06 V and I_0 around 10^{-10} A. Once the diode and the resistor positions replace one other, a *exponential amplifier* appears (Fig. 2.41,b) with the following parameters:

$$U_{out} = I_0 R \exp(U_{in} / U_0).$$

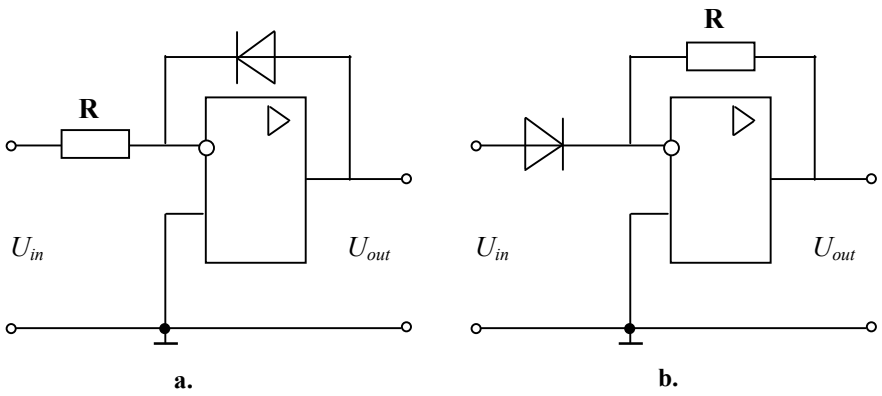


Fig. 2.41

Summary. Unlike the filters, which have an effect upon the frequency response, most of math converters improve the step response of the referred signals.

Summers and subtructers are the simplest math converters. Integrators, differentiators, logarithmic and exponential amplifiers perform more complex operations. The most universal math converters provide full proportional, integral, and differential signal converting as well.

2.4 Switching Circuits

2.4.1 Switches

As distinct from a linear circuit, in which the transistors and IC never saturate under normal operating conditions (class A mode of operation), the switching circuits, however, may re-shape the signal and open the feedback loop during the operation (classes B, C, D) therefore they are more efficient than the above discussed transistor circuits. The major benefit of using such design is its extremely low power consumption and lower heat production that makes it popular for use in calculators, watches, satellites, and power sources. Such circuits are usually much smaller physically. They can provide large load currents at low voltages although they produce more electrical and audible noise. In addition, these circuits are somewhat more costly to produce.

An ideal switch has no on-resistance, infinite off impedance and zero time delay, and can handle large signal and common-mode voltages. Real switches do not meet these criteria fully, but most of limitations can be overcome.

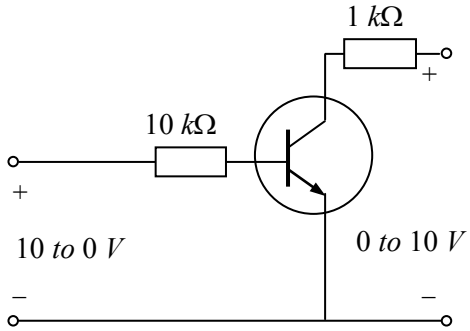


Fig. 2.42

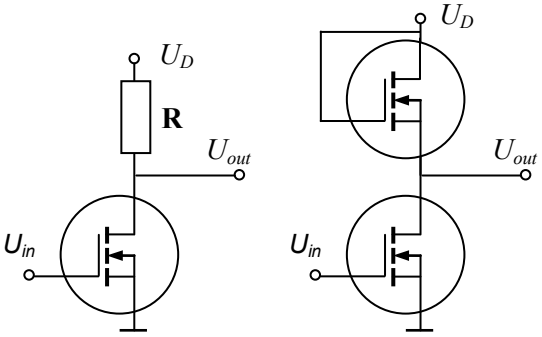


Fig. 2.43

Transistor switches. Transistorized base bias is usually designed to operate in switching circuits by having either low output voltage or high output voltage. For this reason, variations in operating point do not matter, because the transistor remains in saturation or cutoff when the current gain changes. In Fig. 2.42, the transistor is in hard saturation when the output voltage is approximately zero. This means the *Q* point is at the upper end of the load line.

When the base current drops to zero, *Q* point sets to the cutoff. Because of this, the collector current drops to zero. With no current, all the collector supply voltage will appear across the collector-emitter terminals.

Therefore, the circuit can have only two output voltages: 0 or U_{CE}. That is why the switching circuits are often called *two-state circuits*, referring to the low and high outputs, and the operating devices of these circuits are called *switches*.

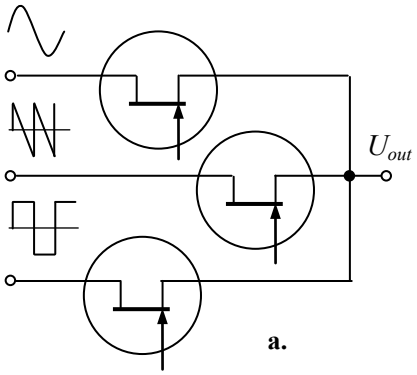
The two-stage transistorized circuits are the class B operation devices in contrast to the earlier discussed class A operation devices. Class B operation means the collector current flows for only one half of ac. For this to occur, the *Q* point is located at cutoff on both the dc and ac load lines.

Inverter switches. Fig. 2.43 shows passive and active *inverter switches* built on the MOSFETs. When the input voltage U_{in} is low (less than the threshold level), the output voltage U_{out} is high (equals the supply voltage) and vice versa if the *passive load* **R** is much greater than the drain resistance R_{DS}. The word "passive" means an ordinary resistor. When using an *active load*, the lower MOSFET still acts as an on-off switch, but the upper one acts as a large resistance with

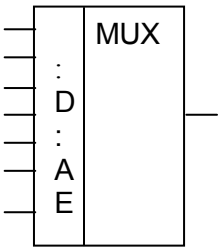
$$R = U_D / I_D.$$

The circuits are called inverter switches because their output voltage is in the opposite polarity to the input voltage.

Multiplexer. Fig. 2.44 shows a *multiplexer*, a multiple switch that steers one of the input signals to the output. Each JFET in Fig. 2.44,a acts like a single-pole single-throw switch, which can transmit data inputs by setting one of the address inputs. The circuit symbol in Fig. 2.44,b has data inputs D, address inputs A, and the blocking input E, which closes the output for a time of switching.



a.



b.

Fig. 2.44

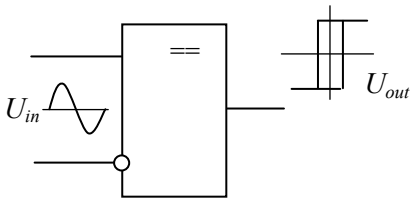


Fig. 2.45

Comparator. A *comparator* may be the perfect solution for comparing one voltage with another to see which is larger. Its circuit symbol is shown in Fig. 2.45. It is the fast differential dc amplifier of high gain and stability, with a logic output that switches to one state when the input reaches the upper trigger point and switches back to the other state when the input falls below the lower trigger point. the first industrial integral comparator μA710 was developed by R.J. Widlar in USA in 1965.

The most common comparator has some resemblance to the operational amplifier as it uses a differential pair of transistors or FETs at its input stage. Nevertheless, unlike an op amp it does not apply external negative feedback, and its output presents a logic level, indicating which of the two inputs is at the higher potential. Op amps are not designed for use as comparators – they may saturate if overdriven and recover slowly. Many op amps have input stages, which behave in unexpected ways when used with large differential voltages, and their outputs are rarely compatible with standard logic levels. When the inverting input is grounded, the slightest input voltage is sufficient to saturate the op amp because the open-loop voltage gain is near 100000. The transfer characteristic of a comparator has almost vertical transition. A *trip point* (also called the threshold, the reference, etc.) of the comparator is the input voltage where the output changes the state (low to high, or vice versa). In the drawn circuit, the trip point is zero. Therefore, the circuit is often called a *zero-crossing detector*.

Comparators need good resolution, which implies high gain (usually, 10 to 300 V/mV) and short switching time (12 to 1200 ns). This can lead to uncontrolled oscillation when the differential input approaches zero. In order to prevent this, *hysteresis* is often added to comparators using a small amount of positive feedback. Hysteresis is the difference between the left and the right trip points. The left is switchback input voltage falling and the right is switchover input voltage rising. Many comparators have some millivolts of hysteresis to encourage the "snap" action and to prevent the local feedback from causing instability in the transition region. As far as the resolution of the comparator can be no less than the hysteresis, the large values of hysteresis are generally not useful.

Latch. Fig. 2.46 illustrates a transistor *latch*. Here, the upper transistor is a *pnp* device and the lower transistor is an *npn* device. The collector of the first transistor drives the base of the second one and backward. Because of a positive feedback, a change in current at any point of the loop is amplified and returned to the starting point with the same phase. For instance, if the bottom base current increases, the top collector current will also increase. This forces a larger base current through the upper device. In turn, this produces a larger collector current, which drives the bottom base harder. This buildup in currents will continue until both transistors are driven into saturation. In this case, the circuit acts as a closed switch.

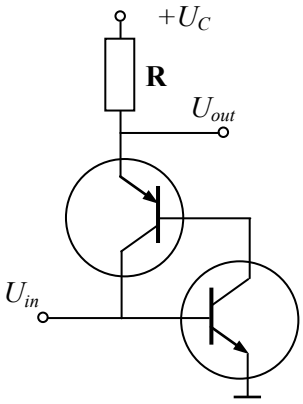


Fig. 2.46

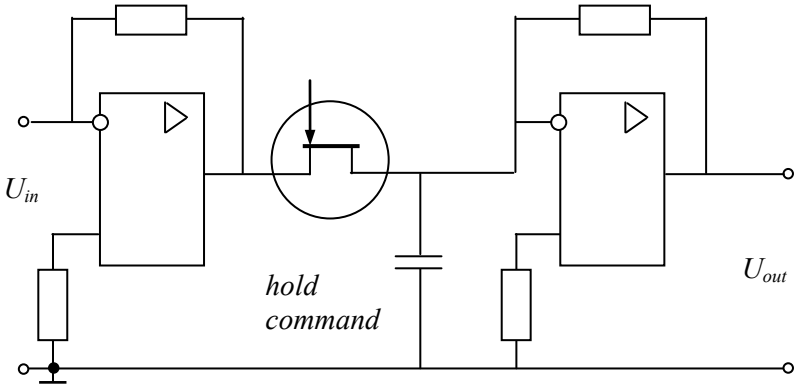


Fig. 2.47

But if something causes the bottom base current to decrease, the bottom collector current will decrease. This reduces the upper base current. In turn, there will be less collector current, which reduces the bottom base current even more. This positive feedback continues until both transistors are driven into cutoff. Then, the circuit acts as an open switch.

One way to close the latch is by *triggering*, that is by applying a sharp pulse to forward bias the bottom base-emitter diode. Once the positive feedback starts, it will sustain itself and drive both transistors into saturation. Another way to close a latch is by breakover. This means using a large supply voltage U_C to break down one of the collector diodes. It ends with both transistors in the saturated state. One way to open the latch is to reduce the load current to zero. Another way is to apply a reverse bias trigger to the bottom base instead of a positive one. This will rapidly drive both transistors into cutoff, which opens the latch.

Sample-and-hold circuit. A *sample-and-hold circuit* (*S/H*), or *sample-and-hold amplifier* (SHA), is the critical part of most *data acquisition* systems. It captures an analog signal and holds it during some operation. When the SHA is in a *hold mode,* the output is closed. When the SHA is in a *sample mode*, also known as a *track mode*, the output follows the input with a small offset equal to the *hold period*.

Regardless of the circuit details or type of SHA, all of such devices have four major components – input amplifier, energy storage device, output buffer, and switching circuit (Fig. 2.47). The input amplifier buffers the input signal by presenting high impedance to the signal source and providing current gain to charge the hold capacitor. The energy storage capacitor is the heart of the SHA. The switching circuit and its driver form the mechanism by which the SHA is alternately switched between track and hold. In the track mode, the voltage on the hold capacitor follows (tracks) the input signal through the closed transistor switch. In the hold mode, the switch is opened, and the capacitor retains the voltage present before it is disconnected from the input buffer. The output buffer offers high impedance to the hold capacitor to keep the held voltage from discharging prematurely.

RS flip-flop. Fig. 2.48 shows a pair of cross-coupled transistors operated as a latch. Each collector drives the opposite base through a resistors $\mathbf{R_B}$. In a circuit like this, one transistor is saturated, and the other is cutoff. Depending on which transistor is saturated, the *Q* output is either low or high. To arrange a momentary pulse between any base and ground, the corresponding transistor closes, and the triggering process starts again. This is an *RS flip-flop*, a circuit that can set the *Q* point to high or reset it to low. Incidentally, a complementary (opposite) output is available from the collector to the other transistor. In a schematic symbol of an RS flip-flop, which latches in either of the two states, a high input *S* sets *Q* to high; a high input *R* resets *Q* to low. Output *Q* remains in a given state until the flip-flop is triggered into the opposite state.

Schmitt triggers. A *Schmitt trigger* is shown in Fig. 2.49,a and its circuit symbol is in Fig. 2.49,b. This is a switching circuit with a positive feedback, the output of which is always flat-topped and steep-sided, whatever the input waveform. This component is a type of a comparator with hysteresis that produces uniform-amplitude output pulses from a random-amplitude input signal. It has applications in pulse systems, for example, converting a sine wave into a square wave.

The input voltage of this non-symmetric device is applied to the inverting input. The positive voltage feedback signal is adding the input signal rather than opposing it. If the input voltage is slightly negative, the trigger will be driven into positive saturation, and vice versa. When the comparator is positively saturated, a positive voltage is fed back to the non-inverting input. This positive input holds the output in the high state. Similarly, when the output voltage is negatively saturated, a negative voltage is fed back, holding the output in the low state. In either case, the positive feedback reinforces the existing output state. The output voltage will remain in a given state until the input voltage exceeds the reference voltage for that state. The transfer characteristic has a useful hysteresis loop. Hysteresis is desirable in the Schmitt trigger because it prevents noise from causing false triggering. Such a circuit is used extensively in electronic sensors with no tendency to "flutter" or oscillation. When the input signal is periodic, the Schmitt trigger produces a rectangular output. This assumes that the input signal is large enough to pass through both trip points, that is

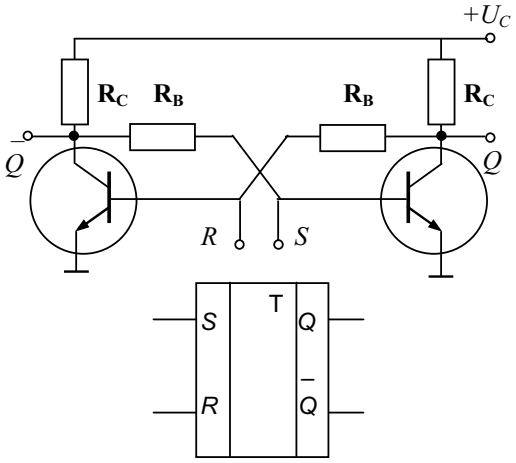


Fig. 2.48

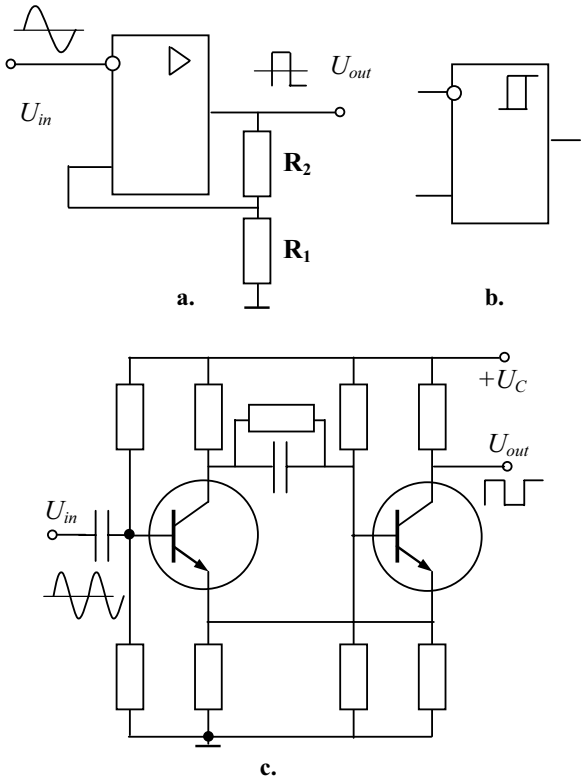


Fig. 2.49

$$U_{in} > U_{out} R_1 / (R_1 + R_2).$$

Another version of the Schmitt trigger is shown in Fig 2.49,c.

Summary. Switching circuits are usually built on BJT and FET transistors. The latter have some advantages, such as low voltage drop in the switch-on mode, high resistance in switch-off mode, low supply power, and good coupling. Both classes of circuits are the primary components of digital devices.

Different kinds of multiplexers play a role of multiple switches that direct one of the input signals to the output line. Comparators are the basic cells of many solutions required when comparing one voltage with another to see which is larger. Sample-and-hold circuits capture analog signals and hold them during some period. RS flip-flops set the output to high or reset it to low level in accordance with the input signals. Schmitt triggers produce uniform-amplitude output pulses from the random-amplitude input signals.

2.4.2 Oscillators

Oscillators (*pulsers* or *signal generators*), produce periodic signals of different shape, usually without an input pulse train. They may be linear and nonlinear devices with or without the input terminals. Some typical non-sinusoidal repetitive signal waves that pulsers generate are given in Fig. 2.50. They are as follows: a – meander, b – rectangular, c – triangle, d – sawtooth, e – pulsating, f – arbitrary signal. Most of the oscillators consist of resistors, inductors, and capacitors. In addition, diodes and transistors are used in nonlinear devices.

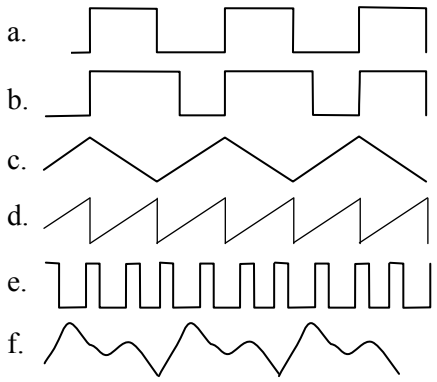


Fig. 2.50

Sine wave generators. Fig. 2.51,a shows a *sine wave generator* built as the *Wien bridge*. Thanks to the feedbacks, while the supply voltage is applied, this circuit generates the oscillations shown in Fig. 2.51,b with a period defined as

$$T = 2\pi RC,$$

where $R = R_1 = R_2$, $C = C_1 = C_2$, and $R_3 = 2R_4$. For instance, if R = 10 kΩ and C = 10 nF, then f = 1,6 kHz, T = 0,628 ms.

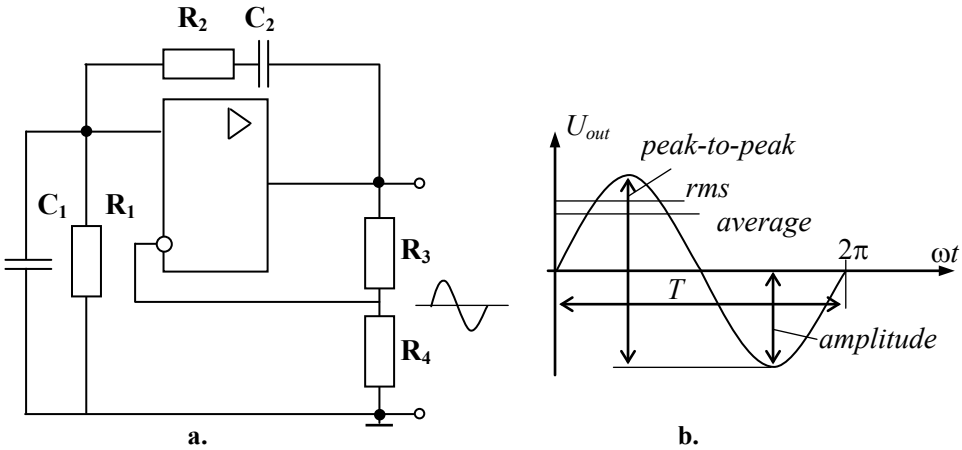


Fig. 2.51

Fig. 2.52 represents the circuits that convert a sinusoidal input signal to the pulsating output voltage. They are called *precision rectifiers* because the rectified diodes are included into the feedback loops. The second op amp in Fig. 2.52,b inserts the missed alternation to the rectified pulse chain.

Push-pull amplifiers. When a transistor is biased for the class B mode of operation, it clips off half a cycle of the input signal. To reduce distortion, two transistors are used in *push-pull* arrangement that is the pair of identical transistors connected so that the signal can be introduced across. Fig. 2.53,a shows a way to connect a class B transistors by linking an *npn* emitter follower to *pnp* emitter follower. The load is connected to the emitters of the transistors, which operate as repeaters.

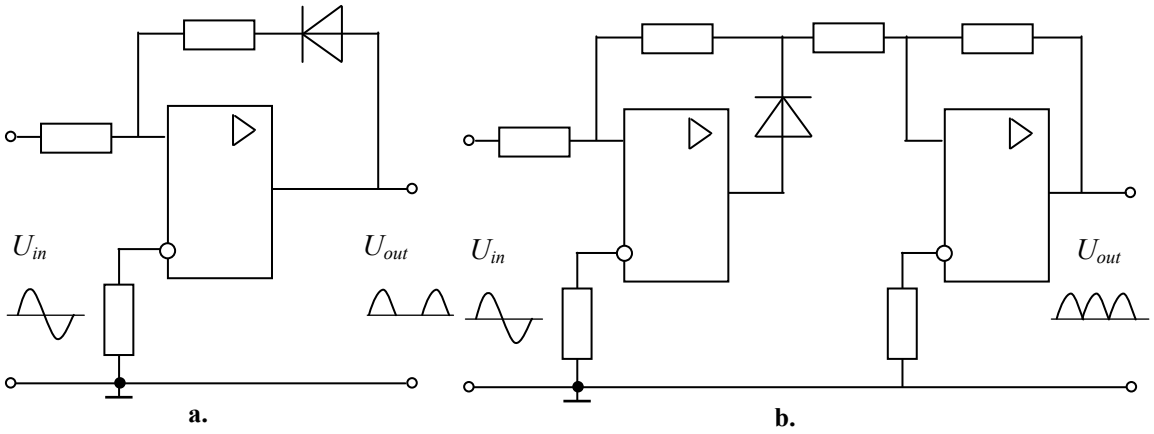


Fig. 2.52

A designer arranges the biasing of the *push-pull amplifier* to set the *Q* point at cutoff. As a result, half the ac supply voltage is dropped across the transistor collector-emitter terminals. The output of the push-pull emitter follower looks similar to the input. This means one of the transistors conducts during half of the cycle, and the other transistor conducts during the other half of the cycle. Unfortunately, because of no operation near zero, the output signal cannot follow the input exactly. Therefore, in the case of the sine input signal the output is no longer a sine wave.

To avoid distortion, diodes are used, which provide the class AB operation in the balanced supplied circuit, as shown in Fig. 2.53,b.

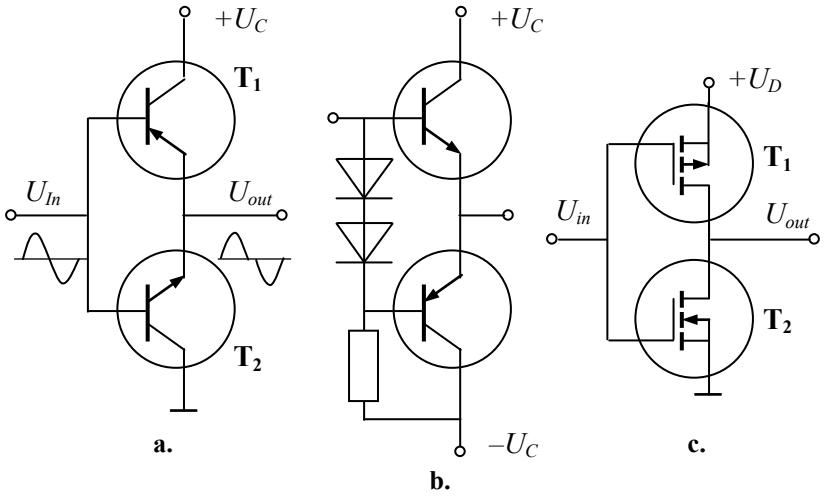


Fig. 2.53

Connecting the *p*-channel and *n*-channel MOSFETs forms the basic bilateral switch shown in Fig. 2.53,c. This combination reduces the forward resistance, improves linearity, and also produces a resistance, which varies much less with the input voltage. The circuit built on the *p*-channel ($\mathbf{T}_1$) and *n*-channel ($\mathbf{T}_2$) MOSFETs is analogous to the class B push-pull bipolar amplifier. When one device is on, the other is off, and vice versa. Push-pull amplifiers are popular in the output stages of the multistage amplifiers.

Astable multivibrators. A *multivibrator* is a rectangle pulse generator with the positive feedback. A circuit diagram of an *astable multivibrator*, which has no stable state, is given in Fig. 2.54,a. It generates non-sinusoidal oscillations of determined frequency. Here, the op amp with positive feedback includes the capacitor **C** that is charged by the op amp output through the resistor **R**. When $R_1 = R_2$, the period of multivibrator is calculated as follows:

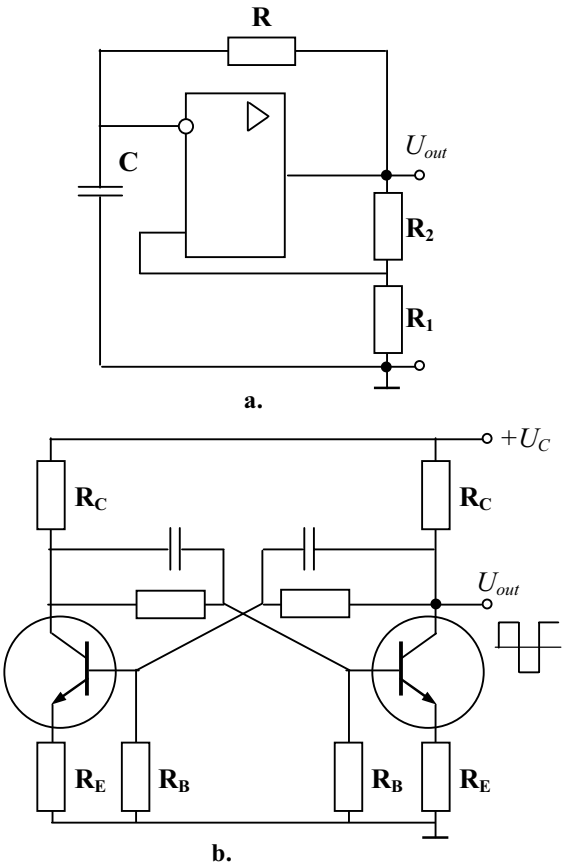


Fig. 2.54

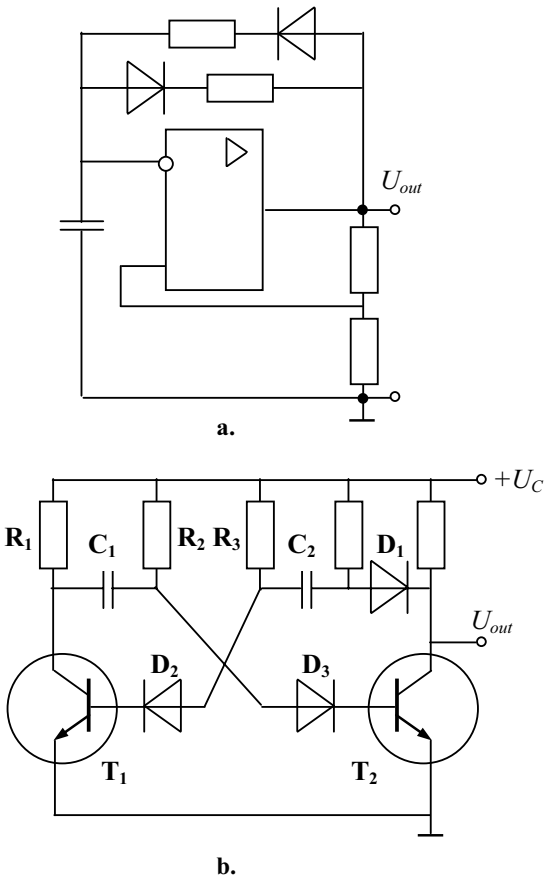


Fig. 2.55

$$T = 2RC \ln 3 = 2{,}2\ RC.$$

For instance, if $R = R_1 = R_2 = 10$ kΩ and $C = 1$ μF, then $T = 22$ ms (45,5 Hz).

The same principle of operation has the astable multivibrator shown in Fig. 2.54,b. The circuit includes two interconnected transistor amplifiers. The input of the first amplifier is the output of the second one. Once the current of one transistor becomes higher the other, the voltage drop grows on the resistor of its collector. This change is transferred through the corresponding capacitor to the base of the other transistor so that the current grows increasingly up to the first transistor saturation and the second transistor closing. After stabilizing the transient, the capacitor discharges and opens the closed transistor. Then the process repeats, and the current of the second transistor becomes higher than in the first one. The oscillation frequency depends on the resistances of resistors $\mathbf{R_B}$ and on the capacitors.

An asymmetrical astable multivibrator, shown in Fig. 2.55,a, includes a pair of diodes that provide different width of positive and negative pulses. The multivibrator shown in Fig. 2.55,b has the same principle of operation. It consists of three diodes. The diode $\mathbf{D}_1$ isolates the collector of the transistor $\mathbf{T}_2$ from the discharge of the capacitor $\mathbf{C}_2$ when $\mathbf{T}_2$ switches off. In this way, a fast-rising waveform can be obtained. The diodes $\mathbf{D}_2$ and $\mathbf{D}_3$ prevent breakdown of the base-emitter junctions when the transistors are turned off. The frequency of operation is given by the formula

$$f = 1 / (T_1 + T_2),$$

where $T_1 = \sqrt{2}\, R_2C_1$, $T_2 = \sqrt{2}\, R_3C_2$. This asymmetric circuit generates the output pulses with different continuation of positive and negative polarity.

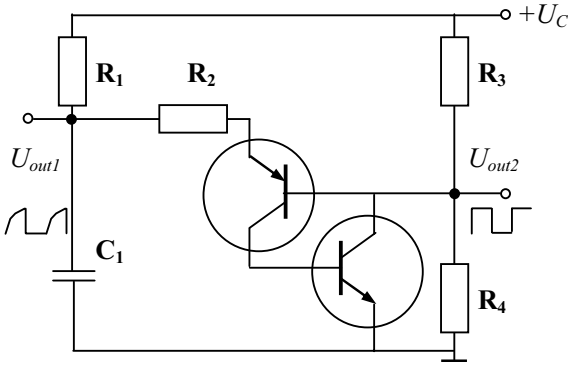
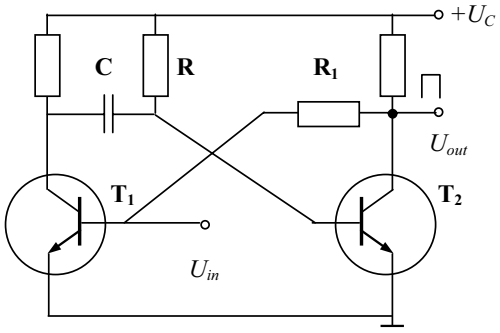
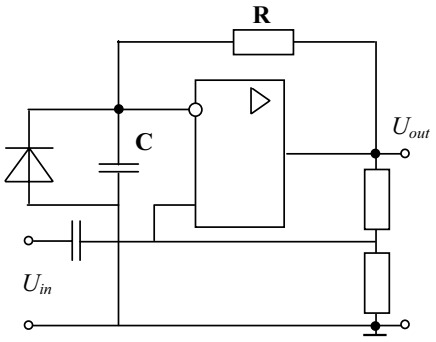


Fig. 2.56



a.



b.

Fig. 2.57

The astable multivibrator in Fig. 2.56 has two different outputs, a sawtooth and a rectangle. Usually, $R_3 = R_4$ and the frequency of both outputs is given by

$$f = 1 / (\sqrt{2}\, R_1C_1).$$

Monostables. When a pulse of a determined or variable width is required, a *monostable* circuit is used. Fig. 2.57,a shows a monostable (*single-shot*, *one-shot circuit*). It generates the only pulse after switching on, and to continue operation, an input signal must enter the circuit. The pulse width of the single-short output signal is determined by **R** and **C** values.

At the initial state, the transistor $\mathbf{T}_2$ passes the current and $\mathbf{T}_1$ is closed. The capacitor is charged. After U_{in} enters $\mathbf{T}_1$ base, the $\mathbf{T}_1$ switches on and the capacitor closes $\mathbf{T}_2$. The capacitor discharges through **R** but $\mathbf{T}_1$ continues conducting thanks to base current from $\mathbf{R}_1$. After the full discharging of the capacitor, $\mathbf{T}_2$ switches on again and $\mathbf{T}_1$ switches off**.** The output pulse width is approximately 0,7 *RC*.

The one-shot shown in Fig. 2.57,b has the same principle of operation. The diode connected across the capacitor provides the state mode of the monostable because the negative output U_{out} cannot recharge the capacitor. The input signal U_{in} is required to continue the operation.

Bistables. Many *bistable multivibrators* with the input terminals are known. These devices with memory are the backgrounds of different triggering circuits, such as RS flip-flops, where the output changes the state at each input pulse. Eccles and Jordan invented this device as early as the mid-1910s. Today, they usually play the role of *timers*.

Blocking oscillator. A *blocking oscillator* shown in Fig 2.58 represents the group of so called *relaxation oscillators* that generate non-sinusoidal oscillations. Unlike a multivibrator, the sharp pulses with broad pauses between them are produced on the output of this circuit. The transformer with hysteresis is an essential component of the blocking oscillator. Originally, the forward biased transistor emits the current to the primary winding of the transformer. The signal passes through the capacitor to the base of the transistor. The capacitor charges and sends the pulse to the transformer. After the transistor saturation, the feedback signal falls, the capacitor discharges, and the oscillation starts again. The oscillation frequency depends on the resistance and capacity.

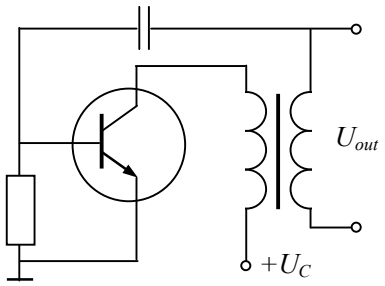


Fig. 2.58

Summary. The oscillators built on **RC** components usually have simple principle of operation, low price, and high reliability. Nevertheless, they are unstable and temperature dependent. Their output waveform has distortions and changes with time. The oscillators, which use **LC** components, have high stability and almost no dependence on the component parameters. Their drawbacks are sufficiently high complexity, size, and cost.

2.4.3 Quantizing and Coding

Analog input variables, whatever their origin, are frequently converted by transducers into voltages and currents. These electrical quantities may appear as:

- fast or slow direct measurements of a phenomenon in the time domain,
- modulated ac waveforms,
- some signal combinations, with a spatial configuration of related variables.

Examples are outputs of thermocouples, potentiometers, and analog computing circuitry; optical measurements or bridge outputs; synchros and resolvers.

Digital levels. Arbitrary fixed voltage levels referred to a ground, either occurring at the outputs of logic gates, or applied to their inputs, normally represent information in a digital form. Unlike linear circuits, in digital processing only two states are present on the outputs of the switching devices: on state and off state. On state is referred to the logical "1" or *TRUE* value. Off state is equal to the logical "0" or *FALSE* value. Most logic systems use positive logic, in which "0" is represented by zero volts or a low voltage, below 0,5 V whereas, "1" is represented by a higher voltage. *S*witching from one state to another is a very fast process. The intermediate values of conductivity do not apply in such conditions.

Groups of levels represented digital numbers are called *words*. The level may appear simultaneously in parallel on a *bus* or groups of gate inputs or outputs, serially (or in a time sequence) on a single line, or as a sequence of parallel bytes. A bus is a parallel path of binary information signals – usually 4, 8, 16, 32, or 64-bits wide. Three common types of information usually found on buses are as follows: data, addresses, and control signals. Three-state switches having inactive, high, and low output levels permit many sources to be connected to a bus, while only one is active at any time.

Quantizing. A unique parallel or serial grouping of digital levels called a *code* is assigned to each analog level, which is *quantized* (i.e., represents a unique portion of the analog range). A typical digital code would be this array:

$$d_7\ d_6\ d_5\ d_4\ d_3\ d_2\ d_1\ d_0 = 1\ 0\ 1\ 1\ 1\ 0\ 0\ 1$$

It is composed of eight bits. The "1" at the extreme left is called a *most significant bit* (MSB), and the "1" at the right is called a *least significant bit* (LSB). The meaning of the code, as a number, a character, or an analog variable, is unknown until the *conversion relationship* has been defined.

A binary digital word, usually 8-bits wide, is called a *byte*. Often, a byte is a part of a longer word that must be placed on a 8-bit bus sequentially in two stages. The byte containing the MSB is called a *high byte*; that containing the LSB is called a *low byte*.

Coding. In data systems, it is the simplest case when the input or the output is a unipolar positive voltage. The use of two logic levels naturally leads to the use of a *scale-of-two* or *binary scale* for counting where the only digits used are "1" and "0" and the position of the "1" indicates what power of 2 is represented. These states are usually stored in the flip-flops that change one state to another when the command pulses enter their input terminals. The most popular code for this type of signal is the *straight binary* that is given in the sheet below for a 4-bit converter:

Base 10	Scale	+10 V full scale (FS)	Binary code	Gray code
15	15/16 FS (+FS–1 LSB)	9,375	1111	1000
14	14/16 FS	8,750	1110	1001
13	13/16 FS	8,125	1101	1011
12	12/16 FS	7,500	1100	1010
11	11/16 FS	6,875	1011	1110
10	10/16 FS	6,250	1010	1111
9	9/16 FS	5,625	1001	1101
8	8/16 FS	5,000	1000	1100
7	7/16 FS	4,375	0111	0100
6	6/16 FS	3,750	0110	0101
5	5/16 FS	3,125	0101	0111
4	4/16 FS	2,500	0100	0110
3	3/16 FS	1,875	0011	0010
2	2/16 FS	1,250	0010	0011
1	1/16 FS (1 LSB)	0,625	0001	0001
0	0	0,000	0000	0000

Another code worth mentioning at this point is a *Gray code* (or *reflective-binary code*), which was invented by E. Gray in 1878 and later re-invented by F. Gray in 1949. In the Gray code, as the number value changes, the transitions from one code to the next involve only one bit at a time. This is in contrast to the binary code where all the bits may change, for example to make the transition between 0111 and 1000. This makes it attractive to analog-digital conversion. Some digital devices produce Gray conversion internally and then convert the Gray code to the binary code for external use.

In many systems, it is desirable to represent both positive and negative analog quantities with binary codes. Either *offset binary, twos complement, once complement,* or *sign magnitude* codes will accomplish this operation. In *binary-coded-decimal* (BCD), each base-10 digit (0…9) in a decimal number is represented as the corresponding 4-bit straight binary word. It is a very useful code for interfacing to decimal displays such as in digital voltmeters.

Summary. Analog variables may be converted into digital words and backward. During the conversion, a quantizing is performed and unique portions of the analog range are composed to the digital codes. The code high byte contains MSB and its low byte contains LSB.

In digital systems, the straight binary code is the most popular. The drawback of this code concerns the transition noise, which sometimes leads to transition errors. The Gray code is free of this disadvantage because its transitions from one code to the other involve only one bit at a time. In some systems, different bipolar codes are used.

2.4.4 Digital Circuits

Logic circuits are built on *digital gates*, which are the elementary components of any digital system. Different kinds of sequential logic circuits may be constructed by using the digital gates by joining them together to assemble many switching devices.

Binary logic. There are several systems of logic. The most widely used choice of levels are those in TTL (*transistor-transistor logic*), in which "1" corresponds to the minimum output level of +2,4 V and "0" corresponds to the maximum output level of +0,4 V. A standard TTL gate has an average power of 10 mW. A TTL output can typically drive 10 TTL inputs. TTL devices are built on BJT, which are supplied with 5 VDC and this value should be kept sufficiently accurately.

Another very popular logic system is CMOS, but its levels are generally made to be compatible with the older TTL logic standard. The basis of the CMOS elements is the MOSFET that operates in a wide range of voltages from 7 to 15 V; its average value is 10 V.

Logic gates. Any required logic combination can be built up from the few basic circuits called *gates*. The three most widespread basic circuits are those of the *AND*, *OR,* and *NOT* gates. Other ones are *NOR*, *NAND,* and *XOR.* The internal circuitry of the logical IC is not usually shown in the circuit diagrams, since the circuit actions are standardized.

The actions of logic gates are usually described by a *truth table* like this one:

U_1	U_2	NOT U_1	U_1 OR U_2	U_1 NOR U_2	U_1 AND U_2	U_1 NAND U_2	U_1 XOR U_2
0	0	1	0	1	0	1	0
0	1	1	1	0	0	1	1
1	0	0	1	0	0	1	1
1	1	0	1	0	1	0	0

Another method of description deals with *Boolean expressions,* (in honor of mathematician G. Boole, 1850), using the symbols '+' to mean *OR*, '×' to mean *AND*, and '–' to mean *NOT*.

NOT gate. In Fig. 2.59, the transistor operates as a *NOT gate* or *inhibitor circuit* because its output is opposite to the input signal. This component inverts, or complements the input signal thus it often is called inverter. If the input is high, the output is low, and vice versa. The symbol of the NOT gate is shown in Fig. 2.59 also. The truth table of the *NOT* gate is stated above. The logical equation of the gate is as follows:

$$U_{out} = \text{NOT } U_{in}.$$

Other expression is possible also:

$$U_{out} = \overline{U_{in}}.$$

OR gate. An *OR gate* is the circuit with a number of inputs and only one output. This component has a high output when at least one input is high. In Fig. 2.60, two inputs are drawn. After the entering the positive voltage to the first input, the first diode begins to conduct. By virtue of the voltage drop on the resistor, the output pulse is generated. The same result should occur after entering the pulse to the second input.

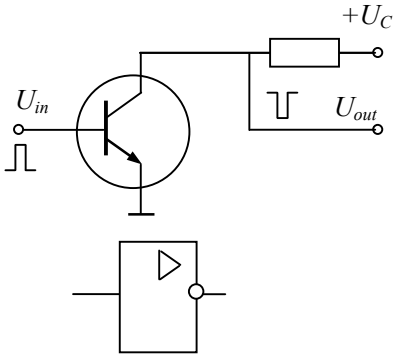


Fig. 2.59

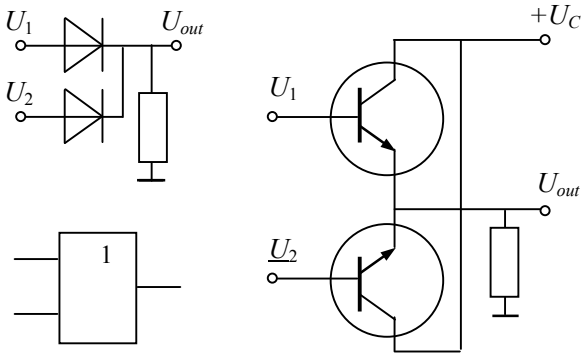


Fig. 2.60

In the transistor *OR* circuit with the common emitter, collectors should be reverse biased. When there are no inputs, the transistors are closed and the output is empty. Once the positive pulse enters an input, the corresponding transistor opens. Its emitter current flows through the resistor, the voltage drop of which is the output signal. The logical equation of the *OR* gate is

$$U_{out} = U_1 \text{ OR } U_2.$$

The operation of the gate can be expressed also as follows:

$$U_{out} = U_1 + U_2.$$

The truth table of the *OR* gate is given above.

NOR gate. The emitter follower is not the only output of the circuit described earlier. The collector output may be used too, if the emitter terminals are grounded as shown in Fig. 2.61. In these circuits, the output signal is inverted regarding to the input. The topology is known as a *NOR gate* (*OR-NOT circuit*). This component is a *NOT OR*, or inverted *OR* gate. Its output is high only when all the inputs are low.

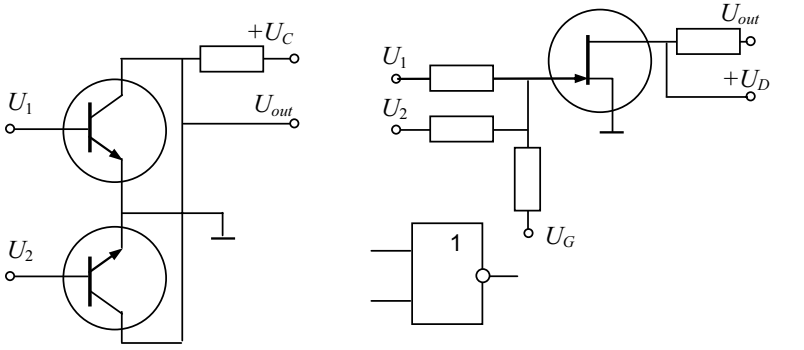


Fig. 2.61

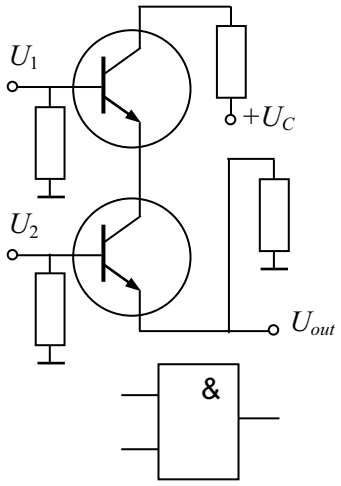


Fig. 2.62

AND gate. The circuit given in Fig. 2.62 is called an *AND gate*. This component has a high output only when all inputs are high. To get the output signal, both input signals should be presented simultaneously. The emitter of the upper transistor is coupled to the collector of the lower transistor. When the transistors are not open together, there is no current flow across the transistors and the output is empty. After the input signals come, each transistor becomes forward biased. Therefore, the collector's currents flow to the output. The *AND* circuit solves the logical equation

$$U_{out} = U_1 \text{ AND } U_2.$$

A similar expression is as follows:

$$U_{out} = U_1 \times U_2.$$

The truth table of the *AND* gate is given above.

NAND gate. This component is a *NOT AND*, or inverted *AND* gate. Its output is low only when all inputs are high. One of the *NAND gates* shown in Fig. 2.63 is the same circuit as in Fig. 2.62 with other output. Another circuit is built on the MOSFET transistors.

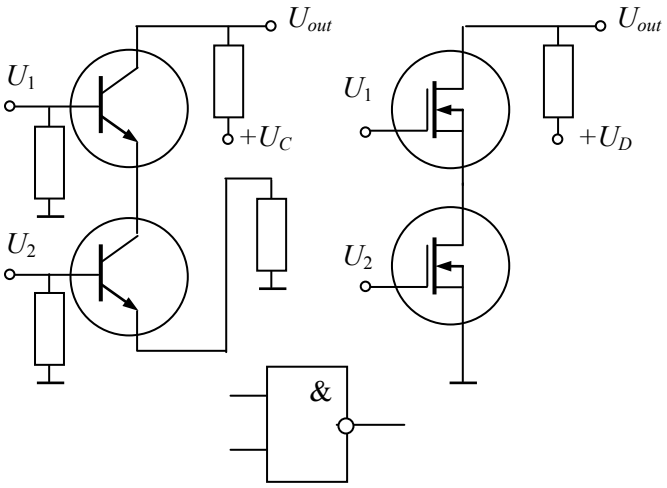


Fig. 2.63

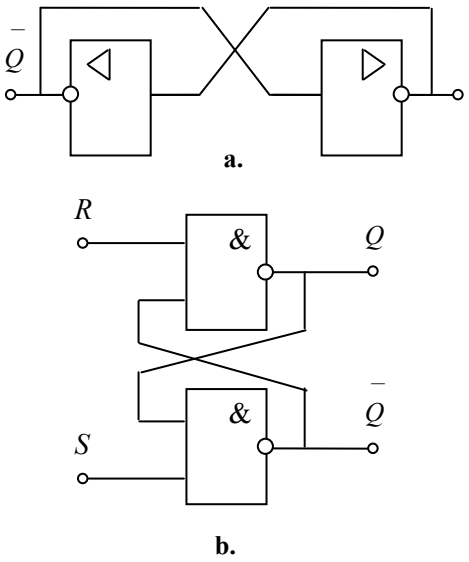


Fig. 2.64

XOR gate. A *XOR gate* (*exclusive-OR circuit*) may be built on the basis of the expression

$$U_{out} = \text{NOT } (U_1 \text{ AND } U_2) \text{ AND } (U_1 \text{ OR } U_2)$$

as a combination of the earlier discussed gates. The truth table of the *AND* gate is given above. This component has a high output when an odd number of inputs (1, 3, 5, etc.) is high. An even number of high inputs generates a low output.

Sequential logic. Using the digital logic, different kinds of switching devices may be constructed. They are known as *sequential logic circuits* that change the output when the correct sequence of signals appears at the inputs. Fig. 2.64 displays a multivibrator and an RS-type flip-flop. Their outputs are in the counter-phase states. The RS-type flip-flop is used to lock information, one RS latch per each bit. The truth table of the RS-type flip-flop is given below:

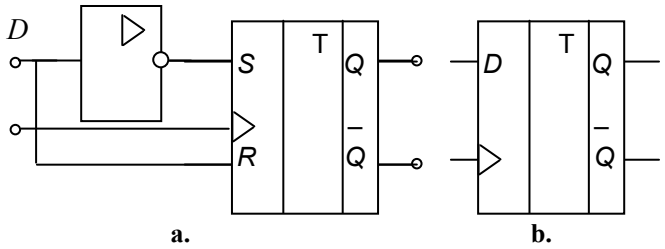
S	R	Q
1	0	1
0	1	0
0	0	Forbidden
1	1	Forbidden

The applications of the simple RS latch are rather limited, and most sequential logic circuits make use of the principle of *clocking*. A clocked circuit has the clock input marked by a triangle to which *clock pulses* can be applied. Unlike the RS-type flip-flop, a *D-type flip-flop* is controlled by the clock input (Fig. 2.65). Its circuit action takes place only at the time of the clock pulse, and may be synchronized to a *leading edge* or a *trailing edge*, thus earning the circuit the name an *edge triggered circuit*. When the clock pulse level changes, the output becomes equal to *D* input, which means *Q* repeats *D*. The truth table of the D-type flip-flop is given below:

Clock	D	Q
0	1	No change
1	1	1
0	0	No change
1	0	0

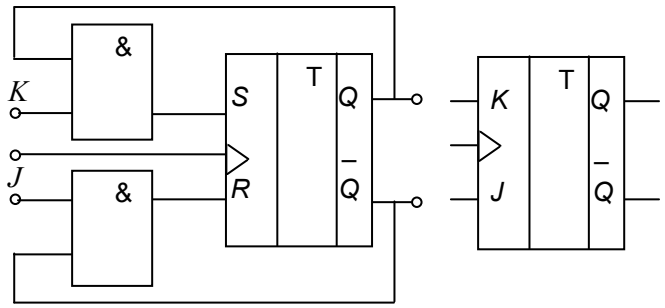
A *JK-type flip-flop* (Fig. 2.66) is a much more flexible design, which uses a clock pulse along with the two control inputs labeled *J* and *K*. The flip-flop changes the state when the clock pulse level is equal to unity. Here, *Q* repeats *J* when *J* is not equal to *K*. While *J* and *K* are in zero, *Q* stores its previous level. If *J* and *K* are equal to unity, *Q* changes its state. The truth table of JK-type flip-flop is given below:

Clock	J	K	Q
1	0	0	No change
1	0	1	0
1	1	0	1
1	1	1	0



a. **b.**

Fig. 2.65



a. **b.**

Fig. 2.66

Encoders and decoders. An *encoder* converts decimal numbers to a binary code. An example of unipolar binary 7/3 encoder is given in Fig. 2.67,a. Here, the decimal digits from 0 to 7 enter the associated inputs of the *OR* gates and the bits appear on their outputs. The circuit symbol of the encoder is shown in Fig. 2.67,b. For instance, when "1" enters the inputs "6", the output code $d_0 d_1 d_2$ becomes equal to "011".

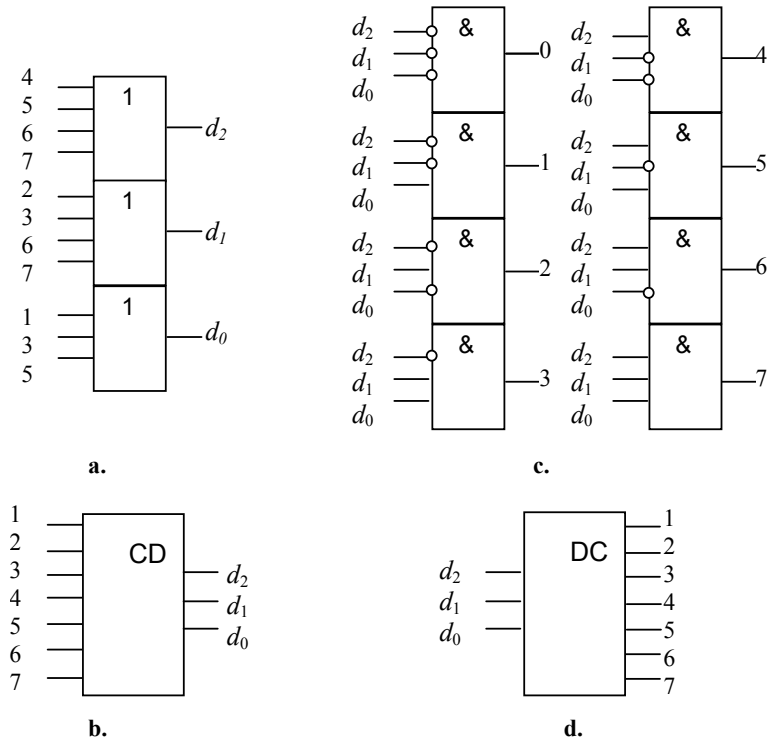


Fig. 2.67

A *decoder* executes the opposite function. Its input signal represents the binary code and the decimal value is on the output. An example of unipolar binary 3/7 decoder is given in Fig. 3,26,c. Here, the binary code enters all inputs of both *AND* gate and "1" appears on one of the digit outputs. For instance, if the input code $d_0 d_1 d_2$ is equal to "110", then "1" appears only on the output "6".

Summary. Logic circuits are built on the TTL and CMOS digital gates. The actions of logic gates are usually described by truth tables. *NOT, OR, NOR, AND,* and *NAND* are the most popular logic gates, which are commonly used in digital electronics. Unlike the simple gates, sequential logic circuits change output when the correct sequence of signals appears at the inputs. Encoders and decoders convert codes from one form to another.