Springer Series in Advanced Microelectronics 61

Andrey D. Grigoriev
Vyacheslav A. Ivanov
Sergey I. Molokovsky

Microwave Electronics

Edited by Professor A. D. Grigoriev

Springer

Springer Series in Advanced Microelectronics

Volume 61

The Springer Series in Advanced Microelectronics provides systematic information on all the topics relevant for the design, processing, and manufacturing of microelectronic devices. The books, each prepared by leading researchers or engineers in their fields, cover the basic and advanced aspects of topics such as wafer processing, materials, device design, device technologies, circuit design, VLSI implementation, and subsystem technology. The series forms a bridge between physics and engineering and the volumes will appeal to practicing engineers as well as research scientists.

More information about this series at http://www.springer.com/series/4076

Andrey D. Grigoriev · Vyacheslav A. Ivanov
Sergey I. Molokovsky

Microwave Electronics

Edited by Professor A. D. Grigoriev

Springer

Andrey D. Grigoriev
Saint-Petersburg Electrotechnical
University "LETI"
Saint-Petersburg
Russia

Sergey I. Molokovsky (deceased)
Saint-Petersburg Electrotechnical
University "LETI"
Saint-Petersburg
Russia

Vyacheslav A. Ivanov
Saint-Petersburg Electrotechnical
University "LETI"
Saint-Petersburg
Russia

ISSN 1437-0387 ISSN 2197-6643 (electronic)
Springer Series in Advanced Microelectronics
ISBN 978-3-319-68890-9 ISBN 978-3-319-68891-6 (eBook)
https://doi.org/10.1007/978-3-319-68891-6

Library of Congress Control Number: 2017957671

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Nature is simple in its laws,
but immeasurably rich and diverse in their applications!

G. W. Leibniz

Preface

Microwave electronics is currently a large and rapidly developing branch of science and technology, which has a huge impact on the country's economy and defense capability. Most telecommunication systems, radio astronomy, accelerative techniques, thermonuclear machines, technology and medicine, and practically all types of weapons use microwave electronics technology to some extent. These achievements are mostly due to the use of new interaction mechanisms of charged particle fluxes with electromagnetic fields, use of new materials, and new manufacturing technologies of microwave electronics devices and systems.

The physical laws and phenomena used in microwave electronic devices, and the operating principles of these devices, along with their design, characteristics and parameters form the subject of the "Microwave Electronics" discipline. These laws, phenomena and devices are studied more deeply in the masters programs of the corresponding field.

Unfortunately, the textbooks and manuals on this discipline existing at the time of preparation of this manuscript were either published a long time ago and are out of date, or do not cover all the discipline's content. In particular, there are no modern manuals considering the processes taking place in a vacuum and in semiconductor microwave devices from the unified position.

The proposed textbook describes in reasonable detail the main interaction mechanisms of the microwave electromagnetic field with charged particles in a vacuum and in a solid. Much attention is paid to the peculiarities of these mechanisms using vacuum and solid-state microwave electronic devices. This theoretical part is supplemented by a description of the design of the devices, their parameters and characteristics.

The authors hope that the proposed textbook will assist students studying microwave electronics and related areas.

The book will also be useful to postgraduates and specialists engaged in the development and application of microwave electronics products.

The textbook is based on the lecture course taught by the authors to bachelors and masters in St. Petersburg State Electrotechnical University "LETI".

The authors express also their gratitude to V. B. Yancevich for valuable advices and assistance in preparing the book for publication. The authors are also sincerely grateful to V. B. Yankevich, the Head of the LETI Radio-technical Electronics Department, for valuable advice and invaluable assistance in preparing the manuscript for publication.

The Introduction, Chapters 1, 3, 5... 8, Chapter 10 and Appendix C was written by A. D. Grigoriev, Chapter 12... 14 and Appendix A was written by V. A. Ivanov, Chapters 2, 4 and 11 was written together by A. D. Grigoriev and V. A. Ivanov, Chapter 9 and Appenix B was written together by A. D. Grigoriev and S. I. Molokiovsky.

Saint-Petersburg, Russia Vyacheslav A. Ivanov

Contents

Notations

Scalar values are denoted by Latin letters, in italic, and Greek letters typed in a direct font: *a, v, A*, φ, ψ, etc. Vectors and tensors are denoted by Latin and Greek letters in bold direct type: $\mathbf{A}, \mathbf{B}, \Psi$. When necessary, cases the notation of vectors, tensors, and matrices are enclosed in direct brackets: $|\mathrm{A}|, |\mathrm{B}|$ or are over lined: $\bar{\varepsilon}, \bar{\mu}$. Complex quantities, when necessary, are marked with a dot above the symbol: $\dot{\rho}, \dot{\mathbf{A}}, \dot{A}$. Constants are typed in a direct font: e, i. The scalar product is denoted by a point: $\mathbf{A} \cdot \mathbf{B}$, and the vector product by a sidelong cross: $A \times B$. To denote the differential operations on vectors, the Hamiltonian operator ∇ is used.

a	Acceleration, m/s^2
$\mathbf{A}$	Vector potential, V·s/m
$\mathbf{B}$	Magnetic induction, V·s/m^2
B	Susceptance, S
$c = 2.9979 \times 10^8$	Speed of light in a vacuum, m/s
C	Capacitance, F
$\mathbf{D}$	Electric displacement vector, A·s/m^2
D	Diffusion coefficient, m^2/s
e = 2.71828	Base of the natural logarithms
$e = 1.602 \times 10^{-19}$	Electron charge absolute value, C
$\mathbf{E}$	Electric field strength, V/m
f	Frequency, Hz
$\mathbf{F}$	Force, N
G	Conductance, S
G	Amplification factor
$h = 6.626 \times 10^{-34}$	Planck constant, J·s
$\mathbf{H}$	Magnetic field intensity, A/m
j	Imaginary unit
i, I	Current, A
j, J	Current density, A/m^2
k, $\mathbf{k}$	Wave number, wave vector, 1/m

$k = 1.38 \times 10^{-23}$	Boltzmann constant, J/K
L	Inductance, H
M	Mutual inductance, H
M	Beam coupling coefficient
n	Concentration, m^{-3}
n_m	Medium refractive index
n_p, n_g	Deceleration of wave phase and group velocity
P	Power, W
$\mathbf{P}_w$	Power density (energy-flux density), W/m^2
$\mathbf{p}$	Momentum, kg·m/s
$\boldsymbol{Q}$	Q-factor
q	Electric charge, C
R	Resistance, Ohm
R_c	Interaction impedance, Ohm
R_e	Resonator equivalent resistance, Ohm
T	Temperature, K
U	Voltage (electric potential), V
$\mathbf{v}$, v	Velocity, m/s
W	Energy, J
w	Energy density, J/m^3
X	Reactance, Ohm
Y	Admittance, S
Y_g	Transmission line wave admittance, S
Z	Complex impedance, Ω
Z_0	Resonator wave impedance, Ω
Z_g	Transmission line wave impedance, Ω
α	Attenuation constant, m^{-1}
β	Phase constant, m^{-1}
$\gamma = \beta - i\alpha$	Propagation constant, m^{-1}
δ	Field penetration depth, m
ε_r	Relative dielectric constant
$\varepsilon_0 = 10^7/(4\pi c^2)$	Dielectric constant, A·s/(V·m)
η	Efficiency
$\eta_0 = 120\pi$	Intrinsic impedance of free space, Ohm
θ	Transit angle, rad
λ	Electromagnetic wavelength in a vacuum, m
λ_c	Cutoff wavelength in transmission line, m
λ_g	Wavelength in transmission line, m
μ	Charge carrier mobility, $m^2/(V·s)$
μ_r	Relative permeability
$\mu_0 = 4\pi \times 10^{-7}$	Magnetic constant, V·s/(A·m)
Π	Poynting vector, W/m^2
ρ	Volume electric charge density, C/m^3
ρ	Characteristic impedance, Ohm
σ	Conductivity, S/m

τ	Relaxation constant, s
φ	Phase, rad
Φ	Scalar potential, V
ψ	Probability function
Ψ	Magnetic flux, Wb
ω	Angular frequency, rad/s

Introduction

The subject of microwave electronics is the study of the physical processes occurring in electronic devices intended for generating, amplifying and converting microwave band electromagnetic oscillations, as well as the development of methods for the design and engineering of these devices and recommendations for their application.

In accordance with the recommendations of the International Electrotechnical Commission the microwave range is the part of the electromagnetic oscillations spectrum from 3×10^8 to 3×10^{11} Hz (300 MHz to 300 GHz), that corresponds to wavelengths in a vacuum from 1 to 1 mm. The microwave range is divided into several bands:

ultrahigh frequencies (UHF or decimeter waves)—the frequency band from 300 MHz to 3 GHz (wavelengths are 1 m to 10 cm);

superhigh frequencies (SHF or centimeter waves)—the frequency band from 3 to 30 GHz (wavelengths are 10–1 cm);

extremely high frequencies (EHF or millimeter waves)—the frequency band from 30 to 300 GHz (wavelengths are 1 cm to 1 mm).

Often the microwave range includes adjacent frequency bands. For the low frequencies it is the very high frequencies (VHF) band or meter waves (MW)—the frequency range is 30–300 MHz (wavelengths are 10–1 m), and for the high frequency side—the hyperhigh frequency band (HHF) or decimillimeter waves (DMMW)—the frequency range is 300 MHz to 3 THz (wavelengths are 1 mm to 100 μm). This range is often called the terahertz range, and waves in this range are called T-rays. Currently researchers are paying special attention to this range.

As can be seen, microwaves occupy a large portion of the spectrum with an extreme frequency ratio of 1:1000 (or 1:100,000, if we add the adjacent bands). They are located on the frequency scale between radio and optical waves. Historically, people began to use the optical range foremost when for the first time an ancient man lit a fire in a cave to warm and light it. At the end of the nineteenth century thanks to the works of H. Hertz, A. S. Popov, G. Marconi and many other scientists, the radio range's turn came.

With the development of radio engineering, the advantages of using higher frequencies became increasingly evident. Therefore, in the mid-1930s the mastering of the microwave range began. Both researchers and equipment designers met with great difficulties, since the methods of generating, amplifying, detecting, and canalization of the electromagnetic radiation developed for neighboring ranges proved to be unsuitable for microwaves. The use of optical methods was problematic because microwave quantum energy is small compared to the energy of thermal motion. The transfer of methods from the radio range was hindered by the long electron-transit time in the active area of devices compared to the period of oscillations, as well as large parasitic capacitances and inductances of the design elements of devices. As a result, it was necessary to develop new mechanisms of charged particles interaction with the electromagnetic field, as well as new designs of devices and wave guiding structures. The first practically realized mechanisms and constructions of microwave devices appeared at the end of the 1930s.

Radar creation and improvement prior to the Second World War and then during the war gave a powerful boost to the development of microwave technology and electronics. After the end of the war, the development of microwave electronics continued at a heightened pace; there appeared new areas of its application—radio astronomy, radio spectroscopy, charged particle accelerators, thermonuclear fusion reactors, medical apparatus, and microwave heating facilities. By the end of the last century, the microwave systems of telecommunications including cellular and satellite communications and global positioning systems began to develop at a rapid pace. The use of microwaves in biology and medicine, in chemistry, in the food manufacturing industry, in new materials production technology, in logistics and in other fields of science and industry is expanding.

The formation and development of microwave electronics became possible thanks to the works of numerous scientists and engineering teams from many countries.

Today, microwave electronics is a synthetic field of knowledge that unites fundamental sciences (electrodynamics, plasma and solid state physics, mathematical modeling), engineering, increasingly complex production technologies, and modern measuring and testing equipment.

Successful work in this area requires deep theoretical knowledge and practical skills.

In this textbook, the authors attempted to consider the processes of electromagnetic field interaction with fluxes of charged particles in a vacuum, in plasma, and in a solid, from a unified position. This approach allowed the identification of common features and differences in the designs, characteristics and parameters of various microwave devices.

The textbook contains three parts. In the first part the mechanisms of individual and collective electron emission, wave and oscillating processes in electron fluxes are considered, and the basic concepts of microwave electronics are introduced.

The second part is devoted to vacuum microwave devices. Electron optical and electrodynamics systems of these devices are considered, and their operation

principles, designs, characteristics and parameters are expounded on. Along with "ordinary" devices, relativistic microwave electronic devices are also considered.

The third part describes the principle of operation, design and parameters of semiconductor microwave devices. Much attention is paid to the devices with new wide-band materials, namely silicon carbide and gallium nitride.

In conclusion, the problems facing the developers of microwave devices and identified and discussed, along with their possible solutions.

Part I
Microwave Electronics Physical Foundations

Chapter 1
Main Stages of Microwave Electronics Development

1.1 Background

The era of radio began with proof of the existence of electromagnetic waves by H. Hertz, professor of Karlsruhe Polytechnic University, in 1888. He created the first primitive spark generator and receiver of these waves. The wavelength of the radiation, which he investigated, was about 3 m (frequency of 100 MHz). In 1890 E. Branly invented a more sophisticated device for receiving electromagnetic waves, it was a tube filled with metal filings. Under the influence of electromagnetic radiation, the resistance of the tube sharply decreased due to micro-breakdowns of oxide films covering the surface of the filings. Essentially, this was the first *solid-state electronic device*. In 1894 O. Lodge improved this device and gave it the name "*coherer*". With its help Lodge showed wireless transmission and reception of Morse code signals, namely the transmission of signals by radio over a distance of about 40 m. It happened on August 14, 1894 at the Royal Institute in London. However, Lodge did not patent his device. Later A.S. Popov and G. Marconi kick started the wide use of radio communication.

In these first experiments, radio waves of comparatively short length were used —lying in the meter wave band. However, the need to increase radio communication coverage, voice and music transmission necessitated the invention of new types of generators—arc oscillators and machine generators, operating in continuous wave mode with a wavelength of several kilometers.

Subsequently, the coherers in receivers were replaced by "crystal detectors "—semiconductor devices with a Schottky barrier, invented by the German professor K. Brown in 1898 (although at that time there were no such concepts as "semiconductor" and "Schottky barrier").

In 1906, the American scientist Lee de Forest invented a three-electrode tube *audion* (triode) capable of amplifying radio signals. At that moment, vacuum electronics began. In the 1920s, amplifiers and generators on vacuum devices

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_1

reached the power that allowed their use in radio stations transmitters. However, radio transmitters still used long and ultra-long waves.

In 1921, an employee of the Nizhny Novgorod Radio Laboratory O. Losev discovered that the contact of zincite (ZnO) with a steel wire has negative dynamic resistance. Essentially, it was the first *tunnel diode* . Using this device, Losev assembled a receiver with high sensitivity, which he called a crystadyne. However, his research did not go any further.

In the late 1920s to early 1930s, radio amateurs found out that using short waves (wavelength of 10–50 m) it was possible to establish communication over long distances using low-power transmitters. At that point, there began a race to master ever higher frequencies, which continues to the present day (Fig. 1.1). During the last 100 years, the top operating frequency of communication systems has increased by almost a million times! This race received a boost in the late 30 s, when the task was to detect fast-flying aircraft, and the development of radar began.

As early as 1895, A.S. Popov noticed the possibility of detecting objects (ships) with the help of radio waves when he noticed a weakening of the signal transmitted by one ship to another as a third ship passed between them. However, no attempts were made to implement this observation. "The use of radio waves to detect remote metal objects" was demonstrated by C. Huelsmeyer in 1904 (finding a ship in dense fog), but the distance to the ship was not determined. The first pulse radar (RAdio Detection and Ranging) was demonstrated in the United States by Robert Page in 1934.

A similar system was developed by Rudolf Kuhnhold in Germany in 1935 and at the same time by Robert Watt in the UK. In 1943, Paige significantly increased the accuracy and interference immunity of radars by proposing a monopulse system which is still in use.

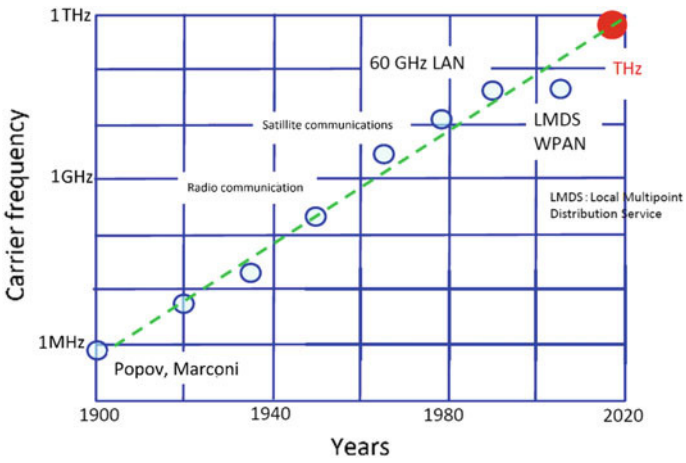


Fig. 1.1 Change in top operating frequency of communication systems

In the USSR, the first radio detector of aircraft "Rapid" was created in the Leningrad Radiophysical Institute in 1934, under the direction of A.I. Merzheevsky. It operated at a wavelength of 5 m and had a transmitter power of about 200 W. The structure of "Rapid" included one transmitting and three receiving antennas. It detected aircraft at a distance of up to 5 km. The first pulse radio locator was created by Y.B. Kobzarev at the Leningrad Physicotechnical Institute in 1938. It allowed detection of aircraft at distances of up to 50 km and simultaneous determination of the distance to the target.

The English were the first to begin the widespread use of radars for defense against air attacks. The invention of the magnetron allowed them to install radar on fighter planes. "The Battle for England" in 1940–1941, when the United Kingdom fought against Germany alone, was to a large extent won thanks to radar. Later, during the Second World War, radar stations (RS) were widely used by all belligerents on land, at sea and in the air. In the course of the war the USSR received 445 detecting and gun-laying RS through lend-lease.

After the war, RS development continued at a rapid pace. New fields of microwaves application appeared: radio spectroscopy, radio astronomy, household and industrial heating plants, plasma heating and diagnostics in thermonuclear fusion machines, terrestrial and space communications with high speed information transition, hidden object detection systems, biology and medicine, and many others.

In the development of the first RS designers faced the problem of lack of sufficiently powerful and high-frequency power supply units and low-noise amplifiers. It quickly became clear that the electron tubes at that time could not work effectively at ultrahigh frequencies (more than 300 MHz). It was necessary to develop devices that were free from limitations of interelectrode capacitance and the transit time of electrons in the interaction distance.

1.2 Microwave Vacuum Electronics

The R. and S. Varian brothers' invention of the **drift klystron** in Stanford University in 1937 was the first leap in this direction. The device had output power up to 10 W at the frequency of 1 GHz. A substantial contribution to the development of these devices was made by Hansen, who designed the first types of rhumbatron (cavity resonator). After the war the klystrons were improved, and today they provide power of up to several tens of MW in the frequency range of 1–30 GHz.

The first example of a *magnetron* with a slit anode was created by A. Hall in 1920. He also proposed the term "magnetron". However, this and subsequent examples of magnetrons with a slit anode were not suitable for practical use.

A multi-cavity resonant **magnetron** was invented by Hollmann in Berlin in 1935, but the Germans did not appreciate the value of this device. In 1939, in the USSR, Alekseev and Mulyarov developed the design of a multi-cavity pumped magnetron capable of generating power of up to 300 W at the frequency of 3 GHz. In 1940 J. Randall and H. Boot at the University of Birmingham created a compact

multi-cavity magnetron, which provided power 100 times greater than any microwave radiation source known at that time. In September 1940, Churchill agreed to Tizard's proposal about the transfer of a magnetron sample to the US in exchange for financial and industrial assistance. During Tizard's mission, the magnetron sample with an output power of 6 kW at the frequency of 3 GHz was transferred to the government of the USA, where the mass production of these devices was developed. The disadvantage of the first multi-cavity magnetrons was instability and frequency "hopping" but in 1941 Randall and Boot solved this problem by introducing straps between the resonators. Other types of sources (klystrons) available in the USA and Germany at that time had the power of not more than 10 W in this range.

Magnetrons became the basis of Second World War radar stations. However, in the radar receivers crystal detectors were still used since the limiting frequency of the vacuum diodes did not exceed 400 MHz.

The rapid development of vacuum microwave electronics in the 1940s and 1950s was characterized both by the devices appearance and by rapid parameters improvement of the already known ones. This period can be called the golden age of vacuum microwave electronics.

The O-type traveling-wave tube—a broad-band low-noise amplifier, which the radar designers lacked, was invented in 1942 in the laboratory of the British Admiralty by R. Kompfner. The theory of this device was developed by Pierce in 1950–1952 in the Bell laboratory (the USA). It allowed improvement of the device parameters significantly.

The O-type backward-wave tube is an electrically tunable low-power generator operating in the centimeter and millimeter wavelength ranges. It was invented by M.F. Stelmach (USSR) in 1948. A similar device was demonstrated by Kompfner in 1951. To date, this is a device capable of generating the record high frequences up to 1 THz.

In the 1950s, the relatives of the magnetron were invented, called **M-type electronic devices**, which also used crossed fields. In 1949, D. Wilbur and P. Peters (USA) developed a mitron, in 1950 Warnecke in France created the M-type TWT, and then in 1952 Epstein in France invented the M-type BWT, and Brown (the USA) in the same year created a platinotron. These devices significantly expanded the capabilities of radar designers.

Electron cyclotron resonance masers were proposed in 1959 independently by Gaponov-Grekhov in the USSR, and Schneider and Pantell in Australia. The first working gyrotrons were created in 1965 at the Institute of Applied Physics of the USSR Academy of Sciences. At present gyrotrons have output power up to several MW in pulsed operation mode and hundreds of kW in continuous operation mode in the millimeter wavelength range with an efficiency of up to 30–40%. Gyroklistrons and gyro-TWTs have also been developed. In most gyroresonance devices, superconducting magnets are used that sharply increase their weight and dimensions.

In the 1960s, a new direction arose—relativistic microwave electronics based on the use of electrons moving with relativistic velocities. However, the first devices of

this type had too low efficiency and bad signal quality which made their practical use impossible.

The situation changed in the 1980s after the creation of an ubitron, a free electron laser (FEL), and relativistic TWT-BWO. These devices (together with gyrotrons) allowed the filling of the so-called "terahertz dip" by generating power from a few mW to tens of W at frequencies of 0.1–1 THz.

Unfortunately, FEL and relativistic TWT-BWT are not essentialy devices but rather facilities that together with the electron accelerator occupy a large production building. They cannot be used on moving platforms. Therefore the task of creating powerful sources of the coherent radiation in the range of 0.3–3 GHz is still relevant.

1.3 Semiconductor Microwave Electronics

In 1947, W. Shockley, D. Bardeen and W. Brattain in the Bell laboratory created the first **bipolar transistor**. From this moment the rapid development of semiconductor electronics began. However, the Brattain transistors had a limiting amplification frequency not exceeding a few MHz.

The requirements for the microminiaturization of radioelectronic aids and the increase in their operating frequency resulted in the development of a number of semiconductor devices and circuits based on them.

Initially these were passive diode elements, in particular, detector and mixer diodes on the Schottky barrier produced using a new technology, as well as varactors and *p-i-n* diodes. On this basis, circuits were created that converted the signal's spectrum: frequency rectifiers, mixers and multipliers, as well as switches, limiters and phase shifters. During this period of time, the creation of active elements for the amplification and generation of microwave oscillations on transistor basis was impossible because of technological difficulties in the fabrication of structures with micron-sized active areas. The way of creating active (amplifying and generating) devices based on diode structures having negative dynamic resistance in the microwave range appeared to be simpler. One of the first devices designed for these purposes was the tunnel diode created by L. Esaki in Japan in 1957.

In 1958, W.T. Read showed that a diode operating under avalanche breakdown conditions can have negative dynamic resistance and, consequently, can be used to generate or amplify electromagnetic oscillations. In 1959 A.S. Tager, A.I. Melnikov and others experimentally discovered the effect of microwave radiation generation at the avalanche breakdown of a semiconductor diode. This was the invention of an impact avalanche and transit diode (IMPATT—IMPact Avalanche and Transit Time). The IMPATT generators allowed hundreds of times more power than tunnel diodes.

In 1963, J. Gunn discovered the effect of negative dynamic resistance appearance in GaAs samples, and the creation of devices based on diode structures with negative differential mobility (Gunn diodes) began. Another name for such devices is devices with intervalley electron transfer, which reflects the physical principle of their

operation. The promising prospects of a simple diode using arsenide-gallium structures led to an unprecedented worldwide increase in the number of studies and developments in this field. However, the practice of using these devices revealed instability of operation, the need to withstand rigid tolerances for manufacturing technology, and the complexity of circuit solutions. All this led to a decrease in interest in Gunn diodes against the background of progress in the creation of microwave transistors. Nevertheless, Gunn diodes are still used today to generate microwave oscillations, especially in the millimeter wavelength range. Prospects of advancement into the terahertz range are opened for Gunn diodes with the use of new materials such as GaN.

In the early 1970s, due to the progress of semiconductor manufacturing technology it became possible to create the first transistors in the lower part of the microwave range. However, the exponential growth in the production of these devices and integrated circuits based on them became possible with the use of a new, more expensive material, gallium arsenide, instead of the widely used cheap silicon.

In the late 1970s, the metal semiconductor field-effect transistors (MESFET) began to be commercially manufactured. The history of the birth and application of MESFET is an example of a discovery that was much ahead of its time. Invented in 1930, it experienced a rebirth in the 1970s. It was the appearance of MESFET that made it possible to create amplifiers and generators in the microwave range. Further progress in the creation of not only discrete devices but also monolithic integrated circuits was based on the use of AlGaAs heterostructures. In the 2000s, successes in the technology of submicron heterostructures created on wide-gap semiconductors allowed the creation of devices operating in the upper part of the microwave range of 100–200 GHz. Such materials include silicon carbide SiC, gallium nitride GaN, and diamond C. It is important to note that the use of new materials and reduction of the active area dimensions to sizes of 0.05–0.1 μm led to the possibility of obtaining the vacuum (ballistic) carrier transfer similar to the transfer process in vacuum devices. The appearance of field-effect transistors with high electron mobility (HEMT) in the 2000s allowed the creation of devices operating in the upper part of the microwave range of 100–200 GHz, and the use of new wide-band materials (GaN, SiC) increased the output power of semiconductor amplifiers and generators in the microwave range by an order of magnitude. Currently GaN HEMT, capable of putting out power up to 100 W in the frequency range up to 10 GHz (Cree Company, USA) are being developed. These transistors can be used to design amplifiers with an output power of up to several kW.

1.4 Comparative Characteristics of Vacuum and Semiconductor Devices

At present both vacuum and semiconductor devices are widely used in microwave devices and systems. The optimal choice of a particular device is determined, first of all, by the required power and frequency. Figure 1.2 shows the attained power levels of different microwave device classes versus operating frequency. The

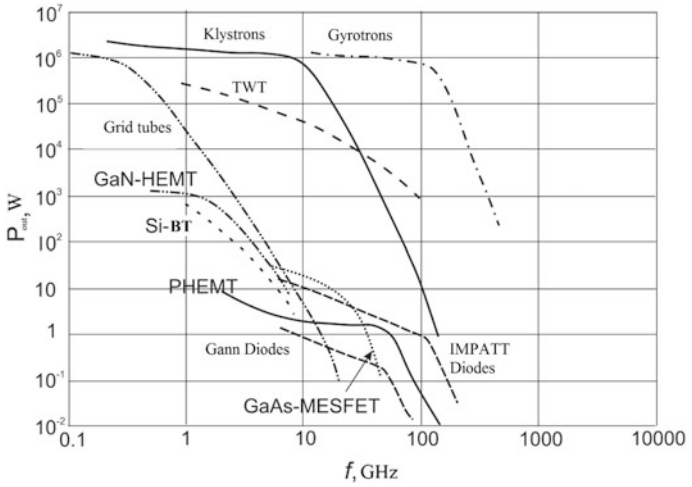


Fig. 1.2 Attained power level of different microwave device classes

notations on the figure are: GaN HEMT—GaN field-effect transistors with high electron mobility, Si BT—silicon bipolar transistors, PHEMT—pseudomorphic field-effect transistors with high electron mobility, GD—Gunn diode, GaAs MESFET—Schottky barrier gallium-arsenide field-effect transistors, and TWT—the traveling-wave tube. Apparently, vacuum devices (klystrons, TWT, gyrotrons) are significantly (by several orders) ahead of semiconductor devices, both in maximum output power and maximum operating frequency.

The main application areas of vacuum devices are radar station transmitters, transmitters of high-speed communication lines, power systems for charged particle accelerators, plasma heating in thermonuclear reactors, and microwave technology facilities. Progress in the development of microwave semiconductor devices is associated, first of all, with the rapid development of radar stations with Active Electronically Scanned Arrays (AESA) and mobile communication systems.

Currently, powerful GaN-based transistors are beginning to replace vacuum devices in the output cascades of AESA radar transmitters since such transmitters do not require much power—it is obtained by adding the powers of hundreds and even thousands of individual radiators in the radio beam.

1.5 Prospects for the Development of Microwave Electronics

At present, the development of microwave electronics is proceeding in several directions. The struggle for mastering ever higher frequencies—the terahertz range—continues. The task here is to develop small-sized powerful sources of coherent radiation,

by promoting "classical" devices—klystrons, TWT, inductive output tubes—to higher frequencies area, which requires the use of completely new design solutions and manufacturing techniques. The parameters of new types of devices (gyrotrons, orotrons) are being improved both in the areas of increasing operating frequency and power, and in increasing the efficiency, reliability and quality of the oscillations spectrum.

At a low power level, the same problem is solved using solid-state devices—resonant tunnel diodes, and Gunn diodes on GaN. From the optical range side, the greatest achievements in operating frequency reduction are associated with the development of quantum cascade lasers.

The areas of preferential use of vacuum and solid-state microwave devices are shown in Fig. 1.2, where the frequency is plotted along the abscissa axis and the average device power is plotted along the ordinate axis. With the development of microwave electronics the curves characterizing the level achieved in vacuum and solid-state electronics are shifting up and to the right. At the same time, the parameters of existing devices improve, namely their efficiency, service life, radiation resistance, and noise properties. Microminiaturization of microwave equipment, the development of group technologies for the manufacture of both individual devices and facilities based on them such as transmit-receive modules, phase shifters, and protective devices is happening rapidly. Simultaneously, the labor intensity of device production and associated costs are reducing.

Chapter 2
Interaction of Charged Particles with an Alternating Electromagnetic Field

2.1 Radiation of Individual and Collective Charged Particles

Individual radiation. As was mentioned in the Introduction, the principle of any electronic device operation is based on the interaction between charged particles (CPs) and the high-frequency electromagnetic field, accompanied by an exchange of energy between them. Electrons are used as CPs in vacuum electronic devices and free electrons and holes in semiconductor devices. Hereinafter, the word "electron" will often be used instead of the term "charged particle".

Let us consider the process of electromagnetic energy emission by a single charged particle. Electrodynamics states that if a CP moves rectilinearly and uniformly in an inertial coordinate system, it does not radiate. However, if it moves in electric field with deceleration, its kinetic energy decreases. The excess energy is radiated in the form of quanta of electromagnetic radiation (photons). This process is called *braking or slowing-down radiation*. The radiation power is defined by the formula

$$P = \frac{1}{4\pi\varepsilon_0}\frac{2q^2a^2}{3c^3}, \tag{2.1}$$

where q is particle charge, and a is its deceleration. When a particle accelerates, it gets energy from the electromagnetic field by absorbing photons.

Another mechanism of radiation arises when a CP moves near an ideally conducting surface. Consider, for example, an electron moving perpendicular to the surface (Fig. 2.1). At a given time, together with its mirror image, it forms a dipole with an electric moment $p = 2ql$, where l—distance to the surface.

The distance l varies with time, leading to change in the electric moment. A dipole with a variable moment is a radiator whose emission power, in accordance with classical electrodynamics, is expressed by the formula

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_2

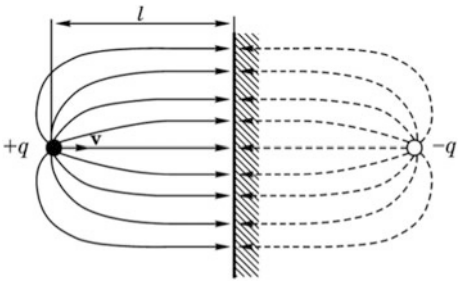


Fig. 2.1 Analysis of transition radiation

$$I(\omega) = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{2\pi c}\left[\frac{1+\beta^2}{\beta}\ln\frac{1+\beta}{1-\beta} - 2\beta\right]\mathrm{d}\omega, \tag{2.2}$$

where $\beta = v/c$ is the relative velocity of the electron. Full radiation energy according to this formula is infinite. Really finite conductance σ of the surface imposes an upper limit on frequency $\omega_{\max}\varepsilon_0 \ll \sigma$ and full radiation energy becomes finite.

This process is called *transition radiation*. As can be seen, this is a short pulse that occurs when the electron hits the surface (when the dipole collapse). Note the difference between the two types of radiation considered. If the power of the braking radiation is proportional to the square of the deceleration, then the power of the transition radiation is proportional to the velocity for small β.

Braking and transition radiation are used in devices with electrostatic control, klystron type devices and most semiconductor devices.

Radiation also arises when a CP moves uniformly with a velocity exceeding the speed of light in a given medium. This is the so-called *Vavilov-Cherenkov radiation* or *Cherenkov radiation*. Consider an electron moving at a constant velocity v in a medium with a refractive index n_m in the z-direction. The moving electron creates a current with a density

$$j_z = ev\delta(x)\delta(y)\delta(z - vt). \tag{2.3}$$

Applying the Fourier integral to (2.3), we obtain

$$j_z(\omega) = \frac{e}{\pi}\delta(x)\delta(y)\mathrm{e}^{-\mathrm{i}\beta_e z},$$

where $\beta_e = \omega/v$—"electronic" phase constant. The recorded expression describes a current wave propagating along the z axis. This wave excites the electromagnetic wave propagating at the speed of light in the surrounding space $u = c/n_m$. In the cylindrical coordinate system ρ, φ, z, the propagation vector of this wave has two components owing to the axial symmetry of the problem:

$$k_z = \beta_e \quad \text{and} \quad k_\rho = \sqrt{k^2 - \beta_e^2} = \beta_e\sqrt{(v/u)^2 - 1}.$$

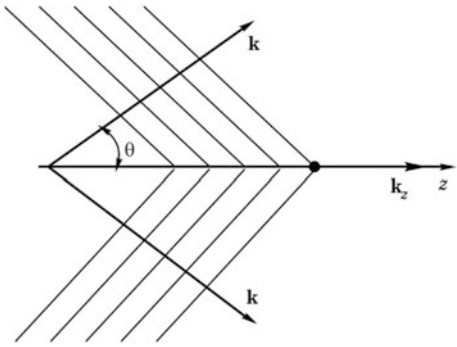


Fig. 2.2 Cherenkov radiation

If $v/u > 1$, the radial wave number k_ρ is real and the conical wave actually diverges from the electron motion trajectory (Fig. 2.2). The wave vector forms angle θ with the z axis, defined by the expression

$$\cos\theta = \frac{\beta_e}{\sqrt{k_\rho^2 + k_z^2}} = \frac{u}{v}.$$

If $v/u < 1$, k_ρ is the imaginary value and radiation is absent.

The energy radiated during the time interval (t_1, t_2) can be calculated using the formula

$$W = \int_V \mathrm{d}V \int_{t_1}^{t_2} j_z E_z \mathrm{d}t,$$

where E_z—the longitudinal component of the electric field of the wave, V—the volume occupied by the current. The radiation energy in the frequency band $\Delta\omega$ from the unit path length of the particle is given by the following

$$\Delta W = \frac{e^2}{4\pi\varepsilon c^2}\left(1 - \frac{u^2}{v^2}\right)\omega\Delta\omega.$$

A similar effect occurs when a CP moves near a periodic structure. In this case, the electric current wave excites in the surrounding space electromagnetic field that can be represented in the form of an expansion in terms of spatial harmonics with longitudinal wave numbers

$$k_{zn} = \frac{\omega}{v} + \frac{2\pi n}{D}, \quad n = 0, \pm 1, \pm 2, \ldots,$$

where D—the period of the structure. By entering angle $\theta_n = \arccos(k_{zn}/k)$, we find that if this angle assumes a real value for certain n, radiation is observed at the

n-th spatial harmonic. This radiation is called the *Smith-Purcell radiation* or *variotron* radiation. Cherenkov radiation is used in such devices as traveling-wave tubes, whereas Smith-Purcell radiation is used in backward-wave tubes.

Note that the presence of the medium in which or near which the charged particle moves is a necessary condition for the existence of Cherenkov radiation or variotron radiation. Therefore, this radiation can be regarded as radiation of the medium in which a passing particle influences the motion of its own charged particles.

Finally, another type of individual CP radiation, termed *oscillator radiation*, is produced when a particle is in periodic motion—vibrations and (or) rotation. It can be regarded as a kind of slowing-down radiation, since the particle accelerates and decelerates during periodic motion.

Consider an electron moving along the z axis at velocity υ and simultaneously oscillating in a plane perpendicular to this axis at velocity $v_\perp$ and angular frequency ω_0. Such an electron can be regarded as an oscillator with frequency ω_0, radiating due to a change in its total energy, which is composed of the energy of longitudinal and transverse motion. In this case, the radiation occurs at the following frequency

$$\omega(\theta) = \frac{\omega_0}{|1 - (v_z/u)\cos\theta|}, \tag{2.4}$$

which depends on the angle between the z axis and the straight line directed from the observation point to the electron. This dependence is called the *Doppler effect*. A normal Doppler effect is observed when $(v_z/u)\cos\theta < 1$. Radiation in this case occurs due to a decrease in the energy of the transverse motion of the particle. An anomalous Doppler effect is observed when $(v_z/u)\cos\theta > 1$. The oscillator in this case increases the energy of the transverse motion. The radiation energy and the increase in transverse motion energy are drawn from the longitudinal motion energy of the particle. Note that when the denominator value in (2.4) is small, the radiation frequency can exceed the frequency of the oscillator several times. This is important to create generators of millimeter and submillimeter sub-bands, as well as in optical and X-ray bands.

A normal Doppler effect is used in cyclotron resonance masers (CRM) and an anomalous Doppler effect is used in free electron lasers (FEL).

Any radiation is formed in a certain region (the formation zone), which is characterized by a coherent addition of radiations from individual points. The Vavilov-Cherenkov radiation has an infinitely extended zone of formation. Transition radiation, slowing-down radiation and oscillatory radiation are characterized by a finite formation zone, which is bounded by points, the radiation from which has a phase shift equal to π radians. The length of the formation zone l_p is maximum for radiation directed along the particle path:

$$l_p = \gamma^2 \lambda,$$

where λ is the radiation wavelength, and $\gamma = (1 - v^2/c^2)^{-1/2}$ is the relativistic factor.

This analysis of individual CP radiation does not consider the retroactive effect of the field on the particle and, therefore, can only be regarded as approximate.

Collective radiation. The radiation of one charged particle is too weak to be used for practical purposes. To increase the radiation intensity, it is necessary to use a number of individual radiators, or as they say, an ensemble of particles. Depending on the population of the ensemble, its density and size, different types of interactions between individual particles can be observed.

Consider, for specificity, an ensemble of particles, part of which has energy W_1, and another part—W_2, whereby $W_1 > W_2$. As they say, particles are located on two energy levels, the upper and the lower levels. When passing from the upper level to the lower level, the particle emits a radiation quantum, and when it tranfser in reverse direction it absorbs the quantum. Let us denote the particle lifetime at the upper level as τ_1. Upon *spontaneous radiation* the particles shift to the lower level independently. Therefore, the radiation power is proportional to the number of particles N: $P = N\hbar\omega_0/\tau_1$ where $\hbar\omega_0$—quantum energy. The radiation in this case is not coherent, as each particle emits in its phase and is distributed evenly in all directions (Fig. 2.3a). Spontaneous transitions also occur from the lower level to the upper level absorbing field energy.

As the concentration of particles increases, spontaneous correlation arises between the radiators due to the exchange of quanta. If the correlation time $\tau_c < \tau_1$, the ensemble can shift to the lower level in time $\tau_c \sim N^{-1}$. As a result, the radiation power

$$P = N\hbar\omega_0/\tau_c \sim N^2\hbar\omega_0$$

is proportional to the square of the number of particles. In addition, the spatial distribution of the radiation changes (Fig. 2.3b) and it becomes coherent. Thus, the

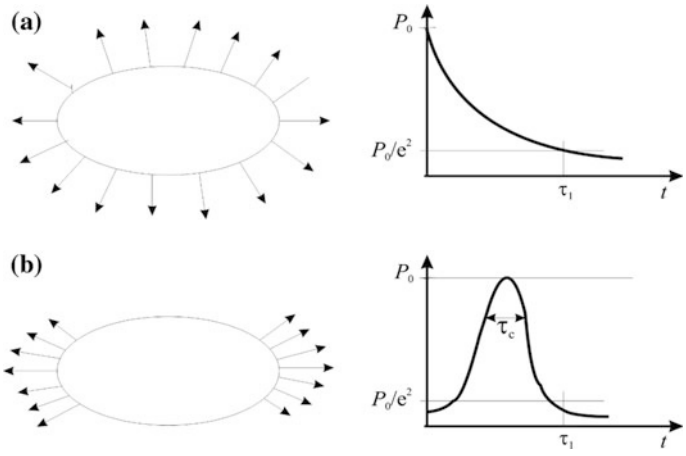


Fig. 2.3 Spontaneous radiation (**a**) and Dicke super radiance (**b**)

radiation intensity increases sharply. This phenomenon is called the *Dicke super radiance*. It is used to create super-powerful pulses of microwave radiation.

Induced radiation. Under the influence of external radiation, the oscillators emit at the same phases (phasing of the oscillators occurs). Such external radiation can also be the radiation of the particles themselves, if it does not wander, but confine in a finite volume (resonator or transmission line) containing an ensemble of particles. To generate induced radiation, it is required, among other conditions, that the maximum ensemble size does not exceed half the wavelength of the radiation. Induced radiation is used in most microwave electronic devices. To implement it, it is necessary to create a small CP ensemble (to perform spatial bunching) and (or) to ensure the phasing of individual oscillators (phase bunching). Electronic devices differ according to types of individual CP radiation and bunching mechanisms they use.

2.2 Macroscopic Equations of Microwave Electronics

A huge number of charged particles used to amplify or generate microwave radiation can be considered as a continuous charged medium. The limit of the ratio of the number of particles of given type ΔN contained in volume ΔV, to this volume when the latter tends to zero, is called the *concentration* of particles $n = \lim_{\Delta V \to 0}(\Delta N/\Delta V)$. In this and other similar equations, a physically infinitesimal volume should be considered as a volume that is small enough in comparison with the characteristic dimensions of the problem. At the same time, it must be large enough to contain in it the number of particles sufficient to ignore statistical fluctuations of their number.

Charge density is determined using a similar equation:

$$\rho(\mathbf{r}, t) = \lim_{\Delta V \to 0}\left(\Delta Q/\Delta V\right).$$

Charge ΔQ, in volume ΔV, is defined using the formula $\Delta Q = q\Delta N$, where q—particle charge. Consequently,

$$\rho(\mathbf{r}, t) = qn(\mathbf{r}, t). \tag{2.5}$$

Suppose that all particles in volume ΔV have the same velocity $\mathbf{v}(\mathbf{r}, t)$. This assumption is called the *hydrodynamic approximation*. The charge, carried across the surface element $\Delta\mathbf{S}$ during period δt is

$$\delta Q = \rho\Delta\mathbf{S} \cdot \mathbf{v}\delta t.$$

Hence, the current passing through the surface element is given as $\Delta I = \delta Q/\delta t = \rho\mathbf{v} \cdot \Delta\mathbf{S}$, and current density as $\mathbf{J} = \rho\mathbf{v}$. This current is called *convective*, since it is due to the motion (convection) of charged particles.

Charged particles in electronic devices emit and absorb photons. A large number of photons can be considered as a continuous substance—electromagnetic field. In this approximation, the laws of classical electrodynamics, that is, Maxwell's equations, describe the electromagnetic field:

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}; \tag{2.6}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0; \tag{2.7}$$

$$\nabla \cdot \mathbf{D} = \rho; \tag{2.8}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{2.9}$$

We can see from (2.6) and (2.8) that the sources of electromagnetic fields are electric currents and charges, i.e. the ensemble of charged particles.

Calculating the divergence of the both sides of (2.6), and substituting the result in (2.8), we obtain the *continuity equation*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0. \tag{2.10}$$

Integrating this equation over an arbitrary volume V, bounded by surface S, we find

$$\frac{\mathrm{d}Q}{\mathrm{d}t} + I = 0, \tag{2.11}$$

where $Q = \int_V \rho \mathrm{d}V$—electric charge contained in volume V, $I = \oint_S \mathbf{J} \cdot \mathrm{d}\mathbf{S}$—electric current flowing through surface S. Obviously, (2.11) expresses the law of charge conservation. Hence, this law is contained in Maxwell's equations (2.6)–(2.9).

Maxwell called the second term on the left-hand side of (2.6) *displacement current density* $\mathbf{J}_d = \partial \mathbf{D}/\partial t$. Using this notation, (2.6) can be rewritten as follows:

$$\nabla \times \mathbf{H} = \mathbf{J}_d + \mathbf{J} = \mathbf{J}_{\mathrm{tot}},$$

where $\mathbf{J}_{\mathrm{tot}}$—total current density, which is equal to the sum of the displacement current density and the convection current density. Calculating the divergence of both sides of this equation, we obtain

$$\nabla \cdot \mathbf{J}_{\mathrm{tot}} = 0. \tag{2.12}$$

This equation is called the *total current law* (Ampere's circuital law). It contends that the total current density has no source, i.e. the total current lines are either closed or begin at infinity and go to infinity.

2.3 Motion Equations of Charged Particles

2.3.1 Motion of a Single Particle in Vacuum

The motion of a charged particle in vacuum is described by the Newtonian equation

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{2.13}$$

where $\mathbf{p} = m\mathbf{v}$ is the particle impulse (momentum), and $q, m, \mathbf{v}$ are its charge, mass and velocity. The right-hand side of the equation is the force with which the field acts on the CP (Lorentz force). The Lorentz force is created by a complete field acting on the particle. This field is the sum of the external electromagnetic field relative to the given particles and the field created by the particles themselves. The total field, acting on the particles, changes their positions and velocities, leading to an amendment in the field radiated by these particles. Thus, in electronics it is necessary to solve the so-called *self-consistent problem*, taking into account the mutual influence of the field and charged particles.

In accordance with the relativistic theory the mass of the particle

$$m = \frac{m_0}{\sqrt{1 - (v/c)^2}}, \tag{2.14}$$

where m_0 is the mass of the particle at rest, v is its velocity in the inertial coordinate system, and c = 2.998 × 10^8 m/s is the speed of light in vacuum. The relativistic increase in the mass of electrons plays an important role in powerful and super-powerful microwave devices operating at high accelerating voltages (more than 30 kV).

The work done by the electromagnetic field when the particle is moved by distance d$\mathbf{l}$ is

$$\mathrm{d}A = \mathbf{F}\mathrm{d}\mathbf{l} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}\mathrm{d}t = q\mathbf{E} \cdot \mathbf{v}\mathrm{d}t.$$

It can be seen that the magnetic field does not perform work on the particle. The power of particle interaction with the electromagnetic field

$$P = \mathrm{d}A/\mathrm{d}t = e\mathbf{E} \cdot \mathbf{v} \tag{2.15}$$

can be either positive (the field gives part of its energy to the particle) or negative (the particle gives energy to the field). The latter is used in microwave electronic devices.

Let small volume ΔV contain the charge $\Delta q = \rho \Delta V$. The power of interaction of this charge with the electromagnetic field

$$\Delta P = \rho \Delta V \mathbf{E} \cdot \mathbf{v}.$$

Dividing the left- and right-hand sides of this equation by ΔV, we obtain the specific interaction power of the current with the electromagnetic field

$$p_i = \mathbf{J} \cdot \mathbf{E}. \tag{2.16}$$

Consider the inner product of (2.13) and $\mathbf{v}$. Neglecting the relativistic dependence of the mass on the velocity, we can write

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \cdot \mathbf{v} = e\mathbf{E} \cdot \mathbf{v} + e(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}.$$

The second term on the right-hand side of this equation is zero, since the two vectors in the vector-scalar product have the same directions. After simple transformations of the remaining terms we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(W_k + W_p\right) = 0, \tag{2.17}$$

where $W_k = mv^2/2$ is the kinetic energy of the particle, $W_p = q\Phi$ is its potential energy, and $\Phi = -\mathbf{E} \cdot \mathbf{r}$ is the field potential at the location of the particle. Equation (2.17) expresses *the energy conservation law* for charged particles moving in electromagnetic field. The energy conservation law makes it possible to determine the velocity of a particle moving in electrostatic field. Let the initial velocity of the particle be zero and the electrostatic potential at the location of the particle equal Φ_1. The electric field causes the particle to start moving and after a while, it falls into a point, where the potential is equal to Φ_2. The change in the potential energy of the particle along this path equals $\Delta W_p = q(\Phi_1 - \Phi_2) = qU$, where $U = \Phi_1 - \Phi_2$ is the the potential difference (voltage) between the start and end points. From the energy conservation law (2.17) it follows that the kinetic energy of the particle must change by the same amount (with an opposite sign). Since at the initial point, the kinetic energy of the particle is zero, we obtain an equation to determine the velocity of the particle in accordance with relativistic correction:

$$\frac{m_0}{2\sqrt{1 - v^2/c^2}}v^2 = qU.$$

Solving this biquadratic equation, we obtain

$$v = \sqrt{2\frac{qU}{m_0}}\left(\frac{\gamma + 1}{2\gamma^2}\right)^{1/2}, \tag{2.18}$$

where $\gamma = 1 + |qU|/(m_0c^2)$ is the relativistic factor. For electrons $(q = -e, m_0 = m_e)$

$$v = 5.95 \times 10^5 \sqrt{U}\left[(\gamma+1)/(2\gamma^2)\right]^{1/2}, \quad \gamma = 1 + 1.96 \times 10^{-6} U. \tag{2.19}$$

Figure 2.4 shows the dependence of electron velocity on the accelerating voltage. The relativistic correction becomes significant for accelerating voltages greater than 25 kV.

During the joint solution of Maxwell's equations and the equation of motion, it is necessary to take into account that Maxwell's equations are written in so-called *Euler variables*, in which coordinates and time are independent. The equation of motion (2.13) is written in *Lagrange variables*, that is for a particular particle with coordinates depending on time. For the transition from Lagrange variables to Euler variables, it is necessary to consider $\mathbf{p}(\mathbf{r}, t)$ as a complex function of time:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \frac{\partial \mathbf{p}}{\partial t} + \frac{\partial \mathbf{p}}{\partial \mathbf{r}}\frac{\partial \mathbf{r}}{\partial t} = \frac{\partial \mathbf{p}}{\partial t} + (\mathbf{v}\nabla)\mathbf{p}. \tag{2.20}$$

2.3.2 The Particles Ensemble Motion in Vacuum

In many cases it is necessary to take into account that for a given element volume at any given time, particles have different velocities (pulses). We shall consider the motion of a particle aggregate (ensemble). If particles of one type only are considered, then a *one-particle distribution function* $f(\mathbf{r}, \mathbf{p}, t)$ is introduced. This function determines the number of particles in the six-dimensional phase space element $\mathrm{d}n(t) = f(\mathbf{r}, \mathbf{p}, t)\mathrm{d}V\mathrm{d}P$ at time t. The phase space element is determined by the volume of the space $\mathrm{d}V(\mathbf{r})$ located around the point with radius vector $\boldsymbol{r}$ and the volume of impulse space $\mathrm{d}P(\mathbf{p})$, containing particles having moments close to impulse $\boldsymbol{p}$. It is obvious that the integral of the distribution function over all possible impulses should be equal to the number of particles per unit volume, i.e. their concentration:

$$\int_p f(\mathbf{r}, \mathbf{p}, t)\mathrm{d}P = n(\mathbf{r}, t).$$

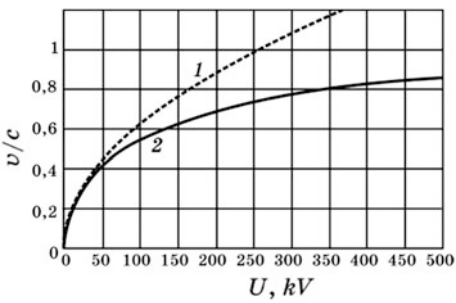


Fig. 2.4 Electron velocity versus accelerating voltage: 1—without relativistic correction; 2—with correction

The charge and current densities respectively, are determined by the following expressions:

$$\rho(\mathbf{r},t) = e\int_{p} f(\mathbf{r},\mathbf{p},t)\mathrm{d}P; \tag{2.21}$$

$$\mathbf{J}(\mathbf{r},t) = \frac{e}{m}\int_{P} \mathbf{p}f(\mathbf{r},\mathbf{p},t)\mathrm{d}P. \tag{2.22}$$

Consider, as an example, the so-called *Maxwellian distribution function*, which in several practically important cases corresponds to the *electron velocity distribution* in the electron beam in vacuum devices with a thermal spread of velocities:

$$f(v) = n\left(\frac{m}{2\pi kT_e}\right)^{3/2} \mathrm{e}^{-\frac{mv^2}{kT_e}}, \tag{2.23}$$

where T_e is the so-called electron temperature, which characterizes the average energy of the thermal motion of electrons.

Let us formulate the equation of motion (transfer) for an ensemble of one kind of particles in vacuum. To do this, we use Liouville's theorem, which asserts that the distribution function remains constant along any trajectory in the phase space, i.e.

$$\frac{\mathrm{d}f(\mathbf{r},\mathbf{p},t)}{\mathrm{d}t} = 0. \tag{2.24}$$

Using the rule for finding the derivative of a function of many variables, we transform (2.24) into:

$$\frac{\mathrm{d}f(\mathbf{r},\mathbf{p},t)}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{r}}\frac{\partial \mathbf{r}}{\partial t} + \frac{\partial f}{\partial \mathbf{p}}\frac{\partial \mathbf{p}}{\partial t} = 0. \tag{2.25}$$

We rewrite (2.25), taking into account the relations

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}, \quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F},$$

where $\mathbf{v}$ is particle velocity in a phase volume element, and $\mathbf{F}$ is the external force acting on these particles. As a result, we get:

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_r f + \mathbf{F}\cdot\nabla_p f = 0, \tag{2.26}$$

where the indices r and p denote operators acting on coordinates and projections of the impulse respectively.

If we add the action force of the self-consistent field, which depends on the distribution function, to force field **F**, we obtain the *Vlasov equation*, describing the motion of particles in a field equal to the sum of the external field and the field created by the particles themselves. This approach is used to analyze the motion of particles in vacuum electron devices in the presence of a thermal spread of velocities in an electron beam.

The terms of (2.26) have a simple physical meaning. The distribution function changes in time $(\partial f/\partial t)$ due to the escape of particles from the element volume $\mathbf{v} \cdot \nabla_r f$ (in the presence of a spatial gradient in the distribution of particles) and the change in momentum under the action of an external force $\mathbf{F} \cdot \nabla_p f$. If there are several types of particles in the space, the equation (2.26) must be written separately for each kind of particles.

2.3.3 The Particles Ensemble Motion in Solid

Contrary to the motion of particles in a vacuum, when analyzing the transport processes in a solid, it is necessary to take into account the processes of particle scattering, recombination, and generation. These phenomena are accounted for by introducing the so-called collision term $(\partial f/\partial t)_c$ in the right-hand side of (2.26):

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \mathbf{F} \cdot \nabla_p f = \left(\frac{\partial f}{\partial t}\right)_c. \tag{2.27}$$

This equation is called the *Boltzmann kinetic equation* (or Boltzmann transport equation (BTE)).

The use of the Boltzmann equation for the analysis of the interaction of the charged particles flow and the field is a rather complicated problem. In practical work, it is important to have equations coupling the integral parameters of the process, such as the average drift velocity, electric current density and energy. In order to obtain such equations, we perform the transformation of (2.27). Take into account the fact that the instantaneous particle velocity consists of two components: a directed drift velocity $\mathbf{v}_d$ and thermal (chaotic) components $\mathbf{u}$, i.e. $\mathbf{v} = \mathbf{v}_d + \mathbf{u}$. Thermal velocity averaged in time is equal to zero, and the velocity distribution is given, for example, by (2.23). Integrating it over the momentum space, we obtain the *particles number conservation law*:

$$\partial n/\partial t + \nabla(n\mathbf{v}_d) = 0, \tag{2.28}$$

where n is the concentration of charged particles, $\mathbf{v}_d = \mu\mathbf{E}$ is the drift velocity of the particle, i.e. the average velocity of its motion under the action of the electric field, μ is the mobility of the particles, which can also be a function of the electric field strength.

Using the relaxation time approximation (see Appendix A), and also taking into account the processes of generation and recombination of charge carriers, we obtain a generalized *particles number conservation equation*:

$$\partial n/\partial t + \nabla(n\mathbf{v}_d) = G - R, \tag{2.29}$$

where G is the particle generation rate, and R is the rate of their recombination.

Multiplication of (2.27) by $\mathbf{v}$ with subsequent integration over the velocity space gives us (after a complicated mathematical transformation) the transport equation or the *density of momentum flow* $\widehat{P} = nm\,\mathbf{vv}$ *equation*:

$$\frac{\partial(nm\mathbf{v}_d)}{\partial t} + \nabla \cdot \widehat{P} - ne(\mathbf{E} + \mathbf{v}_d \times \mathbf{B}) = n\mathbf{F}_{\text{ir}}, \tag{2.30}$$

where $\nabla \cdot \widehat{P}$ is the spatial divergence of the momentum flow, $n\mathbf{F}_{\text{ir}}$ is the effective internal friction force, which reflects the change in momentum due to scattering events, $\mathbf{vv}$—dyadic product of two vectors, called a *dyad*.

Multiplying (2.27) by v^2 and integrating over the entire velocity space gives the *energy conservation equation*:

$$\partial(nW)/\partial t + \nabla q \cdot \mathbf{p} - \mathbf{J} \cdot \mathbf{E} = nQ, \tag{2.31}$$

where $W = mv_d^2/2 + 3kT_e/2$ is the total energy of the particle, formed from the energy of the directed drift of the particle $mv_d^2/2$ and the thermal energy $3kT_e/2$, $\nabla\mathbf{q}$ is the spatial divergence of the energy flux $\mathbf{q} = W\mathbf{v}_d$, $\mathbf{J} \cdot \mathbf{E}$ is Joule heat energy, and Q is the change in energy due to scattering processes.

The obtained equations are called *conservation equations*. The similarity of the forms of the three equations obtained should be noted. They allow us to find the required integral parameters for the motion of carriers in a given medium. Parameters of the medium are hidden in these equations in their right-hand parts G, R, F_{ir}, Q, reflecting the specific nature of the scattering events.

Finding an explicit functional dependence of these values on the parameters of the material and external conditions is a difficult task. In the next section, we will consider approximate solutions of these equations, using parameters of the material, in which the charge carriers move, and the relaxation time approximation for describing the scattering processes.

In many modern nanoelectronic devices, the size of the region where particles interact with an electromagnetic field is so small that it is necessary to take into account the quantum properties of charged particles. In these cases, a wave function is introduced for the particle probability distribution $\Psi(\mathbf{r}, t)$ that satisfies the Schrödinger equation. In the non-relativistic approximation, this equation has the form

$$i\hbar\frac{\partial\Psi}{\partial t} + \frac{\hbar^2}{2m}\nabla^2\Psi - U(\mathbf{r}, t)\Psi = 0 \tag{2.32}$$

where $U(\mathbf{r}, t)$ is the potential energy of the field external to the particle at the point $\mathbf{r}$. The square of the Ψ function module determines the probability of finding a particle in the given location at the given time.

2.4 Material Parameters and Relaxation Processes

The processes of amplification or generation of electromagnetic field occur in a material medium that fills the interaction space. Charged particles in this space can be divided into two types—free and bound. Free particles are not coupled with atoms and can travel through macroscopic distances. Bound particles are part of the composition of atoms, ions, or molecules that fill the interaction space. Their position is fixed in space, but they can shift relative to their equilibrium positions under the action of the field. This leads to polarization and magnetization of the material. We describe these processes by the constitutional equations connecting field intensities with inductions.

In vacuum, the inductions are proportional to the field intensity:

$$\mathbf{D} = \varepsilon_0 \mathbf{E}; \quad \mathbf{B} = \mu_0 \mathbf{H}. \tag{2.33}$$

The values of the proportionality coefficients—the dielectric constant ε_0 and the magnetic constant μ_0—depend on the chosen system of units. In the Gaussian system of units, in particular, $\varepsilon_0 = \mu_0 = 1$, and there is no difference between the field intensity and induction. In the SI-system of units

$$\varepsilon_0 = 10^7/(4\pi c^2) \approx 8.86 \times 10^{-12} A \cdot c/(B \cdot M),$$
$$\mu_0 = 4\pi \times 10^{-7} \approx 1.256 \times 10^{-6}\, B \cdot c/(A \cdot M).$$

As can be noticed, in SI $\varepsilon_0 \mu_0 = c^{-2}$.

In the material medium, (2.33) are modified, taking into account polarization and magnetization processes:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 \varepsilon_r \mathbf{E}; \quad \mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) = \mu_0 \mu_r \mathbf{H}, \tag{2.34}$$

where $\mathbf{P} = \varepsilon_0 \kappa \mathbf{E}$ is the polarization vector of the medium, $\mathbf{M} = \chi \dot{\mathbf{H}}$ is the magnetization vector, κ and χ are the dielectric and magnetic susceptibilities. Thus, in Maxwell's equations, the bounded charges and currents are taken into account by introducing $\mathbf{D}$ and $\mathbf{B}$ vectors so that only free currents and charges appear on the right-hand sides of these equations.

When moving in a medium that fills the interaction space, the free CPs can collide with the atoms and ions of the matter located there. The mean free path, or the mean path length of a particle between two collisions characterizes this process. Obviously, this length depends on the concentration of atoms and ions in the medium. If the mean free path is much larger than the size of the interaction region,

most free CPs do not experience a single collision during their transit in the interaction space. Their motion is determined by (2.13). Such a medium is called a *technical-grade vacuum*. For its creation, vacuum electronic devices have a hermetic shell, from which the air is pumped out to a pressure of 10^{-6}–10^{-9} bar. Residual gases at such pressures do not practically influence the interaction processes, but can affect the service life of the device and its maximum achievable parameters.

At a higher gas pressure inside the device, the mean free path becomes camparable with its dimensions. In this case, due to collisions, the gas is ionized by charged particles and *gas-discharge plasma* can appear in the device. This state of matter is characterized by the presence of several varieties of CPs with charges of opposite sign, as well as neutral atoms, so that overall, the medium remains electrically neutral.

The term "gas-discharge plasma" can be used if in the equations of hydrodynamics describing the system, any arbitrary physically small time intervals are always greater than the relaxation time of the system to a local equilibrium distribution. In this case, the used volumes must not only contain a sufficient number of particles, but must also form quasi-homogeneous statistical subsystems.

Suppose that an excess of particles with a positive charge appeared in a certain volume (Fig. 2.5). Then particles with a negative charge move to this volume due to a force of attraction. As a result, the effect of screening excess charge with charges of the opposite sign arises. The radius of the volume, beyond which the negative charge field completely compensates the positive charge field, is called the *Debye shielding length*. The screening radius in plasma containing J types of charged particles is defined by the formula

$$r_d = \left(\sum_{j=1}^{J} \frac{q_j^2 n_j}{\varepsilon_0 k T_j} \right)^{-1/2}, \tag{2.35}$$

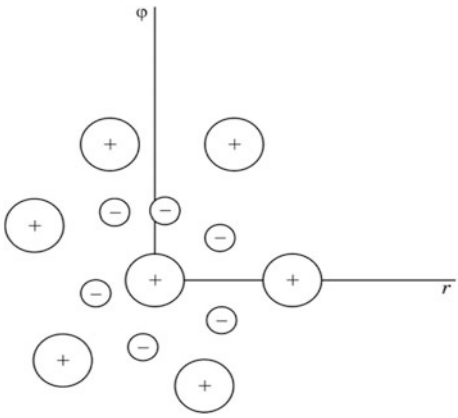


Fig. 2.5 The Debye screening

where q_j, n_j, T_j are the charge, concentration, and temperature of the j-th type of particles (see Appendix A). Please note that all charged particles, regardless of their charge, contribute to the Debye length.

For the existence of plasma, it is necessary that the number of charged particles is sufficient for screening, i.e., that $\boldsymbol{r}_D^3 \sum_{j=1}^{j} n_j \gg 1$. In addition, the internal interactions of particles must predominate over interactions on the boundary of the interaction space. This condition is expressed by the inequality $\boldsymbol{r}_D/L \ll 1$, where L is the maximum size of the interaction space. If the above conditions are satisfied, the collision term in (2.27) plays an important role and cannot be neglected in analyzing the motion of the CPs.

The behavior of charged particles ensembles in semiconductors is similar to the properties of gas-discharge plasma. Together with the ionized atoms of the impurity, they form the so-called *solid-state plasma*, to which many concepts of gas-discharge plasma are applicable, in particular, the Debye length, the mean free path, and some others (for more details see Appendix A). At the same time, the motion laws of charged particles in a solid differ significantly from the laws of their motion in a vacuum. These differences are caused by the following factors.

Free CPs in a solid move in a periodic field of the crystal lattice. The influence of this field in the first approximation can be taken into account by introducing the effective mass of a particle m^* different from its actual mass.

At a temperature greater than absolute zero the atoms of the crystal lattice oscillate, leading to collisions with free CPs and the latter acquire chaotic thermal velocities. At collision the magnitude and direction of the particle velocity vary randomly. These processes are called *scattering processes*. Velocities acquired by CPs under the action of an electric field are usually much smaller or comparable to thermal velocities.

A large number of collision (scattering) events accompanies the motion of charge carriers in a solid or gas plasma. The physical nature of the collision acts is very diverse. These include scattering by charged and neutral components of the medium, scattering by phonons, dislocations of the crystal lattice, electron-electron scattering, and so on. The stochastic character of these processes, the strong dependence on the field strength, particle energy, temperature, and concentration leads to the need to use the Boltzmann kinetic equation to describe the processes of current transfer. An important role here is played by the correct approximation of the collision term in this equation. Often, the relaxation time approximation is used, assuming that the rate of change in time of a given physical quantity is proportional to the difference between its current and certain equilibrium values. Mathematically, this is a differential equation of the form:

$$\frac{\partial a}{\partial t} = -\frac{1}{\tau}(a - a_0), \tag{2.36}$$

where $a(t)$ is any value characterizing the plasma, a_0 is its equilibrium value, and τ is a coefficient called the relaxation constant of a given quantity.

The solution of this equation has the form

$$a(t) = a_0 + (a(0) - a_0)\mathrm{e}^{-\frac{t}{\tau}} \tag{2.37}$$

One can see from the equation that in a time equal to the relaxation constant τ, the deviation from the equilibrium state decreases by 2.72 times.

Let us consider the processes of the relaxation of basic physical values such as the number of particles, their momentum and energy. During the motion of charges in semiconductors relaxation processes associated with the generation and recombination of charges occur. These processes are characterized by the *lifetime of the carriers*: electrons—τ_n and holes—τ_p. Devices working in the microwave band, must have a material with a lifetime of charge carriers much greater than the period of microwave oscillations T. For electrons this condition is written in the form of an inequality $\tau_n \gg T$. As for the length of the active interaction region L_a, it must be much smaller than the diffusion length $l_d = \sqrt{D\tau_n}$.

According to the results of measurements, the lifetime of a charge carrier in semiconductors has a value of the order of 10^{-6} s. This is substantially greater than the period of microwave oscillations.

If we assume that after the collision the particle completely loses its initial momentum, the process of impulse relaxation is described by (2.36) with the relaxation constant $\tau_i = t_c$, where t_c is the mean particle free time between two collisions. In solid-state plasma this time is 10^{-14}–10^{-13} s.

Despite the change in momentum, the energy of the particle after the collision varies insignificantly (almost elastic collisions), and many scattering events are required to allow the particle to lose its excess energy. The energy of the particle is characterized by the *electron temperature* $T_e = 2W/(3k)$. The change in the electron temperature is determined by an equation of the type (2.36):

$$\frac{\mathrm{d}T_e}{\mathrm{d}t} = -\frac{T_e - T_0}{\tau_\varepsilon}, \tag{2.38}$$

where T_0 is the equilibrium electron temperature, usually taken to be equal to the temperature of the crystal lattice, τ_ε is the energy relaxation constant. Typical values of τ_ε are 10^{-12}–10^{-10} s, or two orders of magnitude greater than the momentum relaxation time. Therefore, it is said that the memory for the momentum of electrons is much shorter than the memory for energy.

Equation (2.38) is considerably simplified. In reality, the parameters τ_i, τ_ε are a function of the electron temperature. In (2.38) a certain effective quantity is implied by τ_ε that approximates the energy change with sufficient accuracy. Relaxation processes are considered in depth in Appendix A.

The process of returning the electrons ensemble to an equilibrium state is illustrated in Fig. 2.6, where the equilibrium momentum distribution function of electrons is shown (curve f_0). Suppose that as a result of some impact on the ensemble, its distribution function takes the form f_1. Apparently, the average momentum of the particles (and, hence, their energy) increased, and the scatter in

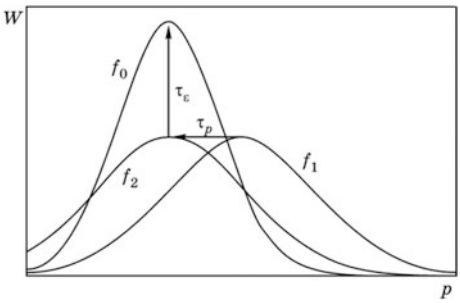


Fig. 2.6 The relaxation of the particles ensemble distribution function into an equilibrium state

the pulses increased (the distribution curve became wider). After the end of the impact, the relaxation processes begin. At first, momentum relaxation ends, as a result of which the mean value of the ensemble pulse becomes equal to the equilibrium value. The distribution curve takes the form f_2. After this, the process of energy relaxation continues, at the end of which the distribution function returns to the equilibrium form (curve f_0).

Along with the relaxation time, it is convenient to use the cooling or heating length of the carriers $l_\varepsilon = v_T \tau_\varepsilon$, where v_T is the average thermal velocity of the particles. At T = 300 K, the average value of the thermal velocity $v_T = 10^5$ m/s, hence the heating length $l_\varepsilon = 0.01{-}1$ μm, which is comparable to the sizes of microwave semiconductor device elements.

Taking into account the introduced relaxation parameters, the conservation equations (2.29)–(2.31) can be written in a simpler form:

$$\frac{\partial n}{\partial t} + \nabla(n\mathbf{v}_d) = -\frac{n - n_0}{\tau_n}; \tag{2.39}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\nabla)\mathbf{v} - \frac{1}{n}\nabla\left(\frac{kT_e}{m}n\right) - \frac{e}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = -\frac{\mathbf{v} - \mathbf{v}_0}{\tau_p}; \tag{2.40}$$

$$\frac{\partial(nW)}{\partial t} + \nabla \cdot \mathbf{p} - \mathbf{J} \cdot \mathbf{E} = -\frac{3}{2}\frac{nk(T_e - T_0)}{\tau_\varepsilon(T_e)}. \tag{2.41}$$

The expressions (2.39)–(2.41) are written for the electrons. Similar equations can also be written for holes. If we exclude relaxation processes, i.e. assume $\tau_n \to \infty, \tau_p \to \infty, \tau_\varepsilon \to \infty$, then (2.39)–(2.41) are transformed into equations for the vacuum:

$$\partial n/\partial t + \nabla(n\mathbf{v}) = 0, \quad nd(m\mathbf{v})/\mathrm{d}t = ne\mathbf{E}, \quad \partial(nmv^2/2)/\partial t = \mathbf{J} \cdot \mathbf{E}.$$

Let us analyze the system (2.39)–(2.41). First of all, let us consider the momentum conservation (2.40), which makes it possible to determine the drift velocity $\mathbf{v}_d$, and hence the current too. Suppose at a first approximation that the spatial

inhomogeneity of the electron temperature T_e and the field $\mathbf{E}$ can be neglected, and the magnetic field is absent. Then for the stationary regime $(\partial/\partial t = 0)$ we obtain:

$$\frac{kT_e}{m}\frac{\nabla n}{n} + \frac{e}{m}\mathbf{E} = \frac{\mathbf{v}}{\tau_p},$$

or

$$\mathbf{v} = -\frac{e}{m}\tau_p\mathbf{E} + \frac{kT_e}{m}\tau_p\frac{\nabla n}{n} = \mu\mathbf{E} - D\frac{\nabla n}{n}, \tag{2.42}$$

where $\mu = q\tau_i/m$ is the mobility of the charge carriers, and $D = kT_e\tau_i/m$ is the diffusion coefficient. These values are called the *kinetic coefficients*. In a system in thermodynamic equilibrium at a temperature T_0, they are related by the Einstein relation:

$$D = \mu kT_0/e \tag{2.43}$$

Accordingly, the concepts of drift velocity are introduced

$$\mathbf{v}_d = \mu\mathbf{E} \tag{2.44}$$

and diffusion speed

$$\mathbf{v}_{\text{dif}} = -D\nabla n/n. \tag{2.45}$$

Using (2.42), it is easy to obtain an expression for the current density:

$$\mathbf{J} = qn\mathbf{v}_d = \sigma\mathbf{E} - eD\nabla n, \tag{2.46}$$

where $\sigma = qn\mu = q^2n\tau_p/m$ is the specific conductivity of the medium, and $qD\nabla n$ is the diffusion current. Keep in mind that the formula (2.46) is obtained without taking into account the spatial derivatives, the role of which is very important in the microwave band, where the active regions of particles and field interaction are commensurable with the indicative spatial scales of energy relaxation.

A quantitative estimate of the drift velocity $\mathbf{v}_d$ can be found by the simultaneous solution of (2.40) and (2.41). For this, first of all, it is necessary to know the dependencies $\tau_i(T_e), \tau_\varepsilon(T_e)$ which can be calculated using appropriate mathematical models or determined experimentally. As a result, we obtain the dependence of the drift velocity on the electric field strength. This dependence is called the *field-velocity chart* (*FVC*).

Electron field-velocity charts $v_d(E)$, obtained experimentally for silicon and gallium arsenide, are presented in Fig. 2.7. Let us note the indicative features of the obtained dependences.

In a solid, the strength of the electric field determines the drift velocity of the particles, and not their acceleration as in vacuum (see 2.13). In the area of low fields, the velocity is proportional to the electric field strength, i.e. $v_d = \mu_0 E$. The

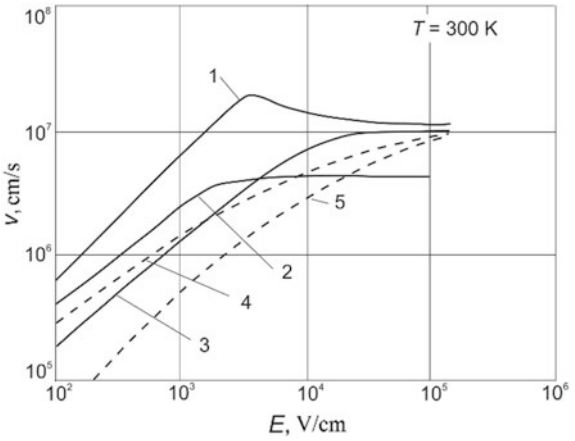


Fig. 2.7 Field-velocity chart for Si, Ge and GaAs. 1–GaAs, electrons; 2–Ge, electrons; 3–Si, electrons; 4–Si, holes; 5–GaAs, holes

mobility μ_0 in this area does not depend on E, but differs for different semiconductors. It is called the *low-field mobility*.

In a sufficiently strong field, the carrier drift velocity reaches a constant value, called the *saturated velocity* v_s. The value of this velocity at a sample temperature of 300 K is practically the same for all semiconductors and is 1–2 $\times$ 10^5 m/s. It is important to note that the root-mean-square thermal velocity of particles for the same conditions also has a value of 10^5 m/s.

In some semiconductors (in this case, in gallium arsenide), the drift velocity decreases with increasing field in a certain range of its values. The differential mobility of electrons $\mu_d = \mathrm{d}v_d/\mathrm{d}E$ in this range of field values is negative. This phenomenon was called the Gunn effect, or the phenomenon of *negative differential mobility (NDM)*. FVC features of other types of semiconductors are considered in Chap. 11.

Comparing expressions (2.19) and (2.42), we observe that in order to reach 10^5 m/s in a vacuum, a charge will need to pass a potential difference of only 0.027 V. Accelerating voltages in vacuum devices are hundreds and thousands of volts. Therefore, the characteristic velocities of the motion of charged particles in a vacuum are two to three orders of magnitude higher than the velocity of motion in semiconductors. This difference determines a significant difference in the size of the interaction region of vacuum and semiconductor microwave devices.

The described parameters refer to the *steady drift velocity*. However, in microwave devices, because of the small size of their active region, such a stationary regime is not always successfully established. In particular, if the active region of the device is less than the mean free path, then collisionless (ballistic) transfer of charge carriers is observed in the region under consideration. In this case, the carrier velocity can substantially exceed the saturated velocity. The transient processes of the drift velocity variation under the action of a time-varying field $E(t)$ can be analyzed qualitatively using the standard FVC and (2.40), (2.41).

Figure 2.8 shows a piecewise linear approximation of the FVC of silicon (solid bold line). Here we also observe a typical dependence of the electron temperature of

charge carriers on the static field $T_e(E)$, which is represented by a dashed line. The field E at the moment of time t_1 changes its value from E_1 to E_2. At the moment of time t_2 the field returns to the initial value E_1. The velocity at the moment of time t_1 must instantaneously increase to a value v_s, if we disregard the finiteness of the momentum and energy relaxation time of the particle. And at the moment of time t_2 the velocity must return to the original value v_1.

However, charge carriers cannot instantly change their velocity and energy. According to (2.44) the velocity begin increasing. Its maximum value, determined by mobility μ_0, is $v_{\max} = \mu_0 E_2$. However, the increase in velocity slows down due to a decrease in mobility due to the "heating" of the carriers, i.e. increasing their electron temperature to T_{e2}. As a result, after the energy relaxation time τ_ε, the drift velocity is set at a level v_s. After the field is reduced at a time $t = t_2$, the carriers gradually "cool" and the velocity assumes a stationary value $v = \mu(T_0)E$.

Thus, in the active region of the devices the drift velocity can substantially exceed the stationary value. This phenomenon was called the *overshooting effect*. An additional argument for the existence of such an effect in real structures, the result of calculating the drift velocity in gallium arsenide under the action of a pulsed electric field is presented in Fig. 2.9. The simulation was carried out using the large particle method. As can be seen, maximum value of the velocity is eight times higher than its saturation value.

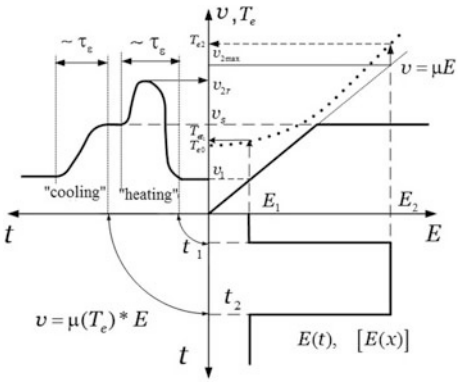


Fig. 2.8 Effect of overshooting. Dotted line–electron temperature

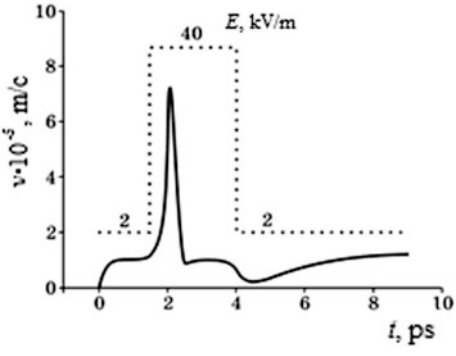


Fig. 2.9 Velocity change in gallium arsenide at the impact of the pulsed electric field

It should be noted that similar processes could also be observed, when charged particles move in a spatially inhomogeneous electric field.

2.5 Noises in Microwave Devices

The information introduction into an electric signal (modulation) is performed by changing the parameters of the signal: amplitude, frequency or phase. Each of these parameters may undergo random changes introduced by the device or by external sources. Let's consider the nature of these changes (or noises) occurring in microwave devices.

The signal is the desired (determined) change in time of the electric oscillation parameters (amplitude, phase, frequency).

Noise is an unwanted (accidental) valid signal distortion. The fundamental sources of noise are energy dissipation and discrete electron charge.

Noises limit the functionality of radio electronic systems because against their background it is difficult to isolate the valid signal if the noise power greatly exceeds the signal power. The sources of noise are both radio electronic components and natural factors. As an example, in Fig. 2.10 the scheme of the communication system is shown with sources of noises indicated. We see that both the transmitting and receiving stations add their own noises to the valid signal. Besides, noises are added during signal propagation. These can be, for example, noises of cosmic origin or noises associated with energy dissipation in the fiber optic route. In accordance with the nature of the noises generation in electronic components they can be divided into three main types: thermal noise, shot noise and low-frequency noise.

Thermal noise occurs due to energy dissipation. It is observed in any element that has an active resistance due to random thermal displacements of the charge carriers inside it.

This mechanism is attributed to the Brownian motion of charge carriers due to the thermal energy in the material. Fluctuations of thermal character can be

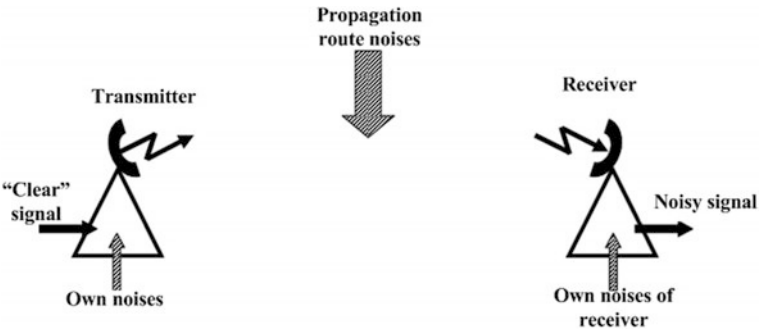


Fig. 2.10 The scheme of the communication system with noises sources indication

considered as a mechanism by which thermal equilibrium is maintained: after a random deviation from the equilibrium state an average return to the initial state follows which is realized by a large number of microscopic interactions resulting in the current or voltage fluctuations in the component.

The thermal noise caused by non-synchronized slowing-dawn emission produced by electrons during their motion and scattering (deceleration) acts on the atoms of the medium. In literature, this noise is called Johnson-Nyquist noise, after the names of two researchers who made a great contribution to the study of this phenomenon. They showed that under thermodynamic equilibrium conditions in the resistor, current fluctuation arises, the root-mean-square (RMS) value of which is described by the formula

$$\left\langle i_n^2 \right\rangle = 4kT_0G_0\Delta f, \tag{2.47}$$

where T_0 is the temperature of the resistor, G_0 is its conductivity, Δf is the pass bandwidth of the noise meter. From this expression it follows that the average square of the noise current in the resistance is directly proportional to the sample temperature T_0, its conductivity $G_0 = 1/R_0$, and also the frequency band Δf in which the noise is measured. The spectral density of the noise current is determined by the Nyquist formula

$$\left\langle i_n^2 \right\rangle_1 = 4kT_0G_0$$

and does not depend on the frequency. Therefore, this noise is called white noise since the spectrum of white light is also uniform.

The power released in the resistor when a noise current flows through it is determined by the formula

$$P_n = \left\langle i_n^2 \right\rangle R = 4kT_0\Delta f,$$

and its spectral density is also independent of frequency. This result, obtained within the framework of the classical fluctuation-dissipation theorem (FDT) cannot be true, since during the integration over the entire spectrum the noise power will appear to be infinite. The application of quantum theory (quantum FDT) makes it possible to correct this error. In accordance with it the spectral density of the noise current

$$\left\langle i_n^2 \right\rangle_1 = 4kT_0G_0\frac{hf/(kT)}{e^{hf/(kT)} - 1}.$$

The dependence of the spectral density of noise current on frequency, calculated according this formula is shown in Fig. 2.11. The cutoff frequency f_b from which the noise intensity decreases rapidly, is determined by the formula $f_b = kT/h$. For a temperature of 300 K this frequency is 6.2 THz, i.e. in the entire microwave band including the submillimeter, thermal noise can be considered as white noise.

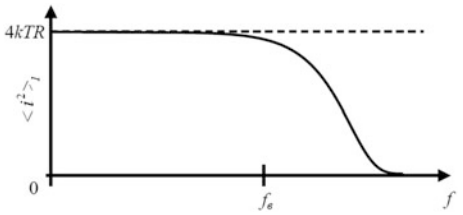


Fig. 2.11 Spectral density of the noise current calculated according to quantum FDT

However, in devices cooled by liquid helium, hydrogen or nitrogen, the dependence of the noise spectral density on the frequency must be taken into account.

Noise current, flowing through the resistor creates on its clamps a noise voltage whose rms value is determined by the formula

$$\left\langle e_n^2 \right\rangle = 4kTR_0\Delta f,$$

where $R_0 = 1/G_0$.

For complex resistance Z or conductivity $Y = 1/Z$ the Nyquist formula takes the form

$$\left\langle i_n^2 \right\rangle_1 = 4kT_0 \mathrm{Re}Y;$$
$$\left\langle e_n^2 \right\rangle_1 = 4kT_0 \mathrm{Re}Z.$$

To perform calculations in circuits with noise sources, the representation of a noisy element in the form of the source of the noise voltage (Fig. 2.12a) or the noise current (Fig. 2.12b) and noiseless ideal resistor is used. Both schemes are equivalent. They are valid under the condition of thermodynamic equilibrium between the resistor and the environment.

In real devices, this equilibrium is disturbed by the presence of "embedded" or applied "heating" electric fields. In this case, for the calculation of the noise current the Van der Ziel formula is used:

$$\left\langle i_n^2 \right\rangle = 4q^2 D(T_e) nS \frac{1}{\Delta l} \Delta f. \tag{2.48}$$

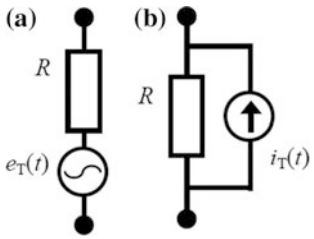


Fig. 2.12 Equivalent circuit of a resistor with noise: **a** with the voltage generator, **b** with the current generator

This formula determines the average square of the noise current $\langle i_n^2 \rangle$, arising from diffusion $D(T_e)$ of the charged particles by a total number n, in a conductive medium element of length Δl and cross section S, with carrier temperature T_e in the frequency band Δf. This formula is more complicated than the Nyquist formula, but it has a wider range of applications for calculating noise and reflects the physics of noise generation.

It is natural that in the case of thermodynamic equilibrium this formula should be transformed into the Nyquist formula. Indeed, in the case of thermodynamic equilibrium there is a relationship between the diffusion coefficient D_0 and carrier mobility μ that is expressed by the Einstein relation: $D_0 = \mu(kT_0/q)$.

Substituting this expression in the Van der Ziel formula and assuming that the temperature is equal for the entire sample, we obtain:

$$\langle i_n^2 \rangle = 4q^2 \frac{kT_0}{q} \mu n S \frac{1}{\Delta z} \Delta f = 4kT_0 \frac{qn\mu S}{\Delta z} \Delta f = 4kT_0 G_0 \Delta f. \tag{2.49}$$

Thus, the Nyquist formula is a special case of (2.48) formula. Let's transform the expression (2.48) multiplying the right-hand side of this expression by the diffusion coefficient D_0 and rearrange the factors:

$$\langle i_n^2 \rangle = 4q^2 \frac{D(T_e)}{D_0} D_0 n S \frac{1}{\Delta l} \Delta f = 4kT_0 \frac{D(T_e)}{D_0} G_0 \Delta f.$$

Introducing the concept of effective noise temperature $T_{\text{ef}} = T_0 D(T_e)/D_0$ and effective noise conductivity $G_{\text{ef}} = G_0 D(T_e)/D_0$ we obtain a generalization of the Nyquist formula for the case of the thermodynamic equilibrium absence:

$$\langle i_n^2 \rangle = 4kT_{\text{eff}} G_0 \Delta f = 4kT_0 G_{\text{eff}} \Delta f. \tag{2.50}$$

The expression $D(T_e)/D_0$ is called the noise ratio, showing how much the device is noisier than the classical Nyquist formula predicts. In practice, the expression (2.49) is used to describe the noise of any one-port network. For nonlinear elements the differential conductivity at the operating point is used as G_0.

The shot noise is associated with the electron charge discreteness and the statistical nature of the charge carrier's transfer through a potential barrier. It was first considered by Schottky, who used an analogy between a large number of electrons transiting through an interelectrode gap and a small shot dropping into a metal container. The shot noise occurs in solid-state devices every time when the current passes through the potential barrier, for example in *p-n* junction. At this, the current consists of a set of individual impulses due to the discrete nature of the charge. Instantaneous current values randomly fluctuate about some mean value. The shot noise also occurs in vacuum devices when electrons leave the cathode that is also associated with overcoming the potential barrier.

When current I_0 passes through the potential barrier, the mean square of the noise current $\langle i_n^2 \rangle$ is determined by the Schottky formula

$$\langle i_n^2 \rangle = 2qI_0\Delta f. \tag{2.51}$$

This formula is true under the following conditions:

- All carriers in the working space are independent of each other, i.e. the space charge is not taken into account.
- Time of carriers' transit through the working space τ_{np} is negligible.
- The average current through the device is not modulated, $I_0 = \text{Const}$.

As can be seen from the expression presented, this shot noise under the mentioned conditions, as well as thermal noise, has the constant spectral density $2qI_0$ independent of frequency.

If the space charge affects the motion of the carriers significantly, the smoothing of the fluctuations occurs and the noise current decreases. In this case, it can be calculated by the formula

$$\langle i_n^2 \rangle = 2qI_0\Gamma^2\Delta f,$$

where $\Gamma^2 = 0.01$–0.1 is the depression coefficient depending on the density of the space charge ρ. The greater ρ is, the lower the depression coefficient is. This effect is used in vacuum devices to reduce the noise of the electron beam. For this, the anode current is set much lower than the cathode emission current, so a dense space charge is formed near it.

As follows from the second condition, formula (2.51) is valid only at comparatively low frequencies $f \ll f_d = 1/(2\pi\tau_{np})$. At higher frequencies it is necessary to take into account the spectrum form of the individual pulses generating the induced current. In this case, the value of the shot noise is calculated by the formula

$$\langle i_n^2 \rangle = 2qI_0M(\theta)\Delta f,$$

where $M(\theta)$ is the energy spectrum of an elementary current impulse produced by a separate carrier.

For triangular impulses produced by electrons moving in a constant accelerating field, this spectrum is shown in Fig. 2.13 as a function of the electron transit angle in the working space $\theta = 2\pi fd/v_{\text{med}}$, $\theta = 2\pi fd/v_{cp}$, where d is the length of the working space, v_{med} is the average velocity of the carrier in this space. For typical vacuum and semiconductor microwave devices, the cutoff frequency of the shot noise lies in the range 0.1–1 THz.

If the current through the device is modulated, the fluctuation equation of shot noise must be used to calculate the noise current.

The flicker noise in electron tubes was first discovered by Johnson in 1925. It was soon discovered that this noise was due to a "flickering" change in the

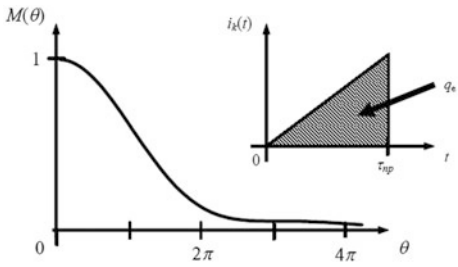


Fig. 2.13 Spectrum of the triangle pulse (on inlet)

emissivity of individual parts of the cathode (hence, the name flicker). Similar phenomena are observed in *p-n* junctions and in the Schottky barriers. They produce excessive noise in addition to thermal and shot noises. The frequency spectrum of this "additive" is located in the low-frequency range, since it is associated with relatively inertial processes. The main part of the flicker noise spectrum is located in the frequency range from 0 to 10^5 Hz, and the spectrum density is in inverse proportion to frequency. Considering the variety of physical phenomena determining the described fluctuations in different devices, this noise has many other names: 1/*f*-noise, flicker noise, excess noise. In generators, these low-frequency fluctuations can result in the appearance of a non-monochromatic generation determined by the term "phase noises".

The spectral density of low-frequency noise S_f is represented by the following expression:

$$S_f = S_0 \left(\frac{f_b}{f}\right)^n \Delta f, \tag{2.52}$$

where f_b is a certain characteristic boundary frequency at which the spectral density of the excess noise power is equal to the spectral power of the white noise $S_0 = 4kT_0$. As practice has shown, this frequency is approximately 10^5 Hz. The coefficient n in most cases is close to 1. Figure 2.14 shows the dependence of the white and flicker noise spectral power density logarithm on the frequency.

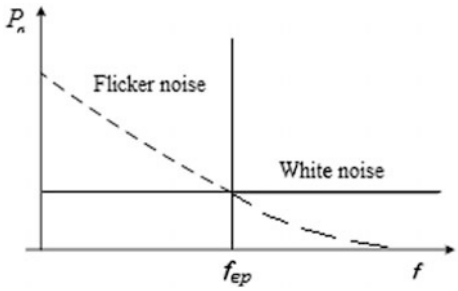


Fig. 2.14 Flicker and white noise powers as functions of frequency

There are other types of noises: generation-recombination noise (GRN), avalanche noise, and explosive noise. A detailed examination of these noises goes beyond the scope of this textbook.

Noise characteristics of the two-ports network. In the previous section noise in two-ports elements, such as resistors and diodes was considered. Many active microwave devices are "noisy two-ports networks" putting in their "mite" to the resulting noise power in the payload. Let us briefly consider the characteristic noise parameters of such elements.

In Fig. 2.15 a noisy two-ports network (the shaded rectangle) and its equivalent circuit are presented. All internal sources of noise in it are placed at the input and are represented by two ideal generators of noise voltage $\langle u_{en}^2 \rangle$ and current $\langle i_{en}^2 \rangle$ in accordance with Thevenin's theorem. In the general case these sources are correlated, i.e. the correlation coefficient

$$C = \frac{\langle u_{en} i_{en} \rangle}{\sqrt{\langle u_{en}^2 \rangle \langle i_{en}^2 \rangle}} \neq 0.$$

Experience has shown that it is often possible to use one source of noise, for example, the voltage generator. To characterize the noise $p_{n\,\mathrm{add}}$, added by the device to an existing signal with noise, the concept of the noise factor $N_{F,}$ is introduced

$$N_F = (p_s/p_n)_{\mathrm{in}}/(p_s/p_n)_{\mathrm{out}} \tag{2.53}$$

where p_s and p_n are the signal power and noise power respectively.

The logarithm of this quantity is called the noise factor

$$K_n = 10 \lg N_F \ (\mathrm{dB}).$$

Let's modify the expression for the noise factor by putting the noise sources to the input of the two-ports network. Suppose that it is possible to manage with one such noise source. Then

$$N_p = \frac{P_{s\,in}}{P_{s\,out}} \frac{P_{n\,out}}{P_{n\,in}} = \frac{1}{G} \frac{P_{n\,in} G + P_{n\,add}}{4kT_0 \Delta f} = \frac{1}{G} \frac{G(4kT_0 \delta f + 4kT_{ef} \delta f)}{4kT_0 \Delta f} = 1 + \frac{T_{ef}}{T_0}. \tag{2.54}$$

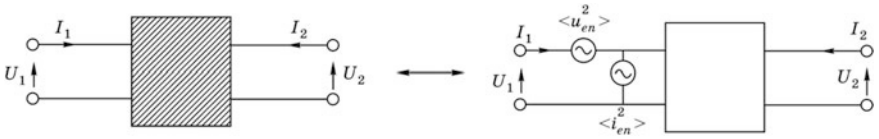


Fig. 2.15 A noisy two-ports network and its equivalent scheme

where G is the amplification coefficient of the two-ports network; T_{ef} is the effective noise temperature of the noise source giving the same output noise power as the real two-ports network; and T_0 is the noise temperature characterizing the "noisiness" of the input signal. As we see, there are used here the expressions for thermal noise in the form of the Nyquist formula. It is seen from equation (2.54) that the noise factor is always greater than one. This means the deterioration in the signal-to-noise ratio at the output as compared to the input.

It follows from equation (2.54) that the effective noise temperature of the two-ports network is

$$T_{\text{eff}} = T_0(N_F - 1). \tag{2.55}$$

At this, the total amplification coefficient G is equal to the product of the coefficients of separate cascades G_i, i.e. $G = G_1G_2 \ldots G_n$. For this case, let us define the effective noise temperature and noise factor. Figure 2.16 shows the connection of two amplifiers with the amplification coefficients G_1, G_2 and noise temperature T_{e1}, T_{e2}.

Let us make alterations to the general formula for the noise factor by transferring the effective noise sources to the input:

$$N_p = \frac{P_{s\,in}}{P_{n\,out}}\frac{P_{n\,out}}{P_{s\,out}} = \frac{P_{s\,in}}{P_{s\,out}}\frac{P_{n\,out}}{P_{n\,in}} = \frac{1}{G_1G_2}\frac{P_{n\,in}G_1G_2 + P_{n\,add\,1}G_1G_2 + P_{n\,add\,2}G_2}{P_{n\,in}} = 1 + \frac{T_{e1}}{T_0} + \frac{1}{G_1}\frac{T_{e2}}{T_0}.$$

Thus, the contribution to the final noise from the second cascade is G_1 times less than the contribution from the first cascade. It is easy to see that for an arbitrary number of cascades this formula takes the form:

$$N_F = 1 + \frac{T_{e1}}{T_0} + \frac{1}{G_1}\frac{T_{e2}}{T_0} + \frac{1}{G_1G_2}\frac{T_{e3}}{T_0} + \cdots \tag{2.56}$$

From this analysis it follows that in the development of low-noise amplifiers careful attention must be paid to the input cascade.

Generator phase noises. In the ideal case, the spectrum of a generator should represent the δ-function at the frequency . However, the device characteristics changing due to noises (thermal noise, low-frequency noise) results in the fact that the real generator power is distributed in a certain frequency range as shown in Fig. 2.17.

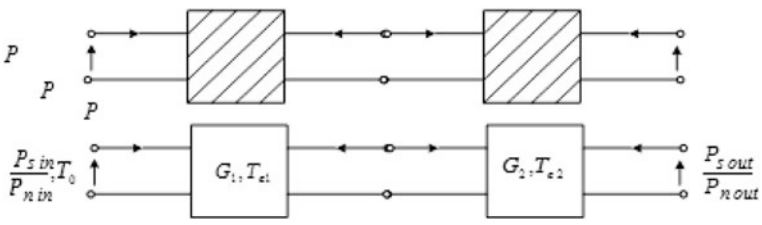


Fig. 2.16 Two in series noisy two-ports networks and their equivalent scheme

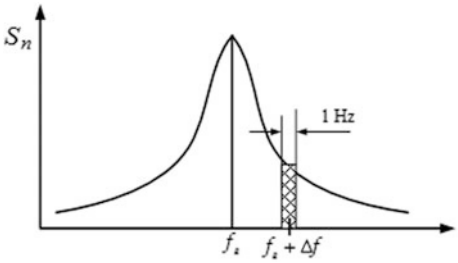


Fig. 2.17 Spectrum of the typical generator output signal

Mathematically such a spectrum can be described by the following simple transformation. Let's represent a generator signal $u_g(t)$ in the form:

$$u_g(t) = U_m[1 + n(t)]\cos\left[\omega_g t + \Theta_n(t)\right]. \tag{2.57}$$

where function $n(t)$ is responsible for the amplitude noise fluctuations, function $\Theta_n(t)$ characterizes the random phase changes, and U_m is the amplitude of the generator signal. The change of the argument in (2.57) can be caused by a change of the frequency ω_g or a change of the phase $\Theta_n(t)$. These two methods are indistinguishable.

In case of small changes in the resultant frequency, let us write the phase $\Theta_{n(t)}$ in the form of a harmonic function f_g with the amplitude $\Delta f/f_m$, where f_m is some modulation frequency, i.e.

$$\Theta_n(t) \approx \frac{\Delta f}{f_m}\cos(2\pi f_g t) = \Theta_p \cos(2\pi f_g t). \tag{2.58}$$

Let us substitute (2.58) in (2.57) and transform the equation for the case of the amplitude noises absence $(n(t) = 0)$ and a small modulation index $\Theta_p = \Delta f/f_m \ll 1$. As a result, we obtain:

$$u_\Gamma(t) = U_m\left\{cos\omega_g t - \frac{\Theta_p}{2}\left[(\cos(\omega_g + \omega_m)t - \cos(\omega_g - \omega_m)t\right]\right\}. \tag{2.59}$$

Equation (2.59) shows that the resultant spectrum is symmetric with respect to frequency as shown in Fig. 2.17. According to this spectral characteristic, the parameter S_p is determined by characterizing the spectral density of the phase noises. This parameter is calculated as the ratio of the generator power $P_{\Delta f}$ at a frequency shifted by Δf in relation to the central frequency f_g to the power at the center frequency P_0. The powers are determined in the 1 Hz band. In decibels

$$S_p = 10\lg(P_{\Delta f}/P_0).$$

Figure 2.18 shows the S_p parameter dependence on the detuning $\Delta f = f - f_g$ for the generator with an operating frequency of 2.2 GHz. As can be seen, in this case,

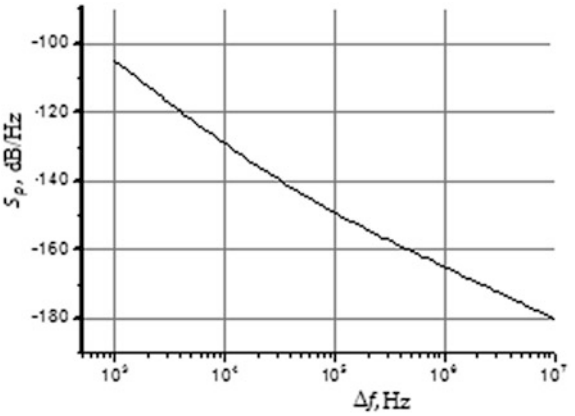


Fig. 2.18 Dependence of spectrum density of generator phase noises on detuning

even a relatively small detuning results in a sharp power decrease of 10–16 orders of magnitude. The signal of such a generator can be considered monochromatic with a high degree of accuracy.

Advancement Questions

1. Name the main types of individual radiation of charged particles. Under what conditions is this radiation is observed?
2. What is the Doppler effect? Name the kinds of this effect. In which microwave devices does the Doppler effect plays a defining role?
3. Under what conditions does Cherenkov radiation occur? Give examples.
4. What are the properties of Dicke superradiance? Under what conditions is it observed?
5. Explain the difference between spontaneous and induced radiation. What conditions are necessary for the occurrence of induced radiation?
6. Describe the conditions of applicability of Maxwell's equations.
7. Can the magnetic field change the kinetic energy of a charged particle?
8. What are Euler variables and Lagrange variables? How is the transition from one variable to the other carried out?
9. Explain the physical meaning of the Boltzmann equation.
10. Explain the relationship of the Boltzmann equation and the system of conservation equations for the number of particles, momentum flux, and energy flow.
11. What is the Debye radius (length)? The plasma frequency?
12. What is the difference between the concepts of plasma frequency for vacuum and semiconductor devices?
13. What does Maxwell's relaxation frequency characterize?
14. What is the difference between solid-state and gas-discharge plasma?

15. What is the mobility of charge carriers? Does negative differential mobility occur in nature?
16. What are the relaxation processes? What is the difference between impulse and energy relaxation processes?
17. What is the nature of noises in radio-electronic systems?
18. List the sources of noises in electronic equipment.
19. How does the spectral density of thermal noise depend on frequency?
20. What formula determines the power of thermal noises in the absence of thermodynamic equilibrium?
21. What factors determine the shot noise power?
22. How does the carriers' space charge influence the shot noise?
23. Explain the nature of flicker noise.
24. What is phase noise and how it is assessed quantitatively?

Chapter 3
Oscillations and Waves in Charged Particle Beams

3.1 Space Charge Oscillations

Let's consider a space uniformly filled with plasma formed by free electrons and positive ions with a charge density of each particle type equal to ρ_0. Such plasma is electrically neutral and there is no macroscopic electric field in it. Suppose that under the influence of an external field, the electron layer in the plasma is shifted by a distance z (Fig. 3.1). As a result, in that place where the layer has shifted, an excess density of electrons occurs, and in the place where it was before, there is a lack of electrons, i.e. an excess charge of positive ions. The surface density of excess charge is $\sigma = \pm\rho_0 z$. Between these charged layers, an electric field with intensity $E_z = \sigma/\varepsilon_0$ arises. Let's write down the equation of the motion of electrons in this field:

$$m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -eE_z. \tag{3.1}$$

Substituting the value of the electric field in this equation, we obtain the equation

$$\ddot{z} + \omega_p^2 z = 0, \tag{3.2}$$

having a solution with the form

$$z = Az_0 \cos(\omega_p t + \varphi),$$

where z_0 is the initial coordinate of the electron, and A and φ are the integration constants. Apparently, the electrons make harmonic oscillations around their equilibrium position with an angular frequency, called the *plasma frequency*

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_3

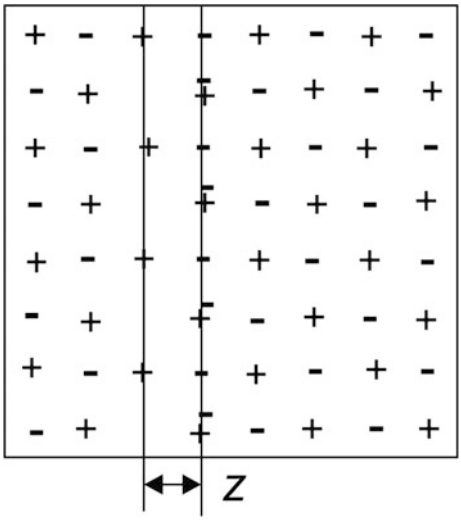


Fig. 3.1 To the analysis of oscillations in plasma

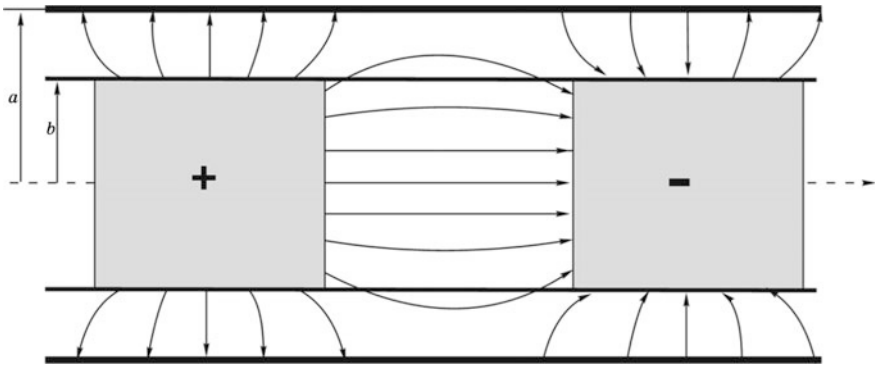


Fig. 3.2 The electric field in a grouped cylindrical plasma column

$$\omega_p = \sqrt{\frac{e\rho_0}{m\varepsilon_0}}. \tag{3.3}$$

If the space filled with plasma is confined by a metal screen, part of the electric field lines is closed on the screen instead of being closed on charges of the opposite sign. As a result, the restoring force affecting the charged particles reduces, leading to a decrease in plasma frequency.

For example, let's consider a cylindrical plasma column of radius b surrounded by a coaxial metallic screen of radius a (Fig. 3.2). The axial field produced by an

electron disk with surface charge density calculated by electrostatic methods is determined by the expression:

$$E_z(r) = \frac{\sigma b}{\varepsilon_0 a} \sum_{m=1}^{\infty} \frac{J_1(v_{0k} b/a) J_0(v_{0k} r/a)}{v_{0k} J_1^2(v_{0k})},$$

where v_{0k} is *k-th* root of equation $J_0(v) = 0$.

Having averaged this field along the column radius and substituting it into (3.1), after a series of simplifications we obtain an equation of the form (3.2), in which instead of ω_p there is a reduced plasma frequency $\omega_q = s\omega_p$:

$$\ddot{z} + \omega_q^2 z = \frac{e}{m} E_z, \tag{3.4}$$

where E_z is the longitudinal electric field affecting the electron and the *reduction coefficient* s is determined by the approximate expression.

$$s = \left[2 \frac{a}{l} \frac{J_1^2(\zeta_{01} b/a)}{\zeta_{01}^3 J_1^2(\zeta_{01})} \right]^{1/2} \approx 0,735 J_1(2,4b/a) \sqrt{\frac{a}{l}}. \tag{3.5}$$

In this expression, l is the distance between neighboring particle bunches, which is determined by the nature of their motion.

Obviously, if we remove the motionless ions, the (3.2) remains valid since the ions produce only the field that is constant in time. Therefore, oscillations are also observed in an ensemble of electrons, the charge of which is not compensated by ions. The equation of oscillations (3.4) for electrons was first obtained by Y.A. Katsman.

3.2 Space Charge Waves in Electron Beams

Let us assume that the electrons move along the z axis with a velocity υ_0. Accordingly, they create an electric current with a density $J_{z0} = -\rho_0 v_0$ (ρ_0 is the absolute value of the electrons' space charge). Suppose that because of an external action the electrons received variable components of velocity and space charge alternating according the harmonic order

$$v = v_0 + \tilde{v} = v_0 + \dot{v} e^{i\omega t}; \quad \rho = \rho_0 + \tilde{\rho} = \rho_0 + \dot{\rho} e^{i\omega t}.$$

Current density in the flux

$$J_z = -(\rho_0 v_0 + \rho_0 \tilde{v} + \tilde{\rho} v_0 + \tilde{\rho} \tilde{v}).$$

(the minus sign on the right-hand side is due to the negative charge of the electron).

Assuming that the amplitudes of the variable components are much smaller than the constant values, the product of the variables in this expression can be ignored and the current density can be represented as a sum of constant and variable components:

$$J_z = J_{z0} + \tilde{J}_z = J_{z0} + \dot{J}_z e^{i\omega t},$$

where

$$\dot{J}_z = -(\rho_0 \dot{v} + v_0 \dot{\rho}) \tag{3.6}$$

Since the electron flux is assumed to be unlimited along the z axis, the amplitude of the total current density in it must be equal to zero:

$$\dot{J}_z + i\,\omega\varepsilon_0 \dot{E}_z = 0, \tag{3.7}$$

where Ez is the amplitude of the variable component of the electric field excited by the electron flux.

Let us write down the electrons' motion equation. From (2.19) it follows that

$$i\,\omega\dot{v} + v_0 \frac{\partial \dot{v}}{\partial z} = -\frac{e}{m} E_z. \tag{3.8}$$

Substituting the value of the field strength from (3.7) into this equation we obtain

$$-\omega^2 \dot{v} + i\,\omega v_0 \frac{\partial \dot{v}}{\partial z} = \frac{e}{m\varepsilon_0} J_z. \tag{3.9}$$

Now we find another relation between the velocity and current density. The equation of continuity in this case has the form

$$\frac{\partial \dot{J}_z}{\partial z} - i\,\omega\dot{\rho} = 0. \tag{3.10}$$

From the current density amplitude estimation (3.6), we find

$$\dot{\rho} = -\frac{\dot{J}_z + \rho_0 \dot{v}}{v_0}. \tag{3.11}$$

Substituting (3.11) into (3.10), we obtain

$$\frac{\partial \dot{J}_z}{\partial z} + \frac{i\,\omega}{v_0} J_z = -\frac{i\,\omega\rho_0}{v_0}\,\dot{v}. \tag{3.12}$$

Equations (3.9) and (3.12) form the equations system for the unknown amplitudes of current density and velocity. We can find $\dot{v}$ from (3.12):

$$\dot{v} = -\frac{v_0}{i\,\omega\rho_0}\frac{\partial \dot{J}_z}{\partial z} - \frac{1}{\rho_0}\dot{J}_z; \tag{3.13}$$

Substitute the value found in (3.9):

$$\frac{\omega\, v_0}{i\rho_0}\frac{\partial \dot{J}_z}{\partial t} + \frac{\omega^2}{\rho_0}\dot{J}_z - \frac{v_0^2}{\rho_0}\frac{\partial^2 \dot{J}_z}{\partial z^2} - \frac{i\,\omega\,\dot{v}_0}{\rho_0}\frac{\partial \dot{J}_z}{\partial z} - \frac{e}{m\varepsilon_0}\dot{J}_z = 0.$$

After reducing similar terms and multiplying the result by ρ_0/v_0^2 we have:

$$\frac{\partial^2 \dot{J}_z}{\partial z^2} + 2i\,\beta_e\frac{\partial \dot{J}_z}{\partial z} - \left(\beta_e^2 - \beta_p^2\right)\dot{J}_z = 0, \tag{3.14}$$

where the electron propagation constant $\beta_e = \omega/\upsilon_0$ and the plasma propagation constant $\beta_p = \omega_p/\upsilon_0$ are introduced.

We shall seek a solution of (3.14) in the form of a current wave

$$\dot{J}_z = \dot{J}_{zm}e^{-i\Gamma z} \tag{3.15}$$

where Γ is an unknown propagation constant. Substituting (3.15) into (3.14), after reducing to J_{zm}, we obtain

$$\Gamma^2 - 2\beta_e\Gamma + \left(\beta_e^2 - \beta_p^2\right) = 0. \tag{3.16}$$

Solving this quadratic equation we find two values of the propagation constant:

$$\Gamma^{\pm} = \beta_e \pm \sqrt{\beta_e^2 - \left(\beta_e^2 - \beta_p^2\right)} = \beta_e \pm \beta_p. \tag{3.17}$$

From here

$$\dot{J}_z = \dot{J}_{z+}e^{-i(\beta_e+\beta_p)z} + \dot{J}_{z-}e^{-i(\beta_e-\beta_p)z}. \tag{3.18}$$

Thus, in the electron flux, two waves with different propagation constants can spread. These waves are called *space-charge waves*. Let's find their phase velocities:

$$v_{p\pm} = \frac{\omega}{\Gamma^{\pm}} = v_0\frac{\omega}{\omega \pm \omega_p}.$$

Usually in electron fluxes $\omega_p \ll \omega$, so one wave (with a lower subscript) propagates a little faster than electrons (it is called *fast*), and the phase velocity of a wave with an upper index is less than the electrons velocity (this wave is called *slow*).

The group velocities of both waves are equal to the constant component of the electron velocity:

$$v_g^{\pm} = \left(\frac{\mathrm{d}\Gamma^{\pm}}{\mathrm{d}\omega}\right)^{-1} = \left(\frac{\mathrm{d}(\omega \pm \omega_p}{v_0 \mathrm{d}\omega}\right)^{-1} = v_0.$$

Let's find the relationship between the velocity amplitudes and charge density. From (3.13)

$$\dot{v}_m^{\pm} = \pm \frac{\beta_p}{\beta_e} \frac{J_{zm}^{\pm}}{\rho_0}. \tag{3.19}$$

Substituting the current density value from (3.19) into (3.10), after simple transformations we find

$$\frac{\dot{\rho}^{\pm}}{\rho_0} = \mp \frac{\beta_e \pm \beta_p}{\beta_p} \frac{\dot{v}\pm}{v_0}. \tag{3.20}$$

Figure 3.3 shows the dependence of the electron velocity amplitudes and the space-charge density of fast (*a*) and slow (*б*) waves from the coordinate. The flux has an accelerating voltage of 10 kV, current density of 0.5 A/mm^2, and the ratio of the alternating current component to the constant component is 0.01. The frequency of excitation of space-charge waves is 10 GHz, and the flux plasma frequency is 2 GHz. It can be seen that in the fast wave density and velocity are in phase. This means that the number of electrons having a velocity greater than the average velocity is larger than the number of electrons with lower velocities.

Therefore, the kinetic energy of the flux in which the fast wave is excited is greater than the energy of the unperturbed flux. It is said that the fast space-charge wave carries positive *kinetic power*.

In the slow wave, the velocity and density are in antiphase. Therefore, a slow wave has the larger number of electrons with velocities less than the average

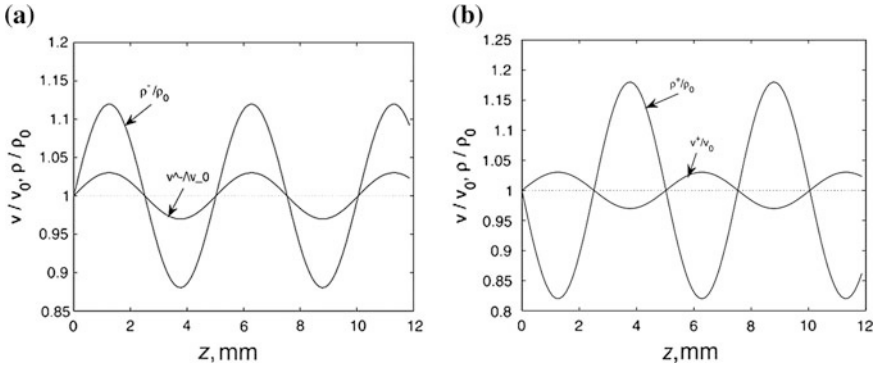


Fig. 3.3 Velocity and charge density distribution in fast and slow space charge waves

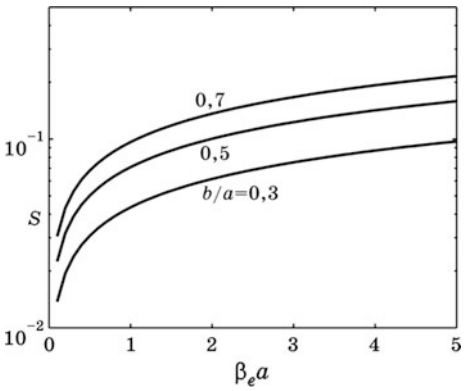


Fig. 3.4 Dependence of reduction coefficient on reduced canal radius

velocity, and a smaller number of electrons with velocities greater than the average. As a result, the kinetic energy of the electrons of the flux in which the slow space-charge wave is excited is less than the energy of the unperturbed flux. Therefore, it is said that the slow wave carries a negative kinetic power. If both space-charge waves with equal amplitudes are excited in the flux then the kinetic energy of the flux does not change.

If the electron flux is limited in the transverse direction, it is necessary to take into account the plasma frequency decrease due to the influence of the screen. In particular, if the flux has the form of a cylinder of radius b and propagates in a cylindrical screen of radius a, we can use formula (3.5) to calculate the reduction coefficient taking the "electronic" wavelength $\lambda_e = 2\pi/\beta_e$ as the distance between the disks. In this case, the formula takes the form

$$s = 0.293 J_1(2.4b/a)\sqrt{\beta_e a}. \tag{3.21}$$

Figure 3.4 shows the dependence of the reduction coefficient $\beta_e a$ for different values of the flux radius to the screen radius ratio. As can be seen, the screen influence leads to a significant decrease in the plasma frequency, especially for small values of $\beta_e a = 2\pi a/\lambda_e$, i.e. when the distance between disks is large compared to the channel radius.

Note that if the electron flux propagates in a longitudinal magnetic field, other types of waves (cyclotron and synchronous) can propagate in it along with space-charge waves.

3.3 Charge Carrier Waves in Semiconductors

Let us consider a homogeneous semiconductor material in which the charge carriers move under the action of the electric field. For simplicity, we consider the motion of charge carriers of one type only. We introduce the coordinate z along which the

charge carriers move and the lines of electric force are directed. In the transverse direction, the material is assumed to be unbounded so all values depend only on the z coordinate and time.

Let the particles be affected by an electric field having both constant and variable components:

$$E_z = E_{z0} + \tilde{E}_z.$$

Under the influence of the field variable component, the variable components of the particle concentration, current density, and velocity in the flux appear:

$$\begin{aligned} n &= n_0 + \tilde{n}; \\ J_z &= J_{z0} + \tilde{J}_z; \\ v_z &= v_0 + \tilde{v}. \end{aligned}$$

We assume that the values of the constant components are much greater than the maximum values of the variable components (the small-signal approximation). Let's write the expression for the convection current of particles using (2.43):

$$J_z = qnv - qD\frac{\partial n}{\partial z}. \tag{3.22}$$

Let us write this equation separately for constant and variable components:

$$J_0 = qn_0v_0 - qD\frac{\mathrm{d}n_0}{\mathrm{d}z}; \tag{3.23}$$

$$\tilde{J} = q\tilde{n}v_0 + qn_0\tilde{v} - qD\frac{\partial \tilde{n}}{\partial z}. \tag{3.24}$$

In expression (3.24), the second smallness order terms are omitted.

Maxwell's equations (2.8) are also separate for constant and variable components and have the form

$$\frac{\mathrm{d}E_0}{\mathrm{d}z} = \frac{q}{\varepsilon}(n_0 - N); \tag{3.25}$$

$$\frac{\partial \tilde{E}}{\partial z} = \frac{q}{\varepsilon}\tilde{n}. \tag{3.26}$$

The joint solution of (3.23) and (3.25) for given boundary conditions allows us to determine the stationary distributions of concentration, field and velocity, and also to calculate the current density J_0. In the following, we assume that these values do not depend on the coordinate z.

To calculate the variable components, we differentiate (3.24) by z and assume that they depend on time according to the harmonic law with angular frequency ω.

Taking into account the fact that the total electric-current density $J_{tot} = J + i\,\omega\varepsilon E$ in any section is equal to zero, we obtain

$$-qD\frac{\partial^2\tilde{n}}{\partial z^2} + qv_0\frac{\partial\tilde{n}}{\partial t} + qn_0\frac{\partial\tilde{v}}{\partial z} + i\,\omega\varepsilon\frac{\partial\tilde{E}}{\partial z} = 0.$$

Using (3.26) and taking into account $\tilde{v} = \mu_d\tilde{E}$, where $\mu_d = dv/dE|_{E=E_0}$ is the differential mobility, we find

$$D\frac{\partial^2\tilde{n}}{\partial z^2} - v_0\frac{\partial\tilde{n}}{\partial z} - (i\,\omega + \omega_m)\tilde{n} = 0, \tag{3.27}$$

where $\omega_m = qn_0\mu_d/\varepsilon$ is the differential frequency of Maxwellian relaxation. The solution of this equation is in the form $\tilde{n} = \dot{n}e^{-i\Gamma}$, where Γ is unknown propagation constant. Substituting the value of n_1 in (3.27), we obtain the equation with respect to Γ:

$$\Gamma^2 - \frac{i\,v_0}{D}\Gamma + \frac{i\,\omega + \omega_m}{D} = 0. \tag{3.28}$$

This equation has two complex roots:

$$\Gamma^{\pm} = \frac{i\,v_0}{2D} \pm \sqrt{-\frac{v_0^2}{4D^2} + \frac{i\,+\omega_m}{D}}. \tag{3.29}$$

Thus, two space-charge waves with different propagation constants can spread in the semiconductor.

Let's transform expression (3.29) rooting out the first term:

$$\Gamma^{\pm} = \frac{i\,v_0}{2D}\left(1 \pm i\sqrt{1 - \frac{4(i\,\omega + \omega_m)D}{v_0^2}}\right).$$

The Maxwellian frequency of relaxation is usually much greater than the frequencies used in the microwave band. In addition, the second term under the root is usually much greater than 1, so the latter can be neglected. As a result, for constant propagation, we obtain the approximate expression

$$\Gamma^{\pm} = \pm\beta - i\,\alpha^{\pm} \approx \mp\sqrt{\frac{4\omega_m}{D}} + i\frac{v_0}{D}. \tag{3.30}$$

Thus,

$$\tilde{n} = \dot{n}_1 e^{\alpha z}e^{-i\,\beta z} + \dot{n}_2 e^{\alpha z}e^{i\beta z}. \tag{3.31}$$

The solutions for other quantities, namely velocity, current density and field strength have the same form. As can be seen, two waves propagate in opposite

directions, with the amplitude of one of them increasing, and the amplitude of one of them decreasing while they spread.

In many cases, the diffusion can be ignored. As a result, (3.27) is simplified. The solution of this equation can also be found in the wave form. The equation for the propagation constant has the form

$$\Gamma = \beta - i\,\alpha = \frac{\omega}{v_0} - i\frac{\omega_m}{v_0}. \tag{3.32}$$

In this case, one wave propagates in the semiconductor with a velocity equal to the constant component of the charge carriers' velocity and with the damping constant, the sign of which depends on the ω_m sign. If the differential charge carriers mobility is positive, $\omega_m > 0$ and the wave decays while propagating. However, if the differential mobility is negative, the relaxation frequency is also negative and the wave amplitude increases exponentially while propagating. This effect can be used to amplify and generate microwaves. For $\omega_m = 0$, the charge wave in homogeneities propagates without damping.

Note that in this section there is presented the linear theory based on the assumption about the smallness of the variable components that characterize the wave in comparison with the constant values. Such an assumption can not always be used, so the efforts of many researchers were aimed at creating the nonlinear theory of interaction that involves a wider range of phenomena and allows more accurate results to be obtained.

Advancement Questions

1. Write down the equation of electron oscillations in the electron flux. What process does it describe?
2. What is the plasma frequency? On what parameters does the plasma frequency reduction coefficient depend?
3. What is the difference between fast and slow space-charge waves in the electron flux? What is their kinetic power?
4. Which types of wave can propagate in the electron flux?
5. Describe the space charge waves in a solid with and without diffusion.
6. What is the physical meaning of the differential Maxwellian frequency in a solid?

Chapter 4
Interaction of Charged Particle Fluxes with a High-Frequency Electromagnetic Field

4.1 Interaction Power

In Chap. 2, it was shown that in order to obtain technically acceptable radiation power of the charged particles ensemble it is necessary to ensure the bunching of particles, i.e. their concentration in a volume with dimensions not exceeding half of the wavelength and/or their phasing, or in other words to provide conditions when individual particles radiate in the same phases. Phasing is achieved by using an external field acting on the charged particle. As noted, the methods of bunching and phasing are different in different types of microwave devices, but there are general properties of the interaction of charged particles with a high-frequency electromagnetic field, and these are discussed in this chapter.

Modern microwave devices use intensive charged particles fluxes containing a huge number of individual radiators. Accordingly, there is an enormous number of radiation and absorption acts. Therefore, it is reasonable to consider the integral flux characteristics such as current (or current density), velocity and charge density.

The electromagnetic field in microwave devices receives energy from the electron flux. Under the action of an accelerating voltage U_0, each electron of the flux acquires the kinetic energy $w_k = mv^2/2 = eU_0$. The density of the beam kinetic energy with an electron density $n_0 = \rho_0/e$

$$W_k = w_k n_0 = \frac{mv^2}{2}\frac{\rho_0}{e} = U_0\rho_0.$$

The flux of kinetic energy transferred by the beam with a cross-sectional area A per unit of time (kinetic power)

$$P_k = W_k A v = U_0\rho_0 vA = U_0 I_0,$$

where I_0 is the beam current. This power can be partially transformed into the power of the high-frequency electromagnetic field.

Let's analyze interaction processes between CP flux and electromagnetic field.

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_4

We suppose that the electric field has constant and vatiable components:

$$\mathbf{E} = \mathbf{E}_0 + \tilde{\mathbf{E}}. \tag{4.1}$$

Convection current density also has constant and variable components:

$$\mathbf{J} = \mathbf{J}_0 + \tilde{\mathbf{J}}. \tag{4.2}$$

Spatial inhomogeneity of the field and current distribution is supposed to be negligible and can be ignored. Then from (2.41), we obtain

$$\frac{\partial}{\partial t}\left[n\left(\frac{mv_d^2}{2} + \frac{3kT_e}{2}\right)\right] - \mathbf{J}\cdot\mathbf{E} = -\frac{3}{2}\frac{nk(T_e - T_0)}{\tau_\varepsilon}. \tag{4.3}$$

For particle motion without collisions, $\tau_\varepsilon \to \infty$, $T_e \to 0$, so we obtain

$$n\frac{\partial(mv^2/2)}{\partial t} - \mathbf{J}\cdot\mathbf{E} = 0, \tag{4.4}$$

i.e. the change in the kinetic energy of the particles occurs as a result of the energy exchange between the field and the particles.

The instantaneous specific power of the interaction in these equations is determined by the term $\mathbf{J}\cdot\mathbf{E}$. To determine the average specific power over the period it is necessary to integrate this component in time,

$$\langle p\rangle = 1/T\int_0^T \mathbf{J}\cdot\mathbf{E}\partial t \tag{4.5}$$

The integration (4.5) over the interaction space ΔV gives the average interaction power for the period in the volume ΔV:

$$\langle \Delta P\rangle = \int_{\Delta V} p\mathrm{d}V = 1/T\int_{\Delta V}\int_0^T \mathbf{J}\cdot\mathbf{E}\mathrm{d}t\mathrm{d}V \tag{4.6}$$

Let's substitute (4.1) and (4.2) into (4.5):

$$p = 1/T\int_0^T \mathbf{J}\cdot\mathbf{E}\partial t = 1/T\int_0^T (\mathbf{J}_0 + \tilde{\mathbf{J}})\cdot(\mathbf{E}_0 + \tilde{\mathbf{E}})\partial t$$

$$= 1/T\int_0^T (\mathbf{J}_0\cdot\mathbf{E}_0 + \mathbf{J}_0\cdot\tilde{\mathbf{E}} + \tilde{\mathbf{J}}\cdot\mathbf{E}_0 + \tilde{\mathbf{J}}\cdot\tilde{\mathbf{E}})\partial t$$

The first term in the resulting integrand determines the specific power transferred by the "unperturbed" flux. The second term determines the power of the interaction of the flux constant current component with the alternating field. The third term determines interaction between the field constant component and the variable component of the convection current. The last term is equal to the product of the variable current and field components.

Under the harmonic law of the changes in the current and field variables components in time, the integrand terms $\mathbf{J}_0\tilde{\mathbf{E}}$ and $\tilde{\mathbf{J}}\mathbf{E}_0$ do not contribute to the average interaction power, i.e.

$$\int_0^T (\tilde{\mathbf{J}} \cdot \mathbf{E}_0 + \mathbf{J}_0 \cdot \tilde{\mathbf{E}})\partial t = 0 \tag{4.7}$$

Thus,

$$p = 1/T \int_0^T \mathbf{J} \cdot \mathbf{E}\partial t = 1/T \int_0^T (\mathbf{J}_0 \cdot \mathbf{E}_0 + \tilde{\mathbf{J}} \cdot \tilde{\mathbf{E}})\partial t \tag{4.8}$$

We note that expression (4.7) is valid only when the field and current in the interaction area do not depend on coordinates.

The integral of the product $\mathbf{J} \cdot \mathbf{E}$ determines the specific power of the alternating current interaction with the time-varying field. Therefore, after averaging (4.3) and (4.4) with respect to the period of oscillations, we obtain

$$\Delta\left[n\left(\frac{mv_d^2}{2} + \frac{3kT_e}{2}\right)\right] + \int_0^T \tilde{\mathbf{J}} \cdot \tilde{\mathbf{E}}\mathrm{d}t = \frac{1}{T}\int_0^T \frac{3}{2}\frac{nk(T_e - T_0)}{\tau_\varepsilon(T_e)}\mathrm{d}t; \tag{4.9}$$

$$\Delta\frac{nmv_d^2}{2} - \int_0^T \tilde{\mathbf{J}} \cdot \tilde{\mathbf{E}}\mathrm{d}t = 0. \tag{4.10}$$

Expressions (4.9) and (4.10) show that in vacuum the charge carrier flux interaction with the alternating electric field leads to the change in the kinetic energy of the charge carriers. In semiconductors, this interaction leads not only to the change in the kinetic energy of the directed motion $nv_d^2/2$ of charges but also to the change in the chaotic (thermal) component of their energy $3/2nkT_e$ due to scattering.

If the charge carriers (for example, electrons) move in a solid under the influence of an alternating electric field with a constant drift velocity, it may seem that their kinetic energy does not change and energy exchange with the electric field does not occur. However, this is not right. The electron velocity is composed of two components: the thermal motion velocity and the velocity acquired by the action of the electric field. Let us consider an electron moving in a crystal lattice. In the absence of an external field, it moves between collisions rectilinearly with a "thermal"

velocity determined by the lattice temperature T. Its energy is $mv_T^2/2 = kT$. Under the action of the field, its velocity between collisions increases and acquires a component directed along the field (the drift velocity). In collisions, the electron gives part of the energy to the ions and the lattice heats up to a higher temperature than in the absence of the field. In the sufficiently strong electric field, the drift velocity remains constant and the field energy is expended mostly for the carriers thermal energy increasing. This process determines the increase of the electron gas temperature T_e which becomes larger than the lattice temperature. Such electrons are called "hot". They give energy to the lattice, which heats up and the sample comes to thermodynamic equilibrium.

If the field contains both constant and variable components, at times when the field is less than the average value, it "warms" the electrons less, and at times when the field is larger than the average value, the heating of the electrons increases. If the number of electrons in the interaction area in moments of the weak field is larger than at times when the field is larger than the average, the temperature of the electron gas and the lattice is lower than in the absence of the alternating field. Thus, part of the constant field energy is transferred to the alternating field that is required for the device's operation. The condition for such an energy exchange is the modulation of the carrier flux density and the correct phasing of the current and field. They must be shifted in phase by more than 90° at least.

Let us assume that the carrier flux with current density $\tilde{\mathbf{J}} = J_z(z,t)\mathbf{e}_z$ and cross section S interacts with a known field $\tilde{\mathbf{E}} = E(z,t)\mathbf{e}_z$. The interaction area is bounded by planes with coordinates $z = 0$ and $z = d$. Instantaneous interaction power:

$$P(t) = \int_{\Delta V} \mathbf{JE}\mathrm{d}V = \int_0^d J_z(z,t)E(z,t)S\mathrm{d}z = \int_0^d I_z(z,t)E(z,t)\mathrm{d}z \tag{4.11}$$

For a short gap length d, when it is possible to take $I_z(z,t) \approx I_z(0,t)$ the (4.11) has the simple form:

$$P(t) \approx I_z(0,t)\int_0^d E(z,t)\mathrm{d}z = I_z(0,t)U(t)$$

where $U(t) = \int_0^d E(z,t)\mathrm{d}z$ is the potential difference (voltage) of the gap. Thus, the power in this case is calculated as in AC circuits.

Consider a flux of charged particles with cross-sectional area S characterized by a convection current

$$I_k = I_0 + \dot{I}\mathrm{e}^{i(\omega t+\varphi)} \tag{4.12}$$

and interacting with an alternating electric field $\mathbf{E} = \dot{E}\mathrm{e}^{i(\omega t+\psi)}\mathbf{e}_z$. We assume that the particles move along the z axis. In accordance with (2.15), the interaction power on the $\mathrm{d}z$ segment is given by

$$dP = qn\mathbf{v} \cdot \mathbf{E}Sdz = 0.5\dot{I}\dot{E}^* dz. \tag{4.13}$$

As can be seen, the power of the interaction is a complex quantity:

$$dP = dP_{re} + i dP_{im}$$

where

$$dP_{re} = \frac{1}{2}\dot{I}\dot{E}^* \cos(\varphi - \psi); \tag{4.14}$$

$$dP_{im} = \frac{1}{2}\dot{I}\dot{E}^* \sin(\varphi - \psi). \tag{4.15}$$

The real part of the power can be either positive or negative, depending on the phase difference between the alternating current and field components. In the first case, the field gives part of its energy to the charged particles. Such a mechanism is used in high-frequency accelerators of charged particles. In the second case, the particles give part of their energy to the field increasing its intensity. It is this interaction mechanism that is used in electronic devices.

The imaginary part of the interaction power (reactive power) becomes equal to zero after averaging over the field changing period. Therefore, it does not change the energy of the field and particles, but changes ratio of electric and magnetic field energes.

Equation (4.13) can also be written in the following form:

$$dP = \frac{1}{2}\dot{I}\dot{E}_z^* \frac{dz}{dt} dt = \frac{1}{2}\dot{I}\dot{E}_z^* \dot{v} d\tau, \tag{4.16}$$

where $d\tau$ is charged particles transit time through the segment dz. When integrating this relationship in the finite range of τ, it should be taken into account that the velocity and time of transit contain variable components that depend on the high-frequency field amplitude.

4.2 Interaction with Quasi-Static Field, the Induced Current. The Shokley-Ramo Theorem

Let's consider a system of electrodes near which a charged particle with charge q moves (Fig. 4.1). This charge induces opposite-sign charges on the electrodes, the sum of which is $-q$. When the particle moves, the charges induced on the electrodes change. The charge inceases on those electrodes, to which the particle approaches and decreases on those from which the particle moves away. Therefore, in the circuit connecting the electrodes, a current arises which is named the *induced current*.

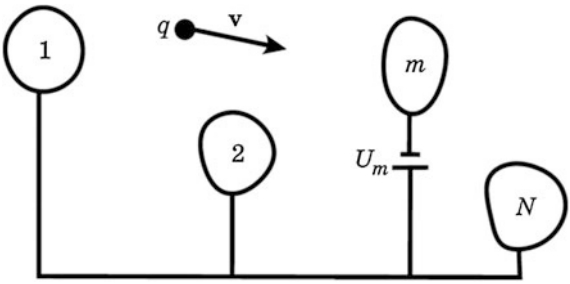


Fig. 4.1 Charge moving in the electrode system

Let's define the current flowing in the electrode circuit with the number m. To do this, let us assume that a potential U_m is applied to this electrode, while the remaining electrodes are at zero potential. The electrode m creates an electric field $\mathbf{E}_m$ at the location of the particle. The work that this field does moving this particle over a distance d$\mathbf{r}$ during the time dt is given by the formula

$$\mathrm{d}A = \mathbf{F} \cdot d\mathbf{r} = q\mathbf{E} \cdot d\mathbf{r} = q\mathbf{E}_m \cdot \frac{d\mathbf{r}}{\mathrm{d}t}\mathrm{d}t = q\mathbf{E}_m \cdot \mathbf{v}\mathrm{d}t,$$

where $\mathbf{v}$ is the particle velocity acquired under the action of the field. On the other hand the current induced in the circuit of this electrode does the work

$$\mathrm{d}A = I_{im}U_m\mathrm{d}t.$$

Equating these two expressions in accordance with the energy conservation law, we obtain

$$I_{im} = q\frac{\mathbf{E}_m}{U_m}\mathbf{v} = q\mathbf{E}_{1m}\mathbf{v}, \tag{4.17}$$

where $\mathbf{E}_{1m}$ is the electric field produced by the m-th electrode when unit potential is applied to it.

Formula (4.17) expresses *the Shockley-Ramo theorem* for calculating the induced current. This theorem allows determination of the current induced in the circuit of any electrode, if the electric field created by it is known. Let us recall that formula (4.17) does not take into account the potential delay and can therefore be used only in those cases when the electric field in the electrode system is close to static. The Shockley-Ramo theorem also does not take into account relativistic effects.

Let's assume that an electron flux with a current density $\mathbf{j(r)} = \rho\mathbf{(r)v(r)}$ propagates in the interelectrode space. Then the induced current produced by electrons contained in the volume dV,

$$\mathrm{d}I_{im} = \mathbf{E}_{1m}\mathbf{v}\rho\mathrm{d}V = \mathbf{E}_{1m}\mathbf{j}_k\mathrm{d}V.$$

The induced current in the circuit of the m-th electrode is found by integrating the obtained expression over the volume of interaction (interelectrode space):

$$I_{im}(t) = \int_V \mathbf{E}_{1m}(\mathbf{r}, t)\mathbf{j}_k(\mathbf{r}(t), t)\mathrm{d}V. \tag{4.18}$$

Note that in the obtained expression, the position of the convection current element depends on time: $\mathbf{r} = \mathbf{v}t$.

4.3 Current in the Flat Interelectrode Gap and Its External Circuit

Suppose that charged particles move along the z axis in a homogeneous alternating electric field with amplitude $\dot{E}\mathbf{e}_z$. The power of this field interaction with the current on a segment with length d and unit area of the cross section is determined in accordance with (4.18) by the integral

$$P = \frac{1}{2}\dot{E}\int_0^d \dot{J}^* \mathrm{d}z.$$

Let's transform this expression:

$$P = \frac{1}{2}\dot{E}d\frac{1}{d}\int_0^d \int_S \dot{J}_k^* \mathrm{d}S\mathrm{d}z = -\frac{1}{2}U I_i^*, \tag{4.19}$$

where $\dot{U} = -\dot{E}d$ is the voltage amplitude of the segment d. The minus sign in this expression appeared because the power transferred from the charged particles to the electromagnetic field is considered positive. As can be seen, in this case, the induced current is equal to the convection current averaged over segment d.

As an example, let's consider the space between two ideally conducting electrodes with area S located at a distance d from each other (*the flat interelectrode gap*) and connected by an external circuit (Fig. 4.2a). The edge effects near the electrode boundaries are ignored. Suppose that in the space between the electrodes, a thin layer of charged particles moves with velocity v.

The width of the layer is $\Delta \ll d$. Convection current in the interelectrode gap

$$I_k = \begin{cases} 0, & z < z_0; \\ \rho v S, & z_0 - \Delta \le z \le z_0; \\ 0, & z_0 < z < d, \end{cases} \tag{4.20}$$

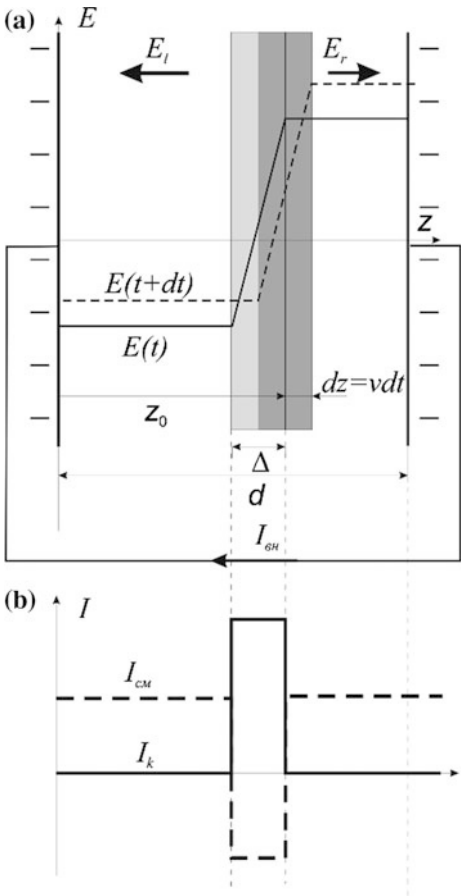


Fig. 4.2 Current in an interelectrode gap and its external circuit

where z_0 is the coordinate of the right boundary of the charges. The charged particles layer induces charges on the electrodes with surface density $\sigma_l = -\rho\Delta(d + \Delta - z_0)/d$ at the left electrode and $\sigma_r = -\rho\Delta z_0/d$ at the right electrode. These charges excite an electric field

$$E = \begin{cases} \sigma_l/\varepsilon, & z < z_0; \\ \left[-\sigma_r + \frac{-\sigma_l+\sigma_r}{\Delta}(z - z_0)\right]/\varepsilon, & z_0 - \Delta \le z \le z_0; \\ -\sigma_r/\varepsilon, & z_0 + \Delta < z \le d. \end{cases}$$

Since the layer of charged particles moves with a velocity v, $z_0 = vt$, if the origin of time is taken as the time when the layer of particles starts leaving the left electrode. As a result, the electric field in the gap varies with time, causing the displacement current.

$$I_{cm} = \varepsilon S \frac{\partial E}{\partial t} = \begin{cases} \rho v S \frac{\Delta}{d}, & 0 < z < z_0 - \Delta; \\ \rho v S \left(\frac{\Delta}{d} - \frac{d}{\Delta}\right) & z_0 - \Delta \le z \le z_0; \\ \rho v S \frac{\Delta}{d}, & z_0 < z < d. \end{cases} \tag{4.21}$$

The distributions of the convection current and displacement current at time t are shown in Fig. 4.2b. In sum, they give the total current that does not depend on the z coordinate.

If we average the total current along the length of the gap, its value will not change:

$$\langle I_{\text{tot}} \rangle = \frac{1}{Sd} \int_0^d \int_S \left(\varepsilon \frac{\partial E}{\partial t} + J_k \right) \mathrm{d}S\mathrm{d}z = \frac{S\varepsilon}{d} \frac{\partial}{\partial t} \int_0^d E \mathrm{d}z + \frac{1}{d} \int_0^d I_k \mathrm{d}z.$$

The first integral is zero, since the electrodes are short-circuited by an external circuit. The second integral is equal to the induced current. This current also flows in the external circuit in accordance with Ampere's circuital law.

Thus, in the interelectrode gap, the convection current and the displacement current flow, and the induced current flows in the external circuit. The induced current begins when the first layer of charged particles leaves the left electrode and ends when the last layer of particles enters the right electrode. The dependence of this current on time is shown in Fig. 4.3. Note that the value of the induced current is the same both when electrons settle down on the rightelectrode and when they go throw it and move further.

If an alternating voltage is applied between the electrodes an additional displacement current flows in the interelectrode gap

$$I_{cm1} = \varepsilon S \frac{\partial E}{\partial t},$$

where $E = -U/d$, and in the external circuit besides the induced current, the capacitance current flows $I_C = C\mathrm{d}U/\mathrm{d}t$, where $C = \varepsilon S/d$ is the interelectrode capacitance.

Let the convection current contain a variable component

$$I_k = I_{km} \cos(\omega t - \beta_e z), \tag{4.22}$$

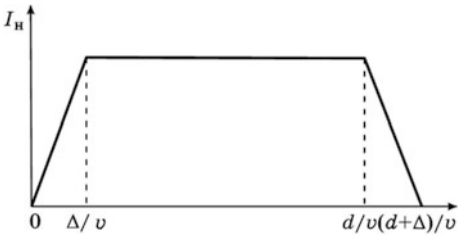


Fig. 4.3 Dependence of induced current on time

where $\beta_e = \omega/v_e$ is the electron phase constant. Let's find the variable component of the induced current:

$$I_i = \frac{1}{d}\int_{-d/2}^{d/2} I_k \mathrm{d}z = I_{km}\frac{1}{\beta_e d}\left[\sin\left(\omega t - \frac{\beta_e d}{2}\right) - \sin\left(\omega t + \frac{\beta_e d}{2}\right)\right] = I_{km}\frac{\sin(\beta_e d/2)}{\beta_e d/2}\cos\omega t. \quad \text{or} \quad \dot{I}_i = M\dot{I}_k. \tag{4.23}$$

The ratio of the induced current amplitude to the convection current amplitude is named the *interaction coefficient M*. From the obtained formula, it follows that for the flat gap with homogeneous field

$$M = \frac{\sin(\theta/2)}{\theta/2}, \tag{4.24}$$

where $\theta = \beta_e d = \omega d/v_e$ is the transit angle of the electron in a flat gap. Expression (4.24) can be easily generalized on gaps with an inhomogeneous electric field with the help of formula (4.18) Thus, the real flat gap of width d can be replaced by an infinitesimally narrow gap located in the middle of the real gap in which the induced current is by M times smaller than the convection current.

Figure 4.4 shows the dependence of the interaction coefficient on the transit angle in the gap. As can be seen, as the transit angle tends to zero, the amplitude of the induced current tends to the amplitude of the convection current. The interaction of the gap field with the electron flux in this case is the most effective. At transit angle of 180°, the interaction coefficient is 0.63, and at transit angle of 360° it vanishes. Therefore, in electronic devices (vacuum and semiconductor devices) using the interaction in narrow gaps, the transit angle should not exceed 120° (M = 0.826).

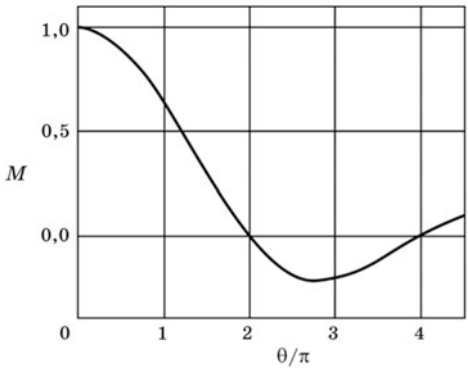


Fig. 4.4 Dependence of the interaction coefficient on transit angle

At large transit angles, the interaction coefficient becomes negative. This means that the current induced in the external circuit changes its direction and flows towards the convection current.

In general case, the field strength in the interaction space depends on the longitudinal coordinate: $E_z = E_z(z)$. We assume that the convection current also depends on the coordinate z. Let us divide the interaction space into sections with length $\mathrm{d}z$ and assume that in each such section the field and the convection current do not depend on the coordinate. Then the voltage in this section is $\mathrm{d}U = -E_z(z)\mathrm{d}z$ and the convection current is $I_k(z)$. In a section with length $\mathrm{d}z$, the convection current causes the induced current

$$\mathrm{d}I_i = E_1 I_k \mathrm{d}z = \frac{E_z(z)}{|E_z(z)|\mathrm{d}z} I_k(z)\mathrm{d}z.$$

To obtain induced current we integrate this expression over the gap length:

$$I_i = \frac{\int_0^d E_z(z) I_k(z)\mathrm{d}z}{\int_0^d |E_z(z)|\mathrm{d}z}. \tag{4.25}$$

Since the variable component of the convection current depends on the coordinate according to law (4.22), the formula (4.25) for the induced current amplitude takes the form

$$\dot{I}_H = I_{km} \frac{\int_{-d/2}^{d/2} E_z(z) \cos \beta_e z \mathrm{d}z}{\int_{-d/2}^{d/2} |E_z(z)|\mathrm{d}z}. \tag{4.26}$$

Until now, we have supposed that the outer circuit has no resistance. Now let's assume that it contains the resistor R. The induced current creates the voltage $U = RI_i$ on the resistor, and this induces an additional electric field in the gap $E_{ext} = -U/d$. This field decelerates positively charged particles, so their kinetic energy at the exit from the gap is less than at the input. The energy lost by the particles is transferred to the electromagnetic field. Thus, the interelectrode gap can cause induced slowing-down radiation from the short layer of charges moving in this gap.

4.4 Electric Gap Field Effect on the Motion of Charged Particles

Consider an electron flux in vacuum passing a flat gap of the resonator formed by two parallel grids located at a distance d_1 from each other. We assume the grids are opaque for the field and transparent for the electrons.

Let us introduce the rectangular coordinate system whose z-axis is normal to the grids. Let the first grid be located in the plane $z = 0$, and the velocity of the particles entering the resonator gap $v_0 = \sqrt{2eU_0/m_0}\mathbf{e}_z$, where U_0 is the accelerating voltage (relativistic effects are not taken into account). High-frequency voltage $U_1 = U_{m1} \sin \omega t$ is applied between the grids. Then, for the electric field strength in the gap, we obtain

$$E = -(U_{m1}/d_1)sin\omega t.$$

The equation of the motion of electrons in this field has the form

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{eU_{m1}}{m_0 d_1}\cos(\omega t) + \varphi.$$

Integrating this equation, we obtain

$$v = -\frac{eU_{m1}}{m_0 \omega d_1}\cos \omega t + C.$$

We define the integration constant C, using the initial conditions. Assuming that the electron enters the interaction gap ($z = 0$) at time $t = t_0$ and its velocity at this moment is $v = v_0$ we find

$$C = v_0 + \frac{eU_{m1}}{m_0 \omega d_1}\cos \omega t_0.$$

As a result, we obtain the expression for the electron velocity

$$v = v_0 + \frac{eU_{m1}}{m_0}\omega d_1(\cos \omega t_0 - \cos \omega t). \tag{4.27}$$

Let's determine the velocity of electrons at the exit from the gap at time $t_1 = t_0 + \tau$, where τ is the electron transit time between the grids of the interaction gap. In the general case, this time is determined by the initial electron velocity v_0 and by the variable velocity v which the electron acquires during interaction with the alternating gap field.

Assuming that the amplitude of the variable voltage is much less than the accelerating voltage, $U_{m1} \ll U_0$ we can neglect the effect of the variable velocity component on the transit time and calculate the transit time by the formula $\tau \approx d_1/v_0$. Then

$$v = v_0 + \frac{eU_{m1}}{m_0 \omega d_1}[\cos(\omega t_1 - \omega\tau) - \cos\omega t_1].$$

Using the well-known relation for the cosine difference of two angles, we obtain

$$v = v_0 + \frac{eU_{m1}}{m_0 \omega d_1} 2 \sin\left(\omega t_1 - \frac{\omega\tau}{2}\right) \sin\frac{\omega\tau}{2}$$

Transforming this expression, we obtain

$$v = v_0 \left[1 + \frac{1}{2}\frac{U_{m1}}{U_0}\frac{\sin(\theta_1/2)}{\theta_1/2}\sin(\omega t_1 - \theta_1/2)\right], \tag{4.28}$$

where $\theta_1 = \omega d_1/v_0$ is the unperturbed angle of the first resonator gap transit. In this formula, the finite width of the gap is taken into account by the factor $\sin(\theta_1/2)/(\theta_1/2)$ equal to the interaction coefficient (2.3.20) introduced in par. 3.2.2 in the analysis of the induced current in the gap. In both cases, the interaction coefficient allows the replacement of the real gap by the ideal infinitely narrow one located in the middle plane of the real gap. Introducing the velocity modulation factor

$$v = \frac{1}{2}\frac{U_m}{U_0}M,$$

we obtain

$$v = v_0 + v_m \sin(\omega t_1') = v_0[1 + v_1 \sin(\omega t_1')], \tag{4.29}$$

where $v_m = v_1 v_0$ is the variable electron velocity amplitude, and $t_1' = t_1 - \tau_1/2$ is the moment when an electron passes the middle of the gap. In the following, we omit the prime at t_1.

We see that the amplitude of the velocity is directly proportional to the amplitude of the alternating voltage U_{m1} and the interaction coefficient in the first gap M_1. Usually, system parameters are selected so that the value of M_1 is in the range 0.7–0.9, that corresponds to the transit angles $\theta_1 = 90{-}150^\circ$.

The motion of charged particles in electronic devices in the one-dimensional approximation is clearly depicted in so-called *space-time diagrams* (Applegate diagrams).

In these diagrams, the abscissa represents the time, and the ordinate represents the coordinate along which the particles move. For each charged particle that enters the interaction area at a certain time, a trajectory is constructed that is its coordinate dependence on time. The particle velocity at the given time is equal to the tangent of the slope of the trajectory. The charge density at the given moment at the point with the given coordinate is proportional to the number of trajectories per unit length (if the particles have equal charges and enter the interaction area at regular time intervals). An example of the space-time diagram for five particles moving in a constant accelerating field is shown in Fig. 4.5.

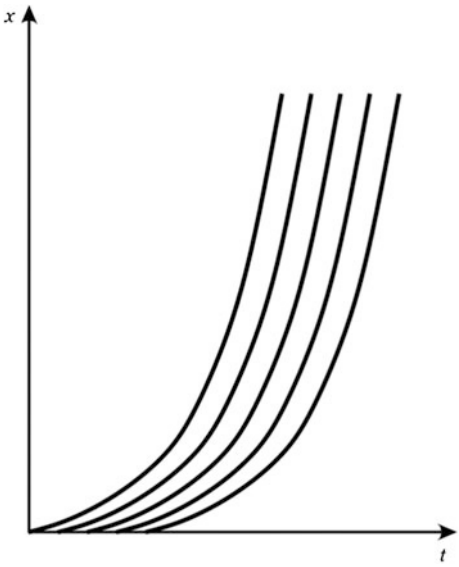


Fig. 4.5 Space-time (Applegate) diagram

4.5 Energy Exchange Between Electrons and the Gap Field

Let us calculate kinetic power of the electron beam at the entrance to the gap. The kinetic energy of one electron is $w_1 = mv^2/2$, and the number of electrons entering the gap per unit time is equal to nvA, where $n = \rho_0/e$ is the electron concentration in the beam, and A is its cross-sectional area. Since at the entrance to the gap all electrons have equal velocities $v = v_0$ the kinetic power at the entrance to the gap is

$$P_{k1} = \frac{mv_0^2}{2} nv_0 A = \frac{mnv_0^3 A}{2}. \tag{4.30}$$

Electrons in the gap undergo velocity modulation, so their velocities are not equal at the exit from the gap. For small modulating voltages, the velocity of electrons is determined by formula (4.29). If electrons transit through the infinitely narrow gap, their velocities change, but their mutual positions do not alter. Therefore, the electron density at the exit of the gap remains constant and the kinetic power of the flux averaged over the period of oscillations is equal to

$$\bar{P}_{k2} = \frac{mnA}{2} \frac{1}{T} \int_0^T v^3 \mathrm{d}t = \frac{mnAv_0^3}{2} \frac{1}{2\pi} \int_0^{2\pi} (1 + v_1 \sin \omega t)^3 \mathrm{d}\omega t.$$

Let us transform the integral expression:

$$\frac{1}{T}\int_0^T (1+v_1\sin\omega t)^3 \mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} (1+3v_1\sin\varphi + 3v_1^2\sin^2\varphi + v_1^3 sin^3\varphi)\mathrm{d}\varphi.$$

Since $v_1 \ll 1$, the third and forth terms in parentheses can be ignored and in this case the written expression is equal to one. Thus, the kinetic powers of the flux at the gap entrance and exit are equal, i.e. the energy exchange between the electrons and the gap field is absent.

The energy exchange in a gap of finite width is different. The accelerated electrons pass through the gap in less time than the decelerated ones. Therefore, the interaction coefficient for the accelerated electrons is larger than for the decelerated ones. As a result, the accelerated electrons acquire higher velocity changes than the decelerated electrons, i.e. in general, the electron flux at the exit from the resonator has a kinetic energy larger than at the entrance. The electrons obtain excess kinetic energy from the electromagnetic field of the gap. This effect can be reflected in the equivalent resonator circuit by inserting parallel to the gap some conductivity $Y_e = G_e + jB_e$, called the *electron conductivity*.

To calculate the electron conductivity we use the expression for electron velocity in the gap (4.28). Integration gives the coordinate of the electron

$$z = v_0(t-t_0) - \frac{eU_m}{\omega^2 md}[\sin\omega t - \sin\omega t_0 - (\omega t - \omega t_0)\cos\omega t_0].$$

To find electron transit time through the gap $\tau = t - t_0$ we substitute $z = d$ in the obtained expression:

$$d = v_0\tau - \frac{eU_m}{\omega^2 md}[\sin\omega t - \sin(\omega t - \omega\tau) - \omega\tau\cos(\omega t - \omega\tau]. \tag{4.31}$$

In order to solve this transcendental equation we suppose

$$\tau = \tau_0 + \delta\tau, \quad \delta\tau \ll \tau_0,$$

where $\tau_0 = d/v_0$ is the "unperturbed" transit time. Such a representation is true if $U_m \ll U_0$. Using the small parameter expansion $\omega\delta\tau$ of the trigonometric functions and ignoring the higher terms of the expansion, we obtain

$$\sin(\omega t - \omega\tau) \approx \sin(\omega t - \theta) - \omega\delta\tau\cos(\omega t - \theta);$$

$$\cos(\omega t - \omega\tau) \approx cos(\omega t - \theta) + \omega\delta\tau\cos(\omega t - \theta),$$

In these expressions, $\theta = \omega\tau_0$ is the unperturbed transit angle.

Let the elementary charge layer $dq = I_{k0}dt_0$ enter the gap through the first grid during time dt_0. In accordance with (2.27) this charge induces the current in the external circuit

$$dI_i = I_0 dt_0 v/d.$$

The summarized induced current at time t is determined by integration over all elementary charges that are currently in the gap:

$$I_H = \frac{I_0}{d}\int\limits_{t-\tau}^{t}\left[v_0 - \frac{eU_m}{\omega md}(\cos\omega t - \cos\omega t_0)\right]dt_0$$

$$= \frac{I_0}{d}\left\{v_0 t - \frac{eU_m}{\omega^2 md}[\omega\tau\,\cos\omega t - \sin\omega t + \sin(\omega t - \omega\tau)]\right\}.$$

Expanding $\sin(\omega t - \omega\tau)$, we obtain

$$I_i = \frac{I_0}{d}\left\{v_0\tau - \frac{eU_m}{\omega^2 md}[\theta\,\cos\omega t - \sin\omega t + \sin(\omega t - \theta)]\right\}.$$

Substituting the value of $v_0\tau$ from (4.31) into this expression after a series of transformations, we obtain

$$I_i = \frac{U_m I_0}{U_0}[F_1(\theta)\,\sin\omega t + F_2(\theta)\,\cos\omega t], \tag{4.32}$$

where

$$F_1(\theta) = \frac{2(1-\cos\theta) - \theta\,\sin\theta}{2\theta^2}, \tag{4.33}$$

$$F_2(\theta) = \frac{2\sin\theta - \theta(1+\cos\theta)}{2\theta^2}. \tag{4.34}$$

As can be seen, the induced current contains the active component I_{i1} being in phase with the gap voltage, and reactive component I_{i2} shifted in phase by 90° with respect to the voltage. Accordingly, it is possible to determine the active and reactive components of the electron beam conductivity (*electron conductivity*):

$$G_e = I_{i1}/U_m; \quad B_e = I_{i2}/U_m.$$

From (4.32) it follows that these conductivities depend on the unperturbed transit angle:

$$G_e = G_0 F_1(\theta); \quad B_e = G_0 F_2(\theta),$$

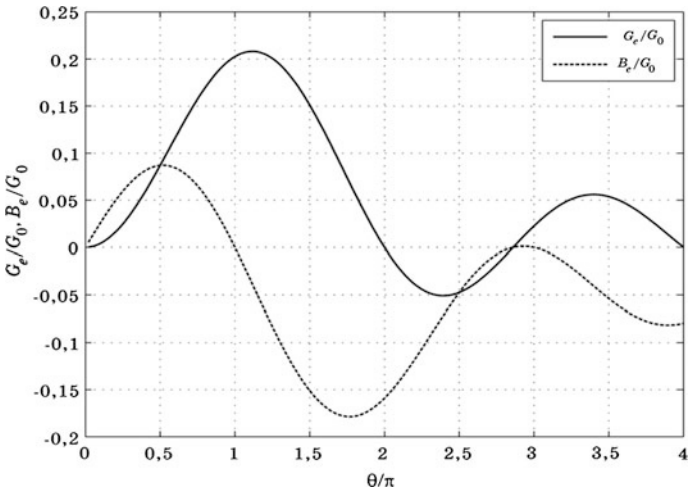


Fig. 4.6 Electron conductivity versus transit angle

where $G_0 = I_0/U_0$ is the conductivity of the beam in direct current. The graphs of the functions G_e/G_0 and B_e/G_0 are shown in Fig. 4.6. It can be shown that

$$\frac{G_e}{G_0} = -\frac{\beta_e}{4}\frac{\mathrm{d}|M^2|}{\mathrm{d}\beta_e}.$$

The average power of the gap field interaction with the electron flux is

$$P_e = 0.5U_m^2 G_e. \tag{4.35}$$

The positive power corresponds to the transfer of energy from the field to the electron flux, while the negative power means that energy is tranfered from the flux to the gap field.

As can be seen in Fig. 2.4.6, for small transit angles the active component of electron conductivity is positive, i.e. the electron beam takes energy from the alternating electric gap field. In other words, the average kinetic energy of the electron beam increases after the passage of the gap. However, in the range of transit angles $\theta = 2\pi$–2.8π the active conductivity component, and hence the interaction power, become negative. This means that electrons give part of their energy to the gap field. If an oscillating circuit tuned to the frequency corresponding to these transit angles is connected to the gap, and if the power transmitted by the electron flux is larger than the power losses in the circuit, self-excited oscillations (*the monotron effect*) will appear in such a system.

The reactive component of electron conductivity at small transit angles ($0 < \theta < \pi$) is positive (the capacitive conductivity) while at large transit angles it is negative. It shifts the resonant frequency of the resonator, part of which is the interelectrode gap.

4.6 Interaction of Charged Particles with a Travelling Wave Field

Let's consider the CP beam with modulated charge density having a cross-sectional area S and velocity v_0. Let the charge density have the constant and harmonic variable components:

$$\rho = \rho_0 + \dot{\rho} e^{j(\omega t - \beta_e z)}, \tag{4.36}$$

where $\beta_e = \omega / v_0$ is the "electronic" phase constant.

The convection current also has the constant and variable components:

$$I_i = -\int_S \rho v_0 dS = I_0 + \dot{I} e^{j(\omega t - \beta_e z)}. \tag{4.37}$$

The integration is carried out over the cross-sectional area of the beam. The electron flux moves in the electromagnetic wave field

$$E_z = \dot{E} e^{j(\omega t - \Gamma_0 z)}, \tag{4.38}$$

where $\Gamma_0 = \beta_0 - j\alpha_0$ is the wave propagation constant. Let's write the power of the interaction on section dz:

$$dP = \frac{1}{2} \dot{I} \dot{E}_z^* e^{-j(\beta_e - \Gamma_0) z} dz. \tag{4.39}$$

The power of an electromagnetic wave is given by

$$P_0 = \frac{|\dot{E}_z|^2}{2\beta_0^2 R_c},$$

where R_c is the *interaction impedance* characterizing the intensity of the wave field longitudinal component. The larger R_c, then the greater the longitudinal component of the wave field for the given transferred power.

The variable component of the current arises when an alternating field is applied to the beam. In the small-signal approximation we can assume that the current amplitude is proportional to the field amplitude: $\dot{I} = a\dot{E}_z$. In this approximation, formula (4.39) can be rewritten:

$$dP = \frac{1}{2} a |E_z|^2 e^{-j(\beta_e - \Gamma_0) z} dz = a P_0 R_c \beta_0^2 e^{-j(\beta_e - \Gamma_0) z} dz. \tag{4.40}$$

Hence, the interaction power on section l:

$$P = aP_0R_c\beta_0^2\int_0^l \mathrm{e}^{-\mathrm{j}(\beta_e-\Gamma_0)z}\mathrm{d}z = \frac{aP_0R_c\beta_0^2}{\alpha_0 - i(\beta_e-\beta_0)}\left[\mathrm{e}^{\alpha_0 l}\mathrm{e}^{-\mathrm{j}(\beta_e-\beta_0)l} - 1\right] \quad (4.41)$$

It can be seen from this expression that the most intense interaction is observed under conditions of synchronism when $\beta_0 = \beta_e$, i.e. when the wave and the charged particles move with identical velocities. Setting $\alpha_0 = 0$ and assuming $(\beta_e - \beta_0)l \ll 1$ from formula (4.41), we obtain

$$P = aP_0R_c\beta_0^2 l,$$

i.e. the power of the interaction is positive (the field transmits the energy to the electrons) and is proportional to the interaction space length. Such a conclusion arises as a result of the assumption made that the amplitude of the convection current variable component is proportional to the amplitude of the wave field.

In the absence of synchronism, a reactive component of the interaction power arises that changes the phase of the electric field. In more detail, the interaction between the CP and the traveling wave is considered in the sections devoted to devices of the corresponding types.

Advancement Questions

1. How is the power of the electromagnetic field interaction with the electron flux determined?
2. What is an induced current? State the Shockley-Ramo theorem of induced current. Under what conditions is it fair?
3. What types of currents flow in the interelectrode gap and its external circuit? Describe the relationship between them.
4. What is the interaction coefficient? What is the physical meaning of the negative interaction coefficient?
5. Describe the phenomenon of electron flux velocity modulation in the interelectrode gap.
6. What is an Applegate diagram? What processes can be analyzed using this diagram?
7. What is electron conductivity? Explain the physical meaning of the real and imaginary components of this conductivity.
8. Describe the process of the interaction of electrons with the traveling wave field.

Chapter 5
A Microwave Device as a Circuit Element

5.1 Microwave Devices Requirements

An electronic device is an integral part (element) of the facility, which, in turn, is a part of the radio electronic system, for example, radar or communication system. Therefore, the device requirements are determined by the facility or system in which it will be used. Historically, the most important applications of microwave devices were radiolocation and communication. Another part of the application is energetic, where due to the energy of microwave oscillations, various technological processes are carried out: heating, sintering and melting of materials, heating of gas plasma, acceleration of charged particles, etc.

As an example, let's consider what requirements are applied to the amplifiers and generators operating in radar and communication systems. For this, let us write two equations: the equation for determining the maximum range of the radar system (radar equation) and the Hartley-Shannon law relating to the information capacity of a signal channel.

The maximum distance of radar is determined by the equation

$$R_{\max} = \left[\frac{P_t G^2 \sigma S_a}{P_r (4\pi)^2}\right]^{1/4} \tag{5.1}$$

where G is the antenna gain factor; P_t is the transmitter power; P_r is the receiver sensitivity; and S_a is the area of the antenna; and σ is the target radar cross-section (RCS) (it is assumed that the same antenna is used for both transmission and reception).

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_5

According to the Hartley-Shannon theorem, the information capacity is determined by the relation

$$C = \Delta f \log_2(1 + P_s / P_n), \tag{5.2}$$

where C is the information capacity; f is the bandwidth of the communication channel; P_s is the power of the useful signal; and P_n is the power of noises and undesirable interference.

Equation (5.1) shows that in order to increase the radar operation distance it is necessary to increase the ratio of transmitter power P_t to receiver sensitivity P_r. Similarly, to increase information capacity (5.2) it is necessary to increase the ratio of the useful signal received from the transmitter P_s to the power of the noise signal received from the outside and added in the receiver P_n. The speed of information transfer is also proportional to the bandwidth of the communication channel Δf. So a MED design must provide necessary output power, bandwidth and noise factor.

In addition, numerous requirements are applied to the devices regarding their weight, dimensions, reliability, durability, resistance to various external factors, cost, and so on.

5.2 Classification of Microwave Devices

All microwave electronic devices use the interaction of charged particles with a high-frequency electromagnetic field. This interaction occurs in three media: vacuum, gas plasma and solids. Accordingly, three large classes of devices are distinguished: *vacuum*, *plasma* and *solid-state*. This does not mean that, for example, for vacuum processes there is no need to take into account the processes in a solid and for solid-state processes—in vacuum and in plasma, etc. However, these processes are of an auxiliary nature, providing, for example, the creation of an electron beam in vacuum devices or the protection of a structure from external effects in semiconductor devices. The device class is determined by the medium in which the interaction processes are carried out. From this point of view, the following definitions for classes of microwave electronic devices can be given.

1. **Vacuum electron devices** (VED) are devices in which the process of field interaction with electron flux occurs in high vacuum. Therefore, when analyzing the interaction processes in these devices, one can ignore collisions of electrons with ions of residual gases. Microwave VEDs appeared first historically and they include the largest number of types and modifications.
2. **Plasma devices**, on the contrary, use gas-discharge plasma in the interaction space. At present, the so-called *microwave gas discharges* (blocking tubes) serving as the protection of receivers from high-power electromagnetic radiation are of the greatest use. Plasma is also used in some analogues of vacuum devices to increase the efficiency of the electron flux interaction with the electromagnetic

field. Such devices have limited application and are not manufactured industrially.

3. In **solid-state** devices, the interaction takes place in a solid—semiconductor, ferroelectric, ferrite, or superconductor. To amplify and generate microwave oscillations of low power (up to 100 W) and for various transformations of the signal spectrum, solid-state devices are used almost exclusively. The most widely used are semiconductor devices, which have the maximum number of types.

In this book, only devices intended for amplification, generation and transformation of the electromagnetic spectrum are considered. Therefore, ferroelectric, ferrite and superconductor devices are not included, although they can be used to construct parametric amplifiers, and superconducting devices can also be used to generate microwave oscillations. Masers (quantum microwave amplifiers) and plasma microwave devices are also not considered.

As noted in Chap. 2, in order to extract energy from the CP ensemble it is necessary to ensure its bunching and (or) phasing. To do this, the electronic devices use electronic flux control facilities. Output power takeoff elements extract power from the bunched (phased) PC flux. In some devices, these facilities can be combined. In addition, constant electric and magnetic fields are used in vacuum and some solid-state microwave devices to ensure the efficiency of interaction.

Microwave devices can be classified according to type of control and power takeoff.

VEDs are distinguished by the type of constant fields used in them and by the mechanisms for controlling and energy takeoff from the electron flux.

In *quasi-static* devices, as the name suggests, the electric field is used to control the electron flux, and the dimensions of the interaction space are much smaller than the wavelength, so that the control field in it at any time practically coincides with the electrostatic field.

In devices with *dynamic* control, the electromagnetic field in the interaction space has a wave character and (or) the bunching process takes place during a time interval comparable with the period of field oscillations.

O-type devices with dynamic control use rectilinear electronic beams. A constant magnetic (or electric) field is used only to form an electron beam of the desired shape. In devices of this type, electrons transmit part of their kinetic energy to the field. Devices of type M in which crossed electric and magnetic fields are used, convert the potential energy of electrons into field energy. There are also devices in which electrons transmit the energy of their rotational motion. These devices are called *gyroresonance* or cyclotron resonance masers (CRM).

In type O devices, both *short-term* and *long-term* interactions of the electron flux with the electromagnetic field are used. In addition, in devices with long-term interaction, it is possible for electrons to interact with both the forward and backward electromagnetic wave. There are also hybrid devices of type O in which both short-term and long-term interactions are used.

Type M devices contain a slow-wave structure, which can be either closed or open. In the first case, the electron flux interacts with the standing wave in the system, and in the second case, it interacts with the traveling wave.

In gyroresonance devices, the interaction with a standing or traveling fast wave is used. They can be distinguished as generator or amplifying types. In the first case, the devices use an internal feedback mechanism.

There is also a class of *relativistic* microwave devices designed to produce ultra-high-power oscillations of the millimeter and sub-millimeter bands.

Semiconductor microwave devices are used to generate and/or amplify electromagnetic oscillations, as well as to control these oscillations. The number of contacts distinguishes them to diodes and transistors. There are diodes with positive and negative dynamic resistance. The first (diodes with a *p-n* junction, with a Schottky barrier, and *p-i-n* diodes) are used in oscillation control circuits (in switches, limiters, phase shifters, detectors, and mixers).

Currently, generators and amplifiers use diodes with negative dynamic resistance such as tunnel diodes (TD), impact avalanche and transit-time diodes (IMPATT) and Gunn diodes (DG).

Transistors are divided into bipolar and field types with further divisions by type of material, type of structure, doping, etc. In the microwave band, the Schottky-barrier field-effect transistors (MESFET) and high electron mobility transistors (HEMT) have the greatest use. Special attention is paid to the development of semiconductor devices based on new materials—gallium nitride, silicon carbide, graphene and diamond.

The foregoing is illustrated in Fig. 5.1. In this figure, the specific types of device are not shown, since this would lead to an excessive figure content.

5.3 The Basic Functional Components of Electron Devices

As already noted, the operation of an electronic device is based on the interaction of the high-frequency electromagnetic field with the charged particles flux in a certain volume called the interaction space. It follows that any electronic device must have a facility for creating the CP flux, giving it the necessary parameters, transporting it through the interaction space and withdraw of exhausted particles.

In vacuum devices, such a facility includes an *electron gun* for creating and forming an electron beam, a system for its transportation (*the focusing system*) that eliminates spreading of the beam due to space charge forces, and a *collector* for gathering exhausted electrons. Together, this components comprise the electron optical system of the device (EOS).

In semiconductor devices, an emitter (*p-n* junction or Schottky barrier or ohmic contact) serves to create the CP flux. As a rule, they contain no special transportation system because the lateral flux dimensions are usually much larger than the longitudinal ones, and the constant component of the space-charge field is zero due to the existance of charged particles of opposite sign (ions) in the crystal lattice.

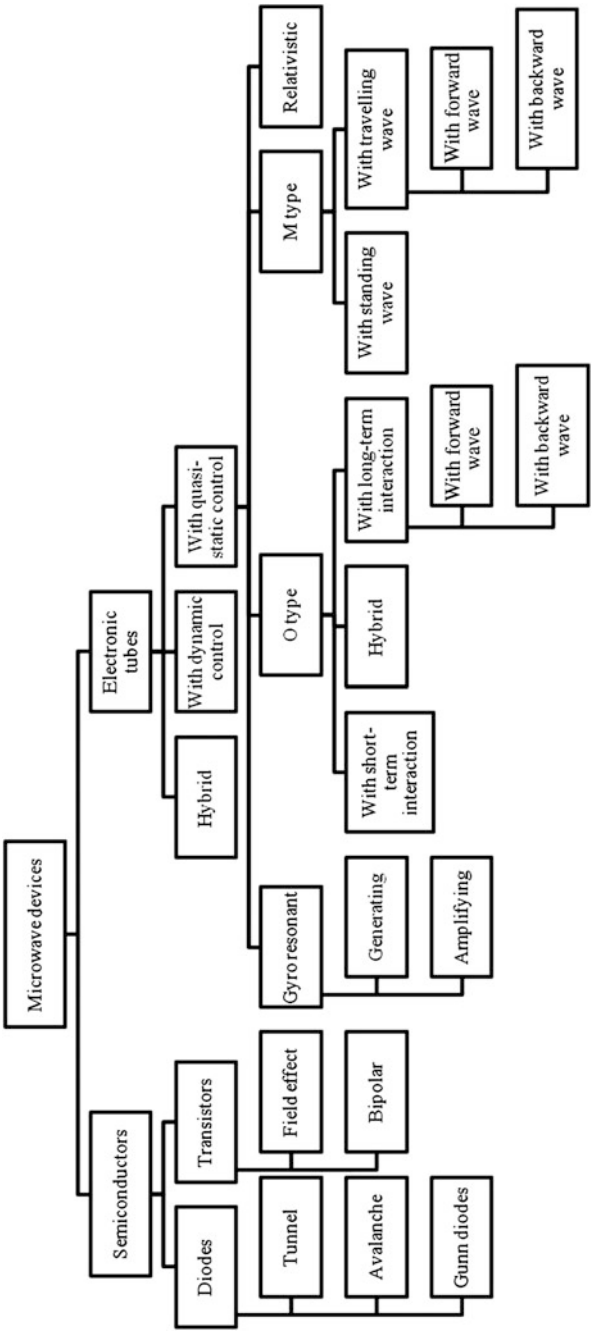


Fig. 5.1 Classification of microwave devices

In the *interaction space,* energy is exchanged between the charged particles and the electromagnetic field. In VED, the electromagnetic field is formed by the *electrodynamics system* of the device (EDS) intended to form the high frequency field of the required frequency and configuration in order to ensure the most effective interaction of this field with the charged particle beam. There are *internal* EDSs that are a part of the device's construction, and *external* EDSs that are not a part of its design. In the latter case, the device is connected to the EDS by special electrodes. The external EDSs are used in vacuum devices operating in the meter and decimeter wavelength band as well as in most semiconductor devices, although some powerful transistors have internal matching circuits. Depending on the device, the EDS creates either standing or travelling electromagnetic waves.

In semiconductor devices, the interaction space is formed by the parts of the device in which there is an alternating electric field.

In addition to these basic elements, electronic devices can have cooling systems, vacuum maintenance systems in the interior, energy input and output systems (often considered as part of the EDS), fastening and buffer systems, and ionizing radiation protection systems.

5.4 Parameters and Characteristics of Microwave Devices

5.4.1 Device Parameters

Like any technical device, microwave devices are characterized by a certain set of characteristics and parameters that allow one to assess the possibility of using the device in a particular system. The parameters of devices are divided into several groups.

The first group is *the design parameters*. It includes the mass of the device, its dimensions, types of connectors, and connecting dimensions.

The second group is *the technical parameters*. Depending on the purpose of the device, it includes various elements. In the main, this book deals with the parameters of amplifying, generating and control devices.

For amplifying devices, the operating frequency and bandwidth of amplified frequencies (measured at -1 or -3 dB level from the maximum power) are indicated. The maximum output power (in continuous or pulsed mode of operation), the minimum permissible duty cycle, and the efficiency are indicated. An important parameter is the power gain factor. For high-power semiconductor devices with a small gain, the added power, i.e. the difference between output and input power, is often indicated. For devices used in low-noise amplifiers, the noise factor or noise temperature is indicated (for more details about these parameters, see the next section).

For generator devices, the operating frequency or frequency tuning band, the output power, the minimum duty cycle, the efficiency and phase noises are

indicated. The short-term and long-term stability of the frequency, as well as parameters determining the electrical and/or mechanical frequency tuning—the tuning band and its differential steepness, the dependence of the output power on the frequency—are of great importance. Usually, the permissible value of the device's load voltage standing wave ratio (VSWR) is indicated.

Detector and mixing diodes are characterized by maximum operating frequency, current sensitivity and breakdown voltage, as well as by a number of other parameters.

Control devices are characterized by the band of operating frequencies, by the maximum power they can control, the actuation time, the insertion loss and isolation.

A large group of *operational parameters* includes nominal values of supply voltages (currents) and permissible deviations from them, the intensity of artificial air or liquid cooling, and reliability (lifetime, meantime between failures, etc.). Allowable intervals of operating temperatures, pressure, humidity, maximum level of ionizing radiation, parameters of vibration resistance, impact resistance and radiation resistance are set. *Electromagnetic compatibility* parameters are often set such as the level of spurious emissions, and susceptibility to external interferences.

Note that the values of many of the listed parameters are determined by the circuit in which the device is employed.

5.4.2 Characteristics of Microwave Devices

Characteristics of the device in graphical or tabular form determine the dependence of its parameters on various external factors. Consider the most important of these dependencies.

The amplitude characteristics of an amplifying device show the dependence of the output power on the input power. Figure 5.2 shows three typical characteristics corresponding to different frequencies.

At low input power levels, the characteristics are linear, i.e. the output power is proportional to the input power. The angle of descent tangent is equal to the power gain factor of the device. With increasing input power, the characteristic begins to deviate from the linear relationship. The value of the allowable deviation (usually – 1 dB) determines the boundary of the linear section of P_l. At a certain value of input power, maximum output power (saturation power) P_s is reached.

At frequencies other than the operating frequency of the device, the amplification factor decreases and the saturation power also changes.

In the absence of an input signal, the power of its own noises P_n is observed at the output of the amplifier. The ratio P_l / P_n is called the dynamic range of the device. Operation within the dynamic range allows minimizing nonlinear and intermodulation distortions, however, the device does not transmit the maximum possible power.

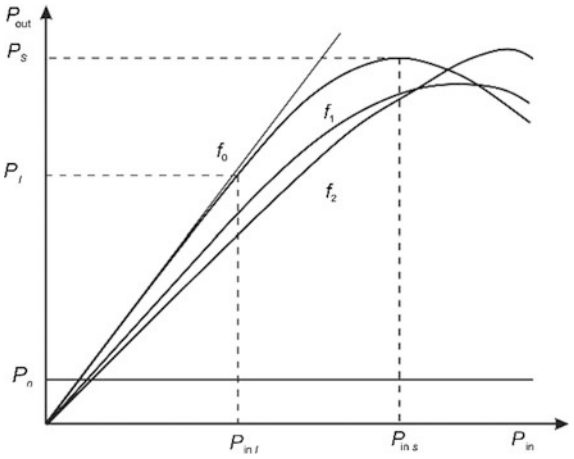


Fig. 5.2 Amplitude characteristic of the amplifier. $f_1 < f_0 < f_2$

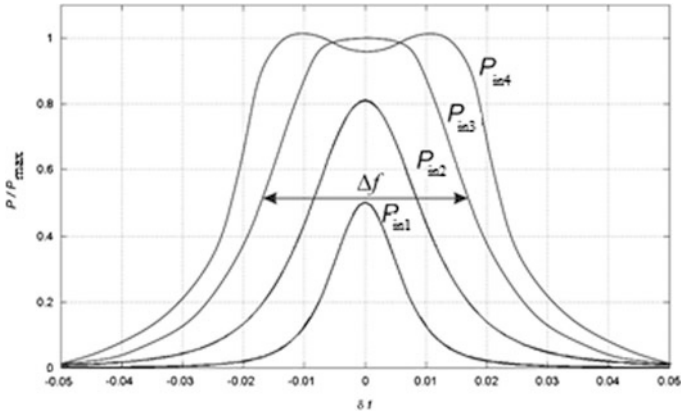


Fig. 5.3 Family of frequency responses; $P_{вх1} < P_{вх2} < P_{вх3} < P_{вх4}$

Therefore, for devices operating in the linear amplification mode (with amplitude modulation) the important parameter is the proximity of P_l and P_s values. Note that the input and output powers are often measured in dBm (decibels relative to 1 mW).

The frequency response (FR) shows the dependence of device output power on frequency at the fixed level of input power. The family of these characteristics constructed for different values of the input power (Fig. 5.3) makes it possible to determine the maximum power of the device and the bandwidth of the amplified frequencies, measured at level −1 ($0.8P_s$) or −3 dB ($0.5P_s$) of the maximum power.

Figure 5.3 shows four characteristics corresponding to different values of input power. Power P_{in1} corresponds to the value at which the maximum output power on

the central frequency (saturation power P_s) is not reached. Input power P_{in4} is larger than the value corresponding to the saturation power. As can be seen, while the power increases, the shape of the FR changes. The width of the working frequency band increases, and at large input powers the curve becomes a double-humped one. Such a change in the FR shape is typical for most amplifiers. In the figure, the relative width of the working frequency band $\Delta f = (f_{max} - f_{min})/f_0$ defined for the curve corresponding to saturation power at the level −3 dB from P_s is shown. As depicted in Sect. 5.1, the maximum possible transmission rate of information by the communication channel is determined together with other parameters by the width of the operating frequency band of the amplifier.

The circuit phase response (**CPR**) displays the dependence on frequency of the signal phase difference at the output and input of the amplifier. The CPR of an ideal amplifier without dispersion in its operating frequency band is linear (Fig. 5.4. Curve 1).

The real CPR deviates from the linear (curve 2 in Fig. 5.4). The maximum deviation of the real characteristic from the ideal in the working frequency band characterizes the *CPR linearity* of the device. This parameter is important for signal amplifiers with phase or frequency modulation.

The frequency response and the circuit phase response can be combined into one **amplitude-phase frequency response** (**APFR**). To do this, a complex gain factor $\dot{\mu} = \mu e^{j\Delta\varphi}$ is introduced, and $\dot{\mu}$ values corresponding to different frequencies are plotted on the complex plane $(Re\,\dot{\mu}, \mathrm{Im}\,\dot{\mu})$. An example of such an amplifier characteristic is shown in Fig. 5.5. It is often called the Nyquist diagram.

A generator performance is characterized by a load chart (Fig. 5.6) showing how the output power and the generation frequency vary depending on the load reflection coefficient. Usually, this characteristic is constructed in the form of frequency and power contour lines on the Smith chart of load impedance (Rieke diagram). This characteristic allows determination of the permissible values of load impedance, at which the device parameters are better than admissible.

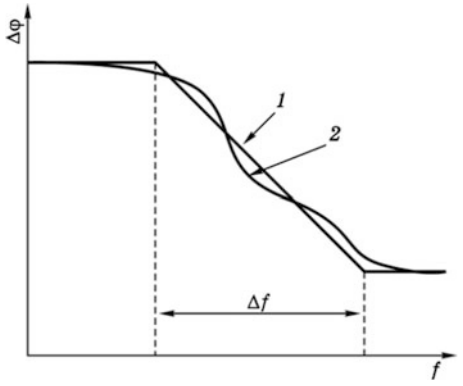


Fig. 5.4 Device phase response

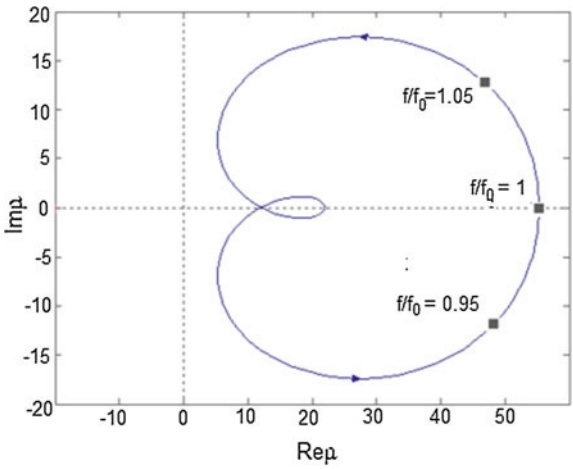


Fig. 5.5 APFR (Nyquist) diagram

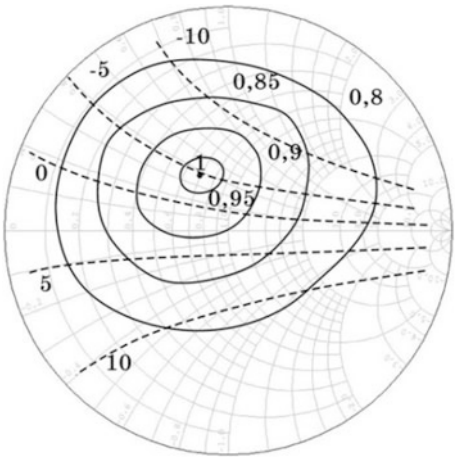


Fig. 5.6 Magnetron performance chart (Rieke diagram) Solid - constant power lines, dashed - constant frequency lines. Figures - deviation from nominal values.

Designers of modern microwave radar and communication systems seek to reduce the transmitter power level by increasing the sensitivity of the receiver. The typical signal level in the receiving antenna is –(100–120) dBm that bespeaks the rigorous requirements regarding noise parameters of these devices. It is important to take into account not only the amplitude noises but also the intermodulation distortions of signals and phase noises.

Advancement Questions

1. List the main types of microwave devices. Describe their common properties and differences.

2. List the main types of vacuum electron devices.
3. List the main types of semiconductor devices.
4. List the main components of the electronic device.
5. List the main groups of electronic device parameters.
6. List the main characteristics of electronic devices.
7. Explain the principle of the construction of a Nyquist diagram.
8. Explain what information is contained in a Ricke diagram.

Part II
Microwave Vacuum Electron Devices

Chapter 6
Devices with Quasi-static Control

6.1 General Characteristics and Parameters of Devices with Quasi-static Control

In devices with quasi-static control, as their name suggests, an electron flux interacts with a quasi-static alternating electric field, i.e., the field in which the curl component of the electric field defined by (2.7) can be ignored by setting $\nabla \times \mathbf{E} = 0$. Then $\mathbf{E} = -\nabla\Phi$, where Φ is the electrostatic potential. Such a field can exist in the gap between the metal electrodes, the distance between which is much smaller than the wavelength of the oscillations.

Quasi-static control is used in *electron tubes* widely applied to amplify and generate electromagnetic oscillations of the radio band. These tubes contain a vacuum bulb B (Fig. 6.1), in which a high vacuum is created, the source of electrons (cathode) C, the anode (plate) A is the electrode to which the constant positive voltage relative to the cathode is applied, and several electron-transparent grids, G_1, G_2,…, G_N, which are under certain constant potentials relative to the cathode. A grid nearest to the cathode is called the control grid, and the high-frequency voltage—the amplified signal is applied to it. Depending on the number of grids, the electron tubes are called *diodes* (no grids), *triodes* (one control grid), *tetrodes* (two grids—control and screen) or *pentodes* (three grids—control, screen and suppressor). There are also multi-electrode tubes—hexodes and heptodes. These types of device are obsolete, were mainly used to convert electromagnetic oscillations in receivers, and are not considered in this book.

An indirectly heated cathode includes a cathode-heater H that heats the cathode to the desired temperature due to the current flowing through the heater. The surface of the indirectly heated cathode is equipotential and this ensures the uniform cathode current density.

As a rule, a hot cathode with direct or indirect heating is used in electron tubes. A directly heated cathode is heated by the flow of electric current through it.

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_6

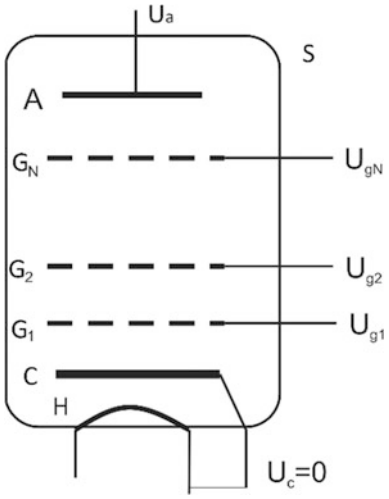


Fig. 6.1 Scheme of a vacuum tube with electro-static control

Any device with grids can be reduced to an equivalent diode, where an anode is located on the site of the first grid, having the effective potential

$$U_{\text{ef}} = \frac{U_{g1} + D_1 U_{g2} + D_1 D_2 U_{g3} + \cdots + \prod_{n=1}^{N} D_n U_a}{1 + k_1 D_1}, \tag{6.1}$$

where U_n is the potential of the n-th grid; U_a is the anode potential; D_n is the penetration factor of the n-th grid; and k_1 is the coefficient depending on the configuration of the electrodes and the distances between them. For example, for plane electrodes $k_1 = (x_{c2}/x_{c1})^{4/3}$. In this formula, x_{c1}, x_{c2} are the distances from the cathode to the first and second grids respectively. The potentials in formula (6.1) are counted from the cathode where the potential is generally assumed to be equal to zero. Penetration factor determines the effect of the corresponding grid (and the anode) on the field near the cathode. The more transparent the grid is, the greater its penetration factor is. Typical values of penetration factor lie in the range of 0.01–0.5.

The cathode in the electron tubes usually works in the mode of current limitation by space charge. In this mode, the cathode current is determined by the three-halves power equation:

$$I_c = 2.33 \times 10^{-6} k_c \frac{A_{\text{aef}}}{x_{\text{aef}}^2} U_{\text{ef}}^{3/2}, \tag{6.2}$$

where A_{aef}, x_{aef} are the area of the equivalent anode and its distance from the cathode; and k_c is the coefficient depending on the electrode geometry. For plane electrodes, $k_c = 1$. It is obvious that the current in the tube can flow only if the effective potential is larger than zero.

Bunching of the electron flux in the tubes is carried out by means of an electric field in the gap between the cathode and control grid affecting the space charge near the cathode. The electrons that pass through the control grid are accelerated by the constant electric field of the anode (or the screen grid) and interact with the alternating electric field of the anode transferring some of their kinetic energy to it. This process can be considered as the induced transition radiation of electrons.

The electro-optical system in devices with quasi-static control is not available in an explicit form. The anode serves to simultaneously take energy from the electrons and to collect the spent electrons. As a rule, the electrodynamics system of VEDs with grid control is external.

A quasi-static (grid) control at low frequencies takes practically zero power from the source of the control voltage, if the voltage on the control grid is negative with respect to the cathode and the electrons do not deposit on it. However, at ultrahigh frequencies, the negative control grid requires the finite power consumption which increase rapidly with rising frequency. This circumstance along with other factors, which will be discussed below, makes it difficult to create amplifiers and microwave oscillators based on quasi-static control devices for frequencies exceeding 1 GHz.

The main parameters of electron tubes are:

- Voltage amplification $\mu = \frac{\Delta U_a}{\Delta U_{g1}}$, at $I_a = \text{Const}$; $D = 1/\mu$.
- Transconductance (mutual conductance) $g_m = \frac{\Delta I_a}{\Delta U_{g1}}$, at $U_a = \text{Const}$;
- Anode resistance $R_a = \frac{\Delta U_a}{\Delta I_a}$, at $U_{g1} = \text{Const}$.

These parameters are coupled by the expression

$$R_a g_m = \mu.$$

Tetrodes, and especially pentodes, have a very high voltage amplification, while triodes—due to the absence of anode screening—have a comparatively low amplification, not exceeding 100. In the first approximation, voltage amplification in triodes is equal to grid penetrability.

At present, low-power electronic tubes have been almost completely replaced by solid-state devices. However, medium- and high-power devices with grid control (triodes and tetrodes) are still competitive. They are widely used as power amplifiers in radio transmitters for various purposes. Compared with other vacuum devices that can be used in this wavelength band, they have the following advantages:

- small size and weight;
- lower supply voltages;
- absence of special focusing systems;
- high average efficiency;
- small nonlinear distortions;
- high phase stability.

A number of factors affect the operation of these devices and limit the possibility of advancing into the higher frequencies region:

- intrinsic reactivity of the devices: the interelectrode capacitance and inductance of the inputs;
- finiteness of the transit time of electrons in interelectrode gaps.

Let's consider the influence of these factors on the interaction processes in microwave vacuum electron devices in more detail.

6.2 The Monotron and Diode Admittance

In Sect. 4.2.4, we showed that under certain conditions an unmodulated electron beam passing through an interelectrode gap can give a part of its kinetic energy to the field of this gap (the monotron effect). If the gap is connected to the oscillating circuit, and if the energy losses in this circuit are less than the power given by the flux, self-generation may occur in the circuit. For their existence, amplitude and phase conditions of self-excitation must be fulfilled:

$$G_c + G_e \leq 0, \quad B_c + B_e = 0,$$

where G_c, B_c are the active and reactive components of the circuit conductivity. The second of these conditions determines the frequency of self-oscillations. Their amplitude is determined by the G_e dependence on the amplitude of the oscillations, i.e., by nonlinear bunching processes in the gap.

A device that realizes this method of generating electromagnetic oscillations is called a *monotron*. Its scheme is shown in Fig. 6.2. Electrons emerging from the cathode are accelerated by the field of the first grid and enter the gap between the grids. This gap is part of the oscillatory circuit (cavity). The distance between the grids, the accelerating voltage and the eigen frequency of the resonator are chosen so that the transit angle in the space between the grids is close to 2.5π. In this case, the active conductivity is negative (Fig. 6.3).

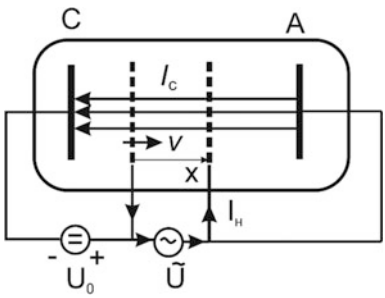


Fig. 6.2 Scheme of a monotron

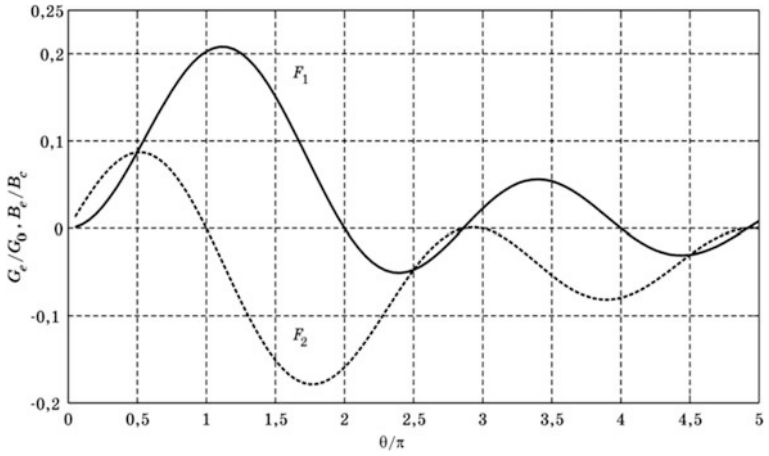


Fig. 6.3 Active and reactive conductivities of the diode versus the transit angle

If it is larger in absolute value than the conductivity of resonator losses, self-oscillations appear in the device, the frequency of which is close to the eigen frequency of the resonator. An experimental monotron sample generating power of up to 100 kW at a wavelength of 5.5 cm was created. Recent developments have shown that monotrons can reach 50% efficiency.

Calculation of electronic admittance of the diode is more complicated since electrons change their velocity on the cathode-anode path. We present without derivation the dependence of the active and reactive components of diode electronic conductivity on the transit angle (Fig. 6.3), which for plane electrodes should be determined by the formula

$$\theta = 3\omega\left(\frac{2\varepsilon_0 mAd}{9eI_0}\right)^{1/3},$$

where A is the cathode area. The active conductivity of the diode in Fig. 6.3 is normalized to its transconductance at the operating point $S = dI_0/dU_0$, and the reactive conductivity is normalized to the conductivity of the diode's "cold" capacitance $B_c = \omega C = \omega\varepsilon_0 A/d$.

As can be seen, the active part of diode conductivity also has a section with negative values; hence, a diode can be used to generate high-frequency oscillations. Llewellyn and Bowen designed such an oscillator in 1938, but its low efficiency (0.1–0.2%) prevents the practical use of these devices. In semiconductor devices such as avalanche, barrier-injection and transit-time diodes, which are considered in Part 3, use the phenomenon of negative conductivity much more effectively.

Let's consider the vector diagram of the currents in the diode. Constant positive voltage U_0 and alternating voltage $\widetilde{U} = \dot{U}e^{i\omega t}$ are applied to the anode. We suppose

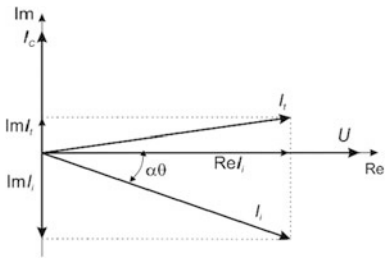


Fig. 6.4 Vector diagram of diode currents

that $\left|\dot{U}\right| \ll U_0$. It is obvious that the convection current in the diode will also have constant and alternating components:

$$I_c = I_0 + \dot{I}e^{j(\omega t - \theta(x))},$$

where x is the coordinate counted from the cathode towards the anode, and $\theta(x)$ is the transit angle of electron from the cathode to the plane with coordinate x. Figure 6.4 shows the vector of anode voltage $\dot{U}$. Induced current $\dot{I}_i$ as averaged over the length of the gap convection current lags the voltage by angle $\alpha\theta$, where $0 < \alpha < 1$. In the external circuit, capacitive current $\dot{I}_c = \omega C^* \dot{U}$ also flows, leading the voltage by 90°. In sum, they form the total current $\dot{I}_t$, which has active and reactive components.

The "hot" capacitance of diode C^* is larger than its "cold" capacitance due to the effect of space charge: $C^* = (4/3)C$. However, the reactive component of the induced current reduces the reactive component of the total current: $\dot{I}_{i\,\mathrm{Im}} = (3/5)\,\omega C\dot{U}$. This result is in agreement with the graph in Fig. 6.3 at transit angles θ not exceeding 90°.

6.3 Operating Modes of Electron Tubes

Depending on the relationship between the value of the constant voltage on the tube control grid and the amplitude of the alternating voltage, the devices with quasi-static control operate in different modes.

Let's consider the main operation modes (classes) of the amplifiers on electron tubes using the triode as an example. The main characteristic of the class is the *cutoff* angle $\theta = \omega\tau/2$, where τ is the part of the alternating voltage period during which the anode current is different from zero.

Class A. A negative voltage (bias) is applied to the grid, which is greater than the cutoff grid voltage, and the amplitude of the alternating voltage is less than the absolute value of the bias voltage (Fig. 6.5a). In this mode, current I_{a0} (the quiescent current) flows through the tube in the absence of the alternating voltage. The grid current is zero since the grid is negatively charged with respect to the cathode.

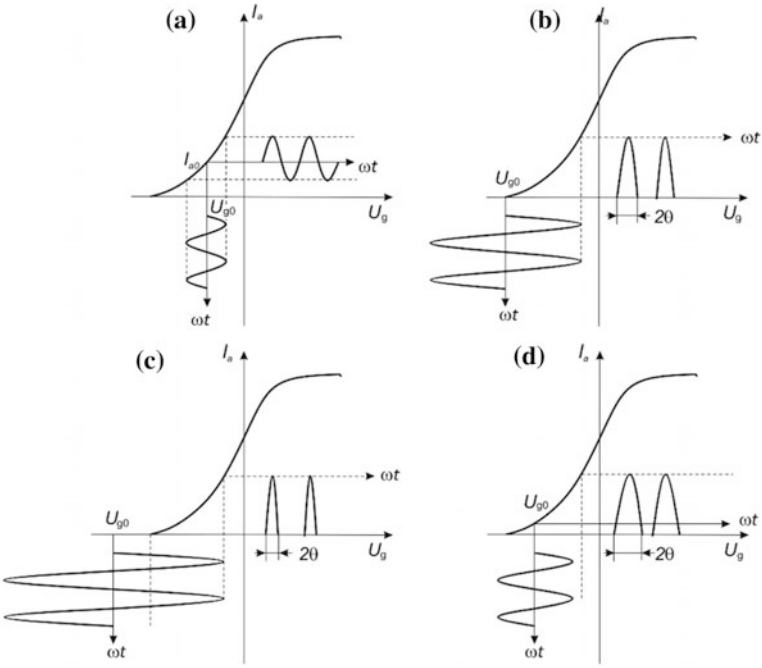


Fig. 6.5 Main operation modes of VED amplifiers

When an alternating voltage is applied to the grid, the anode current begins to alternate, following the grid voltage in the waveform. The cutoff angle $\theta = 180°$. In this mode, the amplifier has low efficiency and small nonlinear distortions. Such a mode is impractical in high-power amplifiers.

Class B. In this mode the bias is chosen equal to the cutoff voltage, so in the absence of a signal, the anode current is zero (Fig. 6.5b). In the presence of the alternating voltage on the grid, the anode current is different from zero only in positive half periods of the grid voltage.

The cutoff angle is 90°. The efficiency in this mode is much larger than in class A, but the anode current waveform significantly differs from the grid voltage waveform. Thus, this mode can be used only in resonant amplifiers.

Class C. The bias voltage is larger than the cutoff voltage, and the amplitude of the alternating voltage is larger than the bias voltage (Fig. 6.5c). The cutoff angle is less than 90°. This class is characterized by high efficiency, however, the output power and the gain decrease while the cutoff angle reduces.

Class AB. Intermediate between classes A and B. The cutoff angle is larger than 90° (Fig. 6.5c). By selecting the cutoff angle it is possible to find the optimal relationship between efficiency and the level of nonlinear distortions.

There are other classes and subclasses of amplifier operation—A+, D, H, G but they are quite complex in implementation and have limited application.

6.4 Amplifier Circuits

Let's consider the inclusion of the triode in the amplifier circuit. The amplifier has two input terminals and two output terminals. The triode has three electrodes. Therefore, one of the electrodes must be common to the input and output circuits. So there exist amplifier circuits with a common cathode, with a common grid and with a common anode (Fig. 6.6).

In the common cathode circuit (Fig. 6.6a) the currents induced in the grid circuit by the charges moving in the cathode-grid I_{cg} and the grid-anode gaps I_{ga} have opposite directions. Therefore, the total grid current in a device with small transit angles in the interelectrode gaps is zero. Small signal power gain

$$\mu_p = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{I_a^2 R_a}{U_g^2/R_g} = g_m^2 R_a R_g,$$

where g_m is the transconductance, can be quite large. Such a connection circuit is the main type for low-frequency amplifiers.

At high frequencies, a phase shift arises between currents induced in the grid circuit, due to the finite transit time of electrons in the interelectrode gaps. Therefore, the sum of the induced currents is different from zero.

To analyze this effect suppose that constant negative bias U_{c0} and a small control alternating voltage $U_c \ll U_{c0}$ are applied to the control grid of the device.

If the grid penetration factor is sufficiently small, then the triode can be considered as a set of two independent plane gaps having one common electrode (grid) permeated by the same electron flux moving from the cathode to the anode. Current I_{cg} induced in the cathode-grid gap lags grid voltage U_g by angle $\xi_{cg} = \alpha_g \theta_{cg}$, where θ_{cg} is the transit angle in the cathode-grid space, $0 < \alpha_c < 1$ is the coefficient that takes into account the averaging of the convection current along the length of the gap. Current I_{ga} induced in the grid-anode gap lags the grid voltage by angle $\theta_{2g} = \theta_{cg} + \alpha_a \theta_{ga}$, where θ_{ga} is the transit angle in the grid-anode space and α_a is the averaging coefficient. The difference in these currents is equal to the grid current I_g flowing in its circuit even in the absence of electron deposition on the grid.

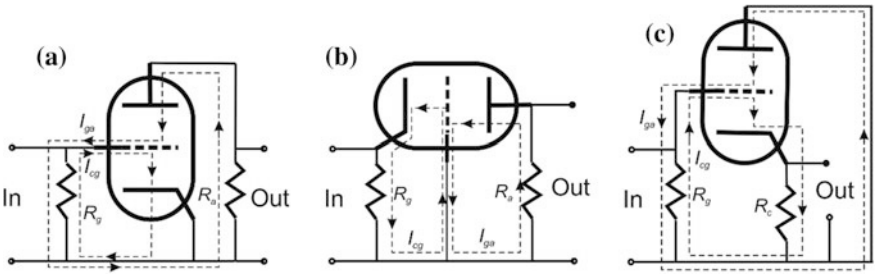


Fig. 6.6 Triode amplifier circuits

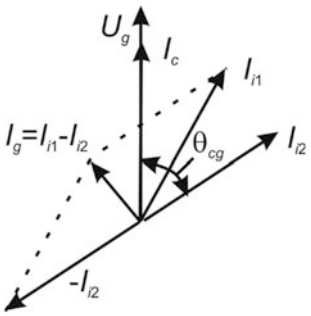


Fig. 6.7 Vector diagram of the currents in the cathode-grid space of the triode

The foregoing is illustrated by the vector diagram shown in Fig. 6.7. The real component of the grid current determines the active component of input conductivity of the grid $G_{\text{вх}} = \text{Re}(I_c)/U_c$. It can be shown that for transit angles which are not very large

$$G_{\text{in}} \approx 0.05 g_m \omega^2 \tau_{cg}^2.$$

Accordingly, additional power $P = 0.5 G_{\text{in}} |U_g|^2$ is dissipated in the amplifier input circuit that reduces the tube gain.

The decrease in the input active conductivity can be achieved by reducing the transit time τ_{cg} of electrons by decreasing the grid-cathode distance d_{cg}. To maintain the permissible value of transit angle θ_{cg}, distance d_{cg} should change inversely with the frequency. Other grid dimensions (wire diameter or thickness of the grid bridges, cell size, etc.) should decrease in approximately the same proportion.

At small values of distance d_{cg}, which are characteristic for decimeter band devices, the further reduction of this distance is limited by a number of factors, in particular tolerances for the manufacture of cathode-grid knot components and deformation of the grid when it is heated by radiation from the cathode and by the grid current. Ensuring stability of the grid shape during heating, the action of accelerations and vibrations requires the use of a complex design and technological solutions.

Besides, the degree of feedback through grid-anode capacitance C_{ga} increases with the rising frequency while capacitive feedback resistance $X_c = -1/(\omega C_{ga})$ decreases. The presence of the feedback can lead to self-excitation of the amplifier. This undesirable effect is practically absent in tetrodes and pentodes in which the control grid is shielded from the anode by other grids.

In the common grid circuit (Fig. 6.6b) feedback is carried out through the cathode-anode capacitance which is significantly smaller than the grid-anode capacitance, since the anode is shielded from the cathode by the grid.

However, unlike the common cathode circuit, in the common grid circuit the current induced in the cathode-grid gap flows in the input circuit, so the input active conductivity of the triode in this circuit is $G_{\text{in}} = g_m$. Accordingly, at low

frequencies, the tube gain in this circuit is significantly lower than in the common cathode circuit. Since the current induced in the output circuit does not pass through the input circuit, the triode input conductance in the common grid circuit coincides with the input conductance of the equivalent diode (Fig. 6.7), and its active component decreases with an increasing transit angle. As the frequency increases, the input conductance of the common cathode circuit begins to rise rapidly, and at the frequency corresponding to transit angle $\theta_{cg} = 5\pi/4$ it becomes equal to the input conductivity of the tube in the common grid circuit. At higher frequencies, the common grid circuit provides higher gain than the common cathode circuit, explaining the wide application of this connection circuit in the decimeter band.

The common anode circuit (Fig. 6.6c) has the same input conductivity as the common grid circuit but because of load resistance in the cathode circuit, there is a negative feedback, which reduces the voltage gain to one. This circuit is often called a "cathode follower" since the output voltage "follows" the amplitude and the waveform of the input voltage. This circuit is used in some special cases.

It should be noted that the considered operation modes and connection circuits of vacuum devices are similar to those for transistor amplifiers.

6.5 The Influence of Cathode Contact Inductance

A triode input active conductivity can also be caused by the inductance of the cathode contact, in addition to the transit angle finiteness in the grid-cathode gap. The input active conductivity caused by these factors worsens the control action of the grid, leading to an increase in the power consumed by the grid circuits.

Let's consider the influence of the cathode contact inductance L_c and the grid-cathode capacitance C_{cg} on the operation of the triode working in the common cathode circuit (Fig. 6.8).

We assume that a negative bias is applied to the triode grid and it does not encompass the electron current.

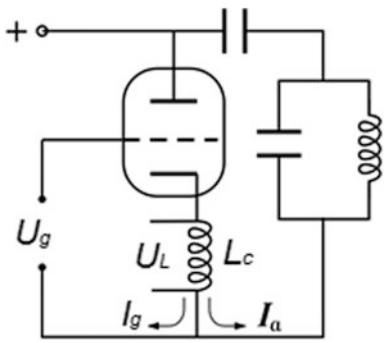


Fig. 6.8 For analysis of the influence of cathode input inductance

Let's calculate the input conductivity of the grid circuit defining it as the ratio of current in the grid circuit I_c to input voltage U_{in}, $Y_c = I_c/U_{\text{in}}$. Voltage $U_{\text{in}n}$ can be represented as the sum of two voltages: the voltage U_c applied directly between the cathode and the grid, and the voltage drop on the inductance U_L, $U_{\text{in}} = U_{\text{c}} + U_L$. Current I_{c} is a capacitive current and is related to voltage U_c by the ratio $I_c = j\omega C_{\text{ск}} U_c$. The vector diagram of the currents in the grid circuit is shown in Fig. 6.9.

For the voltage drop on the inductance, we can write $U_L = (I_c + I_a) j\omega L_{\kappa}$, where I_a is the alternating convection current in the grid-cathode gap which is closed through the anode circuit. The value of this current can be expressed in terms of voltage U_c applied to the gap and triode transconductance g_{m}, $I_a = g_{\text{m}} U_c$. Taking into account the formulas obtained above we find

$$Y_g = \frac{I_g}{U_{\text{in}}} = \frac{j\omega C_{cg} U_g}{U_g + U_g (g_m + j\omega C_{cg}) j\omega L_c} = \frac{j\omega C_{cg}}{1 + (g_m + j\omega C_{cg}) j\omega L_c}.$$

Setting $g_m \gg \omega C_{cg}$ we obtain

$$Y_c = \frac{j\omega C_{cg}}{1 + j\omega L_c g_m}.$$

Dividing the real and imaginary parts in this expression, we obtain

$$Y_g \approx \frac{(\omega^2 L_c g_m + j\omega) C_{cg}}{1 + (\omega L_c g_m)^2}.$$

In real devices $(\omega L_{\kappa} g_m)^2 \ll 1$, so finally we obtain

$$Y_c = \omega^2 C_{cg} L_c g_m + j\omega C_{cg}. \quad Y_g = \omega^2 C_{cg} L_c g_m + j\omega C_{cg}.$$

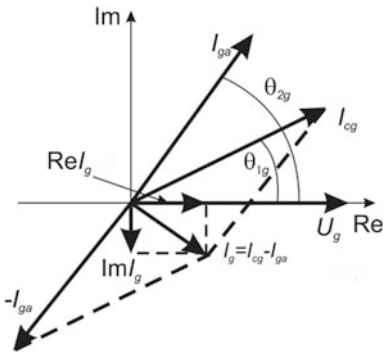


Fig. 6.9 Vector diagram of the currents in the anode-grid space of the triode

From this expression, it follows that the cathode input inductance leads to the appearance of the active component of the grid circuits input conductivity $G_c = \omega^2 C_{cg} L_c g_m$.

Consequently, the grid circuits consume power $P_g = U_{\text{in}}^2 G_c/2$ even when the grid operates with a negative potential to the cathode and does not encompass the electron current. In this case, power P_c is pumped to the anode circuits thanks to inductance $L_к$, which is a common element of the grid and anode circuits. It is important to note that the value of the input active conductivity and, consequently, of the power consumption *increases in proportion to the frequency square*. In modern high-power grid-controlled devices, the disk and cylindrical structures is used to reduce contact inductance.

6.6 The Influence of Space Charge and Displacement Current in the Cathode-Grid Space

In the space between the cathode and the grid, the electrons move relatively slowly, so the influence of the space-charge field in this gap is sufficiently large.

The layer of the space charge leaving the cathode at time t_0 creates an electric field that inhibits the motion of electrons exiting at later time intervals. As a result, the convection current pulse shape in the grid plane does not coincide with the shape of the current pulse at the cathode.

Let's consider, for example, the operation of a triode in class B. The dependence of the effective potential on time has the form

$$U_{\text{ef}} = U_0 \sin \omega t,$$

shown in Fig. 6.10a. At small transit angles the waveform of the cathode current is determined by the three-halves power equation (Fig. 6.10b):

$$I_c(t) = \begin{cases} I_0(\sin \omega t)^{3/2}, & 0 \le \omega t \le \pi; \\ 0, & \pi < \omega t < 2\pi. \end{cases}$$

At large transit angles the process of cathode current take-off becomes more complicated. At the time when the acting potential becomes positive, a layer of space charge exits from the cathode. With unlimited cathode emission, the electric field inside this layer is equal to zero. In the rest of the cathode-grid gap the charges are absent, however, a displacement current equal to the capacitive current flows

$$I_c = \omega C_{cg} U_0 \cos \omega t.$$

At a moment $t = 0$, this current is maximal. According to Ampere's circuital law, the convection current in the space charge layer must be equal to the displacement current. That's why the cathode current pulse takes the shape shown in

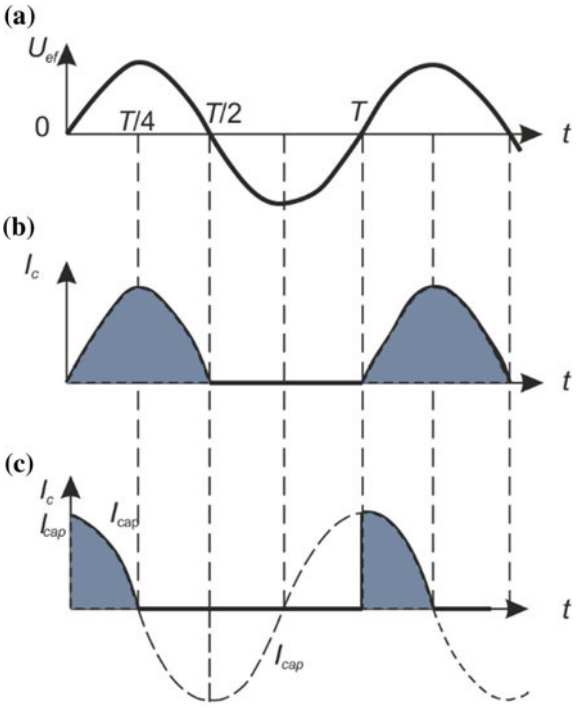


Fig. 6.10 Waveform of the triode cathode current at large transit angles

Fig. 6.10c. As can be seen, the cathode current abruptly assumes the maximum value equal to $I_{\max} = \omega C_{cg} U_0$ and then decreases to zero. The duration of the cathode current pulse is a quarter of the period, not half the period, as for small transit angles. The amplitude of the cathode current pulse $I_{\max}$ in the microwave band is much larger than the value I_0 determined by the three-halves power equation. So one have to use cathodes with the large redundancy of emission current.

Further motion of electrons in the cathode-grid space occurs under the action of the electric field of the grid and the space charge. At time $t = 0$ a dense space charge layer corresponding to the maximum value of the convection current exits from the cathode. The space-charge field of this layer accelerates the electrons located at its front boundary and inhibits the electrons moving behind. Besides, the acting potential varies as shown in Fig. 6.11a (dashed line).

As a result, the trajectories of electrons in the space-time diagram acquire the form shown in Fig. 6.11a. Electron 1 moves at maximum speed and passes the grid plane at time t_1. Electron 2 moves along a similar trajectory. Electron 3 approaches the grid at the time when the field in the cathode-grid space is decelerating, so it completely loses its velocity and turns back to the cathode. At this time, the convection current pulse in the grid plane stops.

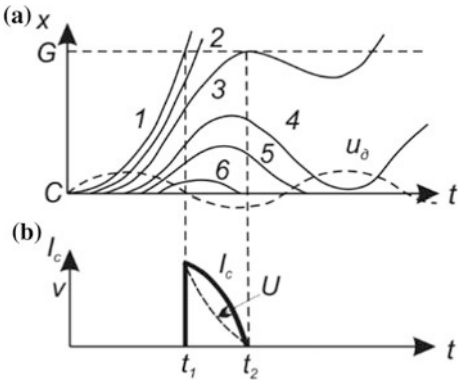


Fig. 6.11 Diagram of movement of electrons in the triode cathode-grid gap

Electron 4 does not reach the grid and, like electron 3, oscillates in the cathode-grid space. After one or several such oscillations, they either transit to the anode or remain on the cathode. Electrons 5 and 6 immediately return to the cathode with non-zero velocities that lead to an additional unintended heating of the cathode.

Thus, the convection current pulse in the grid plane has a longer duration than in the cathode plane and the considerable spread of electron velocities (Fig. 6.11b). The distortion of the cathode current pulse leads to a decrease of the first harmonic content in it and of the efficiency of the device.

6.7 Motion of Electrons in the Grid-Anode Space

Having passed the plane of the grid, electrons get into the grid-anode gap, inducing the current in the external circuit. The motion of electrons in this gap occurs under the action of the anode accelerating field and the alternating voltage on it which has amplitude comparable with the constant anode voltage. Therefore, during the motion, the velocities of electrons change as shown in Fig. 6.12.

The pulse of the induced current begins when the first electron crosses the grid plane and stops when the last electron reaches the anode. As a result, the duration of the induced-current pulse τ_{ai} is much larger than the duration of the convection current pulse in the plane of the grid τ_g.

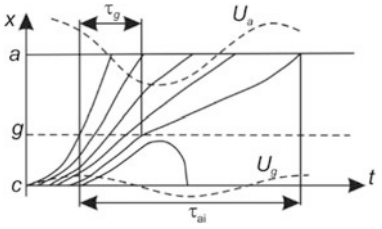


Fig. 6.12 Diagram of motion of electrons in the triode grid-anode gap

This effect reduces the amplitude of the induced current first harmonic that leads to a decrease in amplifier efficiency. Note that the effect of the anode current pulses stretching is practically absent in tetrodes where the electrons are accelerated by the constant field of the second grid before reaching the second grid-anode space.

Figure 6.12 also shows the anode alternating voltage that occurs when the induced current flows through the resonant load tuned to resonance with the voltage on the grid. The minimum voltage corresponds to the maximum of the induced current. At this, due to the transit time, the phase shift between the alternating voltages on the grid and the anode is different from 180° and changes the feedback phase in the circuits of the triode generators and can lead to the failure of self-oscillations.

The noted phenomena limit the use of triode and tetrode amplifiers at the frequencies of the decimeter band.

6.8 Modern Medium and High Power Tetrodes

Medium and high power tetrodes created by new technology remain competitive in the decimeter wavelengths band and are used in the final stages of radio and television transmitters.

In contrast to electron-beam microwave devices, tetrodes use the quasi-static (grid) method of controlling the electron flux. This excludes the need for long drift tubes (in the case of klystrons) or slow-wave structures (in the case of traveling wave tubes). Therefore, the dimensions and weight of tetrodes are relatively small. Besides, complex focusing systems are not required to support the electron flux.

In microwave tetrodes, high levels of output power are achieved due to high currents at relatively low supply voltages.

The fact that in the amplitude modulation mode tetrodes operate in class B and C, i.e. with cutoff of the cathode current, explains their high average efficiency. If there is no high-frequency signal, current does not pass through them. In the klystrons and TWTs in this case, all the electron beam power is exuded at the collector. Therefore, at the amplification of a television signal, for example, the average efficiency of the tetrode transmitters appears to be 2–2.5 times higher than the klystron ones.

The high output signal phase stability of tetrode amplifiers is explained by the small electron transit angles so the delay time of the signal passage and its change at the variation of the signal amplitude and supply voltages are an order of magnitude lower than in klystrons and TWTs.

The lower level of nonlinear distortions in these amplifiers is associated with the features of their amplitude characteristics. If a klystron's amplitude characteristic is close in shape to a segment of the sinusoid with an almost linear initial part, then the similar characteristic of tetrode amplifiers has the form of S-shaped curves with greater linearity in the middle part. It is possible to develop devices with an amplitude characteristic approaching linear.

The above circumstances—the need to reduce the cathode-grid distance and the associated changes in the grid structure, the heating of the grid, its thermal deformation and thermal emission—together constitute the grid problem in high-power devices of the decimeter wavelength band and limit their progress toward higher frequencies.

In high-power devices the control grid can be heated strongly both by radiation from the cathode and by the electron current. The temperature of the grid can reach high values at which its thermionic emission becomes noticeable. The emission can significantly increase if substances evaporated from the cathode deposit on the grid.

In a number of cases, the secondary emission is added to the thermal emission of the grid because of the electron current bombardment of the grid. The emitting grid loses its control properties and becomes a source of electrons itself.

One of the possible ways to solve this is the use of pyrolytic graphite (pyrographite) as a grid material. This is a polycrystalline carbon material obtained by chemical gas-phase deposition during the pyrolysis of hydrocarbons carried out at a temperature range of 750–2400 °C.

This material has an emissivity close to that of an absolutely black body, high thermal conductivity, good mechanical strength, increasing with the rising temperature, increased resistance to heat shocks, lower thermal and secondary emissions compared to materials traditionally used in the manufacture of grids (tungsten, molybdenum, and tantalum).

Grids made of pyrographite are used in high-power tetrodes of ultrahigh frequency (frequency $f \approx 800$ MHz, output power $P_{\text{вых}} \approx 10$–$20$ kW).

The technology for manufacturing pyrographite grids is more costly than the technology for manufacturing grids from traditional materials. Therefore, such materials continue to be used for the high-power electronic devices grid electrodes. Such grids are usually coated with various antiemission materials. These materials include, for example, a complex intermetallic compound of platinum and zirconium Pt_3Zr.

Among the disadvantages of tetrode amplifiers, the low power gain values, relatively low operating frequencies, limited bandwidth and significantly lower reliability and longevity should be noted. The existing problem of cooling the anode which simultaneously serves as the electron collector should also be pointed out.

In modern tetrodes designs, taking into account the peculiarities of operation in the microwave frequency band, the energy losses in the input circuits associated with the transit time are negligibly small.

The inductances of the leads are minimized due to the corresponding constructive technological solutions. The interelectrode capacitances are naturally used as the lumped capacitances of the tetrodes amplifiers and generators oscillator circuits. This makes possible to analyze tetrodes amplifiers (generators) by well-known techniques used in low-frequency electronics.

Figure 6.13 shows a tetrode amplifier assembled according to the common grid circuit. The amplifier uses coaxial resonators. The capacitive part of the input resonator is formed by the first grid—the cathode capacitance C_{cg1}. The output

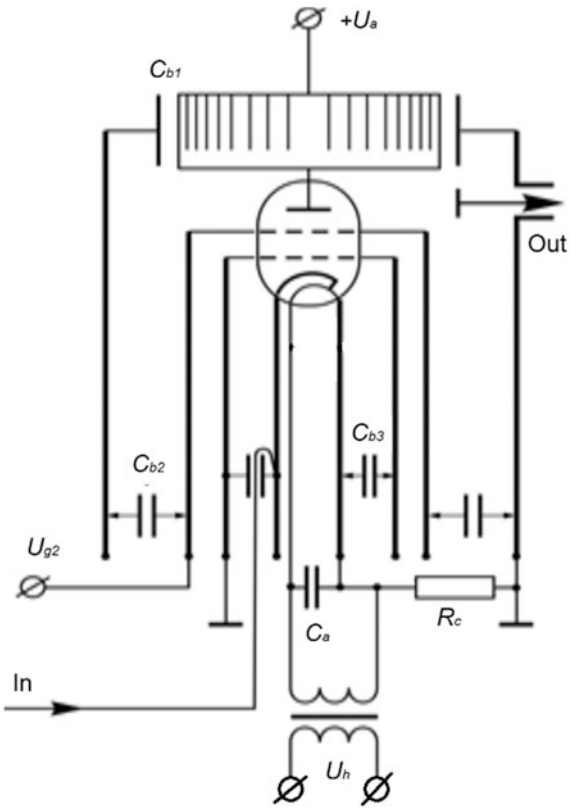


Fig. 6.13 Common grid circuit of a tetrode amplifier

resonator is connected to the second grid—the anode gap, and its capacitive part is formed by the second grid—anode capacitance C_{g2a}. To supply the anode and the second grid with constant voltages U_a and U_{g2}, blocking capacitors C_{b1} and C_{b2}, integrated into the anode grid-circuit, are used. To decouple the cathode and the first grid, the capacitor C_{b3} is used, which is built into the cathode-grid circuit.

The situation is different when the constant voltage of the screen grid is much lower than the anode voltage. In this case, the convection current pulses enter the output gap at a low speed determined by the screen grid voltage. In the interaction gap, they experience the accelerating effect of the constant electric field which is created due to the difference in the potentials of the screen grid and the anode and the braking action of the alternating field of the output resonator. The joint action of these two fields leads to the fact that even with the alternating voltage amplitude close to the difference between the constant potentials of the anode and the screen grid, the electron velocity in the gap increases. This means that not the kinetic energy of the electrons but thier potential energy which they have passing the screen grid, is converted into the energy of the alternating field. This energy conversion is an example of transition radiation of electron bunches.

Several leading electronics companies produce tetrodes and beam power tubes of very high output power for broadcasting and over-the-air television. The modern tetrode for broadcasting with a record output power of 2 MW in continuous mode, developed at the JSC "Svetlana", is shown in Fig. 6.14. It contains an external copper anode with water cooling, disc cathode and grids connectors. The grids of this device are made from pyrographite. At the top of the device we can see heater connectors.

The device is designed to work with an external electrodynamical system.

6.9 Microwave Vacuum Microelectronics Devices

In the late 1960s, low-power microwave electronics became almost completely solid. However, despite impressive successes, solid-state electronics could not fully meet the stringent requirements of radio-electronic systems developers, especially in the field of radiation resistance, resistance to overload, temperature stability, and cutoff operating frequencies. At the same time the development of integrated technologies, the development of stable field-emission emitters created the basis for the emergence and development of *vacuum microelectronics*, a new trend in electronics that uses integrated technologies to create microminiature vacuum devices. Due to their small size, these devices are ideally suited to microwave operation.

Developments in this direction began in the 1960s with the works of K. Shoulders and I. Brodie, but the decisive contribution was made by C. Spindt and his colleagues, who developed the design and technology of the matrix field-emission molybdenum cathode with tips and investigated its properties. Spindt's published article with the results of these studies appeared in 1976. The construction of this cathode, known as the Spindt cathode, is shown in Fig. 6.15.

On a silicon substrate 1, an insulating film 2 of silicon dioxide is formed by the oxidation method onto which a molybdenum coating 3 is applied. Round holes are formed by photolithography and etching in the coating and the film. Molybdenum cones 4 are grown in these holes by the method of vacuum deposition. The dimensions of one of the structures made are: $h = 1.5$ μm, $t = 0.4$ μm, $D = 0.5$ μm. Above it, a solid metal electrode is located at the distance of about 100 μm from the structure (not shown in Fig. 6.15).

It is not difficult to see that the described structure is a triode in which the molybdenum tip plays the role of the cathode, the role of the grid is performed by metallization 3, and the metal electrode performs the role of the anode. When a suffiently large positive in respect to the cathode voltage is applied to the anode, the autoelectronic emission from the molybdenum tip begins. The emission current can be regulated by changing the voltage on the molybdenum coating 3.

The autoelectronic emission current is determined by the Fowler-Nordheim law:

$$J_e = AE^2 \exp(-B\varphi^{3/2}/E),$$

where φ is the work function of the metal; E is the electric field strength near the cathode; and A, B are the constants depending on the shape of the cathode.

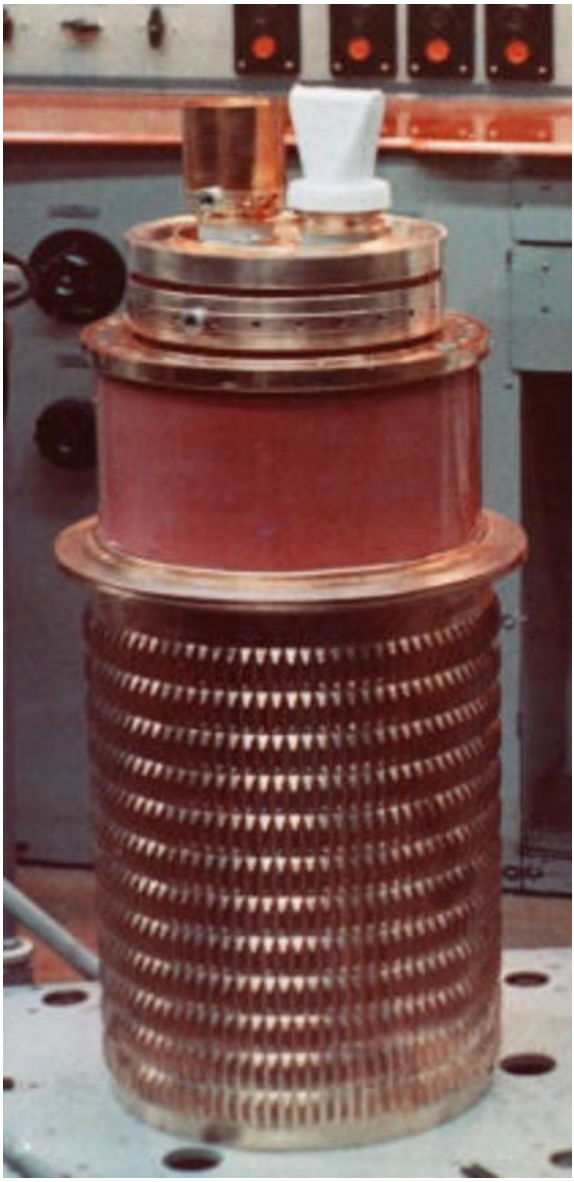


Fig. 6.14 A high-power tetrode

It follows from this formula, that for the cathode material with the work function of 4.5 eV typical for metals, the noticeable emission current appears at field strengths (3–5) 10^7 V/cm. Such fields can be created only on tips having a small bending radius. To characterize such tips, the so-called *intensity factor* is used which is equal to the ratio of the field strength at the tip to the field strength in the area far from it. The intensity factor depends strongly on the bending radius reaching

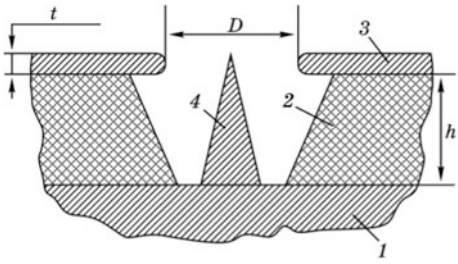


Fig. 6.15 The Spindt cathode

several tens and even hundreds at bending radii of several hundredths of a micrometer. Therefore, it is very important to use materials and technologies to minimize the bending radius. The minimum value of this radius is the size of one atom, but this limit is unattainable, since one atom on the tip will be pulled out of the metal by ponderomotive forces. At present, the greatest intensity factor is provided by cathodes made of carbon nanotubes. However, the technology does not yet allow for the growing of a regular lattice from such nanotubes.

In the Spindt cathode with molybdenum tips, a current of several hundred microamperes is observed at an anode voltage of 200–250 V, and that is a lot for a microelectronic device. The most recent developments made it possible to reduce the voltage on the control electrode to 30 V and the anode voltage to 150 V. At this, the triode has a very small differential transconductance and gain. These parameters could be improved by decreasing the tip height, but if this is done, the anode current sharply decreases. The low gain and output power prevent the introduction of these devices into microwave equipment.

Microvacuum devices with lateral geometry were developed, and Fig. 6.16 shows a diode (Fig. 6.16a) and a triode (Fig. 6.16b) with this geometry. It contains a cathode 1, an anode 2, an isolating plate 3 and a dielectric substrate 4. A triode has an additional control electrode 5. In order to increase the intensity factor, the edge of the cathode electrode is made jagged. Such devices can be easily implemented in the circuits of traveling wave amplifiers (distributed amplifiers). The structure of such an amplifier is shown in Fig. 6.17.

The amplified signal is applied to the input transmission line formed by the heavily doped substrate and the control electrode in the form of a tape. The output transmission line excited by the cathode currents is formed by the control electrode and the anode. For correct amplifier operation, the phase velocities of waves in both transmission lines must be equal. This condition limits the frequency band of the amplifier, as the dispersion laws in the input and output lines are different.

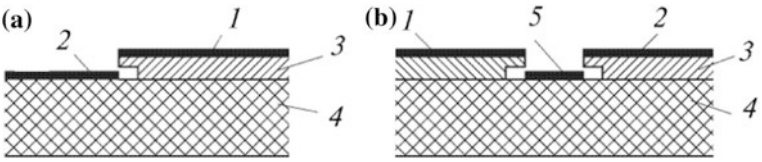


Fig. 6.16 A microvacuum device with a lateral structure

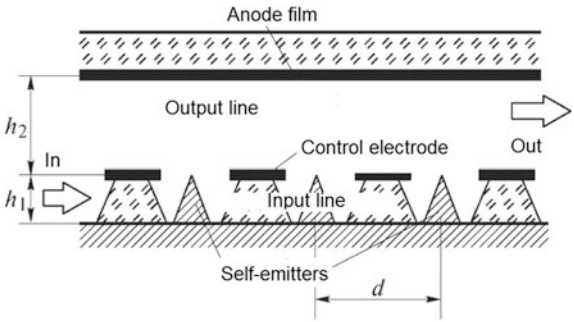


Fig. 6.17 Travelling wave microvacuum amplifier

In the first approximation, the gain of such an amplifier is proportional to its length l:

$$\mu = 20 g_{m1} Z_{02} \beta l,$$

where μ is the power gain, dB; g_{m1} is the specific amplifier differential transconductance (the differential transconductance per length unit); and Z_{02} and β are the wave impedance and phase constant of the output transmission line.

In reality, the gain is reduced due to losses in the transmission lines, which appear to be significant, since the thickness of metal films is usually much less than the depth of penetration. Another reason is the input transmission line excitation by the density modulated cathode current. Since the excited wave is in the antiphase with the cathode current (and with the wave excited by the source of the input signal), the resultant wave decreases beginning from the specific length of the amplifier. Accordingly, the gain also decreases. Thus, there is an optimum amplifier length at which the maximum gain is realized.

The analysis shows that acceptable gain values can be obtained with differential transconductance at one tip not less than 10 μA/V, and this is at the limit of modern technological possibilities.

In recent years, proposals to implement the traveling wave klystron and the monotron on the basis of the tip autoelectronic emission cathodes have appeared, but the implementation of these ideas requires a lot of work by designers and technologists.

Advancement Questions

1. What is an equivalent diode? How do you determine the anode voltage of this diode?
2. In what mode does a hot cathode work in electronic devices with quasi-static control?

3. Define the monotron. What interaction phenomena are used in it?
4. List the classes of operation of electron tubes and the features of the operation of devices in each class.
5. Describe the operation of the amplifier in a common cathode circuit.
6. Describe the operation of an amplifier in a common grid circuit.
7. What is the influence of the cathode input inductance?
8. Describe the shape of the cathode current pulse in tubes with a large transit angle.
9. What problems do designers of grids in high-power devices have to solve?
10. Describe the design and operation of the Spindt cathode.
11. Describe the design of the distributed microvacuum amplifier.
12. What factors limit the use of vacuum microelectronics devices in the microwave band?

Chapter 7
O-Type Microwave Devices

7.1 General Characteristics of O-Type Devices

O-type microwave vacuum electron devices are dynamic control devices with a rectilinear electron beam. This beam interacts with standing or travelling electromagnetic waves. The name of this class of devices originates from the French word *l'onde* (wave). The constant magnetic field in these devices is used only for transporting the beam and does not participate in the interaction process. Klystrons, traveling wave tubes, backward wave oscillators and hybrid devices are among the representatives of this type of device.

The process of energy conversion in O-type devices includes three stages.

1. Formation and acceleration of the electron beam. At this stage, the energy of the power source is transformed into the kinetic energy of the electron beam.
2. Bunching of the electron beam, which consists of the formation of a sequence of electron bunches from the homogeneous beam formed in the first stage. In this case, a variable component of the convention current arises.
3. Extraction of energy from the bunched electron beam by the field of the electrodynamic system. At this stage, the kinetic energy of the electron flux is converted into the energy of the electromagnetic field.

In O-type devices, these stages of energy conversion are separated in time and are taken apart in space.

By the nature of the interaction between the beam and the field, the devices with discrete (short-term) interaction and devices with continuous (long-term) interaction are distinguished. The first type of device includes klystrons, in which resonators (cavities) are used as an electrodynamic system. The interaction between the beam and the field occurs discretely in narrow resonator interaction gaps. Continuous interaction is realized in traveling and backward wave tubes, where the electron beam interacts permanently and continuously with the field of the traveling wave. There are also hybrid devices in which both types of interaction are used.

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_7

7.2 Klystrons

7.2.1 *The Structure and Operating Principle of the Double-Cavity Transit-Time Klystron*

The double-cavity transit-time klystron was invented and built in 1937 by the brothers Russell and Sigurd Varian in the Stanford University laboratory. Toroidal resonators (rhumbatrons) were used in the klystron, and they were also proposed shortly before by W.W. Hansen, who worked at the same university. The name "klystron" is made up of the Greek word klýzo—to strike, to wash with a wave—and the ending (elec)tron. In the USSR, the first amplifying klystrons were developed by N.D. Devyatkov and colleagues in 1940.

The structure of the double-cavity transit-time klystron is shown schematically in Fig. 7.1. The input resonator of the device is excited from an external source, and an alternating electric field appears at its interaction gap. An electron beam formed by the electron gun consisting of the cathode and the anode enters the gap of the first resonator, where it is exposed to the alternating electric field that leads to the electron velocity modulation. The velocity-modulated beam enters the drift space, where the process of bunching the electrons occurs. The electron beam becomes inhomogeneous in density and the alternating convection current arises in the form of a sequence of bunches. Passing the output resonator gap, the bunches excite oscillations in it. The electric field of these oscillations in the resonator gap inhibits electron bunches, part of the kinetic energy of which is transferred to the electromagnetic field of the resonator. According to the theory of radiation, the induced slowing-down oscillation of electrons arises under the influence of the resonator field.

A collector captures the spent electrons. The couplers connect the klystron to the source of the input signal and the load.

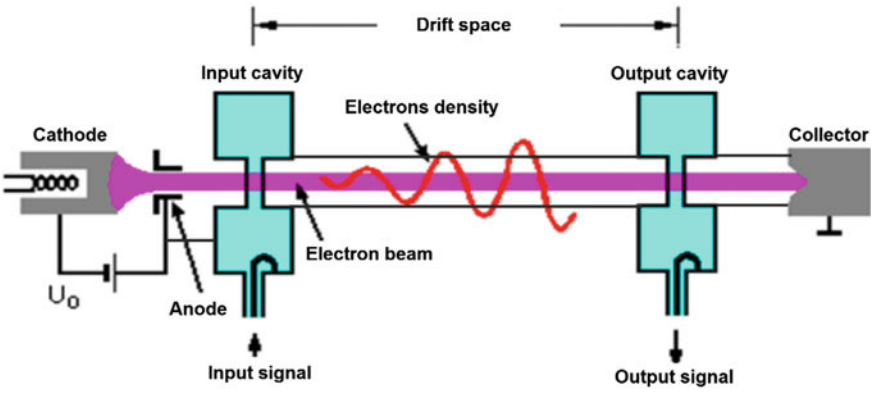


Fig. 7.1 Scheme of a double-cavity klystron

Having considered briefly the main processes that determine the operation of a transit-time klystron we turn to mathematical analysis.

7.2.2 Velocity Modulation in the Interaction Gap

The process of velocity modulation of an electron beam passing through the plane resonator gap is considered in Sect. 4.2.2, where the concept of the interaction coefficient is introduced. The velocity modulation factor is introduced in Sect. 4.2.3.

In high-power klystrons, gridless gaps are used since grids do not withstand the thermal load from electrons bombarding them. Figure 7.2 shows the picture of the gridless interaction gap field calculated using a special program. The figure shows the vectors of the electric field intensity. The length of these vectors is proportional to the intensity modulus, and the direction corresponds to the field direction at the vector origin point. As can be seen, the electric field of the gridless gap is non-uniform and has both longitudinal and radial components. The electric field penetrates into the drift tube, and therefore the actual axial extension of the field l_E is larger than gap width d, and depends on radial coordinate r. For $r \approx a$, where a is the radius of the drift tube, $l_E \approx d$ and on the axis of symmetry ($r = 0$) $l_E \approx d + 4a$.

The dependence of the longitudinal component of the electric field on the radius is given by

$$E_z(r, z) = E_0 I_0(\gamma_e r)/I_0(\gamma_e a),$$

where E_0 is the field intensity on the edge of the drift tube, $E_0 = E_z(a, z)$, where the field is almost homogeneous, $\gamma_e = (\beta_e^2 - k^2)^{1/2}$ is the radial propagation number,

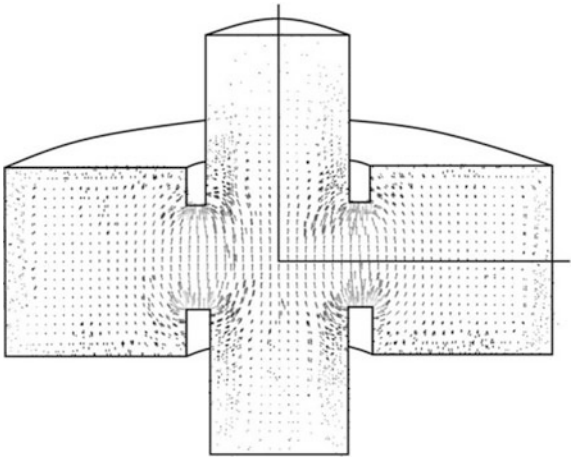


Fig. 7.2 Electric field of the gridless gap

$\beta_e = \omega/v_e$ is the electron phase constant, a is the drift channel radius, and $I_0(x)$ is the modified zero order Bessel function of the first kind.

This dependence leads to an inhomogeneous velocity modulation of the electrons crossing the interaction gap at various distances from the axis of symmetry. The gridless gap interaction factor taking into account the non-uniformity of the electric field longitudinal component along the radius has the following form:

$$M(r) = \frac{\sin(\beta_e d_1/2)}{\beta_e d_1/2} \frac{I_0(\gamma_e r)}{I_0(\gamma_e \mathrm{a})},$$

where r is the radial coordinate of the trajectory along which the electron moves.

As follows from the above formula, near the drift tube ($r \approx a$) the value of the interaction coefficient is the same as in the grid interaction gap. On the axis of symmetry ($r = 0$), the interaction factor is reduced. The electron beam usually occupies only a part of the drift tube cross section determined by the filling factor $\xi = b/a$, where b is the beam radius. Usually, ξ = 0.5–0.8. Thus, the difference in the modulation depth over the cross section of the beam is defined by the formula

$$\frac{v(0)}{v(b)} = \frac{1}{I_0(\gamma_e b)} = \frac{1}{I_0(\xi \gamma_e a)}.$$

The last formula shows that in order for the velocity modulation inhomogeneity to be small, it is necessary to limit the radius of the drift tube and the filling factor. Modulation analysis uses the average cross-sectional interaction factor

$$M_{\mathrm{med}} = \frac{1}{S} \int_S M(x_1, x_2) \mathrm{dS},$$

where S is the cross-sectional area of the beam, and x_1, x_2 are transverse coordinates. For a cylindrical beam

$$M_{\mathrm{med}} = \frac{M(a)}{\pi b^2 I_0(\beta_e a)} \int_0^b I_0(\beta_e r) 2\pi r dr = M(a) \frac{2I_1(\beta_e b)}{\beta_e b I_0(\beta_e a)}.$$

Figure 7.3 shows the dependence of the averaged interaction factor on $\beta_e a$ for different values of filling factor. It is this average interaction factor that must be substituted into the expression for the velocity modulation coefficient. As we see, the average interaction factor increases with an increase in the filling factor. However, the imperfect form of the electron beam limits the possibilities of magnification ξ.

The presence of a radial velocity component in the gridless gap is an undesirable effect, since this can lead to defocusing of the beam and losses of electrons on the wall of the drift tube.

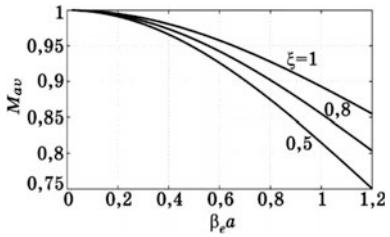


Fig. 7.3 Dependence of the averaged interaction factor on $\beta_e a$ for different filling factors

7.2.3 The Kinematic Theory of Bunching

After leaving the gap of the first resonator, the electrons fall into the drift tube where there is no external electric field. In the first approximation, we can neglect the electrons' disruptive forces and assume that in the drift space they move with the constant velocity which they attained when leaving the gap. Analysis of the motion of the electrons in such an approximation was named the kinematic theory of bunching.

The motion of electrons in the drift space is conveniently studied using the space-time diagram (Applegate diagram) (Fig. 7.4) which depicts the dependence of electron's coordinate on time. It is assumed that the electrons move along the z axis with the origin at the end of the gap.

The diagram shows the dependence of the electron's velocity at the gap exit on time. The dependence of the electron's coordinate on time is determined by function $z = v_e(t - t_0)$, where v_e is the velocity of the electron at the exit from the gap, and t_0 is the time at which the electron leaves the gap. In the space-time diagram, the electron trajectory is represented by a line, the slope of which is equal to the velocity of this electron. For an electron moving with constant velocity, its trajectory is a straight line.

Let's consider the motion of three layers of electrons in the drift space (Fig. 7.4). Layer 1 entered the drift tube at time t_1 with velocity v_{e1} slightly less than v_0:$v_{e1} = v_0 - \Delta v$. Layer 3 entered the drift tube later, at time $t_2 = t_1 + \Delta t_1$, but with

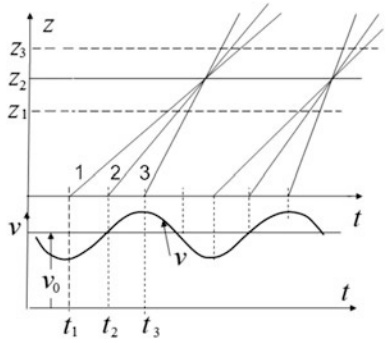


Fig. 7.4 Applegate diagram of electrons motion in the drift space

a higher speed $v_{e2} = v_0 + \Delta v$. Layer 2, occupying an intermediate position, entered the drift tube at time $t_2 = t_1 + \Delta t_1/2$ with velocity v_0.

Let us determine at what time interval layers 1 and 3 intersect plane P located at distance z from the gap. Layer 1 entered into the drift pipe at time t_1 and crosses this plane at time $t_1' = t_1 + z/(v_0 - \Delta v)$. Layer *2* entering the drift pipe at time t_2 intersects plane P at time $t_2' = t_2 + z/(v_0 + \Delta v)$. Then the time interval Δt at which these layers intersect plane P is found as

$$\Delta t = t_2' - t_1' = \left(t_1 + \Delta t_1 + \frac{z}{v_0 + \Delta v}\right) - \left(t_1 + \Delta t_1 + \frac{z}{v_0 - \Delta v}\right)$$
$$= \Delta t_1 + z\left(\frac{1}{v_0 + \Delta v} - \frac{1}{v_0 - \Delta v}\right),$$

or

$$\Delta t = \Delta t_1 - \frac{2\Delta v}{v_0^2 + (\Delta v)^2} = \Delta t_1 - \varphi(z),$$

where the designation is introduced

$$\varphi(z) = \frac{2z\Delta v}{v_0^2 + (\Delta v)^2}.$$

The function φ in this expression has a positive value depending on coordinate z. If $\varphi(z) < \Delta t_1$, then the considered layers intersect plane $z = z_1$ (see Fig. 7.4) in the same order in which they entered the drift tube, but in a shorter time interval $\Delta t < \Delta t_1$. For $\varphi(z) = \Delta t_1$ the layers intersect plane $z = z_1$ simultaneously. If $\varphi(z) > \Delta t_1$, then $\Delta t < 0$ and the layers intersect the considered plane in reverse order. Note that all the layers under consideration move in the same direction, therefore the sign of the convection current does not change when the Δt sign changes.

The layers will intersect plane $z = z_1$ (see Fig. 7.4) in the same order in which they entered the drift tube. The layers will come simultaneously to plane $z = z_2$. The layers will intersect plane $z = z_3$ in reverse order.

Calculation of the convection current in a bunched electron beam is based on the charge conservation law. The value of the charge concentrated between layers 1 and 3 can be found using the current of the unperturbed beam I_0 and time interval Δt_1: $\Delta q = I_0 \Delta t_1$. Charge Δq crosses plane z_1 over time interval Δt and creates a convection current $I = \Delta q/\Delta t = I_0 \Delta t_1/\Delta t$ in this plane. Passing to infinitely small time intervals, we obtain $I = I_0 \mathrm{d}t_1/\mathrm{d}t$.

If $\Delta t < 0$, the formulas for calculating the current should use the absolute value of the time interval $|\Delta t|$ since for a negative value of Δt only the order of this plane intersection by layers changes, but the direction of the electron current does not change.

Thus, the convection current of the beam

$$I = I_0 \left| \frac{\mathrm{d}t_1}{\mathrm{d}t} \right|. \tag{7.1}$$

The expressions obtained above establish the connection between the constant convection current entering the drift tube and the convection current arising in the drift tube in the process of the electrons bunching.

Therefore, to determine the convection current that results from bunching it is necessary to find the derivative $\mathrm{d}t_1/\mathrm{d}t$. The time for electrons to enter the drift tube t_1 and the time of the traversing through the plane located at distance z from its entrance are related by the obvious relation $t = t_1 + z/v$, where υ is the electron velocity at the entrance of the drift tube determined by formula (2.36). Considering this we have

$$t = t_1 + \frac{\tau}{1 + v_1 \sin \omega t_1},$$

where $\tau = z/v_0$ is the static time of the electron's transit.

Setting $v \ll 1$, corresponding to small values of the modulating voltage amplitude ($U_{m1} \ll U_0$) we find

$$t \approx t_1 + \tau(1 - v_1 \sin \omega t_1)$$

or

$$\omega t = \omega t_1 + \omega\tau - \omega\tau v_1 \sin \omega t_1.$$

By introducing the static transit angle $\theta = \omega\tau$ and *bunching parameter* $X = \theta \mathrm{v}_1$, we obtain

$$\omega t - \theta = \omega t_1 - X \sin \omega t_1. \tag{7.2}$$

This equation expresses the dependence of the electron's arrival phase ωt into the given cross-section of the beam with coordinate z on the phase of the electron's exit from the gap of the first resonator ωt_1. The z coordinate defines the static transit angle in the drift space θ and bunching parameter X (along with the velocity modulation factor v_1).

The dependence of $\omega t - \theta$ on ωt_1 for different values of bunching parameter X is shown in Fig. 7.5. As can be seen, for $X < 1$, one, and only one, value of arrival phase ωt corresponds to each value of emission phase ωt_1. The phase trajectories of the electrons on the Applegate diagram in the z cross-section do not intersect. The slope of the curve $\omega t - \theta = f(\omega t_1)$ is always positive so the value of the derivative $\mathrm{d}(\omega t_1)/\mathrm{d}(\omega t)$ is finite and the convection current in accordance with (7.1) is finite at any time.

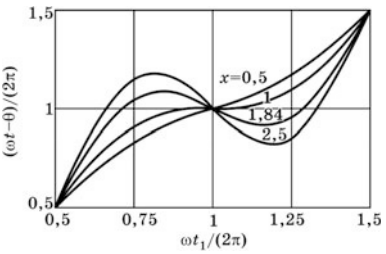


Fig. 7.5 Electron arrival phase versus transit phase from the first resonator gap

For $X = 1$, derivative $\mathrm{d}(\omega t)/\mathrm{d}(\omega t_1)$ at point $\omega t_1 = 2\pi$ becomes equal to zero, corresponding to the intersection of the electron trajectories in the Applegate diagram. Hence, several electrons exited from the first resonator at different moments of time arrive at section $z = \mathrm{Const}$ simultaneously. Finally, for $X > 1$ two such points of the intersection of electron trajectories appeared for each period, i.e. the electrons leaving the resonator later appeared in this section earlier than those leaving at earlier moments (the so-called *surpassing*). At the time corresponding to the intersection of electrons, the infinite pulses of the convection current appear. Let's recall that the effect of space charge is not taken into account.

To calculate the convection current, let us evaluate (7.2):

$$I(\omega t) = I_0 \left| \frac{\mathrm{d}(\omega t_1)}{\mathrm{d}(\omega t)} \right| = \frac{I_0}{|1 - X\cos(\omega t_1)|}.$$

The dependence of the convection beam current on the arrival phase calculated from this formula for different values of X is shown in Fig. 7.6. As can be seen, for $X = 1$ the infinite current peak is observed, and for $X > 1$ there are two infinite current peaks in each period. Let's note that despite the presence of infinite current peaks, the average value of the convection current over the period is finite and equal to I_0.

Thus, the dependence of the bunched beam convection current on time has a complex form which can be expanded in the Fourier series. To do this we introduce new variables $L = \omega t - \theta$ and $E = \omega t_1$. As a result, we obtain

$$L = E - X\sin E.$$

This equation is called the Kepler equation, and its solution is given by Bessel in the following form:

$$E = L + \sum_{n=1}^{\infty} \frac{2}{n} J_n(nX)\sin(nL),$$

where $J_n(z)$ are the first-kind Bessel functions of the n-th order.

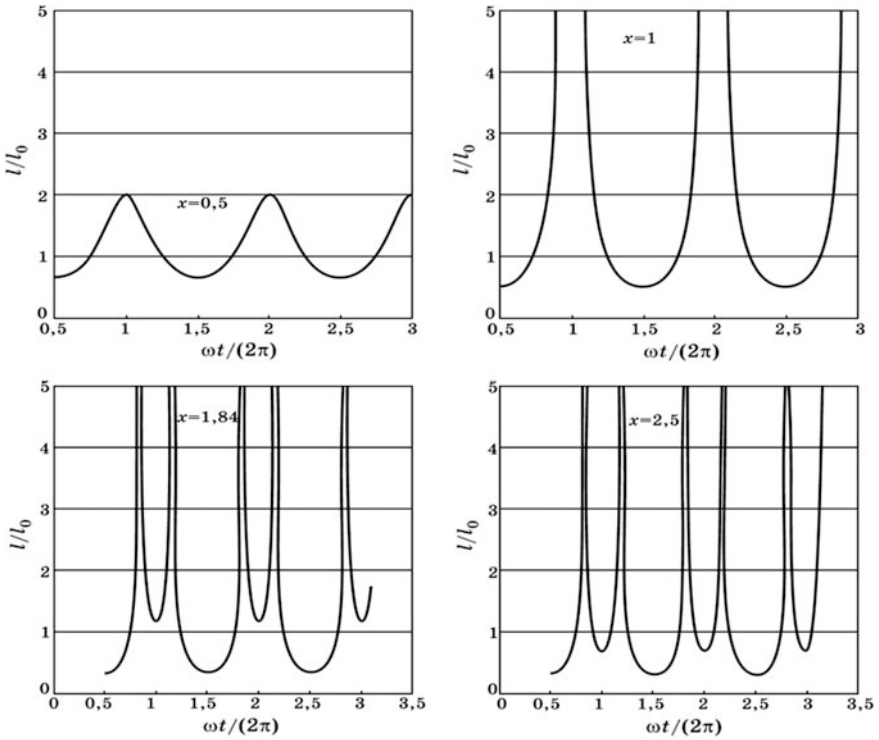


Fig. 7.6 Dependence of beam convection current on time for different values of the bunching parameter

Returning to the original variables, we find

$$\omega t_1 = \omega t - \theta + \sum_{n=1}^{\infty} \frac{2}{n} J_n(nX) \sin[n(\omega t - \theta)].$$

Calculating derivative $\mathrm{d}(\omega t_1)/\mathrm{d}(\omega t)$, we obtain the following expression for the convection current that results from the bunching of electrons:

$$I = I_0 + 2I_0 \sum_{n=1}^{\infty} J_n(nX) \cos[n(\omega t - \theta)].$$

The alternating current component is the sum of the harmonics with frequencies ω_n, multiple to the frequency of the modulating voltage ω and amplitudes $I_{mn} = 2I_0 J_n(nX)$.

In klystron-type devices intended for amplification and generation of microwave radiation, the fundamental harmonic ($n = 1$) with a frequency equal to the modulation frequency $\omega_1 = \omega$ is used as an operative harmonic:

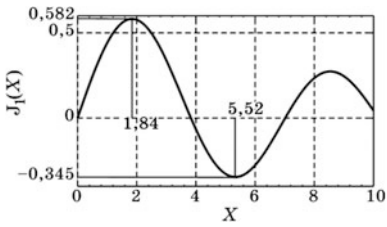


Fig. 7.7 Graph of the first-order Bessel function of the first kind

$$I_1 = 2I_0 J_1(X)\cos(\omega t - \theta).$$

Its amplitude is $I_{m1} = 2I_0 J_1(X)$. The graph of the Bessel function $J_1(X)$ is shown in Fig. 7.7. This function reaches a maximum value $J_1(X)_{\max} = 0.58$ with bunching parameter X equal to 1.84. Consequently, the maximum value of the convection current first harmonic amplitude is $I_{m1\,\max} = 1.16I_0$.

The values of the maximum amplitudes of higher harmonics decrease with increasing harmonic number, but rather slowly, for example, the maximum value of the second harmonic amplitude is $0.973I_0$, the third $-0.869I_0$, the tenth $-0.605I_0$. This fact indicates that klystrons can not only amplify microwave signals, but also efficiently multiply their frequency.

7.2.4 *Effect of Longitudinal Electron Repulsion*

The formation of electronic bunches leads to the local increase of the space charge density, as a result of which longitudinal components of the Coulomb interaction forces appear. They are directed from the center of the electronic group to the periphery, and prevent the approaching of the electron layers. Under the action of these forces, the slow electrons that form the leading edge of the electron bunch accelerate and the fast electrons forming the trailing edge decelerate. This leads to equalization of the electron velocities, as a result of which the bunching process stops and the reverse process—the process of ungrouping—begins. Then new bunches appear and the process repeats. These processes are shown in the space-time diagram (Fig. 7.8).

Under the influence of space charge forces, the electrons also acquire transverse velocities (transverse repulsion). This effect is not taken into account here since it is believed that the focusing system prevents the transverse "spreading" of the electron flux.

The theory of space-charge waves considered in Sect. 3.2 describes such periodic processes in the electron flux well.

In accordance with expressions (3.18) and (3.19), let's write the expressions for the current and velocity of electrons in these waves for the flux of a finite cross-section in the drift tube

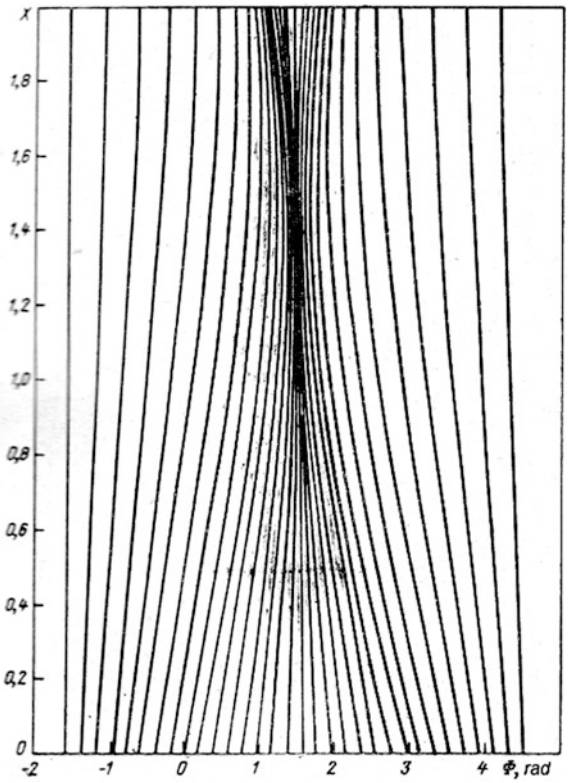


Fig. 7.8 Space-time diagram of the motion of electrons in the drift space, taking account of the space charge

$$\dot{I} = \dot{I}^{+}\mathrm{e}^{-j(\beta_e+\beta_q)z} + \dot{I}^{-}\mathrm{e}^{-j(\beta_e-\beta_q)z}, \tag{7.3}$$

$$\dot{v} = \frac{\beta_q}{\beta_e \rho_0 A}\left(\dot{I}^{+}\mathrm{e}^{-j(\beta_e+\beta_q)z} - \dot{I}^{-}\mathrm{e}^{-j(\beta_e-\beta_q)z}\right). \tag{7.4}$$

The space-charge waves are excited by the field of the resonator, which modulates velocity of the electron flux. At the exit from the narrow gap ($z = 0$)

$$\dot{I}_c(0) = 0, \quad \dot{v}(0) = v_1 v_0. \tag{7.5}$$

From the first condition, it follows that $\dot{I}^{-} = -\dot{I}^{+}$. From the second condition, we find

$$v_1 v_0 = 2\frac{\beta_q}{\beta_e \rho_0 A}\dot{I}^{+}.$$

This formula allows us to find $\dot{I}^+$. Substituting the result into (7.3), we obtain

$$\dot{I} = -jI_0 v_1 \frac{\sin(\beta_q z)}{\beta_q/\beta_e} e^{-j\beta_e z}. \tag{7.6}$$

The above analysis is valid only for small amplitudes of the current and velocity variable components, while the kinematic theory predicts the possibility of the alternating current component appearance with an amplitude larger than that of the constant component. In order to take into account the space-charge forces within this theory, it is necessary to compare the expressions for the convection current obtained by the space-charge wave theory and the kinematic theory for small values of the grouping parameters.

For space-charge waves

$$|\dot{I}| = I_0 v_1 \frac{\sin[(\beta_q/\beta_e)\beta_e z]}{(\beta_q/\beta_e)\beta_e z} \beta_e z = I_0 X \frac{\sin[(\omega_q/\omega)\beta_e z]}{(\omega_q/\omega)\beta_e z}. \tag{7.7}$$

From the kinematic theory of grouping, the amplitude of the convection current first harmonic is given by

$$\dot{I} = 2I_0 J_1(v_1 \beta_e z) e^{-i\beta_e z}.$$

For small values of the grouping parameter

$$|\dot{I}| \approx I_0 X = I_0 v_1 \beta_e z. \tag{7.8}$$

Comparing expressions (7.7) and (7.8) we see that if we replace X in the second of them by

$$X' = \frac{\sin(\alpha_q \theta)}{\alpha_q \theta} X = \frac{\sin(\alpha_q \theta)}{\alpha_q \theta} v_1 \theta, \tag{7.9}$$

where $a_q = \beta_q/\beta_e = \omega_q/\omega$ is the *repulsion parameter*, they coincide. It is proved that such a substitution is also valid for large values of the bunching parameter. Thus, for the alternating component of the convection current, taking into account the longitudinal repulsion of the electrons, the valid expression for the convection current is:

$$I_c = I_0 + 2I_0 \sum_{n=1}^{\infty} J_n(nX') \cos[n(\omega t - \theta)]. \tag{7.10}$$

In the following text, the bunching parameter is always determined by formula (7.9) and is designated as X (without prime).

At finite value of parameter a, the value of the correction taking into account the influence of the Coulomb forces is less than one, $(\sin a_q\theta)/(a_q\theta) < 1$. This means

that the Coulomb forces decrease bunching parameter. To maintain its value, it is necessary to increase the depth of the velocity modulation that ultimately leads to the increase in the required input power and the decrease in the gain of the device.

It follows from (7.9) that the optimal length of the drift tube is equal to one quarter of reduced plasma wavelength. As the length of the drift tube increases, the resulting bunch begins to ungroup under the action of space-charge forces. In practice, the length of drift tubes are chosen as less than a quarter of the reduced plasma wavelength. Such choice reduces the device amplification factor, but also lowers its mass and dimensions.

7.2.5 The Extraction of Energy from the Bunched Electron Beam

The Output Power, Efficiency and Gain.

In the analysis, it is assumed that the output resonator is tuned to the frequency of the input signal, i.e. its resonance frequency ω_{02} is equal to the signal frequency ω and the resonance frequencies of the higher modes do not coincide with the signal frequency harmonics.

The convection current passing through the interaction gap of the output resonator creates an induced current in the resonator, which leads to the excitation of the resonator. The induced current, like the convection current, contains a lot of harmonics. However, for a sufficiently large resonator Q-factor, the field excitation coefficients of all harmonics, except for the fundamental harmonic of the induced current, with frequency ω and equal to the resonance frequency, are very small and can be ignored. Therefore, when analyzing the cavity excitation, it is sufficient to consider the fundamental harmonic of the induced current.

To calculate it, we use relation (2.33) which connects the convection current flowing in the diode gap with the induced current:

$$\dot{I}_{i2} = M_2 \dot{I}_{c1}(l_{12}) \mathrm{e}^{-i\beta_e l_{12}}, \tag{7.11}$$

where l_{12} is the distance between the midpoints of the first and second gaps.

Knowing the value of the induced current, it is possible to determine the alternating voltage at the gap of the output resonator U_{m2}:

$$U_{m2} = -R_{e2} I_{c1} M_2 \cos(\omega t - \theta),$$

where R_{e2} is the equivalent impedance of the output resonator tuned to resonance.

The negative sign in this expression takes into account the fact that the resonator tuned to resonance with the voltage is automatically phased in such a way that the maximum slowing phase of the voltage falls within the maximum of the electron current.

The power of electromagnetic oscillations in the output resonator

$$P_{\text{out}} = 0.5\dot{I}_{i2}^2 R_{e2} = \frac{1}{2} M_2^2 \dot{I}_{c1}^2 R_{e2} = \frac{1}{2} M_2^2 I_0^2 4 J_1^2(X) R_{e2}, \tag{7.12}$$

where R_{e2} is the equivalent impedance of the output resonator.

Electronic efficiency of the klystron η_e is defined as the ratio of the oscillation power in the resonator P_e to the power of the electron flux $P_0 = U_0 I_0$:

$$\eta_e = P_e / P_0.$$

Taking into account the formulas obtained above, the expression for the electronic efficiency can be represented in the following form:

$$\eta_e = \frac{P_e}{P_0} = \frac{U_{m2} I_{c1} M_2}{2 I_0 U_0} = J_1(X) M_2 U_{m2} / U_0. \tag{7.13}$$

Let's estimate its maximum value. The first factor in this expression reaches its maximum value with bunching parameter $X = 1.84$ and is equal to $J_1(X)|_{\max} = 0.58$. It is obvious that the maximum value of the electronic efficiency is achieved when the electrons completely lose their kinetic energy, i.e. when their velocity at the exit from the second resonator gap is zero. At this point, the voltage effective amplitude of the cavity $M_2 U_{m2}$ should be equal to the accelerating voltage U_0. With larger voltage amplitudes, some electrons will receive negative velocities, i.e. they will be rejected from the resonator gap back into the drift tube. This so-called *rejection effect* reduces the output power because the rejected electrons extract energy from the resonator field. Therefore, ratio U_{m2}/U_0 in the maximum efficiency mode should be taken as $1/M_2$. Thus, the estimate gives the following maximum value of the electronic efficiency: $\eta_e|_{\max} = J_1(1.84) = 0.58$, or 58%.

The total klystron efficiency is defined as the ratio of the output power to the power of the electron beam: $\eta = P_{\text{out}}/P_0$ is numerically equal to the product of the electronic efficiency at the resonator efficiency η_r: $\eta = \eta_e \eta_r$. The resonator efficiency takes into account the power losses in the resonator due to the finite conductivity of the resonator walls. Its value can be expressed in terms of loaded Q_l and intrinsic Q_0 of the resonator Q-factor:

$$\eta_r = 1 - Q_l / Q_0.$$

In practice, the resonator efficiency depends on a number of factors, in particular the device operation mode and can be 60–95%.

The klystron gain is defined as the ratio of output power P_{out} to input power P_{in}:

$$\mu_p = P_{\text{out}} / P_{\text{in}}. \tag{7.14}$$

The output power is determined by expression (7.13). The value of the input power can be expressed in terms of the amplitude of the alternating voltage at the resonator gap U_{m1} and its equivalent resistance:

$$P_{\text{in}} = \frac{1}{2R_{e1}} U_{m1}^2.$$

Substituting the output and input power values in formula (7.14), we obtain:

$$\mu_p = \frac{2M_2^2 I_0^2 J_1^2(X) R_{e1} R_{e2}}{U_{m1}^2}. \tag{7.15}$$

Using the definition of the bunching parameter, we find

$$U_{m1} = \frac{2U_0 X}{M_1 \theta}. \tag{7.16}$$

Substituting (7.16) into (7.15), we obtain

$$\mu_p = \frac{J_1^2(X) M_1^2 M_2^2}{2X^2} \frac{R_{e1} R_{e2}}{R_0^2} \theta^2. \tag{7.17}$$

From this formula, it follows that with the transit angle, i.e. the length of the drift tube increasing, the gain increases indefinitely. However, the effect of space charge leads to the fact that the optimal value of the transit angle is $\theta = \pi/2$ which limits the klystron gain.

The amplitude characteristic of the double-cavity klystron is the dependence of output power P_{out} and gain μ_p on the value of input power P_{in} (Fig. 7.9).

In the small signal mode $X \ll 1$ and we can set J_1 (X) = $X/2$. Substituting this value in (7.15) and setting $\theta = \pi/2$, we find

$$\mu_p = \frac{\pi^2}{36} M_1^2 M_2^2 \frac{R_{e1} R_{e2}}{R_0^2} \approx 0.27 M_1^2 M_2^2 \frac{R_{e1} R_{e2}}{R_0^2}. \tag{7.18}$$

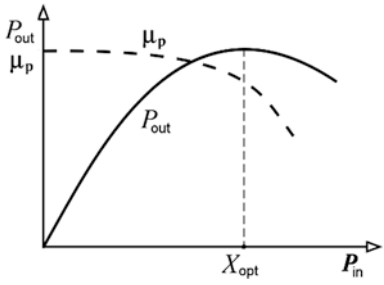


Fig. 7.9 Amplitude characteristic of the double-cavity klystron

In the maximum power mode $J_1(X) = 0.58$, $X = 1.84$. Hence, the gain in the maximum power mode

$$\mu_p|_{P_{\max}} = \frac{0.58^2 M_1^2 M_2^2 R_{e1} R_{e2}}{2 \cdot 1.84^2 R_0^2} \approx \frac{0.12 M_1^2 M_2^2 R_{e1} R_{e2}}{R_0^2}. \tag{7.19}$$

As can be seen, in the maximum power mode the gain is approximately two times smaller than in the low-signal mode. This indicates that the amplitude characteristic of the klystron is nonlinear. Note that usually the power gain of microwave devices is measured in decibels:

$$\mu_p[\text{dB}] = 10 \log \mu_p.$$

Typical values of the power gain of the double-cavity klystron lie in the range 15–20 dB.

The dependence of P_{out} on P_{in} has a linear initial section; the gain retains an approximately constant value in this section. With further increase in the input power, the increase in the output power decelerates (the saturation effect takes place). The maximum value of output power is achieved with optimal bunching of electrons in the drift tube (bunching parameter $X = 1.84$). Further increase of the input power leads to rearrangement of the electrons, destruction of the electron bunch, and a decrease in output power.

The frequency-response characteristic (FRC) of the device shows the dependence of output power on frequency. This dependence is generally determined by the FRC of the klystron cavities, i.e. their resonance frequencies and Q-factors. In double-cavity klystrons, both resonators are tuned to one frequency. At this the loaded Q-factor of the input resonator is usually much larger than the loaded Q-factor of the output resonator, since in order to obtain high circuit efficiency the output resonator must have a strong coupling with the load. Therefore, the relative pass-band of a klystron in the small-signal mode is determined by the loaded Q-factor of the input resonator:

$$\Delta f / f_0 = 1 / Q_{l1}$$

and as a rule does not exceed several tenths of a percent.

The equivalent impedances of resonators are related to their wave impedances and loaded Q-factors:

$$R_{e1,2} = \rho_{1,2} Q_{l1,2}.$$

Taking this expression into account, the formula for the gain of double-cavity klystron can be written as follows:

$$\mu_p = \frac{J_1^2(X)}{2X^2} \frac{M_1 \rho_1}{R_0} \frac{M_2 \rho_2}{R_0} Q_{l1} Q_{l2} \theta^2. \tag{7.20}$$

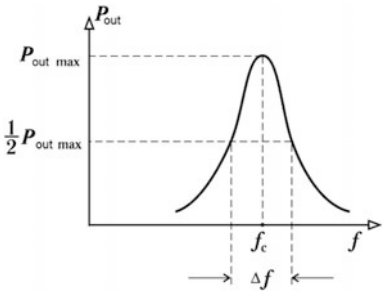


Fig. 7.10 Frequency-response characteristic of a two-cavity klystron

Thus, the expansion of the bandwidth leads to a decrease in the gain of the device. The typical FRC of a double-cavity klystron is shown in Fig. 7.10.

7.2.6 Multi-Cavity Klystrons

The multi-cavity amplification klystron includes several intermediate (idle) resonators in addition to input and output ones (Fig. 7.11). This figure shows cathode C, focusing electrode FE, input resonator C_1, two intermediate resonators C_2, C_3, output resonator C_4 and collector Col. Accelerating voltage U_p relative to the cathode is applied to the resonators and collector.

The simplest version of the device—the three-cavity klystron—contains one intermediate resonator. The number of resonators in modern klystrons is n = 4–8.

The main processes that determine the operation of these devices are the same as in the above double-cavity klystron: velocity modulation, beam bunching, excitation of the resonator and extraction of energy from the electron beam.

Let's trace the behavior of these processes in a multi-cavity device. *The first* (*input*) resonator is excited by input signal P_{in} from an external source, and alternating voltage $U_1 = U_{m1} \sin \omega t$ appears in its interaction gap, creating the primary

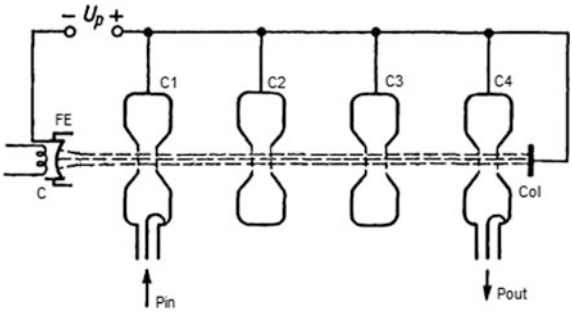


Fig. 7.11 Scheme of the multi-cavity klystron

velocity modulation of the electron beam. It is assumed that the amplitude of alternating voltage U_{m1} is much smaller than that of accelerating voltage U_0. The velocity-modulated electron beam is bunched in the first drift tube T_1. The bunching process leads to the formation of an alternating convection current that contains a number of harmonics, including the fundamental, with a frequency equal to the frequency of the input signal ω. Passing through the gap of the second resonator, the bunched beam induces the current in it, which excites the second resonator. Alternating voltage U_2 appears on the resonator gap, and its amplitude U_{m2} is usually much larger than the voltage amplitude of the first resonator U_{m1}, $U_{m2} \gg U_{m1}$. The electron beam passing through the gap of the second resonator undergoes secondary velocity modulation, the depth of which considerably exceeds the depth of velocity modulation in the first resonator. The bunching of the electron beam in the second drift tube is mainly determined by the secondary velocity modulation.

The alternating electron current produced in the second drift tube excites the third resonator. The electron processes in the subsequent cascades of a multi-cavity klystron are similar to those described above.

Note that the lengths of the drift tubes found from the condition $l = \lambda_q/4$ are usually too large, and it becomes necessary to reduce them based on constructive considerations. At this, the amplitude of the alternating component of the beam current inevitably decreases. Increase of cavities number in the buncher usually compensates this effect.

The output (nth) resonator performs the final stage of the transformation of electron beam kinetic energy into energy of the electromagnetic field—the extraction of energy from the electron beam.

The use of a multi-cavity electrodynamic system opens many possibilities for improving the operational parameters of klystrons such as increasing the gain, increasing electronic efficiency and expanding klystron bandwidth.

There are three main operating modes of the multi-cavity klystron:

- Synchronous tuning mode;
- Increased electronic efficiency mode;
- Broadband amplification mode.

Synchronous Tuning Mode

In this mode, all cavities of the device are tuned to the frequency of the input signal. The amplitude of the alternating voltage at the interaction gaps of the resonators increases with the number of the resonator i ($i = 1, 2, 3, \ldots$). The voltage amplitude in each subsequent resonator significantly exceeds the amplitude of the previous one, $U_{mi} \gg U_{m(i-1)}$. In this case, the device cascades including resonators starting with the first ($i = 1$) to n minus two ($i = n - 2$) can be considered as a voltage amplifier, which provides excitation of the device output section. This section includes n minus 1 (the $n - 1$ cavity), the drift tube T_{n-1} and the output resonator of the device ($i = n$).

The voltage amplitude at the gap of the resonator with number $n - 1$ is considerably greater than the voltage amplitudes in the gaps of the previous cascades and, in

particular, the voltage amplitude at the gap of the previous n minus two resonator $U_{m(n-1)} \gg U_{m(n-2)}$. It determines the deep velocity modulation and bunching process in the output section of the device. The background of electronic processes associated with velocity modulation and bunching in previous cascades has practically no effect on electronic processes in the output section. Therefore, this cascade of the device can be considered as a double-cavity klystron in which the input resonator is excited not by an external source but by the electron beam. In particular, the electronic efficiency of the multi-cavity klystron operating in synchronous tuning mode appears to be the same as for the double-cavity klystron and its maximum value is 58%.

The gain of the multi-cavity klystron operating in synchronous tuning mode depends on the number of resonators and can be estimated from the empirical formula:

$$\mu_p \approx 15 + 20(n-2),$$

where $\mu_p = 10\log(P_{\text{out}}/P_{\text{in}})$ is the device power gain expressed in decibels, and n is the number of device resonators. The value of the gain can reach 60 dB. It should be noted that the increase of multi-cavity klystron gain is achieved due to the decrease of the required value of the input power.

The bandwidth of the klystron is mainly determined by the resonance characteristics of the intermediate resonators and increases with the decrease of their Q-factor. Its relative value is usually a fraction of a percent.

Increased Electronic Efficiency Mode

An increase in electronic efficiency is possible due to an improvement in the degree of electron beam bunching. There are two ways and two methods used to solve this problem.

The first of them assumes a change in the character of the velocity modulation of the electron beam. It can be shown that bunching approaches an ideal with such a periodic change in velocity when within a single period (from $-T/2$ to $T/2$) the velocity alternates according to the law

$$v = v_0 \frac{1}{1 - t/\tau}, \tag{7.21}$$

where v_0 is the velocity constant component, and τ is the static electrons transit time in the drift tube ($\tau > T$).

All the beam electrons that enter the drift tube during one period T are bunched into a thin disk carrying charge Q. Its value is $Q = I_0 T$, where I_0 is the constant current of the electron beam, and T is the velocity modulation period. The waveform of the current is described by the Dirac function:

$$I_c(t) = I_0 T \delta(wt - 2\pi n - \theta), \quad n = 0, 1, \ldots$$

As it is known, when this function is expanded in the Fourier series, a series with equal amplitudes of all harmonics is obtained: $I_{cn} = 2I_0, \quad n = 1, 2, \ldots$. Thus, for

ideal bunching, the amplitude of the convection current first harmonic is $2I_0$ instead of $1.16I_0$ as for the sinusoidal bunching voltage.

If such a bunch leaves the output cavity gap with zero velocity, the theoretical value of electronic efficiency approaches 100%.

Unfortunately, the implementation of velocity modulation with the required velocity change law meets serious practical difficulties. Approach to this ideal velocity dependence on time can be achieved by the use of cavities tuned to the second harmonic of the input signal.

Figure 7.12 shows the addition of oscillations on the first and second harmonics. The resulting curve closely follows the ideal relationship also shown in the figure. The duration of the increasing section of the total curve is 4.2 radians instead of π radians for the sinusoidal voltage waveform.The optimum phase shift angle between the voltages of the first and second harmonics is $\pi/2$ which can be achieved by detuning the second harmonic cavity. However, the voltage on the resonator gap decreases, therefore, this resonator should be placed after two or three cavities tuned to the first harmonic so that the depth of the velocity modulation is sufficient to form the noticeable second harmonic amplitude at the location of the cavity tuned to the second harmonic. Generally, resonators tuned to the first and second harmonics are detuned in opposite directions and this allows the higher second harmonic voltage to be achieved, the optimum value of which is 1/2 of the first harmonic voltage. Powerful multi-cavity klystrons are constructed, containing up to eight cavities, one of which is tuned to the second harmonic of the signal. Multi-cavity klystrons have a total efficiency of up to 85%, but in a rather narrow frequency bandwidth not exceeding 1%.

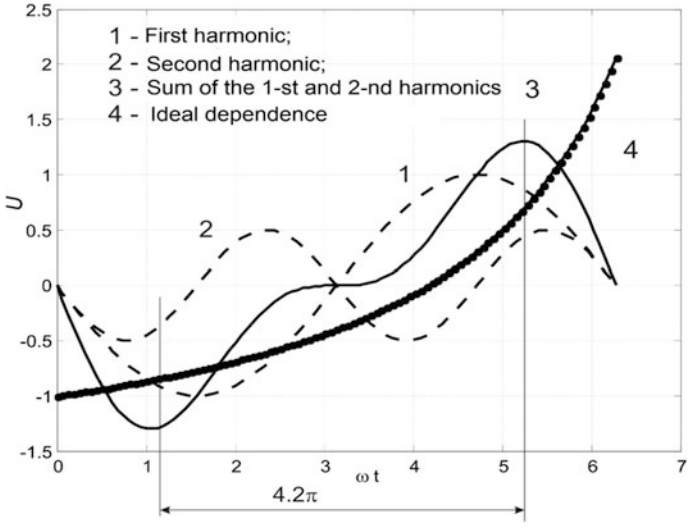


Fig. 7.12 Addition of the first and second harmonics voltages

The second method of increasing electronic efficiency is called *additional bunching* and involves the addition of electronic bunches formed in the output and preceding cascades of the device.

Let's consider the possibilities of this method using the example of a **three-cavity klystron**.

The first cavity of the device is excited from an external source and provides the velocity modulation of the electron beam. The alternating convection current formed in the first drift tube excites the second cavity of the device.

At the interaction gap of this resonator, an alternating voltage U_2 arises with an amplitude U_{m2} that is usually much larger than the amplitude of the first resonator voltage U_{m1}, $U_{m2} \gg U_{m1}$. The gap voltage phase is set so that the maximum decelerating voltage coincides with the center of the bunch formed by the first resonator (if the cavity is tuned in resonance). On the Applegate diagram (Fig. 7.13) 1—time interval of the bunching voltage in the first cavity; 2—the same in the second cavity; 3—the bunch length in the second gap. As can be seen, only half of the electrons from the first bunch enter the new bunch formed by the second resonator. This phenomenon limits the charge of the bunch formed by the second resonator.

In order for the entire primary bunch to be used in the formation of the secondary one, it is necessary to shift the phase of the voltage on the second resonator by $\pi/2$ (the dashed curve in Fig. 7.13). For this purpose, the intermediate resonator is tuned to a frequency slightly larger than the signal frequency $\omega_{02} > \omega$. In this case, its equivalent conductivity becomes complex:

$$Y_r = G_r + jB_r$$

where $G_r = G$, and $B_r = G_r\xi$ are the active and reactive components of equivalent conductivity, $\xi = 2Q(\omega - \omega_{02})/\omega_{02}$ is the generalized resonator detuning, and Q is its Q-factor.

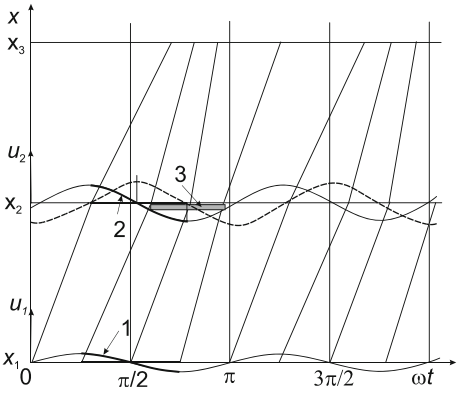


Fig. 7.13 Analysis of the grouping process in a three-cavity klystron

The phase shift between the induced current and the voltage is determined by the expression

$$\varphi = \text{arctg}(B_r/G_r) = \text{arctg}\,\xi.$$

If the intermediate resonator is tuned to a frequency greater than the signal frequency, $\omega_{02} > \omega_0$, the generalized detuning has a negative value. In this case, the reactive conductivity of the resonator is negative, i.e., it has inductive character, and the voltage arising at the resonator gap advances in phase induced current by angle φ, the value of which increases with resonator detuning enlargement (Fig. 7.14).

In this, the zero phase of the alternating voltage arising at the interaction gap of the second resonator shifts to the phase of arrival of the primary electron group to the gap. With a large detuning, $\xi \ll -1$, phase angle φ tends to 90°. In this case,

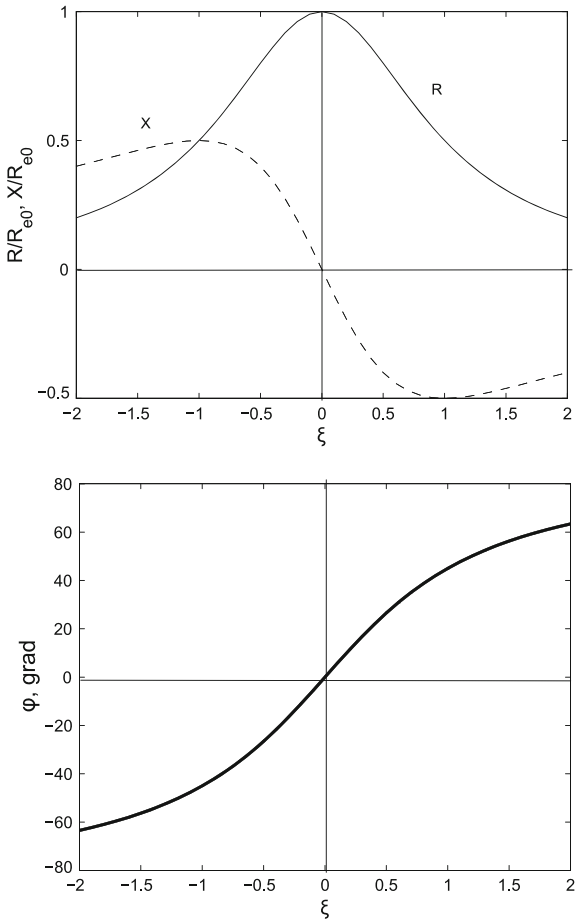


Fig. 7.14 Electron conductivity at the skirtron scheme of cavities tuning

the zero phase of the voltage $U_2(\omega t)$ will coincide with the phase of the primary electronic bunch arrival at the second resonator, and additional bunching due to the second resonator field could occur with respect to the center of the primary bunch.

However, such a mode cannot be realized practically, since at the increase of the resonator detuning, its equivalent impedance

$$R_e = Re(1/Y_e) = \rho Q/(1 + \xi^2),$$

and accordingly the interaction gap voltage $U_2(\omega t)$ both drop. In order to maintain the value of the voltage $U_2(\omega t)$ for detuning corresponding to angles φ = 75°–80° it is required to increase the amplitude of the alternating convection current at the entrance to the second resonator gap.

This, in turn, is achieved by increasing the power of the input signal and the degree of primary bunching of the electron beam. In this case, the electron beam improved bunching mode is actually realized. The calculations give a maximum theoretical value of electronic efficiency for a three-cavity klystron as equal to 74% (for comparison, the maximum electron efficiency of a double-cavity klystron is 58%).

In a multi-cavity klystron with a number of cavities n = 6–8, the mode of additional bunching can be realized not only by the pre-output $(n - 1)$th resonator detuning, but simultaneously by the corresponding detuning of the previous $(n - 2)$ th and $(n - 3)$th resonators. This allows an increase in the bunching degree of the beam arriving at the output resonator and brings the electronic efficiency value to 85% and more.

The mode of additional grouping (the higher efficiency mode) is usually used in devices designed to obtain high power in the narrow frequency band, which is determined by the frequency characteristics of the resonators; its relative value does not usually exceed 1%.

Broadband amplification mode. This mode of klystron operation has the greatest practical application. It is used in devices operating in various telecommunications systems including communication systems using artificial earth satellites and communication systems with space objects, and is also used in devices in radar installations.

To analyze the processes in a klystron, it is convenient to divide it into two sections (systems): the section that produces an alternating convection current in the given frequency band (buncher) and the power extraction section (output system).

The problem of obtaining the specified frequency band is solved separately for the buncher and for the output system. The frequency properties of the buncher are determined by the frequency dependence of the convection current feeding the output system. The parameters of the buncher are selected so that the frequency response of the current has the required bandwidth with an unevenness smaller than that specified (permissible).

Proper choice of resonant frequencies and Q-factors of the buncher cavities can provide nearly flat klystron's frequency response. At the present time, there is no universal prescription for tuning cavities that ensures uniform grouping of the

electron beam in the frequency band, and there is also no unified approach to solving this problem.

One of the possible schemes for frequency-tuning cavities in broadband klystrons which has received practical application is the so-called *skirtron scheme*. According to this scheme, the input resonator is tuned to the lower edge of the frequency band, and all other buncher resonators are tuned to frequencies near its upper edge (Fig. 7.15). The output resonator is tuned to the signal frequency.

In this figure, the relative frequency value f/f_0 is plotted along the abscissa axis, where f_0 is the central frequency of the operating frequency band, and along the ordinate the current alternating component at the buncher output is shown. The dotted curves show the resonance characteristics of the first ($n = 1$), intermediate ($i = 2, \ldots, n-1$) and output resonators ($i = n$).

From this figure, it follows that when the signal frequency change within the frequency band, the "tails" of the impedance characteristics of the cavities are used, where the dependence of the impedance on the frequency is expressed weakly, which, in fact, provides the possibility of obtaining the "flat" characteristic of the buncher alternating current.

Obviously, the use of such tuning scheme is inevitably associated with a loss of amplification and leads to an increase in the number of device cavities.

The skirtron tuning scheme does not provide sufficiently uniform frequency dependence of the convection current first harmonic in the output resonator. Therefore, the modified skirtron scheme is often also used, differing in that the frequency of the second resonator tuning is shifted toward the central frequency.

Figure 7.16 shows as an example the current characteristics of a buncher containing six resonators in the case of the conventional (the solid curve) and the modified skirtron scheme (the dashed curve). The frequency settings of the individual buncher resonators for these two cases are presented in Table 7.1.

As can be seen, the modified scheme allows a flatter current characteristic to be obtained.

It should be emphasized that the considered tuning schemes are just one of many possible systems for broadband tuning of buncher resonators. The common factor in these systems is the tuning of the last two buncher resonators to frequencies near

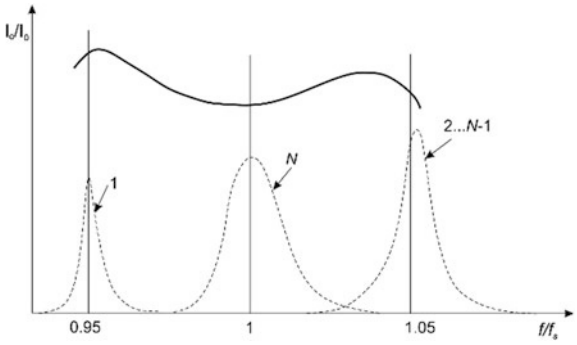


Fig. 7.15 Current characteristics of the buncher in the skirtron-scheme for tuning cavities

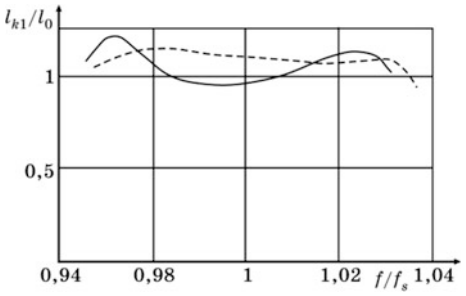


Fig. 7.16 Klystron FRC in the modified skirtron scheme. The block curve is the skirtron scheme, the dashed curve is the modified scheme

Table 7.1 Resonance frequency of klystron cavities for skirtron and modified skirtron tuning scheme

Tuning schemes of resonators							
Res. No.	1	2	3	4	5	6	Tuning scheme
f_0/f_s	0.977	1.039	1.038	1.036	1.036	1.035	Skirtron
f_0/f_s	0.977	1.014	1.037	1.039	1.050	1.057	Modified

the upper boundary of the frequency band, or to frequencies that lie beyond this boundary.

When designing the buncher, it is important to create at the gap entrance of the output cavity a so-called *uniformly gathering bunch*, deviation from the average electron velocity of which is proportional to the distance from the center of the bunch. In this case, at the center of the output gap, the velocities of all the electrons in the bunch due to the action of longitudinal repulsion forces can be practically identical (Fig. 7.17) ensuring the maximum electronic efficiency of the device. To create such a bunch, an optimal scheme of tuning and Q-factors of the buncher resonators is also needed, and sometimes it is also necessary to artificially reduce the Q-factor of some intermediate resonators with the help of an absorber placed in them.

Energy extraction from the electron beam. The bunched electron beam enters the gap of the output resonator, creating an induced current with the amplitude of the first harmonic

$$I_{i1} = M_{\text{out}} \kappa_1 I_0,$$

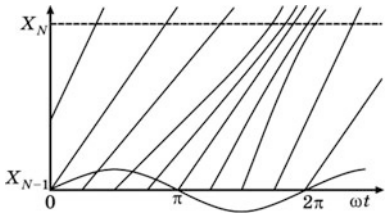


Fig. 7.17 Space-time diagram of the gathering bunch

where coefficient κ_1 depends on the bunching quality and takes typical values of 1–1.5. The induced current creates the voltage on the resonator gap. At the resonance frequency $U = R_e I_{i1} = \rho Q_l M_n \kappa_1 I_0$, where ρ is the cavity wave impedance, and Q_l is its loaded Q-factor.

The most efficient extraction of energy occurs when the electrons of the bunch completely lose their velocity in the gap, i.e., when effective voltage at the gap $U_e = M_n U$ is equal to accelerating voltage U_0. Hence the optimal effective wave impedance of the output resonator

$$\rho_e M^2 = \frac{R_0}{\kappa_1 Q_l} = \frac{R_0}{\kappa_1}\frac{\Delta f}{f_0}.$$

Thus, at operation with maximum electronic efficiency, the bandwidth of the amplified frequencies appears to be directly proportional to the ratio of the effective resonator wave impedance to the beam impedance at constant current $R_0 = U_0/I_0$.

The value of ρM^2 is determined only by the shape of the resonator and does not depend on its resonant frequency and Q-factor. Its maximum value for ordinary toroidal resonators does not exceed 150 Ω. It follows that to expand the bandwidth of amplified frequencies, it is necessary to use output resonators with increased effective impedance and (or) to apply coupled oscillation systems.

Another method of expanding the band is based on reducing the electron beam DC-resistance. These methods can be used both individually and together. Let's consider them in more detail.

7.2.7 Extended Interaction Klystrons

Let us consider two toroidal resonators connected among themselves by a coupler—by a loop or a hole in a common wall. In this system, two types of oscillations can exist—in-phase and anti-phase. In the first case, the phase shift between oscillations in the resonators is zero, and in the second case, it is 180°. Let the distance between the centers of the resonator gaps be chosen so that the electrons transit the gaps in the same phase as the voltage on them. Then the effects of the action on the electrons of each gap field will add up and the total voltage acting on the resonators will be equal to $U_e = 2U$, where U is the voltage on the gap. The stored energy in the two resonators is obviously equal to the doubled energy stored in one resonator (if we do not take into account the energy stored in the coupler, which is usually small). Thus, the effective impedance of the double-gap re-entrant resonator

$$\rho_e M^2 = \frac{|U_e|^2}{2\omega_0 W_e} = \frac{4|U|^2 M^2}{2\omega_0 2W_e} = 2\rho M^2, \tag{7.22}$$

is twice that of ρM^2—the effective impedance of one resonator. Accordingly, by connecting N resonators we get a gain in the impedance by a factor of N. Such an

N-gap resonator can be considered as a short-circuited section of a periodic transmission line having N resonances with phase shifts $0, \pi/N, 2\pi/N, \ldots, \pi$ between oscillations in neighboring resonators.

Usually, a zero or π-mode of oscillation providing the greatest uniformity of the field amplitude in the gaps is used. Multi-gap cavities (resonators with distributed interaction) are particularly effective in millimeter-wave klystrons where technological and thermal limitations prevent the creation of single-gap resonators with high effective impedance. Klystrons with multi-gap cavities are named *extended interaction klystrons* (EIKs).

As the number of gaps increases, the separation of frequencies between individual modes decreases and the share of energy stored in the couplers increases, thereby slowing the increase of the effective impedance. Besides, in each gap the velocity of the bunch decreases, so the distance between the centers of the gaps should be reduced as the gap number increases. In practice, resonators with a number of gaps larger than six are not used.

A high effective impedance (150–250 Ω) is possessed by two-gap quarter-wave resonators. Such resonators are widely used in devices of low and medium power in the decimeter and centimeter bands. Their use in high-power klystrons is hampered by poor heat removal from the central rod.

It should be noted that multi-gap cavities are used not only as output, but also as intermediate ones in the buncher.

The use of coupled resonators. The bandwidth of the energy extraction system can be increased if the active resonator (interacting with the electron beam) is coupled to the output waveguide through the system of coupled (so-called passive) resonators. Such a system has a frequency response of the bandpass filter with a wider bandwidth compared to the single resonator. Thus, the system of two identical connected circuits at the critical degree of coupling between them has a bandwidth $\sqrt{2}$ times larger than that of a single circuit, and at the optimal coupling is 3.1 times greater. This expansion is achieved, however, at the cost of increasing the stored energy in the system and correspondingly reducing the impedance of the active resonator.

However, taking into account the fact that the loaded Q-factors of passive resonators are small, the decrease of impedance is usually insignificant and the expansion of the band appears to be the more significant factor.

The simplest coupled system consists of the active resonators (and one passive resonator) formed by diaphragms in the output waveguide (Fig. 7.18). Inductive diaphragms are used because they do not reduce the electrical strength of the waveguide in practice.

The distance between the diaphragms should be somewhat larger than a quarter of the wavelength in the waveguide at the central frequency, and the distance between the first diaphragm and the coupling hole should be equal to an odd number of wave's quarters. A waveguide segment of this length acts as a transformer that modifies the serial equivalent circuit of the passive resonator into the parallel one.

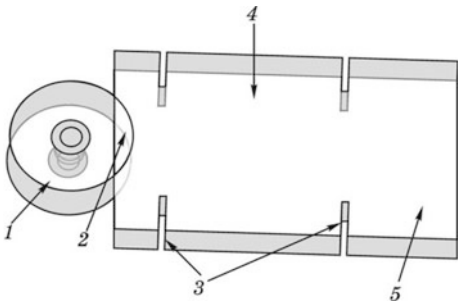


Fig. 7.18 Filter system of energy extraction: 1 Active resonator. 2 Coupling hole. 3 Diaphragms. 4 Passive resonator. 5 Output waveguide

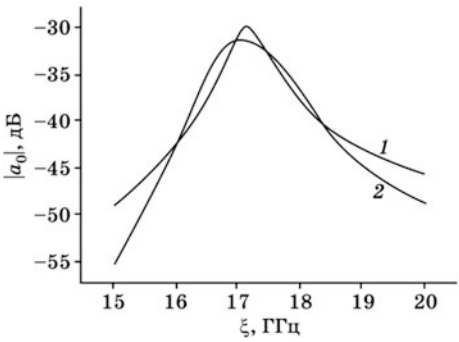


Fig. 7.19 Klystron output system FRC: 1 Without passive resonator. 2 With passive resonator

Depending on the degree of coupling between the active and passive resonators, the frequency response characteristic (FRC) of the output system can be single-humped (with the coupling less than critical) or two-humped (with the coupling more than critical). Figure 7.19 shows the FRC of the output system depicted in Fig. 7.18 in the absence (curve 1) and presence of a passive resonator (curve 2). As can be seen, the introduction of the passive resonator makes it possible to substantially expand the transmission band of the klystron output system.

7.2.8 Multi-Beam and Multi-Barrel Klystrons

Multi-beam klystrons (MBK) opened new possibilities for creating broadband amplifiers. Although the first patents for multi-beam klystrons were obtained back in the 1940s, the prototype of the modern MBK was the device developed by S.V. Korolev in the 1970s.

In a MBK, a third method of expanding the klystron bandwidth is utilized: electron beam DC-impedance reduction. In a single-beam klystron, an increase in the beam current at the given power $P_0 = U_0 I_0$ leads to an increase of the beam perveance with a corresponding increase of space charge forces that prevent the beam bunching and make its transportation difficult. In a multi-beam device, instead

of one beam, several beams are used that run parallel and do not interact with one another (except for the space of high-frequency gaps). Thus, each beam can have a current $I_{01} = I_0/N$, where N is the number of beams, and the repulsive forces in each beam can remain small. Modern MBK contain from 7 to 60 beams. Besides the extended frequency band, MPKs have one more important advantage over single-beam devices: they have a lower supply voltage, which makes it possible to simplify the design of the power supply, reduce its weight and dimensions and increase the reliability of operation.

To form the beams in a MBK, usually one electron gun is used, on the cathode of which there are several emitting surfaces. A beam transport system (usually magnetic) is also common to all beams.

The exhausted beams fall into the common collector and deposit on its surface. For each beam, a separate transit channel is created, but the beams interact with the resonator field in the common high-frequency gap.

Let's consider the relationships between the parameters of a single-beam and N-beam klystrons having the same output power P_{out} and the same beam perveance p. It is assumed that both klystrons have the same efficiency η.

The output power of the MBK is given by

$$P_{\text{out}} = \eta N I_1 U_{0N}, \tag{7.23}$$

where I_1 is the current of one beam, and U_{oN} is the MBK accelerating voltage. Taking into account that

$$I_1 = pU_{oN}^{3/2}, \tag{7.24}$$

we find

$$P_{\text{out}} = \eta p N U_{oN}^{5/2}, \tag{7.25}$$

For a single-beam klystron, the analogous formula is valid

$$P_{\text{out}} = \eta p U_0^{5/2} \tag{7.26}$$

where U_0 is the accelerating voltage of the single-beam device. Equating the right-hand sides of formulas (7.25) and (7.26), we find the relation between the accelerating klystron voltages:

$$U_{0N} = U_0/N^{2/5}. \tag{7.27}$$

Accordingly, the current of one MBK beam

$$I_N = Np(U_0/N^{2/5})^{3/2} = I_1 N^{2/5}.$$

In accordance with formula (7.27), for $N = 7$ the accelerating voltage decreases by 2.18 times, and for $N = 19$, by 3.25 times.

The construction of a multi-beam klystron with a magnetic focusing system based on permanent magnets is shown in Fig. 7.20.

The seven-beam electron flux is formed by the electron gun, which includes the cathode 1, the cathode electrode 2, which is the seven-hole diaphragm, and the anode electrode 3. The formed flux enters the electrodynamic system consisting of resonators 11 and drift tubes. In the drift tubes, each of the beams moves in a separate channel 10. Transportation of electron beams through the electrodynamic system is provided by means of the magnetic focusing system that is formed by annular magnets 4 magnetized in the radial direction, pole tips 5, 6 and magnetic shields 7, 8, 9.

With the correct ratio of the pole tips diameter D to the distance between them l, $D/L \geq 1$, the magnetic field in the drift tubes area is close to uniform. Since the electron beams move in separate channels, for analysis of their transportation the main results and formulas obtained for the single-beam flux can be applied. The exhausted electron flux, which has passed through the electrodynamics system, is deposited on the collector 12.

Let's consider the main advantages that are provided by the use of multi-beam electron fluxes.

1. At a given value of the accelerating voltage U_0, the use of an N-beam flux makes it possible to increase the total flux current by a factor of N. Accordingly, the perveance of the flux $p_N = p_1 N$ increases by N times. In this case, the limitations for the flux perveance associated with the virtual cathode effect are practically removed, since the flux beams move in separate channels.
2. The given power level of the electron flux P_0 can be achieved at a much lower anode voltage.
3. The impedance of the electron flux is substantially reduced:

$$R_N = U_0/I_N = (U_0/I_1)N^{4/5} = R_1/N^{4/5}.$$

where R_N, R_1 are the impedances of the N-beam and single-beam electron flux respectively. This makes it possible to reduce the required values of the equivalent

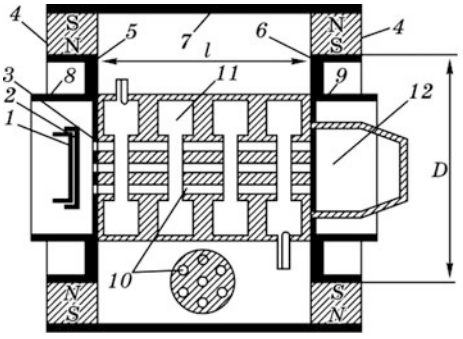


Fig. 7.20 Scheme of a multi-beam four-cavity klystron

impedances of the klystron resonators and provides the possibility of expanding the bandwidth of the multi-cavity klystron.

Disadvantages of multi-beam klystrons are mainly due to the difficulty of creating an electron-optical system for such a device. It is practically impossible to create a multi-beam EOS with converging optics, therefore in the MBK the density of the cathode current is equal to the current density in the beam and reaches several tens of amperes per square centimeter. This limits the service life of the devices. Besides, the large cross-sectional area of the multi-beam flux and, consequently, the interaction area of the resonator, reduce the uniformity of the field along the radius. As a result, the wave impedance of the resonator for the central and peripheral beams is different. In addition, the large interaction area reduces the wave impedance of the resonators that partially compensates the effect of reducing the flux DC impedance. In the millimeter wavelength band, considerable difficulties arise with the alignment of the drift tubes of multi-beam devices.

The so-called multi-barrel klystrons can be considered as a further development of multi-beam devices. These devices combine two or more single-beam or multi-beam klystrons with the common vacuum envelope and magnetic system. They have coupled input, intermediate and output resonators and separate electron guns and collectors. The scheme of such a double-barrel multi-beam device is shown in Fig. 7.21 (the vacuum envelope is not shown). Usually, the prismatic cavities operating at the higher oscillations mode (Fig. 7.22) are used as input and output resonators of multi-barrel klystrons.

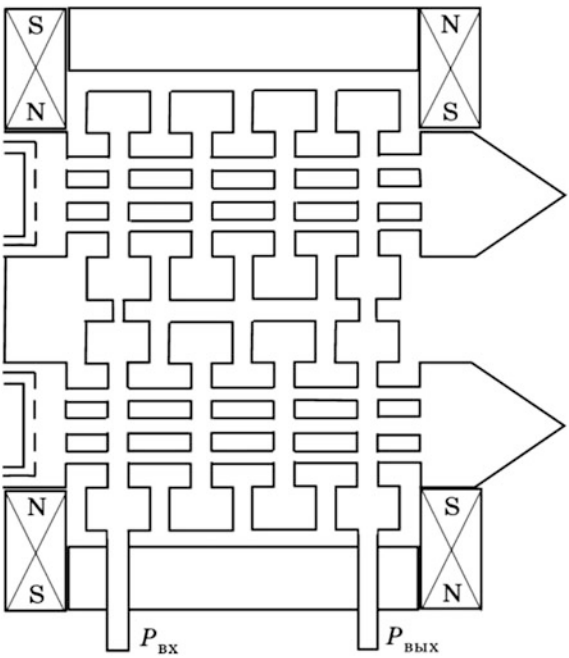


Fig. 7.21 Scheme of a double-barrel multi-beam four-cavity klystron

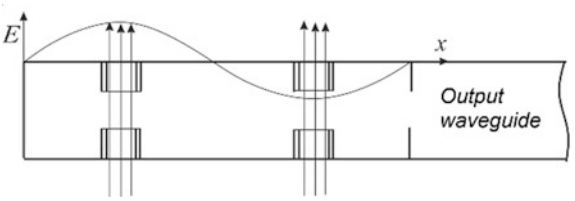


Fig. 7.22 The output resonator of the double-barrel klystron

The klystrons with this design allow an increase of total flux current and output power of the device. The width of the amplified frequency band is not widened in practice, since the resonator operating at the higher oscillations mode has reduced wave impedance.

Multi-barrel klystrons provide a range of opportunities for FRC synthesis with significant reduction in the weight and dimensions of the device in comparison with two separate klystrons, but the complexity of the design limits their use in the millimeter wavelength band.

7.2.9 Sheet Beam Klystrons

An increase in the beam current without increasing the repulsive force of the electrons is also achieved when a sheet electron beam is used. In such a beam, the forces of space charge are determined by the specific perveance of the beam, i.e., by its perveance divided by the width of the beam. The specific perveance value does not depend on the beam width, while the beam current is proportional to its width. Therefore, it is possible to obtain a large current with a small specific perveance, which provides a small beam DC-impedance, good bunching, and high efficiency. A small specific perveance also allows the use of reverse or periodic magnetic systems with permanent magnets to transport the beam. Resonators used for interaction with the sheet beam have their own peculiarities. In particular, their wave impedance is usually less than for resonators designed for interaction with a cylindrical beam. However, a planar structure of such resonators allows the use of modern methods of micromechanics and electrochemistry in their manufacture.

This possibility becomes especially valuable in the development of millimeter and submillimeter band devices.

The shortcomings of sheet beam devices include beam instability in the transverse direction when the cross section of the beam while propagating becomes S-shaped instead of rectangular-shaped. This effect becomes more significant as the beam width increases.

Figure 7.23 shows a klystron with distributed interaction and a sheet beam developed in SLAC. The operating frequency of the device is 94 GHz, the output power is 8 kW, and the beam has transverse dimensions of 0.32×4 mm^2. Filters eliminate the coupling between the resonators through the drift tube.

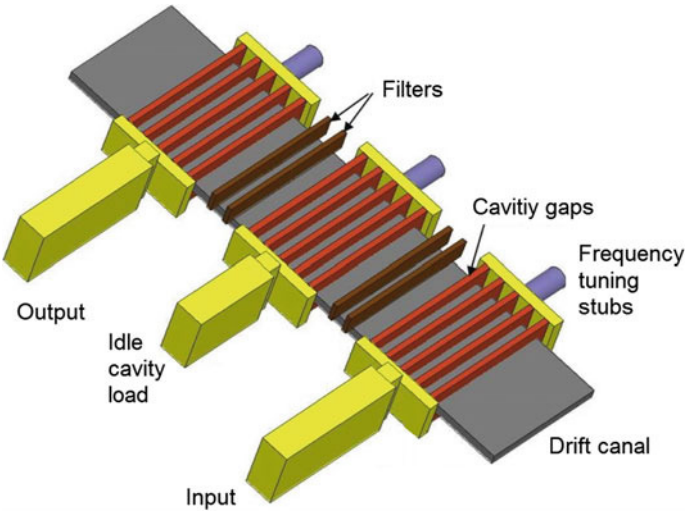


Fig. 7.23 A klystron with a sheet beam

7.2.10 Structure, Parameters and Characteristics of Modern Klystrons

The modern klystron is a complex device, the manufacture of which needs to use the most modern materials and technological processes. Depending on the operating frequency, output power and operating mode, klystron construction has certain features.

In the decimeter band, klystrons are usually designed for use with external resonators. The vacuum envelope of such devices is formed by metallic drift tubes and ceramic rings welded to them. At the ends of the drift tubes there are metal disks forming high-frequency gaps of the re-entrant cavities.

The outer surfaces of these disks located in the air are connected to the non-vacuum inductive part of the resonator. This construction allows a reduction in the volume of the klystron vacuum part and facilitate its pumping and the maintenance of a vacuum during operation.

As a rule, klystrons of the centimeter and millimeter bands are constructed with internal resonators, which form part of the vacuum envelope of the device. The internal resonators do not include ceramic parts, and that increases their wave impedance, while the inner surface of the resonator is protected from the harmful effects of the atmosphere. The absence of mechanical contacts increases the stability of the frequency and other parameters of the resonator.

The electron gun is performed as a separate unit of the device. Many modern klystrons have the option of low-voltage modulation of the beam current, for which the control and shadow grids are introduced into the gun's construction. Electron guns of single-beam devices are performed with beam compression that allows a

reduction in the density of the cathode current and increases the service life of the device. The guns of multi-beam devices are usually used without compression, and the density of the cathode current in them reaches several tens of amperes per square centimeter. In klystrons of the millimeter band, guns that form a sheet electron beam are sometimes used.

As a rule, for the transportation of an electron beam in transit-time klystrons, a magnetic focusing system is used. High-power devices of the decimeter band can have a length of more than 1 m and the diameter of external resonators can be 20–30 cm, therefore solenoids are used as the magnetic system. Higher-frequency devices with internal resonators have smaller dimensions, so they can use magnetic systems with permanent magnets, creating a constant, reverse or periodic magnetic field.

An important part of the transit-time klystron is the collector that collects the exhausted electrons. The main part of the thermal energy, into which the residual kinetic energy of the electrons is converted, is released in the collector. If we take the electronic efficiency of the klystron as equal to 50%, then the thermal power released to the collector is equal to the output power of the device. Therefore, the design of the collector should as much as possible ensure more uniform distribution of the convection current density along its internal surface and good liquid or air cooling. Collectors with energy recuperation are used in klystrons with relatively low efficiency (20–30%) which is typical for low-power klystrons and millimeter band klystrons.

Couplers with input and output circuits provide the power supply to the input resonator and the amplified signal extraction from the output resonator, ensuring the vacuum tightness of the device. To separate the vacuum part from the atmosphere, there are special vacuum windows, represent by the glass or ceramic plate located in the cross section of the transmission line through which the signal is transmitted (see Appendix C). The vacuum window should provide minimum level of signal reflections in the working frequency band of the klystron, and the absence of breakdowns in all permissible modes of device operation.

These requirements are relatively easy to meet for the input coupler, but the design of the output vacuum window of high-power and super-power devices is an independent and complex problem. These windows are complex structures of metal and ceramic including wave converters, waveguide sections, complex-shaped ceramic diaphragms and a number of other elements.

By their design and purpose, amplifying klystrons are divided into several groups.

1. **Super-power klystrons for linear accelerators and feeding colliders**. These devices are characterized by a large output power (several tens of MW in pulse), a large relative pulse duration, and the narrow band of amplified frequencies (tenths of a percent). As a rule, these are single-beam devices, since the large beam current requires the use of electron guns with high compression. The operating frequency of most klystrons of this group lies in the C-band (about 3 GHz), but recently, klystrons operating in the X-band (12 GHz) have been

developed. The frequency increase makes it possible to substantially reduce the length of the accelerator with the given energy of the particles.

As an example, we use the parameters of the klystron XL4, developed in the Stanford Linear Accelerator Center (SLAC). Its appearance is shown in Fig. 7.24.

The basic parameters of the klystrona are:

- Operating frequency—11.424 GHz.
- Accelerating voltage—440 kV.
- Beam current—350 A.
- Microperveance—1.2 μA/V$^{3/2}$.
- Output pulse power—50 MW.
- Pulse duration—1.5 μs.
- Average power—9 kW.
- Number of cavities—7.
- Output system type: four-gap, traveling wave, phase shift to gap—π/2.
- Focusing system—solenoid.
- Induction of magnetic field—0.47 T.

Fig. 7.24 Klystron XL-4

The klystron XL4 serves as a "workhorse" in many of the world's accelerator centers. Its main disadvantage is the large power consumed by the focusing solenoid (about 20 kW). In this connection, an analog of this device was developed – the klystron 75XP3 using periodic focusing by permanent magnets. It appeared to be necessary to reduce the perveance to 0.75 μA/V$^{3/2}$ with the accelerating voltage increasing to 506 kV. One period was added to the output system. As a result, it managed to increase the output pulse power to 75 MW with a pulse duration of 1.6 μs and a repetition rate of 120 Hz. The gain of the described devices in the small signal mode is 45–55 dB.

2. **Broad-band klystrons for radar and communication**. These devices are characterized by a smaller pulsed power and low relative pulse duration. The bandwidth of the devices can reach 10%. Let's give the parameters of the pulsed multi-beam klystron KIU-124 (Fig. 7.25), produced by FGUP "SPE "Thorium", Moscow:

   - Operating frequency—C-band;
   - Accelerating voltage—11 kV;
   - Bandwidth—6%;
   - Pulsed output power—20 kW;
   - Average output power—1.4 kW;
   - Efficiency—25%;

Fig. 7.25 Klystron KIU-124

- Gain—45 dB.
- Focusing system—permanent magnets, reverse.

3. **Television klystrons** are intended to operate in transmitters of terrestrial air TV centers in quasi-continuous mode. They operate in the meter and decimeter wavelength bands. As a rule, these klystrons have outer re-entrant cavities, the non-vacuum inductive part of which is connected to the vacuum capacitive part (gap) by means of flat radial electrodes.
   The parameters of a typical klystron of this kind, the КУ-352, developed by FGUP "SPE "Thorium", Moscow (Fig. 7.26) are:

   - Frequency band—470–630 MHz;
   - Power in sync pulse—13.5 kW;
   - Instantaneous bandwidth—8 MHz;
   - Gain—30 dB;
   - Accelerating voltage—20 kV;

Fig. 7.26 Television klystron KU-352

- Focusing—removable permanent magnets;
- Cooling—forced air;
- Weight—150 kg.

4. **Compact klystrons for airborne radar**. The rigid requirements for the weight and dimensions of such devices and their power supplies determine their construction. As a rule, these are multi-beam and (or) multi-barrel packaged devices with internal resonators. The focusing system is reverse or periodic with permanent magnets. The main parameters of the КИУ-208 device developed at JSC "Svetlana-Electronpribor" (Fig. 7.27) are:

   - Operating frequency—10 GHz;
   - Bandwidth—4.5%;
   - Accelerating voltage—6.5 kV;
   - Output pulsed power—25 kW;
   - Duty rate—30;
   - Number of barrels—2;
   - Total number of beams—54;
   - Cooling—liquid.

The device has low-voltage grid modulation, which required the development of a specially constructed electron gun.

5. **Klystrons of the millimeter and submillimeter bands** are one of the few high-power sources of coherent radiation in specified bands. Their construction is determined by the need to meet the rigid tolerances (in the order of micrometers) for the manufacture of the main parts of the device. For example, the diameter of the drift tube and the resonator gap length of W-band klystrons are a fraction of a millimeter. The complexity and high cost of manufacturing restrict the development and production of this type of klystron. In the technology of their manufacture, both precision machining centers, electro erosion processing plants and X-ray lithography processes are used similarly to those used in the production of integrated chips.

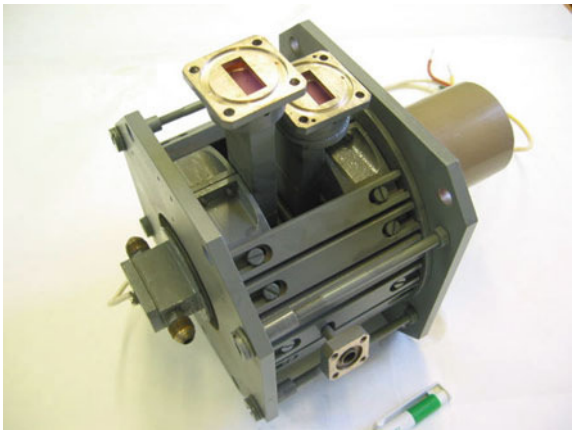


Fig. 7.27 Two-barrel klystron КИУ-208

Table 7.2 Output power level of klystrons produced by CPI

Parameter	35.5 GHz EIK	94 GHz EIK
Pulsed power (W)	3000	2000
Bandwidth (MHz)	200	250
Relative pulse duration	10	10–30
Gain (dB)	46	55
Efficiency (%)	40	33
Weight (kg)	7	6.5
Life duration (h)	>35,000	50,000

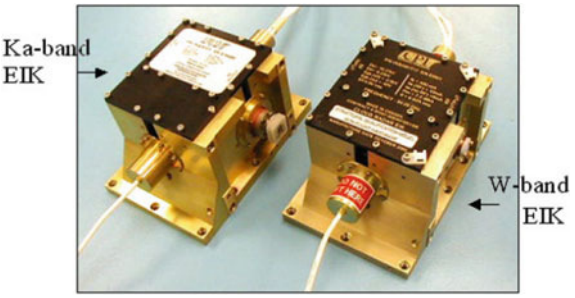


Fig. 7.28 CPI Ka- and W-bands EIKs

Table 7.2 presents millimeter band klystron parameters for satellite meteorological radars produced by CPI, Canada. The devices use multi-gap resonators. Images of these devices are shown in Fig. 7.28.

7.2.11 Other Types of Klystrons

Double-Cavity Oscillator Klystron

Oscillator klystron structure is similar to the structure of double-cavity amplifying klystrons (Fig. 7.1). The difference is that this device has the feedback element from the output resonator to the input resonator. The hole in the wall separating the resonators can serve as such an element (inductive coupling). Part of the electron power given by the electron beam to the output resonator is transmitted through this hole to the input resonator providing the self-excitation mode of the device. In this case, the alternating voltage on the input resonator gap U_1 and the alternating voltage of the output resonator U_2 are related to each other by the relation

$$U_{m1} = kU_{m2}\mathrm{e}^{-j\varphi_{fb}}$$

where k is the voltage feedback coefficient equal to the ratio of the voltage amplitudes of the first and second resonators: $k = |U_{m1}|/|U_{m2}|$ in the absence of the

electron beam, and φ_{fb} is the voltage phase shift in the first and second resonators in the absence of the electron flux.

This condition imposes limitations on the values of the transit angle of electrons in the drift tube (Fig. 7.29). The alternating voltage U_1 performs the velocity modulation of the electron beam. The electron bunch is formed with respect to the electron passing through the interaction gap of the first resonator at the moment when voltage U_1 changes from the decelerating to the accelerating one.

To effectively transfer the energy to the field of the second resonator, the group must enter its interaction gap at the moment of the maximum decelerating phase of the voltage. As can be seen from the curves in Fig. 7.29, this condition will be met if the transit angle of the electrons in the drift tube is

$$\Theta = \frac{3}{2}\pi - \varphi_{oc}, n = 0, 1, 2, \ldots$$

or

$$\theta = 2\pi n + 3/4 - \varphi_{fb}.$$

On the other hand, the value of the transit angle in the drift tube is determined by the formula

$$\theta = \omega l / \sqrt{\frac{2e}{m} U_0},$$

where l is the drift tube length, and U_0 is the accelerating voltage.

Combining the last two formulas, we obtain the necessary phase condition for the oscillator klystron self-excitation:

$$\omega l / \sqrt{2eU_0/m} = 2\pi(n + 3/4) - \varphi_{fb}.$$

The amplitude condition for self-excitation is met if the product is

$$A = Gk > 1,$$

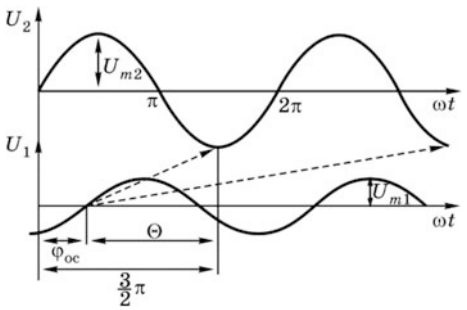


Fig. 7.29 Analysis of the self-excitation conditions of an oscillator klystron

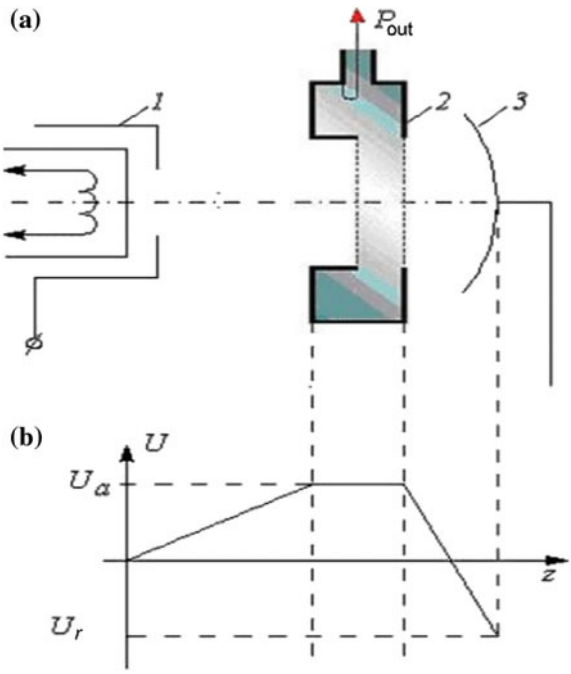


Fig. 7.30 Reflex klystron scheme (**a**) and potential distribution in the interaction space (**b**)

where G is the voltage gain of the klystron in the absence of feedback.

Oscillator klystrons are used as a simple and inexpensive source of microwave energy of small and medium power. However, their release is limited since the stability of the frequency of oscillations and the spectrum purity are often insufficient for specific applications.

Reflex Klystron

The reflex klystron (RK) was invented by R.V. Sutton (England) and independently by N.D. Devyatkov and V.F. Kovalenko (the USSR) in 1940. The scheme of its structure and the axial distribution of the potential are presented in Fig. 7.30.

The electron beam is formed and accelerated by an electron gun 1 and enters the interaction gap of the cavity 2. Assume that the resonator is excited and there is alternating voltage $U = U_m \sin \omega t$ at its gap. After passing through the interaction gap, where it undergoes the velocity modulation, the electron beam enters the resonator-reflector space. Here, the beam electrons are bunched.

In the space between the cavity and the reflector, all the electrons are decelerated by the reflector field to a complete stop and then they return to the cavity.

The accelerated electrons possessing more energy in the resonator come closer to the reflector, their path and time of motion in the bunching space are greater than those of the decelerated electrons.

The space-time diagram presented in Fig. 7.31 illustrates the bunching process and allows us to determine the phase conditions under which the electrons returning

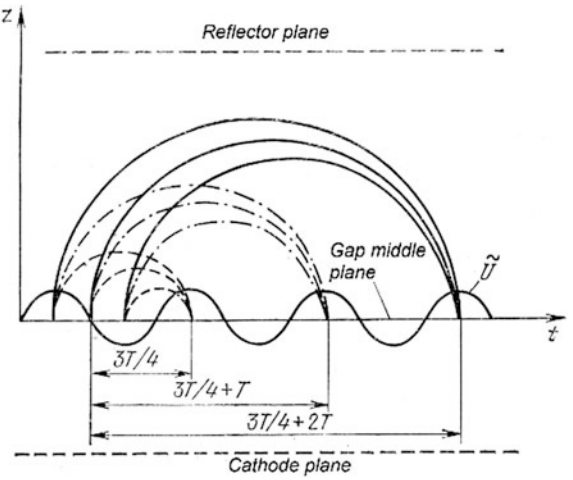


Fig. 7.31 Generation zones of a reflex klystron

to the interaction gap transmit their energy to the cavity field and support the oscillations arising in it.

As can be seen from the figure, the electrons that have passed the resonator gap at different moments of time, form bunches while returning to the resonator. The centers of the electron bunches are the electrons passing through the interaction gap at the moment of zero alternating voltage at its transition from accelerating to decelerating. The electron beam initially having constant density is transformed into a sequence of electron bunches. The repetition frequency of the bunches corresponds to the frequency of the resonator gap alternating voltage modulating the beam.

To maintain the oscillations in the resonator, it is necessary that the bunches return to the resonator gap in the half-period of the alternating voltage which decelerate them. Here it should be remembered that the voltage half-period, which was accelerating for the electrons moving in the forward direction, is decelerating for electrons returning to the interaction gap from the resonator-reflector space.

When the electronic bunch is decelerated, the energy is transferred from the electron beam to the resonator field. Since the bunches contain most of the beam electrons, the energy transferred to the field exceeds the energy loss for the acceleration of the electrons that are not part of the bunch. These electrons transit through the resonator gap during the accelerating half-period. From the Applegate diagram, we can see that the condition for maintaining oscillations in the resonator can be achieved at different reflector voltages for which the optimal transit time of electrons in the grouping space τ_{opt} is related to the period of the alternating voltage T in the following way:

$$\tau_{opt} = (3/4)T + nT, \quad n = 0, 1, 2, \ldots$$

The transit time of electrons in the bunching space depends on the geometric and electrical parameters of the device and is determined by the formula:

$$\tau = \frac{l\sqrt{8mU_0/c}}{U_0 + |U_r|},$$

where τ is the static transit time of electrons in the resonator-reflector space; l is the resonator-reflector distance; and U_0 is the accelerating potential, and U_r is the reflector potential.

Combining the last two formulas, we obtain an equation determining the phase conditions under which the transfer of energy from the electron beam to the resonator field will be at a maximum:

$$\frac{3}{4}T + nT = \frac{l\sqrt{8mU_0/c}}{U_0 + |U_r|},$$

or, substituting $T = 2\pi/\omega$ we find

$$n + \frac{3}{4} = \frac{\omega}{2\pi}\frac{l\sqrt{8mU_0/c}}{U_0 + |U_r|}.$$

Consequently, the reflex klystron has discrete generation zones, each of which corresponds to a discrete value of n.

For fixed values of distance l and accelerating voltage, U_0 the number of the zone in which generation occurs depends on the reflector voltage. The reflector voltage values corresponding to the centers of different generation zones (different values of n) are determined by the equation given above.

The current in the cavity gap adds its conductance to the cavity value. Beam conductance has real and imaginary parts. The value of the imaginary part depends on the phase shift between current and gap voltage. Hence, the generating frequency depends on the reflector voltage. The appearance of the generation zones with the change of reflector voltage, as well as the change of the output power and frequency within each zone are shown qualitatively in Fig. 7.32.

The alternating convection current arising in the process of bunching contains a lot of harmonics with frequencies that are multiples of the alternating modulating voltage frequency: $\omega_i = i\omega, \quad i = 1, 2, \ldots$. The first (fundamental) harmonic of the convection current ($i = 1$) entering the resonator interaction gap is defined by the expression

$$I_1 = 2I_0 J_1(X)\cos(\omega t - \theta - \xi/2),$$

where I_0 is the constant component of the convection current; $J(r)$ is the first-order Bessel function of the first kind, its argument is the bunching parameter $X = \frac{1}{2}\frac{U_m}{U_a}M\Theta$; and Θ is the electrons transit angle in the bunching space.

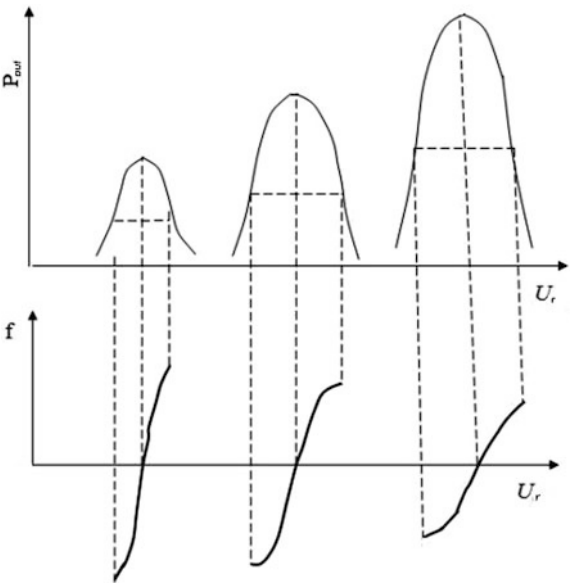


Fig. 7.32 Generation zones of a reflex klystron

The amplitude of the induced current first harmonic is equal to the amplitude of the convection current multiplied by interaction coefficient M, $I_{in} = 2I_0J_1(X)M$. The power transmitted by the electron beam to the resonator (the electronic power) will be determined as

$$P_e = 0.5U_mI_{im} = U_mI_0J_1(X)M.$$

In the reflex klystron, the alternating voltage at the resonator gap provides both the energy extraction from the bunched electron beam and its velocity modulation. The amplitude of the alternating voltage U_m is included in the formula of the bunching parameter. To determine the dependence of the electronic power on the bunching parameter, it is necessary to express the voltage amplitude through the value of the bunching parameter $U_m = 2XU_0/(M\theta)$. Substituting this expression in the formula of electronic power, we obtain

$$P_eI_0U_02XJ_1(X)/\theta.$$

The transit angle of electrons in the resonator-reflector space is determined by the optimal transit time for the given generation zone τ_{opt}. $\theta = \omega\tau_{\text{opt}} = 2\pi(n+3/4)$. Taking this into account, we obtain

$$P_e = I_0U_0\frac{XJ_1(X)}{\pi(n+3/4)}.$$

This expression allows us to find the electronic efficiency of the reflex klystron:

$$\eta_e = \frac{P_e}{P_0} = \frac{P_e}{I_0 U_0} = \frac{XJ_1(X)}{\pi(n+3/4)}.$$

As expected, electronic efficiency depends on the value of bunching parameter X, and this dependence is more complicated than in the straight-transit double-cavity klystron. The product $XJ_1(X)$ reaches a maximum of 0.4 at the value of $X = 2.41$.

Electronic efficiency depends on the generation zone number decreasing with the growth of the zone number (Table 7.3).

According to the calculation, the maximum value of electronic efficiency is in the zero zone, however, the amplitude of the alternating voltage exceeds the accelerating voltage, and such a mode cannot be realized in practice. In reality, reflex klystrons operate in the second and third generation zones, therefore their full efficiency does not exceed 3–5%.

The advantage of a reflex klystron as a generator of microwave radiation is the electronic tuning of the frequency. As the reflector voltage changes within one generation zone, the frequency of the oscillations generated also changes (Fig. 7.32).

This change in frequency is accompanied by a change of output power. Electronic tuning is characterized by the bandwidth and differential steepness.

The tuning bandwidth means the interval of the frequency change Δf, within which the output power changes no more than for a half, or, as they say, the tuning band is determined at the power level 1/2 from maximum. The differential steepness of the electronic tuning depends on the zone number and the electrical mode. For the middle of the band, it is defined by the formula:

$$S_f = f_0 \frac{\pi(n+3/4)}{Q_l(U_0 + |U_r|)},$$

where f_0 is the frequency corresponding to the middle of the generation zone; and Q_n is the loaded Q-factor of the resonator.

Typical values of voltage on the RK resonator are 250–300 V, the beam current is 20–50 mA, the output power is 10–50 mW, and the electronic tuning band is 0.2–0.5%. The operating frequency of reflex klystrons lies in the band from decimeter to millimeter waves. Figure 7.33 shows the appearance of a reflex klystron with an external resonator (a) and an internal resonator (b).

In recent decades, much work has been done to create subminiature reflex klystrons, which have been termed “minitrons”. These devices have very low weight and low operating voltages.

Table 7.3 Reflex klystron efficiency in various generation zones

n	0	1	2	3
$\eta(\%)$	53.1	22.7	14.5	10.6

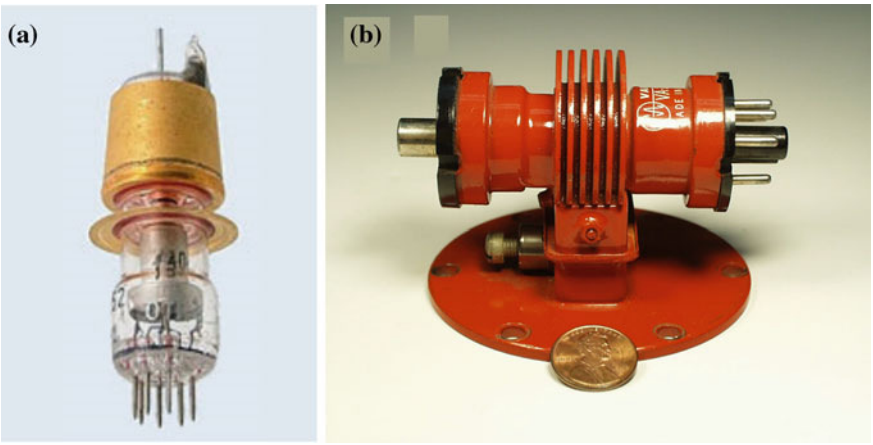


Fig. 7.33 Reflex klystrons with external (**a**) and internal (**b**) resonators

It is enough to say that researchers have managed to reduce the mass of the centimeter band low-power generators to a tenth of a gram (!) and to lower operating voltages to tens of volts. Such devices can compete with solid-state devices with the same output parameters, especially in extreme conditions of high temperatures and high levels of radiation.

7.3 Travelling Wave Tubes

7.3.1 Operating Principle of Travelling Wave Tubes

By the early 1940s, triode and klystron amplifiers, capable of operating in the microwave band were developed and produced. However, these amplifiers were narrowband, because they used high Q-factor resonators. A high Q-factor was necessary in order to increase the weak voltage of the input signal and to ensure efficient control of the electron flux during short-term interaction with it (the interaction time should be shorter than the transit time). It should be noted that the resonator increases not only the signal, but also the noise, so the resonance amplifiers have a relatively large noise factor. In connection with the rapid development of radar during the Second World War, it became necessary to create broadband low-noise amplifiers.

Rudolf Kompfner (an architect by education) solved this problem in 1942, when in his work on creating of low-noise amplifiers at the University of Birmingham, he proposed to get rid of resonators and instead use the interaction of the electron beam with the electromagnetic field of the traveling wave when they propagate with the same velocity. By doing so, the weakness of the field in the transmission line was

compensated by the duration of the interaction, which can amount to dozens of field periods, rather than the period fractions as in the resonator gap. Since the electron velocity is always less than light speed, Kompfner proposed decelerating the electromagnetic wave using the helical slow-wave structure (SWS) to facilitate the long-term interaction. The device invented by Kompfner was called *the traveling wave tube* (TWT). Sometimes, to emphasize that this is an O-type device, it is called *the traveling wave tube of O-type* (TWTO).

The first Kompfner tubes were very unstable and self-excited when there were noticeable reflections from the output of the SWS or from the load. J. Pierce, a Bell Telephone Labs employee, estimated all the advantages of the TWT as a broadband amplifier, developed the TWT theory and formulated the conditions for its stability. His fundamental work in this area dates back to 1946. L.A. Vainshtein (USSR) made the largest contribution to the nonlinear theory of the TWT.

Let's consider the operating principle of a traveling wave tube by means of its scheme shown in Fig. 7.34. Electron gun 1 creates electron beam 2, which interacts with an electromagnetic wave excited in slow-wave structure 3 by an input signal. The amplified signal is extracted from the slow-wave structure. The exhausted electrons deposit on collector 4. All parts of the tube are in a vacuum envelope 5. A focusing system (not shown) supports the beam shape throughout its entire length.

The electromagnetic wave propagates in the SWS with phase velocity v_p. Let us choose the accelerating voltage in such a way that electron velocity v_e is equal to the wave velocity. In this case, the electrons are at rest in the coordinate system moving with the phase velocity of the wave.

A longitudinal component of the wave electric field perturbs the motion of the electrons. Those of them that are in the accelerating field begin to move somewhat faster, and those that are in the decelerating field are slowed down, i.e., the velocity modulation of the electron flux occurs. The faster electrons begin to catch the slower ones up, and as a result a bunch is formed, i.e., bunching of the electron flux occurs. This process is illustrated in Fig. 7.35, which shows the dependence of the wave electric field longitudinal component on the coordinate at some fixed time. The arrows represent the accelerations the electrons (black circles) receive under the action of the SWS field. As can be seen, the bunch is formed around the electron, which is in the zero field that alternates from accelerating to decelerating. While moving toward the collector, the bunch becomes more and more dense but cannot

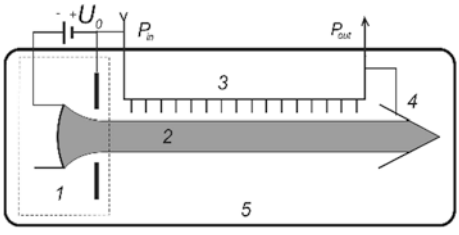


Fig. 7.34 Scheme of a TWT

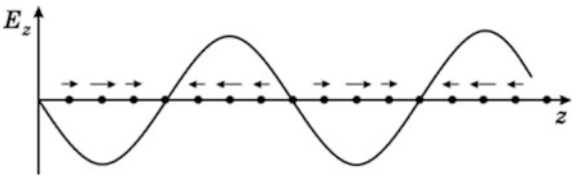


Fig. 7.35 Electron bunching in a TWT

transmit its kinetic energy to the wave field since it is in the region where the field is absent or very weak.

In order for the formed bunch to transmit part of its energy to the wave field, the electron velocity must be slightly larger than the wave velocity. Then the bunch, as it forms, will shift forward relative to the wave, i.e., it will get into the region of the decelerating field and be able to transmit part of its energy to it. At this, the wave amplitude will increase, i.e., signal amplification will be observed. However, the difference in velocities should not be too large, otherwise bunches will "slip through" the decelerating field region. This process can be regarded as the induced Cherenkov radiation of electrons. The induced nature of the radiation provides the electromagnetic field of the wave in the SWS.

The mechanism of energy extraction is illustrated in Fig. 7.36 which shows the distribution of the wave electric field longitudinal component in the output section of the SWS at a certain fixed time. The gray rectangles indicate the electron bunches and the arrows indicate the direction and magnitude of the velocity of bunches relative to the wave.

Bunch 1 has not yet formed completely, and it moves forward relative to the wave. Bunch 2 is already denser and it is in the maximum decelerating field but continues to move forward. Bunch 3 is the densest and it has already transmitted some of its kinetic energy and decelerated. Its velocity is equal to the wave velocity but it continues to transmit energy and to slow down while being in the decelerating field. Bunch 4 has already expanded under the action of space charge forces, and its velocity is less than the wave velocity but it continues to transmit energy being in the decelerating field. On further motion, the bunch will leave the decelerating field and enter the accelerating field where it will take energy from the wave. Therefore, the SWS length must be chosen so that bunches have no time to get into the accelerating field.

Comparing the amplification mechanisms in traveling wave tubes and klystrons, it is easy to see that they are based on the same processes: velocity modulation of

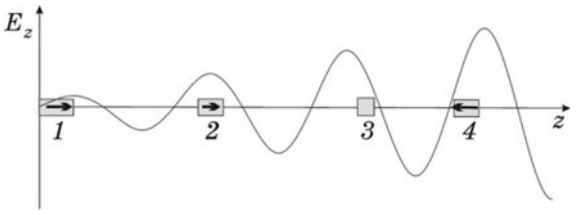


Fig. 7.36 Energy extraction from the electron beam in a TWT

the electron beam, and bunching and energy extraction. However, these processes in klystrons, unlike TWTs, are spatially spread and separated in time. This makes it possible to optimize them separately and allows, in particular, the obtaining of a higher coefficient of electron beam energy conversion to field energy (electronic efficiency) in klystrons.

7.3.2 The Linear Theory of O-Type TWTs

Dispersion Equation of a TWTO

In a TWT, there is an interaction between the wave in the slow-wave structure and the space-charge waves in the electron beam. Because of this interaction, the wave propagation conditions change. We will analyze this interaction in three stages. First, let's assume that the field in the SWS is specified, and find the alternating current excited by this field in the electron flux. Then let's assume that the alternating convection current is specified, and find the SWS field excited by it. Lastly, we find a self-consistent solution using the results of the previous analysis.

In the analysis, we assume that all quantities characterizing the SWS field and the electron beam depend only on time and the longitudinal coordinate z (the one-dimensional model). We represent also all quantities as a sum of a constant and a variable component, the maximum value of the variable component being much smaller than that of the constant (small signal approximation):

$$a = a_0 + \tilde{a}, \quad \max(\tilde{a}) \ll a_0, \tag{7.28}$$

where a is any quantity characterizing the field in the SWS or the electron beam. We postulate that all variables are of the form of the wave propagating along the z-axis, which coincides with the longitudinal TWT axis:

$$\tilde{a} = \dot{a}\mathrm{e}^{-i(\omega t - \Gamma z)}, \tag{7.29}$$

where Γ remains to an unknown propagation constant.

We also assume that the constant component of the electric field in the interaction area is absent ($E_{z0} = 0$). Let's recall that symbol e denotes the absolute value of the electron charge, and symbol ρ denotes the absolute value of the space charge density of the electrons ($\rho = en$, n is the electron concentration).

As it is known, the electromagnetic field in a periodic transmission line, which is any SWS, and can be represented as a sum of spatial harmonics. Each harmonic has its own phase velocity. In this analysis, we take into account only the "working" spatial harmonic with a phase velocity close to the electron velocity. Therefore, hereafter all the quantities characterizing the field in the SWS (E_z, v_{p0}) refer to the operating spatial harmonic.

The first stage. Let's find the convection current excited by the specified SWS field. To do this we write down the equation of the motion of the electrons:

$$\frac{\mathrm{d}\tilde{v}}{\mathrm{d}t} = -\frac{e}{m}\tilde{E}_z \tag{7.30}$$

(relativistic effects are ignored). Expanding the total velocity derivative, we obtain

$$\frac{\mathrm{d}\tilde{v}}{\mathrm{d}t} = \frac{\partial\tilde{v}}{\partial t} + \frac{\partial\tilde{v}}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial\tilde{v}}{\partial t} + \frac{\partial\tilde{v}}{\partial z}v_0 + \frac{\partial\tilde{v}}{\partial z}\tilde{v}. \tag{7.31}$$

Here we take into account that $\mathrm{d}z/\mathrm{d}t = v = v_0 + \tilde{v}$. The last term in expression (7.31) is the second-order quantity of smallness, since it contains the product of two variable quantities, and can be ignored. Substituting expression (7.31) and the variable velocity component in the form of (7.29) into (7.30), we find

$$j\omega\dot{v} - j\Gamma\dot{v}v_0 = -(e/m)\dot{E}_z.$$

Introducing the *electron propagation constant* $\beta_e = \omega/v_0$, we obtain

$$\dot{v} = \frac{j(e/m)\dot{E}_z}{v_0(\beta_e - \Gamma)}. \tag{7.32}$$

Formula (7.32) determines the amplitude of the variable velocity, which the electrons acquire under the influence of the specified SWS field.

In order to find the variable component of the space charge density, we use the continuity equation. In the one-dimensional case, taking into account sign ρ, it has the form

$$\frac{\partial\tilde{J}}{\partial z} - \frac{\partial\tilde{\rho}}{\partial t} = 0, \tag{7.33}$$

from where

$$-j\Gamma\dot{J} - j\omega\dot{\rho} = 0.$$

This relation allows us to find the amplitude of the charge density:

$$\dot{\rho} = -\frac{\Gamma}{\omega}\dot{J}. \tag{7.34}$$

Current density is related to charge density:

$$J = J_0 + \tilde{J} = -\rho v = -(\rho_0 + \tilde{\rho})(v_0 + \tilde{v}) = -\rho_0 v_0 - \tilde{\rho}v_0 - \rho_0\tilde{v} - \tilde{\rho}\tilde{v}.$$

The last term containing the product of small quantities is ignored. Thus, passing to complex amplitudes, we obtain

$$\dot{J} = -\dot{\rho} v_0 - \rho_0 \dot{v}. \tag{7.35}$$

We eliminate charge density from (7.35) by substituting formula (7.34):

$$\dot{J} = -\frac{\rho_0 \beta_e}{\beta_e - \Gamma} \dot{v} \tag{7.36}$$

Formula (7.32) relates the electron velocity to the electric field where the alternating electric field should be considered as a sum of two components: the specified SWS field E_{sw} and the space charge field E_ρ arising at the bunching of electrons:

$$\dot{E}_z = \dot{E}_{sw} + \dot{E}_\rho. \tag{7.37}$$

To determine the space-charge field, we use Ampere's law

$$\tilde{J} + \varepsilon_0 \frac{\partial E_\rho}{\partial t} = \text{Const}. \tag{7.38}$$

The electron beam is assumed to be infinitely long. With respect to the alternating current, it is open-circuit, therefore we can set Const. $= 0$ in (7.38). Passing to complex amplitudes, we obtain

$$\dot{E}_\rho = \frac{j\dot{J}}{\omega \varepsilon_0}. \tag{7.39}$$

Substituting (7.39), (7.37) and (7.32) into (7.36), then after simple transformations, we obtain

$$\dot{J}\left[1 - \frac{e\rho_0}{m\varepsilon_0} \frac{1}{(\beta_e - \Gamma)^2 v_0^2}\right] = -j \frac{e\rho_0 \beta_e}{m v_0 (\beta_e - \Gamma)^2} \dot{E}_{sw}. \tag{7.40}$$

In Sects. 3.1 and 3.2, the concepts of *plasma frequency* $\omega_p = \sqrt{e\rho_0/(m\varepsilon_0)}$ and *plasma propagation constant* $\beta_p = \omega_p / v_0$ are introduced. Using these notations, we rewrite (7.40):

$$\dot{J}[(\beta_e - \Gamma)^2 - \beta_p^2] = -j \frac{e\rho_0 \beta_e}{m v_0} \dot{E}_{sw}.$$

We multiply the numerator and denominator of the right-hand part of this expression by υ_0 and take into account that $\rho_0 v_0 = -J_0$, and $v_0^2 = 2(e/m)U_0$, where U_0 is the accelerating voltage. As a result, we obtain

$$\dot{J}[(\beta_e - \Gamma)^2 - \beta_p^2] = j \frac{J_0 \beta_e}{2U_0} \dot{E}_{sw}. \tag{7.41}$$

The resulting relation allows us to find the amplitude of the convection current density excited in the electron flux by the specified field in the SWS.

Let's consider the particular case when there is no field in the SWS, i.e., the right-hand part of (7.41) is identically equal to zero. Then from this equation, it follows that either $\dot{J} = 0$ (the trivial solution corresponding to the unperturbed flux), or

$$(\beta_e - \Gamma)^2 - \beta_p^2 = 0.$$

Hence, we find an unknown propagation constant:

$$\Gamma = \beta_e \mp \beta_p.$$

This solution corresponds to two space-charge waves—fast and slow—propagating in the electron beam with constant amplitude (see Sect. 2.2).

The second stage. At this stage, we use the energy method. The complex power transferred across the cross section of the TWT interaction space is:

$$P = \frac{1}{2}\int_S (\mathbf{E} \times \mathbf{H}) d\mathbf{S} = P_0 \mathrm{e}^{-2\mathrm{j}(\omega t - \Gamma z)}.$$

The power variation in the interaction space section with length dz:

$$\mathrm{d}P = -2j\Gamma P_0 dz.$$

In a "cold" tube (in the absence of an electron beam) the expressions for the power and its differential have the analogous form:

$$P_c = P_0 \mathrm{e}^{-2j(\omega t - \Gamma_0 z)}, \quad \mathrm{d}P_c = -2j\Gamma_0 P_0,$$

where Γ_0 is the wave propagation constant in the "cold" SWS.

The power of the interaction between the SWS field and the electron beam in section dz is obviously equal to the difference in power variations in this tube section with and without electron flux:

$$\mathrm{d}P_e = \mathrm{d}P - \mathrm{d}P_c = -2j(\Gamma - \Gamma_0)P_0 \mathrm{d}z. \tag{7.42}$$

On the other hand, Sect. 4.1 shows that the interaction power is

$$\mathrm{d}P_e = \frac{1}{2}\dot{I}\dot{E}^*_{sw}\mathrm{d}z, \tag{7.43}$$

where $\dot{I}$ is the amplitude of the convection current variable component. Thus,

$$-2j(\Gamma - \Gamma_0)P_0 = \frac{1}{2}\dot{I}\dot{E}^*_{sw}.$$

Let's introduce the SWS *interaction impedance*

$$R_c = \frac{\dot{E}_{sw}\dot{E}^*_{sw}}{2\beta_0^2 P_0} \tag{7.44}$$

and with the help of this definition express the ratio $\dot{E}^*_{sw}/P_0$:

$$\frac{\dot{E}^*_{sw}}{P_0} = \frac{2\beta_0^2 R_c}{\dot{E}_{sw}}. \tag{7.45}$$

In this expression $\beta_0 = \omega/v_{p0}$ is the phase constant of the cold SWS, and v_{p0} is the phase velocity of the wave in the cold SWS. Substituting relation (7.45) into (7.43), we find

$$\dot{E}_{sw} = j\frac{\beta_0^2 R_c}{\Gamma - \Gamma_0}\dot{I}. \tag{7.46}$$

The formula obtained allows us to determine the longitudinal component amplitude of the electric field of the wave excited in the SWS by the specified convection current.

The third stage. In the self-consistent solution, the current excited by the SWS field must be equal to the current that excites this field. To satisfy this condition we integrate expression (7.40) over the cross section of the interaction space and substitute (7.46) into it:

$$\dot{I}[(\beta_e - \Gamma)^2 - \beta_q^2] = -\frac{I_0\beta_e\beta_0^2 R_c}{2U_0(\Gamma - \Gamma_0)2U_0}\dot{I}.$$

Here, we take into account the fact that in the beam with the finite cross section, plasma frequency ω_p should be replaced by depressed plasma frequency $\omega_d = s_d\omega_p$, where s_d is the depression coefficient. Accordingly, in the first approximation, β_p changes to $\beta_q = \omega_d/v_e$. In most cases we can assume that s_d is equal to the plasma reduction factor s, see (3.5).

Let's transform the resulting expression reducing $\dot{I}$:

$$\frac{\Gamma - \Gamma_0}{\beta_e}\left[\left(\frac{\beta_e - \Gamma}{\beta_e}\right)^2 - \left(\frac{\beta_q}{\beta_e}\right)^2\right] + \frac{I_0 R_c}{4U_0}\left(\frac{\beta_0}{\beta_e}\right)^2 = 0. \tag{7.47}$$

Let's introduce *the Pierce gain parameter*:

$$C = \sqrt[3]{\frac{I_0 R_c}{4U_0}\left(\frac{\beta_0}{\beta_e}\right)^2} \approx \sqrt[3]{\frac{I_0 R_c}{4U_0}} \tag{7.48}$$

and divide (7.47) by C^3:

$$\frac{\Gamma-\Gamma_0}{\beta_e C}\left[\left(\frac{\beta_e-\Gamma}{\beta_e C}\right)^2-q\right]+1=0. \tag{7.49}$$

In this expression, *the space-charge parameter* is introduced

$$q=\left(\frac{\omega_q}{\omega C}\right)^2. \tag{7.50}$$

(in the literature, parameter q is often denoted by $4QC$ for historical reasons).

Propagation constants Γ and Γ_0 included in (7.49) are complex:

$$\Gamma=\beta-j\alpha,\quad \Gamma_0=\beta_0-j\alpha_0.$$

We assume that phase constant β_0 and attenuation constant α_0 in the cold SWS are known, and that constants β and α in the system with the beam are subject to determination.

Let's introduce the dimensionless quantities

$$\delta=x+jy=j\frac{\beta_e-\Gamma}{\beta_e C},\quad \theta=d+jb=-j\frac{\beta_e-\Gamma_0}{\beta_e C}. \tag{7.51}$$

From these definitions, it is not difficult to find that

$$x=-\frac{\alpha}{\beta_e C},\quad y=\frac{\beta_e-\beta}{\beta_e C}=\frac{v_p-v_0}{v_p C},\quad d=\frac{\alpha_0}{\beta_e C},\quad b=\frac{\beta_0-\beta_e}{\beta_e C}=\frac{v_0-v_{p0}}{v_{p0}C}. \tag{7.52}$$

Parameter b is called *the parameter of non-synchronism*. It is proportional to the difference between the electron velocity and the wave velocity in the "cold" SWS. *Attenuation parameter* d is proportional to the attenuation constant in the cold SWS. Parameter y is proportional to the difference between the wave velocity in the system with the beam and the phase velocity of the wave in the "cold" SWS. Finally, growth parameter x is proportional to the attenuation constant in the system with the beam taken with the opposite sign. For $x > 0$, the attenuation constant is negative, i.e., the wave grows in the propagation process. These growing waves are of the greatest interest, since they provide amplification of the input signal. At $x \leq 0$, the wave amplitude wanes while propagating, or remains constant.

Substituting (7.51) into (7.49), we obtain *the TWT dispersion equation* first obtained by J. Pierce:

$$j(\delta+\theta)(\delta^2+q)-1=0. \tag{7.53}$$

This is a cubic complex equation relative to δ, which has three roots. This means that at given values of parameters b and d, three waves with different phase

velocities and attenuation constants can propagate in the SWS with the beam, and the waves with positive values of parameter x are of greatest interest. The amplitude of these waves increases during propagation, i.e., these waves amplify the input signal. The non-growing waves must also be taken into account, since the energy of the input signal is distributed among all the waves.

Analysis of The Dispersion Equation Solutions

To analyze the dispersion (7.53) let's separate the real and imaginary parts of it:

$$x + d = -\frac{2xy}{(x^2 - y^2 + q)^2 + 4x^2y^2}, \tag{7.54}$$

$$y + b = -\frac{x^2 - y^2 + q}{(x^2 - y^2 + q)^2 + 4x^2y^2}. \tag{7.55}$$

It is seen from (7.54) that in the absence of attenuation in the cold SWS ($d = 0$), for one of the roots $x = x_3 = 0$. The corresponding value $y = y_3$ is found from the equation

$$(y_3 + b)(y_3^2 - q) - 1 = 0.$$

The values of two other roots are obtained from (7.55) supposing $y_1 = y_2 = y$:

$$x_{1,2} = \pm\left[(4qy^2 - 2y)^{1/2} - (q + y^2)\right]^{1/2}.$$

Two x values correspond to each given value of y: x_1 and $x_2 = -x_1$. This means that both waves have the same phase velocities; one of them is increasing, and the other is attenuating while propagating.

For arbitrary values of b, d and q, dispersion (7.53) can only be solved numerically. The result of this solution for space-charge parameter $q = 1$ and attenuation parameter $d = 0$ is shown in Fig. 7.37. The two dashed horizontal lines on this graph correspond to fast ($y > 0$) and slow ($y < 0$) space-charge waves propagating in the electron beam without slow-wave structure. The oblique dotted line corresponds to the wave propagating in the cold SWS.

Three waves propagating with constant amplitude ($x = 0$) and different phase velocities correspond to the large absolute values of parameter b. As $|b|$ increases, they asymptotically approach the unperturbed space-charge waves and the wave in the cold SWS. For small absolute values of the parameter of non-synchronism, one of the waves still propagates without changing its amplitude ($x_3 = 0$). Its phase velocity is greater than the electron velocity ($y_3 > 0$), therefore it is called a fast wave. The other two waves have the same phase velocity, which is less than the electron velocity (slow waves). The amplitude of one of them increases ($x_1 > 0$), and the other decreases ($x_2 < 0$) while propagating.

In the literature, there is biquadratic dispersion equation, which takes into account the wave in the SWS propagating from the collector to the cathode. This

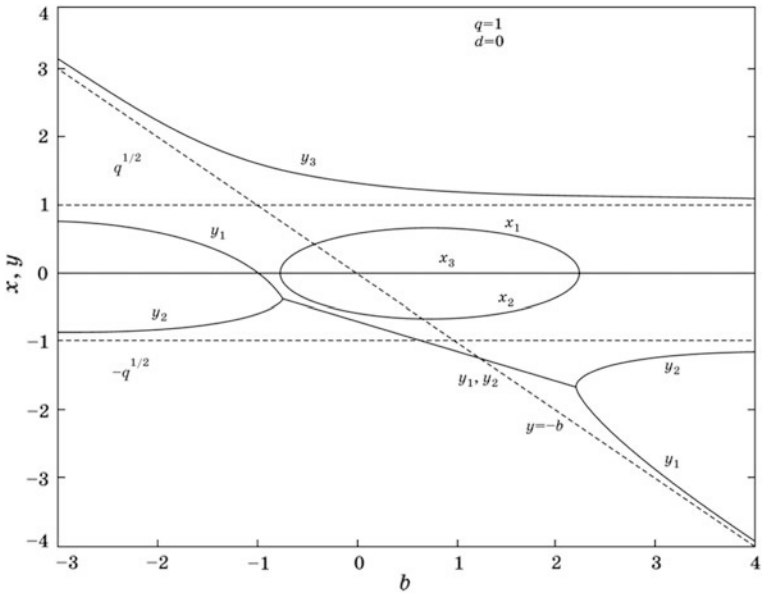


Fig. 7.37 Dispersion equation roots versus the parameter of non-synchronism

wave is interacting very weakly with the electron flux because of the absence of synchronism, so it does not make sense to take it into account.

The range of b values in which $x_1 > 0$ is called the amplification region of the TWT. In more detail, the dependencies of the x and y parameters on b for different values of q and $d = 0$ in the amplification region are shown in Fig. 7.38. As can be seen, as the space-charge parameter increases, the maximum values of parameter x_1 decrease (the wave amplitude growth rate decreases), and the amplification region becomes narrower and shifts toward higher values of parameter b.

For small values of q, the maximum value of parameter x_1 is:

$$x_{1\max} = \sqrt{3}/2, \tag{7.56}$$

and the corresponding values b and y_1 are:

$$b_{\max} = 0, \quad y_{1\max} = -1/2. \tag{7.57}$$

For large values of q, the parameter values corresponding to the maximum x_1 value are:

$$x_{1\max} \approx 1/(\sqrt{2}q), \quad b_{\max} \approx -y_{1\max} = \sqrt{q}. \tag{7.58}$$

The following relations determine the boundaries of the amplification region for large q:

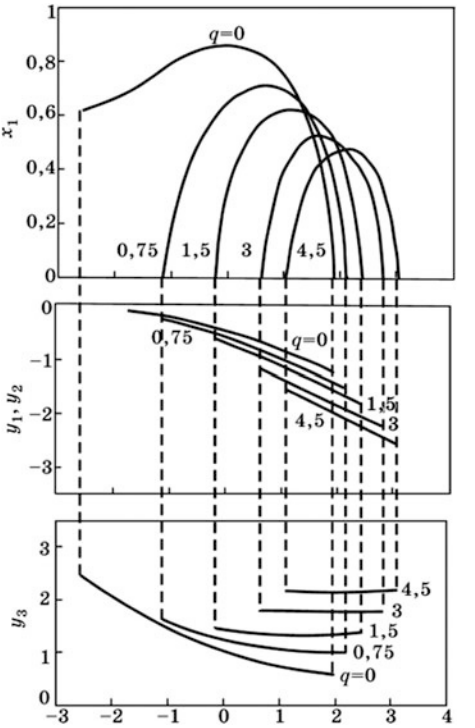


Fig. 7.38 Dependence of dispersion equation roots on b in the amplification region

$$b = b_{\max} \pm 2x_{1\max} \approx q^{1/2} \pm \sqrt{2}/q^{1/4}.$$

The Effect of Distributed Losses

The attenuation in the cold SWS is taken into account by attenuation parameter $d > 0$. As the numerical analysis of the dispersion equation shows, this parameter has little effect on wave propagation velocity but leads to a change of parameter x. The fast wave when $d > 0$ also attenuates ($x_3 < 0$) and the values of growth parameter x_1 in the amplification region decrease. The analysis shows that for small values of q, the change of the growth parameter $\Delta x_1 = -d/3$ and for large q $\Delta x_1 = -d/2$. Thus, if the cold SWS has the small attenuation constant α_0, then the maximum value of the rise constant decreases to the value

$$\alpha_1 = \alpha_{1\max} - m\alpha_0,$$

where $1/3 \leq m \leq 1/2$.

The value of m appears to be less than one since most of the energy taken from the electrons passes only a part of the SWS length. Therefore, for the wave in the TWT, the distributed losses are weaker than in the SWS without the beam where the energy passes through the entire length of the system.

Boundary Conditions

In the TWT, three waves propagate from the cathode to the collector, the amplitude of one of which increases, one decreases, and the third remains constant (distributed losses in the SWS are not taken into account).

In order to find the initial amplitudes of these waves, it is necessary to use boundary conditions. The unmodulated electron beam and the input signal arrive at the SWS input, therefore the boundary conditions are written as follows:

$$\dot{I}(0) = \sum_{n=1}^{3} \dot{I}_n(0) = 0; \quad \dot{v}(0) = \sum_{n=1}^{3} \dot{v}_n(0) = 0; \quad \dot{E}_z(0) = \sum_{n=1}^{3} \dot{E}_{zn}(0) = \dot{E}_{sw}. \tag{7.59}$$

Relations (7.32), (7.41) and the dispersion equation relate the currents and wave velocities to the electric field. Using these relations it is possible to obtain expressions for the current and velocity of electrons:

$$\frac{\dot{I}_n}{I_0} = j\dot{E}_{zn} \frac{1}{2\beta_e C^2 U_0} \frac{1}{\delta_n^2 + q}, \quad n = 1, 2, 3, \tag{7.60}$$

$$\frac{\dot{v}_n}{v_0} = -j\dot{E}_{zn} \frac{1}{2\beta_e C U_0} \frac{\delta_n}{\delta_n^2 + q}. \tag{7.61}$$

These relations, together with the third condition (7.59), form the system of equations allowing us to find the initial field amplitudes of the three waves. For the increasing wave that is of the greatest interest, the ratio of its initial amplitude to the amplitude of the total field is

$$A_1 = \left| \frac{\dot{E}_{z1}(0)}{\dot{E}_{\text{zin}}} \right| = \left| \frac{\delta_1^2 + q}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)} \right|. \tag{7.62}$$

Values δ_1, δ_2, δ_3 are found from the dispersion equation for the specified values of b, d and q. Figure 7.39 shows the dependence of A_1^2 on b for several values of q and $d = 0$. As can be seen, these values lie in the range from −15 to 0 dB.

For $d > 0$, $A_1^2 \approx -10$ dB is practically independent of b. For the particular case $d = 0$, $q = 0$, $b = 0$ value $A_1 = 1/3$, or $A_1^2 = 1/9 = -9.54$ dB. Parameters A_2 and A_3 have the same value. It follows that the electromagnetic field of three waves transfers only 1/3 of the input power. The rest of the input power is spent on electron flux modulating.

The Effect of the Lumped Attenuator

In the TWT, another wave can propagate which is not described by dispersion (7.53). This wave propagates in the SWS from the output to the input. Since

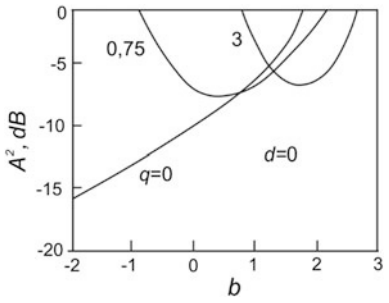


Fig. 7.39 Initial attenuation of the increasing wave versus b

synchronism of this wave with the electron beam is absent, it propagates in practice as in the cold SWS, i.e., there is no attenuation of this wave, or it is very small.

This wave occurs if the TWT output is not ideally matched to the SWS and (or) the tube load is not matched to the output. On reaching the SWS input, this wave partially reflects from it and spreads towards the collector in synchrony with the electron flux. As a result, feedback occurs, which can lead to self-excitation of the device. The first Kompfner tubes operated unstably for that very reason.

J. Pierce proposed the introduction of a lumped absorber (attenuator) into the TWT structure, which weakens the reflected wave and prevents self-excitation of the TWT. The amplitude condition for the stability of the TWT with a concentrated absorber can be written as follows:

$$\mu|\Gamma_{\text{out}}|^2|\Gamma_{\text{in}}|^2 D < 1,$$

where μ is the TWT power gain; $\Gamma_{\text{in}}, \Gamma_{\text{out}}$ are the reflection coefficients from the SWS input and output; and D is the attenuation coefficient of the attenuator. Assuming the maximum values of the reflection coefficients from the input and output are equal to one, we obtain that the attenuation coefficient of the absorber should be larger than the TWT gain:

$$D > \mu$$

In practice, the attenuator is the SWS section, with length $l_a = z_{a2} - z_{a1}$, onto which the absorbing coating is applied. It is also possible to use two sections of the SWS, the output of the first section and the input of the second section being connected to the matched loads. Such a circuit corresponds to the infinitely large attenuation of the "absorber".

Naturally, the introduction of the absorber reduces the TWT gain but this decrease is much less than the absorber attenuation. The explanation is that although the electromagnetic wave in the SWS is almost completely absorbed by the absorber, the waves remain in the electron beam and at the exit from the absorber they excite new waves in the SWS, one of which increases. The analysis shows that the decrease in gain $\Delta\mu_p$ depends on the length of the attenuator,

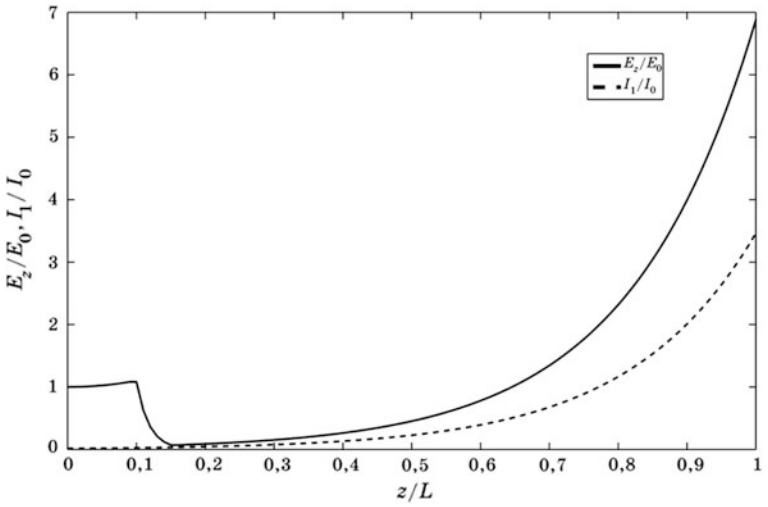


Fig. 7.40 SWS field and convection current first harmonic versus the z coordinate

parameters b and q. Upon the average with the optimal value of parameter b $\Delta\mu_p \approx -6$ dB. Figure 7.40 shows the distribution of the electric field longitudinal component and the first harmonic of the convection current along the length of the tube with parameters $C = 0.02$; $q = 0$; $d = 0$; $b = 0$. The attenuator is in the region $0.1 \leq z/L \leq 0.15$, where L is the length of the tube. As can be seen, the amplitude of the current does not decrease in the region of the absorber. After leaving it, this current raises a new increasing field wave.

TWT Gain

Using the results obtained above, it is possible to calculate the TWT power gain, which is determined by the formula

$$\mu = 10 \lg(P_{\text{out}}/P_{\text{in}}).$$

Substituting here the above values of the output and input power we find

$$\mu = 10 \lg \left| \frac{\dot{E}_{z1}(L)}{\dot{E}_{z1}(0)} \right|^2 = 20 \lg \left| \frac{\dot{E}_{z1}(L)}{\dot{E}_{z1}(0)} \right|.$$

Taking into account the relation between the initial amplitude of the growing wave and its amplitude at the end of the slow-wave structure, we obtain

$$\mu = 20 \lg A + 20 \lg D + 20 \lg[\exp(\beta_e C x_1 L)].$$

After simplifying this formula, we have

$$\mu = A + D + 8.68\beta_e C x_1 L, \text{ dB}$$

where l is the total length of the SWS. Taking into account that $\beta_e = \omega/v_0$ and setting $\beta_e \approx \beta_0 = 2\pi/\lambda_{sw}$, where λ_{sw} is the wavelength in the cold SWS, we obtain

$$\mu = A + D + 54.5 x_1 CN,$$

where N is the number of wavelengths that fall on the SWS length. All quantities in this formula are expressed in decibels. For the particular case $b = 0,\ q = 0,\ d = 0$

$$\mu = -9.54 + D + 47.3CN.$$

In real TWTs, the typical values of gain parameter C lie in the interval from 0.02 to 0.2. The evaluation of the gain by the foregoing formula shows that for N ~10–30 the amplification of the helical TWT can reach 40–60 dB.

Frequency and Phase Response

A typical form of the frequency response characteristic of the O-type TWT is shown in Fig. 7.41. For the quantitative description of this TWT characteristic, the following parameters are introduced:

- Bandwidth Δf is the frequency band within which the output power (gain) reduces by no more than the specified value (usually −3 or −1 dB) compared to maximum power (gain) and the relative frequency band $\Delta f/f_0$, where f_0 is the center frequency of the band;
- Minimum and maximum band frequencies $f_{\min}$ and $f_{\max}$;
- Frequency band overlap factor $k_f = f_{\max}/f_{\min}$ and frequency band overlap factor expressed in octaves $k_{fo} = \log_2(f_{\max}/f_{\min})$.

The dependence of TWT output power (or gain) on frequency in the weak-signal mode is determined mainly by the dependence of x_1 on parameter b and the SWS dispersion characteristic. Theoretically, in the absence of SWS dispersion, the band

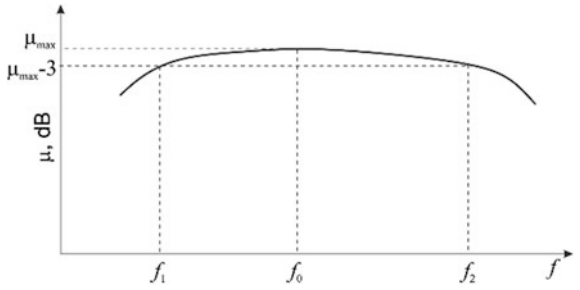


Fig. 7.41 TWT frequency response

of the amplified frequencies can be arbitrarily broad. In this case, the dependence of the gain on the frequency is determined only by the change of N, which is related to the wavelength change. Parameter b, and hence parameters A and D, also remain constant at the frequency change.

In the real SWS, dispersion exists so parameter b depends on the frequency. Accordingly, parameter x_1, and hence the gain, depend on the frequency. Besides, the interaction impedance of the operating spatial harmonic depends on the frequency that leads to the dependence of gain parameter C and space-charge parameter q on the frequency. As a result, the calculation of the frequency response characteristic, even in the linear approximation, appears to be a very complicated problem.

The TWT phase response characteristic is mainly determined by SWS dispersion. If the attenuator is not taken into account, the phase incursion of the increasing wave at SWS length is $\Delta\varphi = \beta_1 L = \beta_e L(1 - Cy_1)$. The value of Cy_1 in the amplification band is much less than 1 and varies relatively weakly, so we can assume that $\Delta\varphi \approx \omega L/v_0$, i.e., the phase shift is proportional to the signal frequency.

The dependence of the output signal phase on the accelerating voltage is of great importance. When this voltage changes, the velocity of the electrons changes, leading to a change of the TWT output signal phase. Since this change is undesirable in most cases, strict requirements are imposed on the stability of TWT power supplies. To evaluate the phase shift change at the accelerating voltage change, this formula can be used

$$\Delta\varphi = -0.58\pi N \Delta U/U_0,$$

where $\Delta\varphi$ is the change in the phase shift between input and output signals, N is the slow-wave structure length in wave lengths, and ΔU is the accelerating voltage change. The phase shift in this formula is measured in radians. It follows from the formula that the phase distortions in the TWT depend linearly on the length of the slow-wave structure. In the TWT with helical SWS, phase distortions are generally negligible.

TWT Noise Parameters

The main source of noise in the TWT, as in other vacuum electron devices, is the shot noise of the electron flux. However, unlike the klystron, the high Q-factor resonator does not enlarge this noise, so TWTs are relatively low-noise devices. Fluctuations in the density and velocity of electrons excite the space-charge waves in the space between the gun cathode and the SWS input. It is possible to select the distance between the gun anode and the SWS input, so that at the SWS input there is minimum of noise current. In this case, the effect of the cathode shot noise is substantially reduced. Modern TWTs have a noise value of less than 3 dB at 3 GHz and less than 4 dB at 10 GHz.

7.3.3 Elements of the Nonlinear Theory of TWTs

Basic Equations

Within the framework of the linear theory, TWT gain does not depend on the input signal amplitude, so TWT output power is unlimited, and that contradicts common sense and the results of the experiment. Therefore, in order to calculate the TWT output power and efficiency as well as its other characteristics, such as nonlinear distortions, and side radiation, etc., it is necessary to create the TWT nonlinear theory. J. Pierce, J. Rowe, L.A. Vainshtein, V.N. Shevchik, and many other researchers made a major contribution to this theory.

Let's derive the basic equations of the TWT nonlinear theory following the Vainshtein method. Let's write the equation of the motion of electrons

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{e}{m}E, \tag{7.63}$$

assuming that the electrons move along the z axis and the electric field lines are also directed along this coordinate. Current time t, and time of the electrons entering the interaction space t_0 are independent variables in (7.63) and the current electron coordinate z is the function of these variables. It is more convenient to use z and t_0 as independent variables, since in this case, the difficulties associated with the non-uniqueness of function $z = z(t, t_0)$ disappear. Assuming

$$t = t(z, t_0) \tag{7.64}$$

we have

$$v = \left(\frac{\mathrm{d}t}{\mathrm{d}z}\right)^{-1}, \quad \frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{\partial^2 t/\partial z^2}{(\partial t/\partial z)^3}.$$

Function (7.64) is single-valued in the absence of backward motion of the electrons, while function $t_0(z, t)$ can be multi-valued, since electrons that exit the $z = 0$ cross-section at different moments of time can enter the z cross-section in reverse order (the overtaking phenomenon). Equation (7.63), when these expressions are substituted becomes

$$\frac{\partial^2 t}{\partial z^2} = \frac{e}{m}\left(\frac{\partial t}{\partial z}\right)^3 E. \tag{7.65}$$

The phase of the electron at the given moment can be defined as follows:

$$\omega t = \omega t_0 - \beta_e z + \theta(\xi, t_0),$$

where θ is the perturbation caused by field E, $\xi = C\beta_e z$. In view of the smallness of the gain parameter, function θ weakly depends on coordinate z.

The charge conservation law regarding variables z, t_0 is written in the following way:

$$I_0 \mathrm{d}t_0 = I(z, t_0)|\mathrm{d}t|, \tag{7.66}$$

where $\mathrm{d}t_0$ is the time interval over which the layer of electrons passes cross-section $z = 0$, and $\mathrm{d}t$ is the time interval over which the same layer passes cross-section z. Absolute value of this interval is taken because the order of the arrival of electrons to cross-section z can be reversed (in the case of overtaking). From here

$$I(z, t_0) = I_0 \left| \frac{\mathrm{d}t_0}{\mathrm{d}t} \right|. \tag{7.67}$$

In the nonlinear mode, the convection current can be expanded in the Fourier series. Limiting ourselves to the current first harmonic, we write

$$\begin{aligned} \varphi(\xi, t_0) &= \omega t + \beta_e z = \omega t_0 + \theta(\xi, t_0) = \varphi_0 + \theta(\xi, t_0), \\ I(z, t) &= I_0 + \mathbf{Re} \sum_{n=1}^{\infty} \dot{I}_n(z) \mathrm{e}^{jn\omega t} \approx I_0 + \mathbf{Re}\left(\dot{I}_1(z) \mathrm{e}^{j\omega t}\right), \end{aligned} \tag{7.68}$$

where

$$\dot{I}_n(z) = \frac{1}{\pi} \int_0^{2\pi} I(z, t) \mathrm{e}^{-jn\omega t} d\omega t.$$

With function

$\varphi(\xi, \varphi_0) = \omega t + \beta_e z = \omega t_0 + \theta = \varphi_0 + \theta$, the expression for the current first harmonic can be written as follows:

$$\dot{I}_1(z) = \frac{I_0}{\pi} \mathrm{e}^{-j\beta_e z} \int_0^{2\pi} \mathrm{e}^{-j\varphi(\xi, \varphi_0)} d\varphi_0 = I_0 \dot{I}(\xi) \mathrm{e}^{-j\beta_e z}, \tag{7.69}$$

where $\dot{I} = \dot{I}_1 / I_0$ is the normalized amplitude of the convection current first harmonic.

The field of the slow-wave structure excited by the electron beam is determined by expression (7.46). Assuming that there is no attenuation in the SWS, this expression can be rewritten as

$$\frac{\mathrm{d}\dot{E}_{sw}}{\mathrm{d}z} + j\beta_0 \dot{E}_{\mathrm{sw}} = \beta_0^2 R_c I_0 \dot{I}.$$

Let's introduce the normalized amplitude of the field in the SWS:

$$\dot{\mathcal{E}}(\xi) = \dot{E}_{\text{sw}}(z)\frac{\mathrm{e}^{-j\beta_e z}}{\beta_e U_0 C^2}. \tag{7.70}$$

Taking into account that

$$\frac{\mathrm{d}}{\mathrm{d}z} = \frac{\mathrm{d}}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\mathrm{d}z} = \beta_e C\frac{\mathrm{d}}{\mathrm{d}\xi},$$

equation (7.70) takes the form

$$\beta_e U_0 C^3\frac{\mathrm{d}\dot{\mathcal{E}}}{\mathrm{d}\xi} + j\beta_e U_0 C^2(\beta_e - \beta_0)\dot{\mathcal{E}}(\xi) = \beta_0^2 R_c I_0 \dot{I}(\xi).$$

After simple transformations, this equation is reduced to the final form

$$\frac{d\dot{\mathcal{E}}(\xi)}{d\xi} + jb\dot{\mathcal{E}}(\xi) = (1 - \mathrm{bC})^2\dot{I}(\xi). \tag{7.71}$$

Let's transform the equation of motion (7.65), taking into account that

$$\begin{aligned}
t &= t_0 - \frac{z}{v_0} + \frac{\theta(\xi, t_0)}{\omega},\\
\frac{\mathrm{d}t}{\mathrm{d}z} &= \frac{1}{v_0} + \frac{\beta_e C}{\omega}\frac{\mathrm{d}\theta}{\mathrm{d}\xi} = \frac{1}{v_e}\left(C + \frac{\mathrm{d}\theta}{\mathrm{d}\xi}\right),\\
\frac{\partial^2 t}{\partial z^2} &= \frac{(\beta_e C)^2}{\omega}\frac{\partial^2\theta}{\partial\xi^2},\\
v &= \left(\frac{\partial t}{\partial z}\right)^{-1} = \frac{v_0}{1 + C\partial\theta/\partial\xi},\\
\frac{\partial\varphi}{\partial\xi} &= \frac{\partial\theta}{\partial\xi}, \quad \frac{\partial^2\varphi}{\partial\xi^2} = \frac{\partial^2\theta}{\partial\xi^2}.
\end{aligned}$$

Using these relations and the formula for the dimensionless field amplitude (7.70), we obtain the equation of motion in the form

$$\frac{\partial^2\varphi}{\partial\xi^2} + \left(1 + C\frac{\partial\varphi}{\partial\xi}\right)^3 Re\left[\dot{\mathcal{E}}(\xi)\mathrm{e}^{j\varphi}\right] = 0. \tag{7.72}$$

The system of (7.69), (7.71) and (7.72) together with the boundary conditions

$$\theta = \varphi_0, \quad \frac{\partial\varphi}{\partial\xi} = 0, \quad \dot{\mathcal{E}} = \dot{E}_{in}\frac{1}{\beta_e U_0 C^2} \text{ at } \xi = 0 \tag{7.73}$$

describes the nonlinear bunching in an O-type TWT. We recall that it was obtained on the assumption that attenuation in the SWS and the space charge are absent. It can be shown that in the absence of these constraints, the nonlinear TWT equations acquire the form

$$\frac{d\dot{\mathcal{E}}}{\mathrm{d}\xi} + j(b - jd)\dot{\mathcal{E}} = (1 - bC)^2 \dot{I}, \tag{7.74}$$

$$\dot{I}(\xi) = \frac{1}{\pi} \int_0^{2\pi} \mathrm{e}^{-j\varphi} d\varphi, \tag{7.75}$$

$$\frac{\partial^2 \varphi}{\partial \xi^2} + \left(1 + C\frac{\partial \varphi}{\partial \xi}\right)^4 \mathbf{Re}\left[\dot{\mathcal{E}} + jq\dot{I}(\xi)\right]\mathrm{e}^{j\varphi} = 0. \tag{7.76}$$

The system of (7.74)–(7.76) with the boundary conditions (7.73) can be solved only by numerical methods. Some results of this solution are given below.

Output Power and Efficiency

Figure 7.42 shows the typical TWT amplitude response (AR) calculated by the numerical solution of (7.74)–(7.76).

Within the framework of the linear theory, the TWT gain is independent of the input signal amplitude, therefore, in the region of a weak signal (region 1), AR is linear. With a further increase of the input power, the tube gain begins to decrease, and the amplitude response deviates from a linear pattern (region 2). At a certain value of input power, output power reaches a maximum (saturation region 3 in Fig. 7.42), then begins to decrease (region 4). This form of amplitude response is due to the peculiarities of electron beam bunching in the TWT (see Sect. 7.1).

In the linear region, the input signal is too weak to form the densest bunches, even close to the SWS output. Therefore, the density of the bunches (and the energy they transmit) is proportional to the amplitude of the input signal.

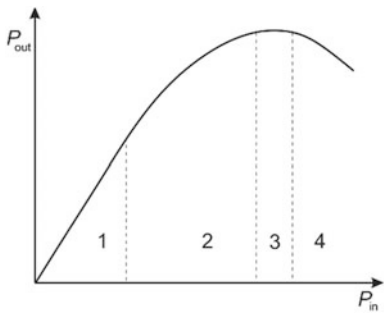


Fig. 7.42 TWT amplitude response

Further increases in input power (region 2) lead to the fact that the maximum density of bunches is reached within the SWS. In this case, bunches give energy to the field, the amplitude of which increases and reaches a maximum near the SWS output. In this region, the electron bunches give energy to the field, since they move faster than the slow wave ($y_1 < 0$). As the energy of the bunches decreases, their velocity diminishes. The slowing of bunches intensifies with increasing input power, so they begin to transmit a smaller part of their energy, and output power growth slows down.

In saturation region 3, the velocity of bunches initially becomes equal to, and then becomes less than, the slow wave velocity. Bunches stop transmitting energy when they leave the decelerating field region. Accordingly, the increase of output power stops. In region 4, bunches are in the accelerating field region and take energy from it, and the output power reduces.

The process of the formation of bunches and the extraction of energy from them is shown in Fig. 7.43 which is drawn according to the results of the numerical solution of (7.74)–(7.76).

The figure shows how the weak perturbations of the electron beam density while propagating are transformed into complex-shaped bunches. These bunches are in the decelerating field region and transfer their energy to it, and as a result the wave amplitude increases. Then, a part of the bunch exits the decelerating field region and the wave amplitude growth stops, i.e., saturation occurs. At this point, the interaction of the wave with the electrons must be stopped. With further movement, most of the bunch falls into the accelerating field and takes energy from the wave, the amplitude of which begins to decrease.

The numerical solution of the nonlinear theory equations makes it possible to calculate the TWT efficiency for any values of the tube parameters. However, we

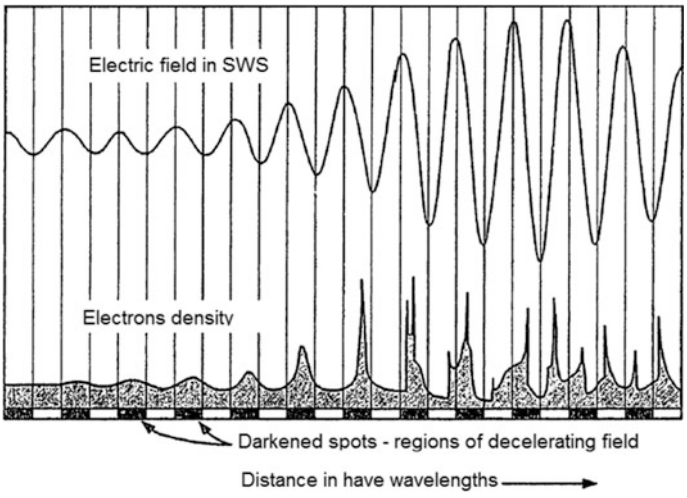


Fig. 7.43 Field and space charge distribution in a TWT

can make an approximate estimate of the electronic efficiency of the travelling wave tube on account of the linear theory equations.

The electronic efficiency is defined by the expression

$$\eta_e = \frac{P_{\text{out}}}{U_0 I_0}, \tag{7.77}$$

where P_{out} is the power transmitted by the electron flux, i.e., the wave power at the SWS output, if the amplification is considered sufficiently large to avoid taking into account the input power. We assume that the amplitude of the convection current first harmonic $\dot{I} = \chi I_0$. Coefficient χ depends on the convection current waveform near the SWS output, but its value cannot be more than two.

Parameter χ is equal to the Fourier series first harmonic amplitude of the curves shown in the graph. In practice, $1 \le \chi \le 2$. The output power is related to the interaction impedance by formula (7.44). Using the relations (7.48) and (7.60), we obtain

$$P_{\text{out}} = \frac{|\dot{E}_{z1}|^2}{2\beta_0^2 R_c} = \chi^2 \frac{C U_0 I_0}{2} \left|\delta_1^2 + q\right|^2. \tag{7.78}$$

For the maximum amplification mode, we find from formulas (7.57) and (7.58)

$$\left|\delta_1^2 + q\right| = \begin{cases} 1, & \text{for small } q, \\ \sqrt{2q}, & \text{for large } q. \end{cases} \tag{7.79}$$

Substituting (7.78) and (7.79) into (7.77), we obtain

$$\eta_e = \chi^2 C \begin{cases} 1/2, & \text{for small } q, \\ \sqrt{q}, & \text{for large } q. \end{cases} \tag{7.80}$$

As can be seen, the electronic efficiency is proportional to gain parameter C and increases with space-charge parameter q. However, the space charge prevents the electrons grouping, so with q increasing, coefficient χ decreases. Therefore, there is an optimal value of the space-charge parameter, at which the electronic efficiency is maximal. Assuming $\chi = 1.3$ and $q = 2$, we obtain $\eta_e = 2.4C$. For $C = 0.2$, we obtain $\eta_e = 48\,\%$, and this is close to the maximum experimental values obtained for high power TWTs.

It should be noted that the electronic efficiency is determined by the difference between the electron velocity and the slow wave velocity. Therefore, to increase efficiency, it is necessary to increase the parameter of non-synchronism b, to values close to the amplification region boundary. However, this causes a reduction in gain. Thus, the b values that are optimal for the gain and for electronic efficiency are different.

Maximum TWT efficiency can be estimated using the energy method. Let's set $q = d = 0$. For these parameters, the optimal value of $b_{\text{opt}} = 0$, and $y_{1\text{opt}} = -1/2$.

Let the velocity of the electrons forming the bunch center be equal to v_e. The velocity of the slow wave interacting with the beam, following from formula (7.52), is equal to $v_{p1} = v_0/(1 - Cy_1) = v_0(1 + C/2)$. The difference in the electrons and wave velocities is

$$\Delta v = v_0 - v_{p1} = v_e\left(1 - \frac{1}{1 + C/2}\right)$$

The deceleration of electrons occurs until the bunch leaves the decelerating field. At this point, its velocity is less than the wave velocity, but as a rough estimate we can assume that it is approximately equal to wave velocity υ_{p1}. The kinetic energy lost by the electron in the process of deceleration is determined by the formula

$$\Delta W_c = \frac{mv_0^2}{2} - \frac{mv_{p1}^2}{2} = \frac{mv_0^2}{2}(1 - \Delta v^2) = \frac{mv_0^2}{2}\left(1 - \frac{1}{(1 + C/2)^2}\right).$$

Taking into account the smallness of gain parameter C in comparison with 1, we obtain approximately

$$\Delta W_c \approx \frac{mv_0^2}{2}C.$$

In this case, the "electronic efficiency" of the electron will be equal to

$$\eta_e = \frac{2\Delta W_c}{mv_0^2} \approx C. \tag{7.81}$$

As can be seen, formulas (7.80) and (7.81) for a small value of q differ only by the numerical factor, and for $\chi = \sqrt{2}$ the difference disappears. For a large value of q, the formula (7.80) takes into account the dependence of the electronic efficiency on the space-charge parameter.

It should be emphasized that both of these formulas are suitable only for a rough estimate of the efficiency of the electron beam energy conversion into SWS field energy. For engineering calculations, it is necessary to use the results of numerical simulation.

7.3.4 Methods of Increasing TWT Efficiency

The efficiency of the traveling wave tube depends on the level of the output power of the device, the type of slow-wave structure used and the mode of device operation, and in particular on the parameter of non-synchronism b among many other factors. The efficiency of a conventional TWT is 10–30% with the upper figure

referring to high-power devices on the coupled cavity chain slow wave structure (CCC type).

The relatively low efficiency value is determined by the specificity of the conversion of the electron beam kinetic energy to the energy of the traveling wave field, which requires synchronism between the wave in the SWS and the electrons. Therefore, electrons cannot move with a velocity much higher than the wave velocity and, accordingly, can transfer only a small part of their energy to it.

There exists several ways to increase the overall efficiency of the traveling wave tubes:

- Direct increase of electronic efficiency due to correction of the wave phase velocity or electron beam velocity;
- Use of "transparent" TWTs;
- Increase of the overall efficiency due to the return of a part of the exhausted electron beam residual energy to the power source, i.e., electron beam energy recuperation.

The increase in electronic efficiency due to correction of the wave phase velocity is achieved through the use of isochronous (tapered) slow-wave structures in which the phase velocity decreases with increasing distance from the SWS input (with the growth of coordinate z) according to a certain law. The required law of phase velocity variation is found through solving the nonlinear TWT equations by computer. To reduce the phase velocity the SWS period is decreased.

Let's note that the chosen law of phase velocity variation is optimal only for a certain value of input power, since the law of electron velocity variation is determined by the field amplitude in the SWS which depends on the input power. Therefore, tapered TWTs have a highly nonlinear amplitude response and can be used effectively only in the pulsed mode. In this connection, a compromise variant of an "isochronous" SWS is often used. The slow-wave structure is divided into sections (usually there are two sections) having different slowing factors. The first (input) section provides a wave phase velocity close to the electron beam velocity; the second section has a greater deceleration and a wave phase velocity smaller than the beam velocity. In the first section, the processes of electron velocity modulation, the formation of electronic groups and the partial transfer of energy from the electron beam to the wave field occur. The main energy extraction is carried out in the second section of the SWS. Optimum values of the section's lengths and their decelerations are determined by computer simulation. The considered method of increasing electronic efficiency was called "the method of the phase velocity hop".

The use of SWS sectioning also makes it possible to implement another means of increasing electronic efficiency—"the voltage step method". In this case, the increased potential is supplied to the second section of the system which has the same phase velocity as the first. This makes it possible to increase the velocity of electron bunches in the channel of the second section and ensures that the velocity of the electrons exceeds the wave phase velocity, which is necessary for effective energy transfer from the electron beam to the SWS field. In principle, a combination

of both methods of increasing electronic efficiency is possible. The application of the methods described above makes it possible to increase the value of electronic efficiency by 10–20%.

Another way to increase the efficiency is based on the use of a "transparent" TWT. Such devices do not have an attenuator, so the SWS in them is short and the gain does not exceed 10–15 dB. Besides, these TWTs have increased requirements regarding the quality of tube input and output matching in order to ensure the absence of self-excitation. The low gain of the transparent TWTs determines their use as a final power amplifier, to the input of which a sufficiently powerful signal is applied. The peculiarities of signal interaction with the electron flux in the absence of the absorber lead to the fact that the electronic efficiency of the transparent TWTs is two to three times higher than the efficiency of a conventional tube. Transparent TWTs are included in the composition of the amplification chain, which uses a conventional TWT or a solid-state amplifier as a preamplifier. This solution allows not only an increase in the efficiency of the chain, but also the adjusting of the output power level within wide limits.

In particular, two-mode amplifiers are designed. In low power mode, no supply voltage is applied to the transparent TWT (or TWT final section), and the signal from the preamplifier passes through it with a slight attenuation. This is an economical mode with low output power. When the voltage is applied to the transparent TWT, the output power of the amplifier increases by 10–15 dB, but the energy consumption also increases.

As is well known, the interaction impedance of the slow-wave structures tends to infinity near the boundary of the transmission band and past the boundary becomes purely reactive. The increasing of the interaction impedance leads to an increase in the gain parameter and, consequently, the efficiency of the device. In this connection, it was proposed to use the TWT at and beyond the bandwidth boundary of the "cold" SWS. Such devices were named *beyond-cutoff TWTs*. The analysis shows that taking into account two spatial harmonics having phase velocities close to the electron velocity makes it possible to eliminate the infinite values of the interaction impedance. One of these harmonics belongs to the direct wave, and the other belong to the opposite one. The SWS interaction impedance is determined, taking into account the sum of the two harmonics fields near the cutoff and appears to be active and independent of the frequency both inside and outside the bandwidth. Therefore, the TWT can amplify the signal even outside the bandwidth and is highly efficient. The beyond-cutoff TWTs (or last sections of the SWS in multi-section TWTs) allow an increase in tube efficiency of 5–10%, however, the bandwidth of such tubes is relatively small.

Even in the case when electronic efficiency reaches 30–50%, the residual power of the electron beam is 50–70% of the total beam power. This power must be absorbed by the electron collector. In this regard, the process of recuperation of energy from the the exhausted electron beam has great practical importance. In addition to increasing the device's overall efficiency, recuperation of the energy of the exhausted electron beam energy allows reduction of the thermal load of the collector, and thereby simplifies the problem of heat extraction. The use of

multi-stage collectors allows an increase in tube efficiency of from 40 to 60%, i.e., to approach to klystrons in this factor.

7.3.5 TWT Design

The Electron Optical System

The TWT electron-optical system consists of an electron gun, a beam transportation system and the collector. At present, both single-beam and multi-beam TWTs are produced. The periodic magnetic focusing with permanent magnets is most often used as the transportation system supporting the beam cross-section shape. The choice of this type of system is due to the large tube length that makes the application of systems with a homogeneous magnetic field ineffective. In high-power TWTs, a solenoid is sometimes used as the transportation system.

The collector serves to collect the exhausted electrons. Since TWT efficiency is relatively small, multi-stage collectors with electron recuperation are often used.

The Slow-Wave Structures

The tube output power, the frequency band and the operating mode determine TWT design. These parameters determine the choice of the slow-wave structures, on which, in turn, the other design elements depend. The main properties of the SWS are described in many books. Some information about them is contained in Appendix B. Therefore, here we briefly consider only the features of those SWSs that are most often used in TWTs.

The helix SWS was first used in Kompfner tubes and since then has remained the most commonly used SWS in TWTs of small and medium power. This SWS provides the deceleration of the fundamental spatial harmonic $n = 5–10$ and has a relatively small dispersion and interaction impedance dependence on frequency, therefore TWTs with such a SWS have an amplification bandwidth of 1–2 octaves. In a helix SWS, with a conductor of circular cross section, the amplitudes of the higher spatial harmonics are small in comparison with the amplitude of the fundamental, zero order harmonic. The disadvantages of this SWS type include the poor heat sink from the helix, and the inability to obtain small slow-down values. Therefore, helix SWSs are used mainly in small and medium power TWTs. The ultimate power is limited by the design of the helix supports. The maximum heat sink is provided by support rods with diamond elements. The design of the supports and the external shield also affect the shape of the SWS dispersion characteristic.

Modifications of helical SWS are structures with a sheet helix, and SWSs of a ring-bar type; SWSs with a sheet helix have a slightly improved heat sink in combination with increased dispersion. These types of SWS are used in TWTs of medium power.

SWSs of the coupled cavity type have a rigid design providing a good heat sink. The CCC used in the TWT works on the minus first spatial harmonic having a

relatively weak dispersion and sufficiently large interaction impedance. The band of these SWS deceleration factors lies within the limits of 1–6. SWSs of the CCC type are used in high power TWTs.

Comb-type SWSs have a simple construction and are used in TWTs of the millimeter band. Unfortunately, the main spatial harmonic of this type of SWS has a strong dispersion which limits the bandwidth of the TWT.

Power Input and Output

The important elements of TWT design are the **input and output couplers**. These elements couple SWS with outer input and output transmission lines. They must ensure the matching of the wave impedances of the outer line and the slow-wave structure in the operation frequency band, the wave impedance of such SWS as CCC and comb-type SWS, being highly frequency dependent. Besides, the energy input and output must contain vacuum windows that separate the internal tube volume from the surrounding space. In addition, the output vacuum window of high-power a TWT must transmit the output power without breakdowns and appreciable absorption.

For the energy input and output from a TWT with a helix SWS, the coaxial line is most often used, the inner conductor of which is connected to the conductor of the helix SWS. Stepped or smooth impedance transformers are used for matching.

The energy input and output from a TWT with a CCC is performed using the rectangular waveguides connected to the first and last resonators by the diaphragm (Fig. 7.44). The design of the outlets is the same, so let's only consider the output of energy.

The last SWS resonator 1 is connected to the impedance transformer 2 and coupled with the rectangular waveguide through the coupling hole. After the

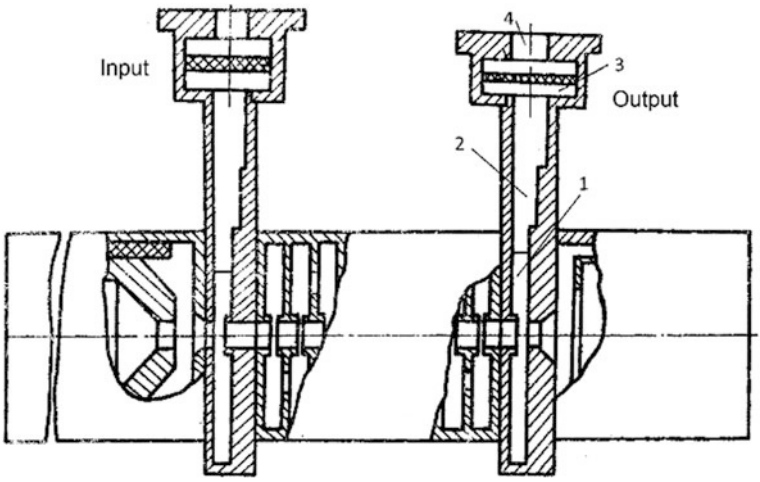


Fig. 7.44 Energy input and output into a TWT

transformer, there is the vacuum window 3 after which the energy enters the output rectangular waveguide 4. The maximum VSWR of such energy outlets does not exceed 1.5 in the tube bandwidth.

7.3.6 Parameters and Application Regions of TWTOs

Travelling wave tubes are designed for the entire microwave band, from decimeter to millimeter waves. This is the most common type of vacuum microwave device, and the share of TWTs in the total production of microwave VEDs is more than 50%.

TWTs are divided into small, medium and high power tubes. The first group includes devices with an output power of up to 1 W. As a rule, they use a helix SWS. These tubes have a wide bandwidth (up to two octaves or more), large gain (45–60 dB) and a low noise factor (3–6 dB). Their efficiency does not exceed 10%, therefore they are used as preliminary amplifiers in telecommunication systems with high data transmission speed, in broadband radar receivers, and in electronic warfare systems.

TWTs of medium power are designed on the basis of the helix SWS or a modification thereof. Their efficiency is 15–30% (usually with recuperation), the gain is 40–50 dB, and the bandwidth is 40–60%. These devices are used in transmitters of radio relay lines, in transponders of communication satellites, and in measuring equipment. Currently, transistor amplifiers actively replace TWTs of small and medium power.

High-power pulsed TWTs have a pulse output power of up to several megawatts, an efficiency of 30–45% and a gain of 35–45 dB. As a rule, a slow-wave structure of the CCC type is used in these devices. The bandwidth of high-power TWTs is 3–6%, which is more than the bandwidth of klystrons of the same power level. They are used in stationary and airborne early warning and control electronic warfare systems. Their use is sensible in cases where klystrons cannot provide the required bandwidth of amplified frequencies.

High-power continuous mode TWTs with an output power of several hundred watts are used in airborne equipment of communication satellites including satellite television systems for the direct transmission of a television signal to individual antennas. TWTs for space equipment have been developed for separate authorized frequencies in the range 1.8–12.7 GHz.

There are two-mode TWTs that can work in both pulsed and continuous modes, their power in pulsed mode being 5–10 dB higher than their power in continuous mode. The transition from one mode to another can be carried out in various ways, in the simplest form by changing the current of the electron beam, for which electron guns with control electrodes are used. The dual mode is also achieved by using a multi-sectioned SWS.

Table 7.4 gives an idea of the typical parameters of a high-power TWT with pulsed and continuous modes. These devices are shown in Fig. 7.45. A UV-45

Table 7.4 Parameters of some TWTs

TWT type	Mode	SWS type	f_0 (GHz)	$\Delta f/f_0$ (%)	P_{out} (kW)	η (%)	μ (dB)	U_0 (kV)	I_0 (A)
UV-45	Cont.	Helix	7.4–12	–	5 mW	5	30	1100	
UV-332	Cont.	CCR	14–15	3.5	0.3	18	35	10–12	0.15
VTX-5783	Pulsed	CCR	8.9–9.4	6	50	18	48	32–34	8.5
TW2-21	Pulsed	CCR	8	7.5	900	37	33	100	24

TWT of the low power type is used in radar receivers. A UV-332 TWT of medium power is intended for use in satellite transponders and radar equipment. These tubes are manufactured in Russia. The high-power TWT VTX-5783 is produced in Canada and is intended for operation in radar transmitters. The high-power TWT developed in China (the last line of the Table) has the same purpose. Figure 7.45d also shows the multi-beam transparent TWT with CCC developed at the JSC "Istok".

Table 7.5 shows the parameters of a TWT with the helix SWS developed for communication satellites by the L3-Communications Company.

The use of multi-stage collectors in these tubes allowed an increase in the efficiency of devices to 60–70% while maintaining the wide frequency band.

In addition, the mean life time of millions of hours is worth mentioning. Such an indicator was achieved due to the careful design of devices and the use of modern technologies in their manufacture.

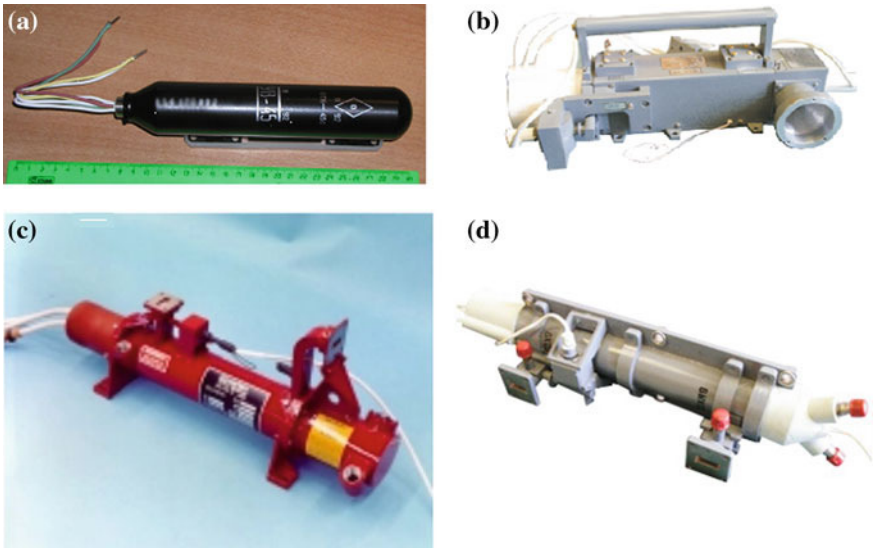


Fig. 7.45 Types of TWT: **a** low power TWT UV-45; **b** millimeter band middle-power TWT; **c** high-power TWT VTX-5783; **d** transparent TWT

Table 7.5 Parameters of TWT with helix SWS

Band	Power (W)	Efficiency (%)	Number of devices on orbit (units)	Mean life time between errors (mln. hours)
L (1.4–1.6 GHz)	80–250	60–70	0	–
S (2−4 GHz)	100–300	62–73	330	10.7
C (5.15–5.35 GHz)	20–120	66–68	237	2.6
X (10–12 GHz)	25–160	63–66	115	1.8
Ku (11–18 GHz)	30–155	60–70	449	2.7
K (18–27 GHz)	30–130	58–66	0	–
Total	–	–	1131	–

7.4 Backward-Wave Oscilators

7.4.1 Operating Principle of Backward-Wave Tubes

M.F. Stelmakh (USSR) got the patent on an O-type backward-wave oscillator (BWO) in 1947, and production of BWOs began in 1951 in both UK and USSR. Another name for a BWO is *carcinotron*. The operation of this device, as well as the TWT, is based on the synchronous interaction of the wave in the slow-wave structure and the electron beam. However, unlike the TWT, the BWO uses the interaction with the negative (inverse) spatial harmonic, the phase and group velocities of which have opposing directions. The direction of the phase velocity coincides with the direction of the motion of electrons (from the cathode to the collector), and the group velocity (and the energy transfer) are directed in the opposite direction. Interacting with the wave field, the electrons undergo velocity modulation and bunching. Hence, they give some of their energy to the wave. This energy is transferred by the wave to the tube output, as a result of which internal feedback arises and under certain conditions self-excitation occurs. In reality, BWOs are always used in the self-excitation mode, i.e. as generators. Hence, their English name of backward wave oscillator (BWO). The transfer of energy from the electrons to the wave in the BWO can be regarded as induced Smith-Purcell radiation.

The scheme of the backward-wave tube is shown in Fig. 7.46. The electron flux is formed and accelerated by the electron gun 1 and enters the transit channel of the slow-wave structure 3. The cathode end of the SWS is connected to the energy output 2 and its collector end is connected to the reflection absorbing termination 4. The beam transportation through the channel is performed by means of the uniform magnetic field that is created by the solenoid 6. The exhausted electron beam goes to the collector 5.

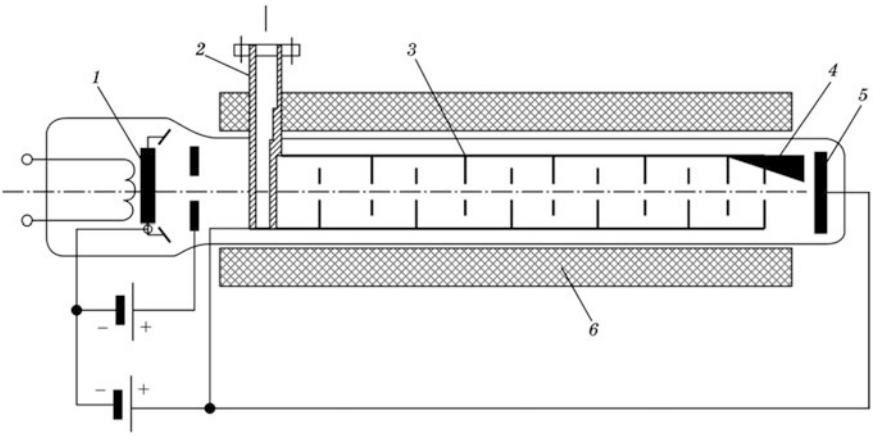


Fig. 7.46 Scheme of a backward-wave oscillator

Let's suppose that there is a traveling wave in the SWS, the working spatial harmonic of which has negative dispersion and a phase velocity approximately equal to the velocity of the electrons. The field of this harmonic modulates the velocity of the electron flux. This velocity modulation leads to the bunching of the beam. Bunches form near the SWS collector end and transmit part of its kinetic energy to the wave. This energy is transferred by the wave to the SWS cathode end and then transfers to the load by means of the output coupler 2. The absorber 4 provides the traveling wave mode in the SWS, absorbing the waves reflected from the output coupler and the load.

The electron beam interacts with the SWS field, which has the form of the traveling wave. The main processes that determine BWO operation are the same as in the traveling wave tube: the velocity modulation of the electron beam by the wave field, the bunching process and the transfer of the electron beam's kinetic energy to the wave field. The basic difference is that the electromagnetic wave transfers energy in the opposite direction with respect to the motion of the electron beam, i.e. from the SWS collector end to the cathode.

Let's assume that oscillations of the electromagnetic field appeared at the SWS collector end, and propagate toward the SWS cathode end in the form of the traveling wave. Under certain conditions, the interaction of this wave with an electron beam can lead to the transfer of the part of the electron beam's kinetic energy to the wave field and to the growth of its amplitude, just as takes place in the TWT. The resulting and amplified electromagnetic wave acts on the beam and leads to its velocity modulation and subsequent bunching.

The bunches formed near the SWS collector end transfer energy to the wave field. This energy propagates to the SWS cathode end, increasing the amplitude of the field in this region, leading to a more intensive velocity modulation and bunching.

Thus, internal positive feedback occurs. As a result, the BWO operates as a regenerative amplifier, and under certain conditions, it self-excites and turns into a generator. In reality, BWOs are almost always used in the generation mode, so they have only one power output from the SWS cathode end. Its collector end is loaded with the reflection absorbing termination to prevent reflections that lead to undesirable feedback on the slow-wave structure.

7.4.2 Linear Theory of BWOs

The BWO linear theory can be constructed on the basis of the TWT linear theory by directing the z axis from the collector to the cathode. In this case, the direction of the wave group velocity will be positive and the direction of the phase velocity and the velocity of the electrons will be negative. Accordingly, we obtain

$$\beta'_e = -\beta_e < 0, \quad \beta'_0 = -\beta_0 < 0, \quad d' = \frac{\alpha'_0}{\beta'_e C} < 0, \quad \alpha'_0 < 0,$$

where the BWO parameters are denoted by a prime. Substituting these quantities into (7.53), we obtain the dispersion equation of the BWO:

$$\mathrm{j}(\delta' - \theta')[(-\delta')^2 + q] + 1 = 0, \tag{7.82}$$

the parameter δ' has the same meaning as in (7.51). In the following, we omit the primes of the BWT parameters.

This equation has three roots corresponding to the three waves propagating in the BWO. Comparing (7.53) and (7.82) we see that if we replace b by $-b$, the δ parameter sign also changes to its opposite. Thus, the complex propagation constants of three waves in the BWO coincide up to a sign with propagation constants of waves in a TWT having the same parameters C, q and d. The amplitudes of the waves at the input ($z = L$) and at the output of the tube ($z = 0$) are related by

$$\dot{E}_{zn}(L) = \dot{E}_{zn}(0)\mathrm{e}^{-j\delta_n \beta_e CL}, \quad n = 1, 2, 3. \tag{7.83}$$

The total field created at the BWO-amplifier input by an external generator is:

$$\dot{E}_z(L) = \sum_{n=1}^{3} \dot{E}_{zn}(L). \tag{7.84}$$

The similar relation is valid at the tube output:

$$\dot{E}_z(0) = \sum_{n=1}^{3} \dot{E}_{zn}(0). \tag{7.85}$$

For $z = 0$, the wave amplitudes are related to the equation analogous to (7.62):

$$\frac{\dot{E}_{zn}(0)}{\dot{E}_z(0)} = \frac{\delta_n^2 + q}{(\delta_n - \delta_{n2})(\delta_n - \delta_{n3})}, \tag{7.86}$$

where

$$n_2 = \begin{cases} n+1, & n<3; \\ 1, & n=3. \end{cases} \quad n_3 = \begin{cases} n+2, & n<2; \\ n-1, & n\geq 2. \end{cases}$$

Substituting (7.86) in (7.84) and (7.85) we find the BWT gain in terms of power:

$$\mu = \left|\frac{\dot{E}_z(0)}{\dot{E}_z(L)}\right|^2 = \left|\sum_{n=1}^{3} \frac{(\delta^2 + q)\mathrm{e}^{-j\delta_n \beta_e CL}}{(\delta_n - \delta_{n2})(\delta_n - \delta_3)}\right|^{-2}. \tag{7.87}$$

For self-excitation of the BWO, it is necessary that its gain turns to infinity. From here

$$\sum_{n=1}^{3} \frac{(\delta^2 + q)\mathrm{e}^{-j\delta_n \beta_e CL}}{(\delta_n - \delta_{n2})(\delta_n - \delta_3)} = 0. \tag{7.88}$$

At given parameters q and d, (7.88) allows us to find the quantities $\theta = -C\beta_e L\,\theta$ and b. If these parameters are found, the conditions for self-excitation of the BWT can be written as follows:

$$(\beta_e - \beta_0)L = b\theta, \tag{7.89}$$

$$I_0 = I_{\text{start}} = \frac{4U_0\theta^3}{R_c|\beta_e L|^3}. \tag{7.90}$$

Equation (7.88) has a number of solutions, each with its own values of starting current and generation frequency. The solution, to which the smallest starting current (the fundamental mode of oscillations) corresponds, is of the greatest interest. When the tube is turned on, it is the one that is excited first and suppresses the others with larger values of starting current. For the basic mode of oscillations and $q = 0$ we have $\theta = 1.98$ and $b = 1.52$. From here

$$(\beta_e - \beta_0)L \approx \pi, \quad I_{\text{start}} \approx \frac{32U_0}{R_c|\beta_e L|^3}. \tag{7.91}$$

The last formula defines the starting current of the BWO, i.e. the value of the current at which the tube begins to generate oscillations. The dependence of the starting current on the accelerating voltage is shown in Fig. 7.47. The same figure

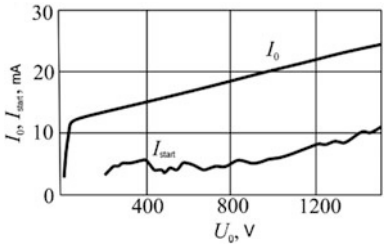


Fig. 7.47 Starting current and beam current versus accelerating voltage

shows the dependence of the beam current on the accelerating voltage. As can be seen, the real beam current at all values of accelerating voltage is much higher than the starting current, so the BWO generates the oscillations in the entire operating range of the accelerating voltages.

Since the operating spatial harmonic has dispersion, i.e., β_0 depends on the frequency, formula (7.89) determines the generation frequency as a function of electron velocity, i.e., the accelerating voltage. Thus, the BWO has the electronic tuning of the generation frequency.

Figure 7.48 shows the distribution of the variable components of the electric field, the beam current, and the electron velocity along the tube length for different values of the space charge parameter. It is assumed that $d = 0$ and there are no reflections from the SWS collector end. Although the graphs are plotted according to the linear theory for $I_0 = I_{\text{пуск}}$, the shape of the curves does not change significantly even at $I_0 \gg I_{\text{пуск}}$. As can be seen, the BWO does not have the section of exponential growth of field density which is characteristic for the TWT.

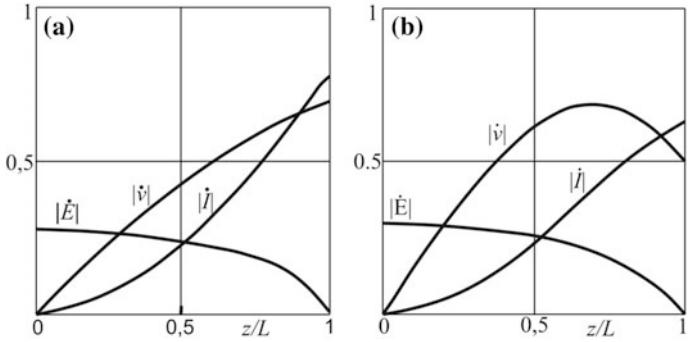


Fig. 7.48 Distribution of field, current and electron velocity along BWO length: $a - q = 0$, $b - q = 0.5$

7.4.3 Electronic Tuning of BWOs

As noted in the previous section, the BWO generation frequency depends on the accelerating voltage. This BWO property is called electronic tuning. Electronic tuning is characterized by the electronic tuning band determined by the relation

$$\frac{\Delta f}{f_0} = 2\frac{f_{\max} - f_{\min}}{f_{\max} + f_{\min}},$$

where $f_{\min}$ and $f_{\max}$ are the minimum and maximum frequencies of the tuning band, and f_0 is the average frequency of the band. The width of the BWO electronic tuning band can reach an octave.

The electronic tuning steepness

$$S_f = \frac{\mathrm{d}f}{\mathrm{d}U_0}$$

determines the rate of the frequency change with the changing of the accelerating voltage. It also determines the stability of the generation frequency which is related to the stability of the accelerating voltage source. In the linear approximation at $d = 0$ from the first formula (7.91), we obtain

$$S_f = \frac{f_0}{2U_0}\frac{1}{1 + |v_0|/v_g},$$

The change of the accelerating voltage is also accompanied by irregular changes of the output power, which are explained by changes in the conditions of SWS matching with the output coupler and the reflection absorbing termination, and by changing the basic parameters of the tube. The experimental dependence of the frequency and the BWT output power on the accelerating voltage is shown in Fig. 7.49.

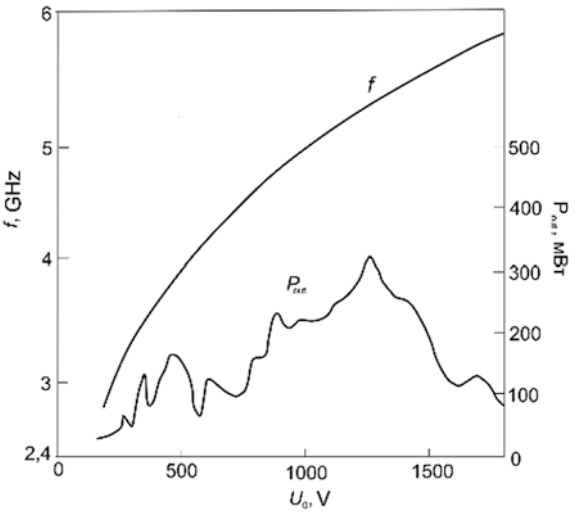


Fig. 7.49 Frequency and BWT output power versus accelerating voltage

7.4.4 Electronic Efficiency of BWOs

The analysis of the dependence of BWO efficiency on its parameters and operating mode is possible only within the framework of the nonlinear BWO theory. This theory is constructed in a manner analogous to the nonlinear theory of TWT, taking into account the opposite directions of the phase and group velocities in the SWS. Without considering the nonlinear BWO theory in detail, we confine ourselves to considering the physical processes of energy extraction and show some results obtained with the aid of the nonlinear theory.

First, let us note that, unlike the TWT, where the maxima of the alternating current and field components coincide, the alternating current in the BWO is maximal near the collector end of the tube and the field is maximal near the cathode end. Therefore, in the BWO, the interaction of the electron beam and the wave in the SWS is less intense than in the TWT.

Further, to achieve maximum efficiency in a TWT, the electrons should move much faster than the wave, so the values of the non-synchronism parameter b corresponding to the maximum coefficient and the maximum efficiency do not coincide: $b_{\text{opt}\eta} > b_{\text{opt}\mu}$. In the BWO, while trying to increase b by increasing the accelerating voltage, the generation frequency changes so that parameter b remains optimal on the gain factor. This is the second reason for the lower efficiency of the BWO compared to the TWT. Calculations show that the electronic efficiency of the BWO can be estimated from the formula

$$\eta_e \approx 0.7\chi^2 \frac{\omega_q}{\omega}\left(1 - \frac{I_{start}}{I_0}\right).$$

The value of $\chi = |\dot{I}_{\max}|/I_0$ in the BWO is smaller than in the TWT due to the lower quality bunching. As a result, the electronic efficiency of the BWO is four to five times less than the electronic efficiency of the TWT.

The total efficiency of the BWO is equal to the product of the electronic efficiency and the efficiency of the electrodynamics system. Typical values of the latter lie in the band 0.8–0.9, with smaller values corresponding to millimeter band devices. As a rule, the total efficiency of the BWO does not exceed 10%.

7.4.5 Resonance BWOs

Resonance BWOs are devices in which the SWS collector end is short-circuited or open (let's recall that in a conventional BWO it is matched). The output of energy is also made unmatched. As a result, a wave close to the standing wave is established in the SWS, and instead of the continuous dispersion characteristic in the SWS bandwidth a number of oscillations modes arise, each of which has its own frequency and phase shift between neighboring periods $\varphi_n = \pi/N_p$, where N_p is the number of

SWS periods. The electron flux interacts with that oscillation mode for which the synchronism condition $\beta_e D = \varphi_n$ is satisfied. As a result, the continuous band of frequency variation breaks up into a number of generation zones, each of which has its own frequency. Since in each zone the generation frequency is fixed, it is possible to select the accelerating voltage to obtain the optimal non-synchronism parameter in terms of efficiency. In addition, the distribution of the electric field along the tube length in the resonance BWO is more uniform than in the conventional BWO, leading to more efficient energy extraction. As a result, the efficiency of the resonance BWO is several times larger than that of the conventional BWO. However, this is achieved by the loss of the main advantage of the BWO—the wide band of electronic tuning (the narrow band of electronic frequency tuning within each generation zone is preserved).

7.4.6 *Design and Parameters of BWOs*

The design of the BWO and its parameters are largely determined by the type of slow-wave structure system. SWSs with "interdigital pins" (Fig. 7.50a) in combination with a sheet electron beam are the most commonly used in the BWO. In the decimeter band, a bifilar helix is sometimes used (Fig. 7.50b), while in the millimeter and submillimeter bands, a ladder SWS (Fig. 7.50c) and folded waveguide are used (Fig. 7.50d).

For beam transportation, an external solenoid, and a magnetic focusing system with permanent magnets (MFS) and sometimes the periodic electrostatic focusing

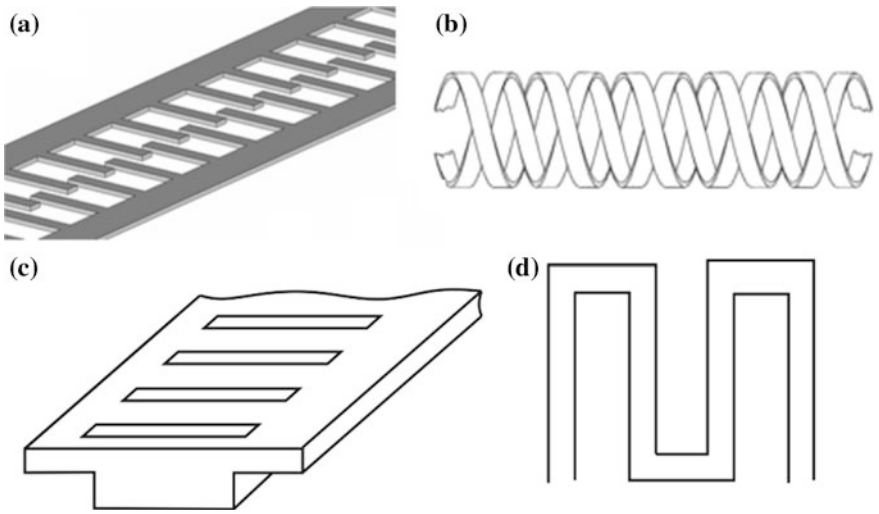


Fig. 7.50 Types of SWS used in BWOs

system (PES) are used. As a rule, BWOs with MFS have the packaged design, i.e., the magnetic system and the tube body are structurally connected.

BWOs are used as low-power generators with electrical frequency tuning. They are produced for the entire microwave band, from decimeter to submillimeter wavelengths. The leading position in this area is occupied by JSC R&D "Istok" where unique sub-millimeter BWOs were developed and are still manufactured. Typical parameters of BWOs produced by this enterprise are given in Table 7.6.

As an example, we show the parameters of a packaged millimeter band BWO (Fig. 7.51a) in Table 7.7. Another example is the terahertz BWO with an external

Table 7.6 Parameters of typical BWOs

Main parameters	F (GHz)	$\Delta f/f$	P_{out} (mW)	U_0 (kV)	The focusing system type
Decimeter and centimeter bands	0.1–20	5–30	20–50	0.2–2	MFS
Centimeter and millimeter bands	1–40	20–40	50–1000	0.1–0.8	PES
Millimeter band	35–260	15–60	6–40	0.4–1.9	Packaged with MFS
Sub-millimeter band	258– 1400	15– 60	0.5–10	1–6	Unpackaged

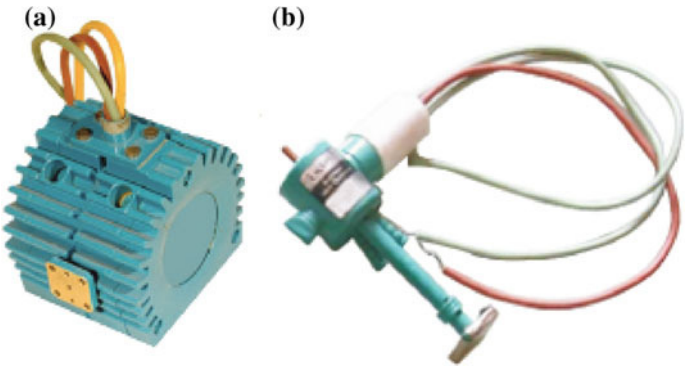


Fig. 7.51 External view of two BWOs

Table 7.7 Parameters of the packaged mm band BWO

Operation frequencies band (GHz)	118–178
Output power (mW)	6–20
Acceleration voltage (V)	500–1500
Beam current (mA)	10–20
Total mass (kg)	<1
Output waveguide size (mm)	0.8 × 1.6

Table 7.8 Parameters of the Terahertz band BWO

Frequency band (THz)	1.176–1.4
Output power (mW)	0.5–2
Acceleration voltage (kV)	1.5–6
Focusing magnetic field (T)	1.1

magnetic system, also developed in JSC R&D "Istok" (Fig. 7.51b). The parameters of this tube are given in Table 7.8.

7.5 O-Type Hybrid Devices

7.5.1 Hybridization Advantages

The "classical" O-type devices considered in the previous sections—the amplifying klystrons, TWTs and BWOs—have advantages and disadvantages. Thus, in the klystron the effective mechanism of grouping and the energy extraction are used, it allows obtaining the high electronic efficiency. However, the bandwidth of the klystron is limited by the use of the resonator for the energy extraction. The TWT has the wide band of amplified frequencies, but it is characterized by lower values of the output power and efficiency. Devices with electrostatic control are characterized by the simple design, the absence of the focusing system, but they have limitations on the maximum operating frequency and the gain.

In connection with the foregoing, the idea to combine the most effective interaction mechanisms in one device arose, thereby getting rid of the limitations inherent in classical O-type devices. A number of such "hybrid" devices have been proposed and investigated, but only some of them have practical applications.

7.5.2 The TWYSTRON

The twistron was proposed in the early 1960s of by employees of the Varian company in the USA. The name of the device (TWYSTRON) consists of the first letters of the words "Traveling Wave" and the end of the word "KLYSTRON". The twystron consists of a klystron buncher containing several resonators and an SWS output section operating in the traveling wave mode (Fig. 7.52). The cathode 1 emits electrons that are accelerated by the anode 3 and form the electron beam 2. The amplified signal from the input source 5 is applied to the resonator 4 which modulates velocity of the electron beam. The intermediate resonator 6 bunches the beam.

The bunched beam excites the wave in the SWS 7 in which velocity is chosen to be somewhat smaller than the electron velocity. The SWS is matched from the side

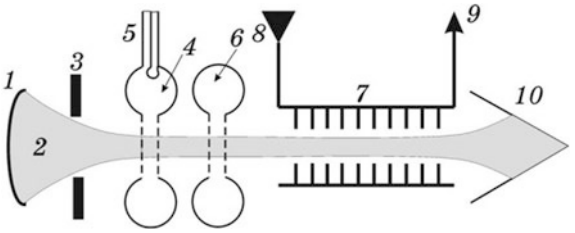


Fig. 7.52 Scheme of a twistron

of the resonators by an absorber 8. The electron bunches automatically appear in the decelerating field of the wave and give it some of their energy, which is trasmitted to the payload by the output coupler 9. The exhausted electrons are deposited on the collector 10. Since the buncher can be made broadband and there are no resonance elements in the energy extraction system, the twistron bandwidth can be larger than that of the klystron and is comparable with the TWT frequency band. The efficiency of the twistron, however, is less than that of the klystron, since the bunch does not transfer energy to the wave when its velocity becomes less than the velocity of the wave. In the klystron output resonator, bunches can ideally be decelerated to a complete stop.

A CCC is usually used as the slow-wave structure in the twistron. The use of a coupled resonators system made it possible to increase the device bandwidth to 7–15% of the average frequency. Among microwave devices with a large pulsed power, twystrons are the commonest wideband amplifying device. They are produced for operation in the centimeter band wavelengths (from 5 to 10 cm), with a pulsed power of 3–8 MW, a gain of 35–50 dB, and an efficiency of 35–40%. Twystrons find general use in the transmitters of high-power ground-based and marine radars.

At present, in connection with the development of multi-beam klystrons with multiple-gap resonators, the difference between the klystron and twystron parameters is reduced, and since twystrons have a structure that is more complex, their production is somewhat reduced.

7.5.3 The Klystrode

The history of the klystrode began in 1939 when A.V. Haeff proposed an "inductive output tube" (IOT). The tube represents the triode, which has a resonator in place of the anode. Electron bunches formed by the control grid are accelerated by the constant field of the resonator to which the anode voltage is applied. While transiting the resonator gap bunches excite the oscillations in it and transfer part of their energy. The exhausted electrons settle on the collector. Thus, in the Haeff tube, the electrostatic method of beam modulation and the dynamic (klystron) method of energy extraction are combined. Since electrons enter the resonator gap already having been accelerated, the length of this gap can be made quite large. For a while,

Haeff tubes were produced for use in television power amplifiers, but in 1941 their production was stopped.

Forty-three years later, a tube with an inductive output was recalled and developed again for high-power TV transmitters of the decimeter band. Its creators were D.H. Preist and M.B. Shrader from the Varian American company. The device, called a klystrode, consists of several parts which are characteristic, on the one hand, of the klystron, and on the other hand, of the triode, which explains the name of the device.

Its scheme is shown in Fig. 7.53. The cathode with the heater 1, the control grid 2, and the anode 4, form the triode part of the device. The amplified signal is applied to the input resonator 3, part of the shell of which is formed by the cathode and the grid. In the resonator tuned to the frequency of the input signal, oscillations arise, which leads to the voltage appearance at the cathode-grid gap. The bias voltage U_c, which is negative with respect to the cathode, is also applied to the grid. Electrons emitted from the cathode are modulated in density by an alternating voltage on the grid and are accelerated by the field of the anode 4. The modulated electron beam 7 passes through the gap of the output resonator 5 exciting the oscillations therein, just as it does in the output resonator of the klystron. The phase of these oscillations when the output resonator is tuned to the signal frequency is automatically set so that the electrons transfer their kinetic energy to the resonator field. The exhausted electrons are deposited on the collector 6.

To provide the instantaneous band of 6–9 MHz required for the transmission of the TV signal, energy extraction in the klystrode is usually performed by the filter system consisting of active and passive resonators. The device has short drift channels before and after the output resonator, and the electron beam is kept from defocusing by the magnetic field of the solenoid or permanent magnets.

Evaluation of the electronic efficiency of the klystrode. Typically, these devices operate in the "B" class with a cut-off angle of 90°. The ratio of the convection current first harmonic amplitude to the constant component of the current in this regime is $\chi = |\dot{I}_{c1}|/I_0 = 1.57$.

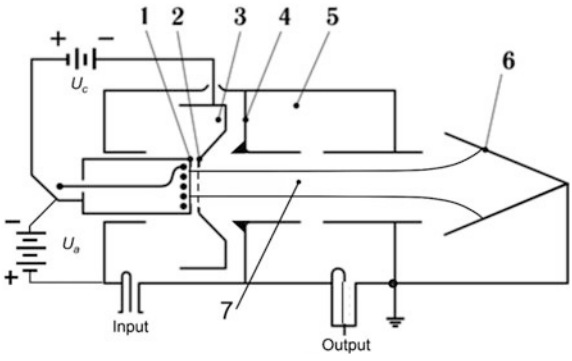


Fig. 7.53 Scheme of a klystrode

The amplitude of the alternating voltage at the output cavity gap U_2 is limited by the ejection effect, and its maximum value is determined by the expression $M_2 U_2 = U_0$ where M_2 is the interaction coefficient of the output resonator. Hence the maximum power transmitted by the electron flux to the resonator field is:

$$P_{e\max} = 0.5|\dot{U}_2||\dot{I}_{c1}| = \frac{U_0 \chi I_0}{2M_2}.$$

The power of the anode voltage source $P_0 = U_0 I_0$. Hence the maximum electronic efficiency is

$$\eta_{e\max} = \frac{P_{e\max}}{P_0} = \frac{s}{2M_2}\eta_{e\max} = \frac{P_{e\max}}{P_0} = \frac{\chi}{2M_2}.$$

Substituting χ into this formula and assuming $M_2 = 0.9$, we obtain a maximum electron efficiency of 87%, which is substantially higher than that of klystrons and TWTs. The total efficiency of the klystrode is equal to the product of the electronic efficiency at the efficiency of the circuit. Assuming the latter equal to 90%, we get a total efficiency of 78%. Such a high efficiency is due to the good shape of bunches provided by the electrostatic control method, but this method limits the maximum operating frequency of klystrodes, which does not exceed 1–1.5 GHz. The low gain, not exceeding 20–22 dB, is also a disadvantage of klystrodes. This means high-power preamplifiers must be used in the transmitter circuits, which increases the cost of the equipment.

Design, parameters and applications of klystrodes. Currently a number of firms are producing klystrodes, which are used mainly in analog and digital television transmitters in the decimeter band (470–810 MHz). The advantages of klystrodes are determined by two factors. First, the klystrode amplitude characteristic is much more linear than that of the klystron. This feature makes it promising to use klystrodes to transmit high-definition digital television where linearity is particularly important.

Secondly, since klystrodes operate in the "B" class, the power they consume depends on signal level. In a conventional klystron, the consumed power does not depend on the signal level and must be permanently high to maintain peak output power. Therefore, the replacement of klystrons by klystrodes brings substantial energy savings.

Table 7.9 shows the parameters of two klystrodes: the K2D110W of the Eimac comapny, and the klystrode developed at JSC R&D "Istok" which received the trade name of Istron. This is a multi-beam (18 beams) device, which has a lower supply voltage and an increased gain.

Further increase of klystrode efficiency is possible due to the use of recuperation, the use of resonators tuned to the second harmonic of the signal, and the transition to the "C" class operating mode.

The devices listed in the table are shown in Fig. 7.54a, b respectively. The same devices mounted in a transmitter are shown in Fig. 7.54c, d.

Table 7.9 Parameters of some klystrodes

Type	Frequency band (MGz)	Output power (kW)	Gain (dB)	Anode voltage (kV)	Anode current (A)	Number of beams	Weight (kg)	Focusing type
K2D110 W	470–815	110	22	34	2.1	1	11.4	Solenoid
Istron	470–810	60	23–24	25	–	18	–	Solenoid

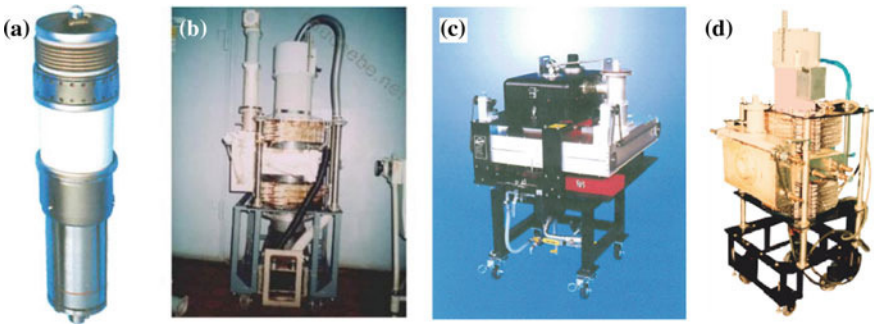


Fig. 7.54 The klystrodes K2D110 W (**a**), Istron (**b**). Klystrodes in transmitters K2D110W (**c**) and Istron (**d**)

7.5.4 The Orotron

The orotron is a self-oscillator with a linear electron beam. This is one of the first devices in which the open mirror resonator was used, which explains its name. The open resonator, due to its dimensions, is large in comparison with the wavelength, and this allows the orotron to work effectively in millimeter and submillimeter wavelength bands

The author's certificate for the orotron was given to F.S. Rusin and G.D. Bogomolov in 1966. Other less commonly used names for this device are the diffraction radiation generator (DRG), the ledatron, and the electronic device with Fabry-Perot resonators. The orotron resonator (Fig. 7.55) consists of flat (1) and spherical (4) mirrors. On the surface of the flat mirror, there is a comb-shaped periodic structure above which the electron beam 2 passes.

Waveguide 3 transmits the energy of oscillations into the output load. The presence of the periodic structure, which can be considered as a slow-wave structure and open resonator make it possible to consider the orotron as a hybrid O-type device, combining interaction elements characteristic of both the TWT and the klystron.

To analyze the operation of the orotron, we assume that the certain oscillation mode was established in its resonator, in which the field has the form of the

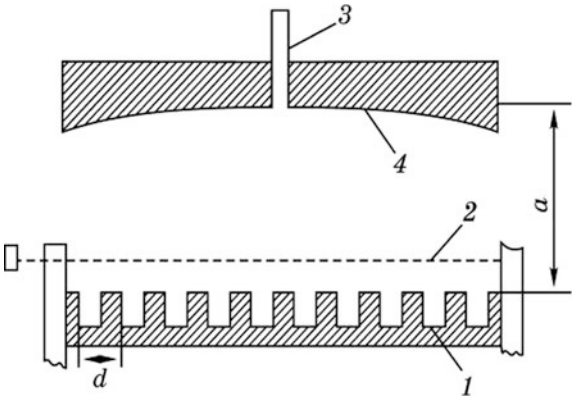


Fig. 7.55 Scheme of an orotron. 1 Flat reflector with SWS. 2 Electron beam. 3 Energy output. 4 Spherical mirror

standing wave existing between the mirrors. On the periphery of the mirrors, the field is limited by the caustic surface S (Fig. 7.56). Inside this surface, the distribution of the electric field along the z axis is described by some function $E_z(z)\psi(z)$, where $E_z(z)$ is a periodic function of the z coordinate, and ψ is a slowly varying function determined by the oscillation mode of the resonator. For the oscillation mode having one maximum on the flat reflector, the field varies according to a law close to Gaussian

$$\psi(z) = A\exp[-z^2/(2r^2)],$$

where r is the "radius" of the field on the surface of the flat reflector (Fig. 7.56).

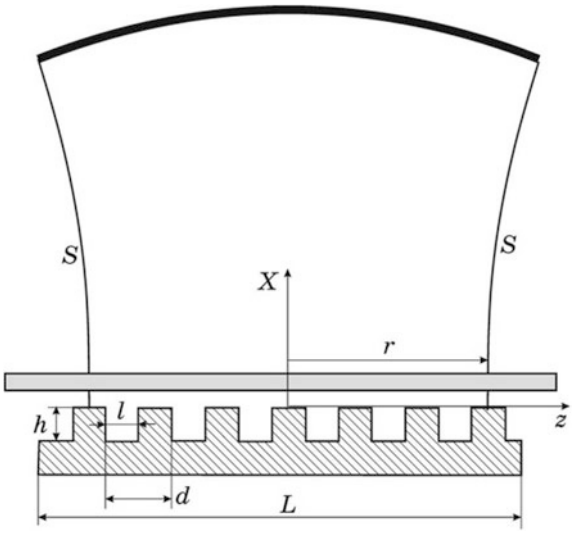


Fig. 7.56 Open resonator of an orotron

The comb on the surface of the flat reflector distorts the picture of the field near it. Assuming the comb is infinite in the direction of the x-axis, the field near the comb can be represented as the expansion in terms of spatial harmonics:

$$\dot{E}_z(x,z) = \sum_{n=1}^{\infty} \dot{E}_{zn} \mathrm{e}^{-jk_{xn}x} \cos(k_{zn}z), \tag{7.92}$$

where

$$k = \omega/c = \sqrt{k_{zn}^2 + k_{xn}^2}, k_{zn} = 2\pi n/d, n = 0, \pm 1, \pm 2, \ldots$$
$$\dot{E}_{zn} = \int_{-d/2}^{d/2} \dot{E}_z(0,z) cos(k_{zn}z) dz$$

is the amplitude of the mth harmonics. Hence

$$k_{xn} = 2\pi \sqrt{\left(\frac{1}{\lambda}\right)^2 - \left(\frac{n}{d}\right)^2},$$

where λ is the wavelength in the free space. Since the condition $d \ll \lambda$ is always kept in the orotron, the transverse wave number for all harmonics is purely imaginary: $k_{zn} = \pm j\gamma_n$, where γ_n are real numbers.

Substituting this expression in (7.92), and taking into account the fact that the perturbed field disappears at infinity, we have

$$\dot{E}_z(x,z) = \sum_{n=-\infty}^{\infty} \dot{E}_{zn} \mathrm{e}^{-j\gamma_n x} \cos(k_{zn}z). \tag{7.93}$$

It follows from (7.92) that each term in the expansion (7.92) consists of two spatial harmonics traveling in opposite directions. For the appearance of Smith-Purcell radiation at the motion of electrons above the periodic structure, it is necessary that the phase velocity of the harmonic and the velocity of the electrons are approximately similar:

$$v_e = v_{pn} = \frac{\omega}{k_{zn}} = \frac{\omega d}{2\pi |n|}. \tag{7.94}$$

The orotron usually uses the negative first spatial harmonic ($m = -1$) which has the largest amplitude and a negative dispersion. Therefore, the orotron is similar to the BWO. However, feedback in the orotron is achieved not only due to oppositely directed phase and group velocities, but also due to the field of the high-Q open resonator. For this, the frequency of radiation should be close to the eigen frequency of one of the resonator modes ω_m, $m = 1, 2, \ldots$. Under the action of the resonator

field, the electrons gather into bunches of length $d \ll \lambda$, which produce the induced radiation. Thus, for the excitation of oscillations in an orotron, two conditions must be fulfilled:

1. synchronous conditions in space (7.94);
2. synchronous conditions in time $\omega \approx \omega_n$.

As the linear orotron theory shows, the minimum value of the starting current is achieved when the space-charge parameter $q_p = \beta_q L$ is close to zero and the synchronism parameter is $\Phi_0 = 2(\beta_e - k_{zn})r = -\sqrt{2}$.

For a thin beam passing near the comb tips

$$I_{start} = 3\frac{\omega_m V}{\beta_e r^3}\frac{N_m}{Q_m|\dot{E}_{zn}|^2},$$

where V is the resonator volume; $N_m = \int_V \varepsilon_0|\dot{E}_m|^2 dV$ is the norm of the *m-th* mode of the resonator oscillation; and Q_m is its loaded Q-factor.

Calculation of orotron efficiency is possible only by numerical methods on the basis of the nonlinear theory. Practice shows that total orotron efficiency in the millimeter wavelength band is 15–30%.

To increase orotron efficiency, the same methods as for the TWT and the BWO can be used: isochronism, i.e. the change of the periodic structure step, and recuperation. In addition, resonators in which the field distribution over the surface of the flat reflector is close to uniform are used. For this purpose, metal reflectors are placed at the edges of the mirror. Calculations show that optimization of the function of the field distribution over the mirror can increase the efficiency of the orotron up to 58%.

An essential increase of efficiency and output power is achieved when using the two-row periodical structure in which the electron beam passes not over, but between the combs.

Table 7.10 shows the parameters of some orotrons developed in the Fryazino branch of the Institute of Radio Engineering and Electronics of the Russian Academy of Sciences, and in Fig. 7.57 the image of one of these devices is shown.

A number of modifications to the classical orotron design is proposed: an orotron with a coaxial open resonator, and an orotron with crossed fields among others. There are prototypes of relativistic orotrons generating tens and hundreds of megawatts in the centimeter wave band.

Table 7.10 Parameters of some orotrons

f (GHz)	$P_и$ (kW)	Efficiency (%)	$Q_н$	U0 (kV)	I_0 (A)
10	53	35	800	13	11.6
37.5	40	17	2000	19	12
90	1.2	6	5000	20	11

Fig. 7.57 Orotron

Advancement Questions

1. Describe the main phases of the energy conversion process in O-type devices.
2. List the main parts of the double-cavity drift klystron and describe the principle of its operation.
3. What is the average interaction coefficient and how does it depend on the radius of the drift channel and the beam radius?
4. What is the grouping parameter? What is the optimal value of this parameter for the amplifier klystron?
5. How is the disruption of longitudinal electrons taken into account when analyzing the grouping process?
6. What is the electronic efficiency of the klystron and on which device parameters does it depend? What is the maximum value of the electronic efficiency of the double-cavity drift klystron?
7. How do you determine the gain of the double-cavity drift klystron?
8. What tuning circuits are you aware of for multi-cavity klystrons?
9. What determines the electronic efficiency of the multi-cavity klystron?
10. What determines the bandwidth of the multi-cavity klystron?
11. What methods for the frequency band expansion for the multi-cavity klystron do you know?
12. What other types of klystrons are you aware of besides transit-time amplifiers? Describe the principle of their operation.
13. Describe the history of the creation, and the operating principle of O-type TWTs.

14. Write down the TWT dispersion equation and analyze the dependence of its roots on the tube parameters.
15. What determines the gain in the low-signal approximation?
16. What determines electronic efficiency of a TWT?
17. What methods of increasing TWT efficiency do you know?
18. What types of SWS are used in the TWT? What is the difference between the TWT on the helix and the TWT on coupled resonators?
19. What are the advantages of hybrid devices?
20. Describe the construction and operating principle of the twistron.
21. Describe the construction of the klystrode. What are its differences from the klystron?
22. Describe the design and operating principle of the orotron.

Chapter 8
M-Type Microwave Electron Devices

8.1 General Characteristics of M-type Devices

Electrons in M-type electronic devices move in crossed, time-constant electric and magnetic fields (there is, of course, also the high-frequency field with which the electron flux interacts). Unlike O-type devices, the constant magnetic field in M-type devices plays an important role in the process of the interaction of electrons with the high-frequency electric field. This role is emphasized by the fact that the name of the devices is derived from the word "magnetic". In M-type devices, electrons transfer not kinetic but potential energy to the electromagnetic field. Their kinetic energy and consequently the motion velocity remain practically constant during the interaction. This feature allows the keeping of synchronism between the wave and the electrons for a long time and provides high electronic efficiency of M-type devices.

The interaction space in M-type devices is formed by the slow-wave structure having two conductors isolated from each other, between which a constant voltage U_0 is applied. The sheet electron beam propagates between these conductors in the magnetic field produced by the external magnetic system and directed perpendicular to the constant electric field.

M-type devices are divided into two groups: devices with a closed electron beam and devices with an open electron beam. In the first group, the electron flux is created by the electron gun separated from the interaction space. The exhausted electrons that do not hit the anode or the cold cathode are deposited on the special collector. Devices with an open electron flux have, as a rule, an open slow-wave structure. Typical representatives of this group include the M-type traveling and backward wave tubes (TWTM and BWOM), as well as dematrons. The devices of this group have a linear or quasi-linear design (i.e., circinate design with a ratio of the inner electrode radius to the outer electrode radius close to 1).

The closed circinate electron beam characterizes the second group. The electrodynamics system of these devices can be either closed or open. Devices with a

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_8

closed SWS include self-oscillators—the magnetron and the mitron. The platinotron and its variations have an open SWS. This group of devices has a cylindrical structure, with the emitting electrode (cathode) being located in the interaction space.

The interaction mechanism in M-type devices is more efficient than in O-type devices, therefore M-type devices, as a rule, have higher efficiency. The high level of intrinsic noise, which is due to the peculiarities of the motion of electrons in the crossed fields, is among the disadvantages of these devices.

8.2 Interaction of Electrons with the High-Frequency Field in M-type Devices

8.2.1 Motion of Electrons in Constant Crossed Fields

Let's consider the motion of an electron in the electric and magnetic fields directed perpendicular to each other (in crossed fields). Figure 8.1 shows the motion of an electron in the space between two flat electrodes, the cathode and the anode. A power supply with a constant voltage U_a is applied to the electrodes. This source creates the electric field $\mathbf{E} = E\mathbf{e}_x$ in the space between the electrodes. The electrode system is placed in the constant magnetic field with induction $\mathbf{B} = B\mathbf{e}_y$. The equations of the motion of an electron in these fields have the form (relativistic effects are not taken into account)

$$m\ddot{x} = -eE - eB\dot{z}, \tag{8.1}$$

$$\ddot{y} = 0, \tag{8.2}$$

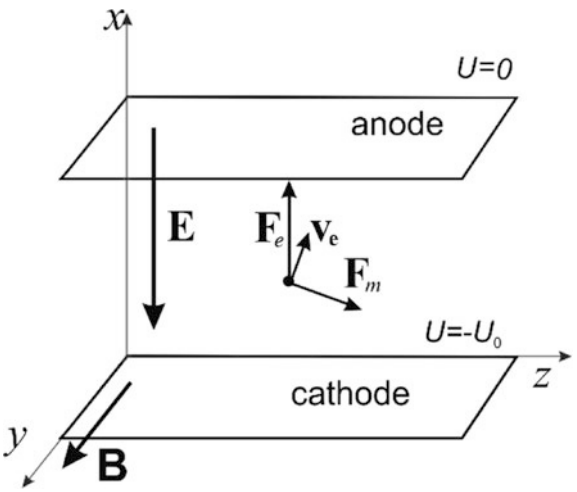


Fig. 8.1 Electron motion in crossed fields

$$m\ddot{z} = eB\dot{x}. \tag{8.3}$$

Let's integrate (8.3):

$$\dot{z} = \frac{eB}{m}x + C_1. \tag{8.4}$$

Suppose that at time $t = 0$, the electron had coordinates (x_0, y_0, z_0) and initial velocity $\mathbf{v}_e(0) = v_{0x}\mathbf{e}_x + v_{0y}\mathbf{e}_y + v_{0z}\mathbf{e}_z$. Then

$$C_1 = v_{0z} - \frac{eB}{m}x_0.$$

Therefore,

$$\dot{z} = \frac{eB}{m}(x - x_0) + v_{0z}. \tag{8.5}$$

Let us substitute formula (8.5) into (8.1):

$$\ddot{x} = -\frac{e}{m}E - \left(\frac{eB}{m}\right)^2(x - x_0) - \frac{eB}{m}v_{0z},$$

or

$$\frac{\mathrm{d}^2(x - x_0)}{\mathrm{d}t^2} + \left(\frac{eB}{m}\right)^2(x - x_0) = -\frac{e}{m}E - \frac{eB}{m}v_{0z}. \tag{8.6}$$

The general solution of the non-homogeneous equation (8.6) can be found as the sum of the general solution of the homogeneous equation and the particular solution of the non-homogeneous equation. The general solution of the homogeneous equation (8.6) (with zero right-hand side) has the form

$$x - x_0 = C_2\mathrm{e}^{\mathrm{j}\omega_c t} + C_3\mathrm{e}^{-\mathrm{j}\omega_c t}, \tag{8.7}$$

where $\omega_c = eB/m$ is the cyclotron (Larmor) frequency. We will find the particular solution of the non-homogeneous equation setting the x coordinate to be constant: $x - x_0 = C_4$. Then

$$C_4 = \frac{1}{\omega_c^2}\left(\frac{e}{m}E - \omega_c v_{0z}\right) = \frac{1}{\omega_c}(v_c - v_{0z}).$$

Thus, the general solution of the non-homogeneous equation (8.7) has the form

$$x - x_0 = C_2 e^{j\omega_c t} + C_3 e^{-j\omega_c t} + \omega_c^{-1}(v_c - v_{0z}), \tag{8.8}$$

The integration constants C_2 and C_3 are found from the initial conditions. At $t = 0$, $x = x_0$. Hence,

$$C_2 + C_3 = -\omega_c^{-1}(v_c - v_{0z}), \tag{8.9}$$

where $v_c = E/B$ is the so-called drift velocity. This name will be explained later.

Differentiating (8.8) and setting t = 0, we obtain

$$C_2 - C_3 = -jv_{0x}/\omega_c. \tag{8.10}$$

Equations (8.9) and (8.10) allow us to find the integration constants:

$$C_2 = -\frac{1}{2\omega_c}(v_c - v_{0z} + jv_{0x}),$$

$$C_3 = -\frac{1}{2\omega_c}(v_c - v_{0z} - jv_{0x}).$$

Substituting these expressions into (8.8), we obtain the general solution (8.6):

$$x - x_0 = \frac{1}{\omega_c}[(v_c - v_{0z})(1 - \cos\omega_c t) - v_{0x}\sin\omega_c t]. \tag{8.11}$$

Substituting this solution into (8.5), integrating the resulting expression and using the initial condition $z(0) = z_0$, we find

$$z - z_0 = \frac{E}{B}t - \frac{1}{\omega_c}[(v_c - v_{0z})\sin\omega_c t + v_{0x}(1 - \cos\omega_c t)]. \tag{8.12}$$

Finally, by integrating (8.2) twice, we find

$$y = y_0 = v_{0y}t. \tag{8.13}$$

Expressions (8.11), (8.12) and (8.13) describe the motion of an electron in the crossed fields under arbitrary initial conditions. Analyzing the solutions obtained, first of all we note that the fields do not influence the electron motion in the direction parallel to the magnetic field. Hence, the electrons keeps coasting in this direction. Therefore, in the following, we will assume that $v_{0y} = y_0 = 0$ and the electron moves in the xOz plane.

Let's consider several particular cases.

1. The electric field is absent, $E = 0$. In this case, from (8.1) and (8.3), we obtain

$$x - x_c = -r_c \sin(\omega_c t + \varphi_0), \quad z - z_c = r_c \cos(\omega_c t + \varphi_0). \tag{8.14}$$

where

$$x_c = x_0 - \frac{v_{z0}}{2\omega_c}, \quad z_c = z_0 + \frac{v_{x0}}{2\omega_c}, \quad r_c = \frac{\sqrt{v_{x0}^2 + v_{z0}^2}}{\omega_c} = \frac{v_c}{\omega_c}, \quad \varphi_0 = \arctan\frac{v_{0x}}{v_{0z}}.$$

This solution describes the motion of the electron along the circle of radius r_c with the center coordinates x_c, z_c, with angular velocity ω_c, and linear velocity v_0. At this, the center of the circle remains stationary. The electron rotates clockwise, if $\omega_c > 0$, and counterclockwise, if $\omega_c < 0$. The sign of ω_c is determined by the direction of the magnetic field.

2. A constant electric field **E** acts on the electron. In this case, the motion of the electron is described by (8.11) and (8.12). This motion can be considered as the rotation along a circle of radius r_c, the center of which moves along the z axis with constant velocity v_c. The linear velocity of electron rotation $v_r = \omega_c r_c$ is usually much larger than v_c. Therefore, velocity v_c can be regarded as the average velocity of the motion of the electron in the time interval $T \gg 2\pi/\omega_c$. The electron seems to drift with velocity v_c along the z-axis, performing rapid rotational motion. Therefore, velocity v_c is called the drift velocity, and the center of the circle along which the electron rotates is the guiding center.

The radius of the circle along which the electron rotates depends on the initial conditions. Thus, if $v_{0x} = 0, v_{0z} = v_c$, it follows from (8.11) and (8.12) that the electron moves rectilinearly along the z axis with a drift velocity, i.e., radius $r_c = 0$.

If the motion of the electron starts at zero velocity ($v_{0x} = v_{0z} = 0$), it will move along the cycloid. This movement is made by a point located on the spoke of a rolling wheel with radius r_c. At $c_{0x} = 0, v_{0z} = v_c$ the point is located on the wheel axis (r = 0) and moves rectilinearly. At ($v_{0x} = v_{0z} = 0$), the point is on the rim of the wheel ($r = r_c$). For $v_{0x} = 0$, $0 < v_{0z} < v_c$, the radius at which the point is located, $0 < r < r_c$, and for $v_{0x} = 0, v_{0z} < 0$ the point is located on radius $r > r_c$. The corresponding trajectories are shown in Fig. 8.2, lines 1, 2, 3 and 4, correspondingly. The electrons also move along the same trajectories in a more general case, when $v_{0x} \neq 0, v_{0z} \neq 0$. All these trajectories are related to trochoidal curves (the cycloid is a special type of trochoid). It is important to note that the center of the circle, along which the electron rotates, in any case moves perpendicular to the direction of the electric field strength, i.e., it is perpendicular to the field force lines.

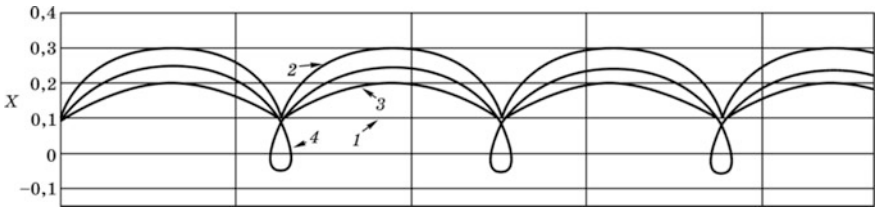


Fig. 8.2 Trochoidal trajectories

8.2.2 *Interaction of Electrons with the Slow Wave*

Suppose that the anode in Fig. 6.1 is made in the form of a periodic structure along which a slow wave of the TM-type propagates with phase velocity v_0. The field of this wave is described by the electric Hertz vector $\mathbf{\Gamma}^e = \Gamma_z \mathbf{e}_z$ which satisfies the homogeneous equation for harmonic time dependence

$$\nabla^2 \dot{\Gamma}_z^e + k^2 \dot{\Gamma}_z^e = 0, \tag{8.15}$$

where $k = \omega/c$ is the wave number in free space. After solving this equation, the electric field is found by formula

$$\dot{\mathbf{E}} = \nabla(\nabla \cdot \dot{\Gamma}^e) + k^2 \dot{\Gamma}^e. \tag{8.16}$$

To solve (8.15) we use the method of variables separation presenting the desired function as a product of two functions: $\dot{\Gamma}_z^e(x, z) = X(x)Z(z)$. Let's represent the wave number in the form of the sum: $k^2 = k_x^2 + \beta_0^2$, where $\beta_0 = \omega/v_0$. v_0 is the wave velocity in the "cold" SWS. Setting the boundary condition $\dot{\Gamma}_z^e(0, z) = 0$, after simple computations, we obtain

$$\dot{\Gamma}_z^e(x, z) = A \sinh(\gamma x) \mathrm{e}^{-\mathrm{j}\beta_0 z}, \tag{8.17}$$

where $\gamma = \sqrt{\beta_0^2 - k^2}$. For large decelerations, $\beta_0 \gg k$, and the second term in (8.16) can be ignored. Introducing the scalar potential $\dot{\Phi} = -\nabla \dot{\Gamma}_z^e$ we find the electric field in the quasi-static approximation: $\dot{\mathbf{E}} = -\nabla \dot{\Phi}$. The guiding centers move perpendicular to the force lines of the field, i.e. along the equipotential surfaces. In accordance with formula (8.17)

$$\dot{\Phi} = U_m \frac{\sinh(\gamma x)}{\sinh(\gamma d)} \mathrm{e}^{-\mathrm{j}\beta_0 z}, \tag{8.18}$$

where U_m is the potential amplitude on the SWS surface. The potential amplitude

$$\Phi = \mathrm{Re}\,\dot{\Phi} = \mathrm{Re}\{-\mathrm{j}\beta_0 A\,\sinh(\gamma x)[\cos(\beta_0 z) - \mathrm{j}\sin(\beta_0 z)]\} = -U_m \frac{\sinh(\gamma x)}{\sinh(\gamma d)}\sin(\beta_0 z).$$

The equipotential lines of this function are shown in Fig. 8.3. The guiding centers of the electron trajectories move along these lines forming so-called spokes. In this case, the equation of the motion of the leading centers in the coordinate system moving with velocity $\mathbf{v} = v_c\mathbf{e}_z$, as follows from (8.1) and (8.3), in the quasi-static approximation takes the form

$$\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{z}'}{\partial z'} = 0,$$

where $z' = z - v_c t$. This equation describes the motion of an incompressible fluid from which it follows that the density of the guiding centers remains unchanged during the interaction, consequently, the space-charge density remains constant.

Let's consider in more detail the interaction of the electrons with the traveling wave field. Figure 8.4 shows the force lines of the high-frequency electric field. In the coordinate system moving with the wave, these lines are motionless. In addition, crossed constant fields $\mathbf{E}_0$ and $\mathbf{B}_0$ are shown. It is assumed that the high-frequency field intensity is much less than the constant field intensity $\mathbf{E}_0$. The electrons moving in rectilinear trajectories with velocity $\mathbf{v}_0 = v_c\mathbf{e}_z$, which is equal to the wave velocity, are injected into these fields.

The electron 1 is in a field which intensity $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_x$ is somewhat larger than $\mathbf{E}_0$ so it shifts forward. The longitudinal velocity of the electron 2 remains constant, since the transverse component of the field in which it is located does not change. This electron remains motionless in the moving coordinate system. The longitudinal velocity of the electron 3 decreases, since it is in a field in which the transverse component of intensity $\mathbf{E} = \mathbf{E}_0 - \mathbf{E}_x$ is less than $\mathbf{E}_0$. Therefore, it shifts backwards.

The longitudinal component of the wave electric field changes the direction of the resultant field vector. Since the electrons move perpendicular to this field, they receive the velocity component directed upward, toward the anode. These electrons

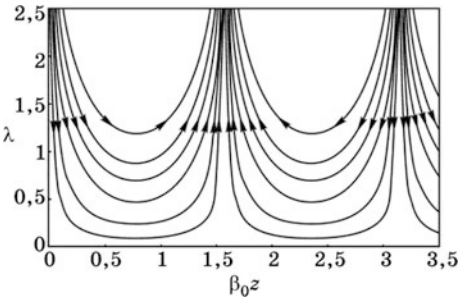


Fig. 8.3 Equipotentials of the quasi-static field in a TWTM

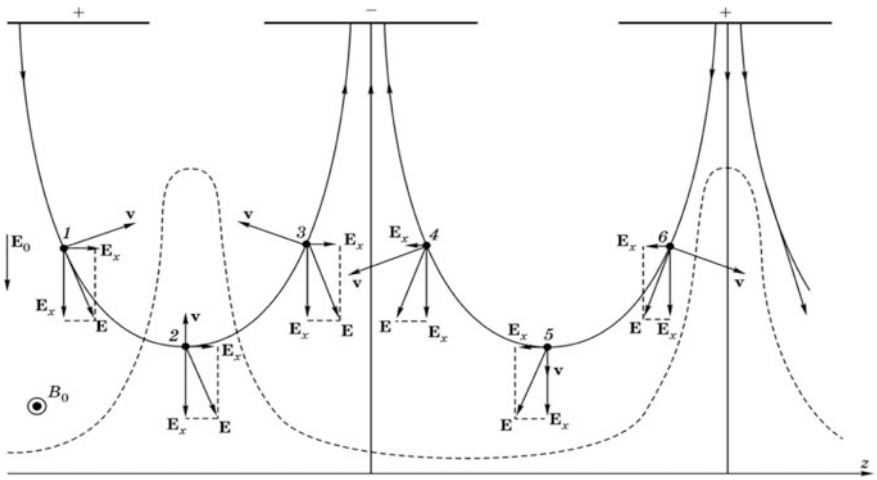


Fig. 8.4 Forces acting on the electrons in the interaction space

form the spoke which is shown in the figure by a dashed line. Moving towards the anode, the electrons in the spoke lose their potential energy transferring it to the wave field.

The electrons 4, 5 and 6, as can be seen in the figure, shift downward increasing their potential energy due to the field energy. However, the total energy balance appears to be in favor of the wave field for two reasons. First, the harmful electrons, while moving toward the cathode, fall into an increasingly weak wave field and their interaction with it becomes weaker, while the useful electrons, rising to the anode, find themselves in an increasingly stronger field and their wave interaction effectiveness increases. Secondly, since the electron beam is injected into the interaction space near the cathode, the harmful electrons quickly hit the cathode, having no time to extract the appreciable part of energy from the field. The useful electrons continue to move to the anode for a long time, transferring a considerable part of their energy to the field. This mechanism is termed "sorting".

In addition to sorting, the interaction effectiveness also increases due to bunching. The electrons are assembled into spokes located in the region of the maximum decelerating field. In addition, although the space-charge density in the spoke, as it will be shown, does not increase, its volume increases and, consequently, the number of "useful" electrons increases.

8.2.3 Linear Interaction Theory in M-type Devices

The interaction of the slow wave with the electron beam is qualitatively considered in the previous section. Here, we will construct the approximate linear interaction

theory for M-type devices, which is necessary for the quantitative assessment of the roles of various factors and for the calculation of the device's parameters. In contrast to the TWTO linear theory, where one-dimensional motion of electrons was assumed, in M-type devices the electron beam changes its shape forming the spokes during the interaction. Therefore, the theory must be at least two-dimensional.

Let's consider the interaction space consisting of the plane negative electrode (cathode) and the positive electrode (anode) located at distance d from each other. We assume that the cathode is under zero potential, and a constant voltage U_0 is applied to the anode. In addition, in the interaction space there is the crossed constant magnetic field with induction $\mathbf{B}_0$. The electron beam of the rectangular cross-section with area $S = hl$ moving parallel to the z axis between the planes $x = x_0$ and $x = x_0 + h$, $h \ll x_0$ is injected into the interaction space.

The anode has the form of the periodic structure along which the slow electromagnetic wave propagates. We'll find the electric field of this wave using the expression for the potential (8.18):

$$\tilde{E}_x = -\frac{\partial \Phi}{\partial x} = \gamma U_m \frac{\cosh(\gamma x_0)}{\sinh(\gamma d)} \mathrm{e}^{\mathrm{j}(\omega t - \Gamma z)}, \tag{8.19}$$

$$\tilde{E}_z = -\frac{\partial \Phi}{\partial z} = -\mathrm{j}\Gamma U_m \frac{\sinh(\gamma x_0)}{\sinh(\gamma d)} \mathrm{e}^{\mathrm{j}(\omega t - \Gamma z)}. \tag{8.20}$$

Here, $\Gamma = \beta - \mathrm{j}\alpha$ is the wave propagation constant taking into account the influence of the electron flux. The unperturbed wave (without the electron flux) has the propagation constant $\Gamma_0 = \beta_0 - \mathrm{j}\alpha_0$.

The equations of the motion of an electron in the interaction space have the form following from expressions (8.6) and (8.3):

$$\ddot{x} = -\frac{e}{m}\left(E_0 + \tilde{E}_x\right) - \omega_c \dot{z}, \tag{8.21}$$

$$\ddot{z} = -\frac{e}{m}\tilde{E}_z + \omega_c \dot{x}. \tag{8.22}$$

Let's set

$$x = x_0 + \tilde{x}, \quad \dot{x} = \tilde{v}_x, \quad z = z_0 + \tilde{z}, \quad \dot{z} = v_c + \tilde{v}_z.$$

In the approximation of the small signal $\tilde{x} \ll x_0$, $\tilde{z} \ll z_0$, $\tilde{v}_z \ll v_c$.
Let the electron enter the interaction space ($z = 0$) at time $t = t_0$. Then

$$t = t_0 + \tau, \quad \frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}\tau}, \quad z_0 = v_c \tau.$$

Substituting expressions for the field (8.19), (8.20) into the equations of motion and dividing the constants and variables, we obtain equations for the variable components of the acceleration:

$$\frac{\mathrm{d}^2\tilde{x}}{\mathrm{d}\tau^2} = \Gamma C \mathrm{e}^{-\mathrm{j}\omega t_0}\mathrm{e}^{-\mathrm{j}(\omega-\Gamma v_c)\tau} - \omega_c \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\tau}, \tag{8.23}$$

$$\frac{\mathrm{d}^2\tilde{z}}{\mathrm{d}\tau^2} = -\mathrm{j}\Gamma D \mathrm{e}^{\mathrm{j}\omega t_0}\mathrm{e}^{\mathrm{j}(\omega-\Gamma v_z)\mathrm{t}} - \omega_c \frac{\mathrm{d}\tilde{x}}{\mathrm{d}t}, \tag{8.24}$$

where

$$C = \frac{e}{m} U_m \frac{\cosh(\gamma x_0)}{\sinh(\gamma d)}, \quad D = \frac{e}{m} U_m \frac{\sinh(\gamma x_0)}{\sinh(\gamma d)}.$$

Let's integrate (8.23):

$$\frac{\mathrm{d}\tilde{x}}{\mathrm{d}\tau} = -\mathrm{j}\frac{\Gamma C}{\omega - \Gamma v_c}\mathrm{e}^{-\mathrm{j}\omega t_0}\mathrm{e}^{-\mathrm{j}(\omega-\Gamma z)\tau} - \omega_c \tilde{z}. \tag{8.25}$$

Substituting (8.25) into (8.24) we'll obtain the second-order differential equation:

$$\frac{\mathrm{d}^2\tilde{z}}{\mathrm{d}\tau^2} = -\mathrm{j}\Gamma\left(D + \frac{C\omega_c}{\omega - \Gamma v_c}\right)\mathrm{e}^{-\mathrm{j}\omega t_0} - \omega_c^2 \tilde{z}.$$

Taking into account the initial conditions, the solution of this equation has the form

$$\tilde{z} = -\frac{\mathrm{j}\Gamma}{\omega^2 - (\omega - \Gamma v_c)^2}\left(D + \frac{C\omega_c}{\omega - \Gamma v_c}\right)\mathrm{e}^{\mathrm{j}(\omega t - \Gamma z)}. \tag{8.26}$$

Performing similar operations with (8.24) and (8.23), we obtain

$$\tilde{x} = -\frac{\Gamma}{\omega^2 - (\omega - \Gamma v_c)^2}\left(C + \frac{D\omega_c}{\omega - \Gamma v_c}\right)\mathrm{e}^{\mathrm{j}(\omega \mathrm{t} - \Gamma z)}. \tag{8.27}$$

The electron velocity is found by differentiating the variable components of the coordinates:

$$\tilde{v}_x = \frac{\mathrm{d}\tilde{x}}{\mathrm{d}\tau} = \mathrm{j}(\omega - \Gamma v_c)\tilde{x}, \quad \tilde{v}_z = \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\tau} = \mathrm{j}(\omega - \Gamma v_c)\tilde{z}. \tag{8.28}$$

Taking into account the fact that for large decelerations $\gamma = \sqrt{\Gamma^2 - \kappa^2} \approx \Gamma$ and setting $\alpha \ll \beta$, we can write

$$|\omega - \Gamma v_c| \approx \left|\omega - \frac{\omega}{v} v_c\right| \ll \omega_c.$$

Using this inequality, we obtain from (8.26) and (8.27) the following approximate expressions:

$$\tilde{x} = \frac{U_m}{U_a} \frac{\Gamma d}{\beta_e - \Gamma} \frac{\sinh(\beta_0 x)}{\sinh(\beta_0 d)} e^{j(\omega t - \Gamma z)}, \tag{8.29}$$

$$\tilde{z} = -\frac{U_m}{U_a} \frac{j\Gamma d}{\beta_e - \Gamma} \frac{\cosh(\beta_0 x)}{\sinh(\beta_0 d)} e^{j(\omega t - \Gamma z)}, \tag{8.30}$$

The expressions obtained show how the motion of electrons varies under the action of the wave field in the SWS. Let's now solve the inverse problem and find the change in the wave propagation constant under the action of the electron flux. The longitudinal component of the beam current can be found as follows:

$$\dot{I} = \rho v_z S - I_0, \tag{8.31}$$

where I_0 is the constant component of the current. The density of space charge is determined from the continuity equation:

$$\nabla(\rho \mathbf{v}) = -\partial\rho/\partial t.$$

Taking into account the fact that $v_{0y} = 0, v_{0z} = v_c$ and not taking into account the products of variables as values of the second order of smallness, we obtain from the continuity equation

$$v_c \frac{\partial \tilde{\rho}}{\partial x} + \rho_0 \left(\frac{\partial \tilde{v}_x}{\partial x} + \frac{\partial \tilde{v}_z}{\partial z}\right) = -\frac{\partial \tilde{\rho}}{\partial t}.$$

Substituting in this equation the values of the velocities from (8.29) and (8.30), we obtain

$$\tilde{\rho} = 0,$$

which confirms the conclusion made earlier that the density of the space charge remains constant during the interaction. The variable component of the current arises not due to the change in charge density but due to the change of its cross-sectional area.

The complex power transmitted by the wave in the absence of the electron flux:

$$\overline{P} = P_0 e^{-2j\Gamma_0 z}$$

The wave power in the presence of the flux will be written in the same way:

$$\overline{P} = P_0 \mathrm{e}^{-2\mathrm{j}\Gamma z}$$

The change of the transmitted power in section $\mathrm{d}z$ is obviously equal to the power of the interaction of the electrons with the wave in this section:

$$\mathrm{d}P_e = \frac{\mathrm{d}(P - \overline{P})}{\mathrm{d}z}\mathrm{d}z = -2\mathrm{j}(\Gamma - \Gamma_0)P. \tag{8.32}$$

On the other hand, the interaction power in section $\mathrm{d}z$ is

$$\mathrm{d}P_e = -\frac{1}{2}\dot{I}\dot{E}_z^*\mathrm{d}z, \tag{8.33}$$

where $\dot{I}$ and $\dot{E}_z$ are the amplitudes of the variable components of the convection current and the electric field.

Equating expressions (8.32) and (8.33) and also taking into account the expression for SWS interaction impedance

$$R_c(x) = \frac{\dot{E}_z(x)\dot{E}_z^*(x)}{2\beta_0^2 P}$$

we find

$$\mathrm{j}(\Gamma - \Gamma_0) = \frac{\beta_0^2 R_c}{\dot{E}_z(x_0)\dot{E}_z^*(x_0)}\frac{\mathrm{d}P_e}{\mathrm{d}z}. \tag{8.34}$$

Since the cross-sectional area of the flux varies, and field E_z depends on coordinate x, in order to calculate $\mathrm{d}P_e/\mathrm{d}z$ by (8.33) it is necessary to integrate with respect to x. Taking into account the fact that the charge density in the beam does not change, $\rho = \rho_0$

$$\frac{\mathrm{d}P_e}{\mathrm{d}z} = \frac{l}{2}\int_{x_0+x_1}^{x_0+h+x_2} \rho_0 v_z(x)\dot{E}_z^*(x)\mathrm{d}x, \tag{8.35}$$

where values x_1, x_2 determine the deviation of the lower and upper beam boundaries from the equilibrium values. To integrate, we must put into (8.35) the values of coordinate $\tilde{x}$ from expression (8.30) and of field E_z from (8.20) and also take into account that $h \ll x_0, x \ll \tilde{x} \ll x_0, v_z = v_c$. Integrating and eliminating the terms of the second and higher order of smallness from the result, we obtain after a series of computations

$$-(\Gamma - \Gamma_0) = \frac{\beta_0^2 R_c I_0}{U_m} \frac{\beta_0 d \coth(\beta_0 x_0)}{\beta_e - \Gamma} = \frac{\alpha_m^2}{\beta_e - \Gamma}, \tag{8.36}$$

where

$$\alpha_m^2 = \frac{\beta_0^2 R_c I_0}{U_a} \beta_0 d \coth(\beta_0 x_0). \tag{8.37}$$

Equation (8.36) is the quadratic equation with respect to Γ, consequently, two waves can propagate in the slow-wave structure with the electron flux. To determine the propagation constants of these waves, we substitute $\Gamma = \beta - j\alpha$ into (8.36):

$$\beta - \beta_0 - j(\alpha - \alpha_0) = -\frac{(\beta_e - \beta) + j\alpha}{(\beta_e - \beta)^2 + \alpha^2} \alpha_m^2.$$

Let's divide the real and imaginary parts of this expression:

$$(\beta - \beta_0)\left[\alpha^2 + (\beta_e - \beta)^2\right] = (\beta_e - \beta)\alpha_m^2, \tag{8.38}$$

$$(\alpha - \alpha_0)\left[\alpha^2 + (\beta_e - \beta)^2\right] = \alpha\alpha_m^2. \tag{8.39}$$

In analyzing these equations, we note first of all that for $\beta = \beta_e$, i.e., when the velocity of the electrons and the phase velocity of the perturbed wave are equal, it follows from (8.38) that $\beta = \beta_0$. This means that the perturbed wave in this case has the same phase velocity as the unperturbed wave. With this, (8.39) takes the form

$$(\alpha^2 - \alpha\alpha_0) = \alpha_m^2,$$

from which

$$\alpha_{1,2} = \frac{1}{2}\alpha_0 \pm \sqrt{\alpha_m^2 + 0.25\alpha_0^2}. \tag{8.40}$$

It follows from this formula, that for any value of attenuation constant in the cold SWS, there is always one increasing wave for which $\alpha_1 < 0$; the second wave then dies out ($\alpha_2 > 0$).

At $\beta_e \neq \beta_0$ in the system, both active and reactive power is exchanged, therefore the velocities of the perturbed and unperturbed waves in the SWS do not coincide. In particular, if we ignore the attenuation in the cold SWS, i.e., we set $\alpha_0 = 0$, then from (8.39) it follows that $\alpha^2 + (\beta_e - \beta)^2 = \alpha_m^2$. Substituting this expression into (8.38), we find

$$\beta = 0.5(\beta_e + \beta_0) \tag{8.41}$$

Thus, in this case the phase constants of both waves are the same and are equal to the average arithmetic of the electronic phase constant and the phase constant of the unperturbed wave in the SWS.

Substituting (8.41) and $\alpha_0 = 0$ into (8.39), we obtain

$$\alpha_{1,2} = \mp\alpha_m \sqrt{1 - \left(\frac{\beta_e - \beta_0}{2\alpha_m}\right)^2}. \tag{8.42}$$

If the difference between the velocities of the wave and the electrons is large, $|\beta_e - \beta_0| > \alpha_m$, then under the root there is an imaginary number, i.e., $\alpha_{1,2} = 0$ and the amplification of the waves is absent.

At lower values of this difference, the increasing wave is present regardless of whether the velocity of the electrons is larger or smaller than the velocity of the unperturbed wave. The physical explanation for this effect is that the electrons transmit to the wave their potential energy rather than their kinetic energy. The dependence of the propagation constants of the two waves on the difference in the phase constants is shown in Fig. 8.5. The described interaction theory does not take into account the effect of the space charge, which plays less of leading role in M-type devices than in O-type devices since the charge density remains constant.

Nevertheless, at a larger space charge density, its influence on the motion of the electrons must be taken into account.

The linear theory is valid for all M-type devices, but for many of them, especially for self-oscillators, the nonlinear theory (the large signal theory) has more significance. The elements of this theory are presented with the description of the specific types of device.

The individual radiation of electrons in M-type devices is of a complex nature. Since the amplification occurs even at the exact equality of the velocities of the wave and the electrons, Cherenkov radiation is absent in this case, or plays a secondary role. The slowing-down radiation occurs when the electrons change the direction of their motion forming the spokes (although the absolute value of the

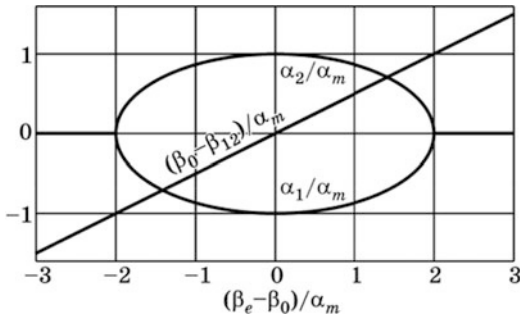


Fig. 8.5 The roots of the TWTM characteristic equation

velocity remains virtually unchanged). The main role plays the transition radiation arising when the electrons approach the anode surface (of the slow-wave structure). This radiation is induced, since it occurs in the SWS field created by the electrons themselves. In addition, the electrons moving in circular orbits create high-frequency oscillator radiation. Since the SWS of the device has no appreciable interaction impedance at this radiation frequency, it is spontaneous, and therefore has very low power. This radiation is usually considered as one of the types of noise in M-type devices.

8.3 M-type Devices with an Open Electron Beam

8.3.1 The Traveling-Wave Tube of M-type

In 1950, a group of French scientists led by R. Warnecke created an M-type TWT. The scheme of this device is shown in Fig. 8.6.

The electron gun consisting of the cathode and the control electrode creates a rectilinear sheet electron beam. This beam enters the interaction space between the cold cathode and the slow-wave structure where the crossed electric and magnetic fields $\mathbf{E}_0$, $\mathbf{B}_0$ exist. The amplified signal is applied to the SWS input which excites the slow wave in the SWS.

The electron drift velocity $v_c = E_0/B$ must be equal to the phase velocity of the operating spatial harmonic in the SWS. Under the high-frequency field action, the electron flux changes its shape, and spokes appear in it.

The electrons in the spokes move to the SWS surface transmitting their potential energy to the field. The amplitude of the wave increases, which leads to a growth of the spokes and this, in turn, increases the intensity of the energy transfer from the electrons in the spokes to the electromagnetic field (Fig. 8.7). The amplified wave is extructed from the tube by means of the output device. The exhausted electrons move to the collector. The localized absorber is necessary to eliminate the feedback that arises from the non-ideal matching of the input and output devices.

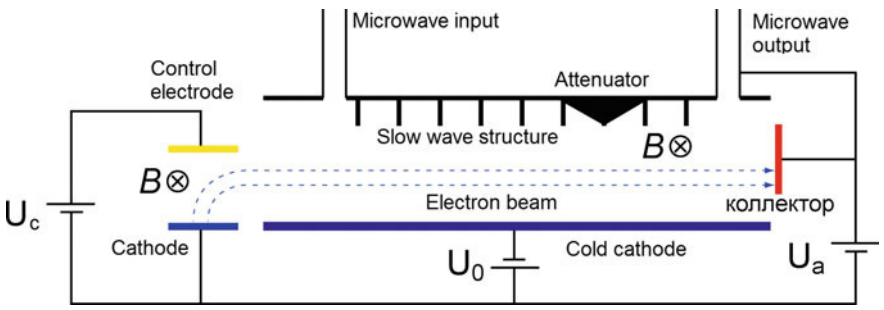


Fig. 8.6 Scheme of a TWTM

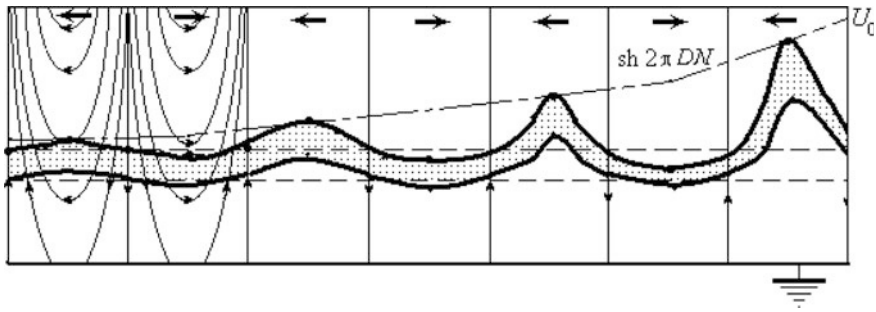


Fig. 8.7 Electron beam shape in a TWTM

The linear theory of TWTM operation is considered in the previous section. The gain in the weak signal mode can be found by the formula

$$\mu = A + D - 8.68\alpha_1 L, \tag{8.43}$$

where L is the length of the interaction space; A is the part of the input power accounting for the increasing wave; and D is the gain loss due to the localized absorber. In the optimal mode, $\beta_e = \beta_0, \alpha_1 = -\alpha_m, A = -6$ dB, $D \approx -6$ dB. In this case

$$\mu = -12 + 54.5 C_m N, \tag{8.44}$$

where N is the number of wavelengths that fit along the SWS length,

$$Cm = \frac{\alpha_m}{\beta_0} = \sqrt{\frac{R_c I_0}{U_a} \beta_0 d \coth(\beta_0 x_0)}$$

is the quantity that plays the role of the gain parameter.

In real devices, the gain is 20–30 dB. By this parameter, the TWTM is inferior to the traveling-wave tubes of O-type and multi-cavity klystrons.

Let's consider the construction of the main parts of the TWTM. The *electron gun* produces the electron flux and injects it into the interaction space. The principle of the gun's operation and its construction are described in Appendix B. The *slow-wave structure* in the TWTM has a flat construction, which is necessary for interaction with the sheet electron beam. It consists of a smooth electrode (the cold cathode) and a periodic structure parallel to it. In the latter, a comb structure and associated modifications are often used. The electron beam interacts with the working spatial harmonic of this structure.

For the comb SWS, the zero spatial harmonic with positive dispersion and high interaction impedance is usually chosen. Input and output couplers are located at the beginning and end of the SWS, and in the middle of it there is a localized absorber (attenuator). The *collector* is designed to collect the exhausted electrons.

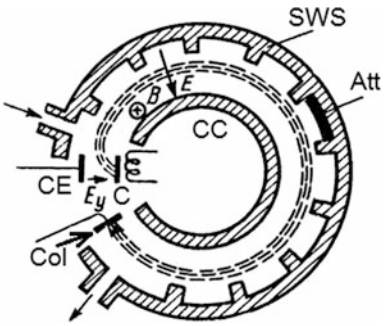


Fig. 8.8 Rolled up TWTM. *C* Cathode, *CC* cold cathode, *CE* control electrode, *Att* attenuator, *Col* collector

In addition, the TWTM has a magnetic system usually consisting of permanent magnets and pole tips. To reduce the weight and dimensions of the magnetic system, the SWS is often folded into a ring (Fig. 8.8).

Let's estimate the efficiency of the TWTM using the following assumptions. The electrons are injected into the interaction space along optimal trajectories and move rectilinearly there (in the absence of the high-frequency field). The electron beam is thin and transits at distance x_0 from the cold cathode having a zero potential. All electrons deposit on the collector with potential U_0. The accelerating voltage U_0 determines the initial energy of the electron at the entrance to the interaction space:

$$W_1 = eU_0.$$

At the moment of striking an electrode, the electron only has kinetic energy

$$W_2 = \frac{mv_c^2}{2} = eU_a \frac{x_0}{d} \tag{8.45}$$

It is taken into account in this expression that the drift velocity of the electron is determined by the equivalent constant voltage, i.e., by the potential at the electron's entry point. Thus, in the process of interaction, the electron transfers the energy $W_i = W_1 - W_2$. The electronic efficiency is defined as the ratio of the energy transferred to the energy that the electron possessed at the moment of entering the interaction space:

$$\eta_e = \frac{W_l}{W_1} = 1 - \frac{W_2}{W_1} = 1 - \frac{x_0}{d}. \tag{8.46}$$

Efficiency increases with decreasing distance from the electron beam to the cold cathode, but the tube gain also decreases, since the SWS interaction impedance decreases with x_0 decreasing tending to zero at $x_0 \rightarrow 0$.

Voltage U_a determines the velocity of electrons, and this must be equal to the phase velocity of the wave in the SWS. Replacing v_c by v_p in (8.45) and substituting the result into (8.46), we find

$$\eta_e = 1 - \frac{mc^2}{2eU_a}\left(\frac{v_p}{c}\right)^2 = 1 - \frac{25.6 \cdot 10^4}{U_a}\left(\frac{1}{n}\right)^2,$$

where $n = c/v_p$ is the deceleration of the phase velocity of the working spatial harmonic. As can be seen, with increasing deceleration, the tube efficiency also increases. However, it is necessary to increase the magnetic induction, since

$$n = \frac{c}{v_p} = \frac{c}{v_c} = \frac{cdB_0}{U_a}.$$

In practice it is not possible to fulfill the ideal conditions for the introduction of electrons into the interaction space, and, therefore, they also produce rotational motion in addition to translational motion. If the electron hits the SWS at the apex of its cycloidal trajectory, its velocity is two times higher than the translational velocity and the energy is four times higher. Accordingly, formula (8.46) must be written in the form

$$\eta_e = 1 - 4x_0/d.$$

Different beam electrons get to the anode at different points of the trajectory, so the electronic efficiency of the device must be determined by the formula

$$\eta_e = 1 - k_e\frac{x_0}{d},$$

where coefficient k takes values from 1 to 4. In practice, the total efficiency of the TWTM is 40–60%. Table 8.1 shows the parameters of some TWTM.

From the table it follows that according to efficiency and bandwidth, the TWTM competes favorably with O-type devices. The small supply voltage and dimensions should also be added to the advantages of these devices.

The contradiction between the large gain and high efficiency can be eliminated by changing the height of the interaction space. Usually the step change of this parameter is used. The circuit of such a two-cascade TWTM amplifier is shown in Fig. 8.9. The first section provides the amplification of the signal, since the flux passes close enough to the SWS surface, which is matched at the collector end. The second section, matched at the cathode end, is excited by the flux bunched in the first section and provides high efficiency, since the beam enters it close to the cold cathode. The two SWS sections are not connected to each other, so there is no

Table 8.1 Parameters of the high power TWTM

Device type	Average wave length (cm)	Output power (kW)	Bandwidth (%)	Gain (dB)	Efficiency (%)	Operation mode
L-3913	3.75	1.5	20	23	30	Continuous
L-4215	15	8000	17	15	60	Pulsed

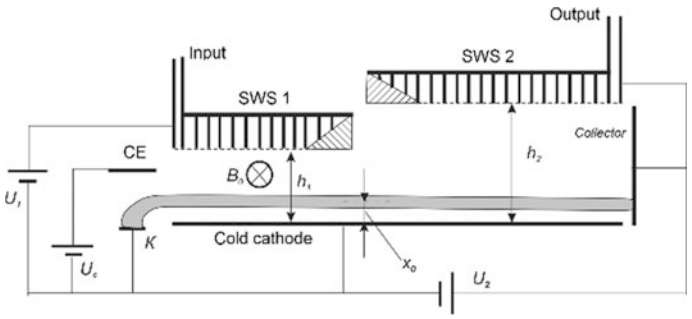


Fig. 8.9 A two-cascade TWTM

danger of self-excitation in the device. Such a two-stage amplifier has a gain of 20 dB more than a conventionally designed TWTM with a localized absorber.

The dematron. In 1962, J.F. Hull and his colleagues invented an M-type device which they called the "dematron" (Distributed Emission Magnetron Amplifier + 'tron'). The dematron can be considered as a TWTM modification where the part of the cold cathode length is covered by the emission surface. Usually, secondary emission coatings are used for this. Unlike the TWTM, in dematrons there is no need to support the fixed wave phase velocity in the SWS, since at the cathode there are always the electrons emerging in the necessary phase of the field. Therefore, dematrons have a wide instantaneous band of amplified frequencies. However, the rate of wave increase in a dematron is lower than in a TWT and to obtain the required amplification it is necessary to use a longer interaction space. When a signal is applied to the input of the dematron, the "random" electrons in the interaction space return to the cathode under the action of the signal field and expell the secondary electrons out of it. If the signal is strong enough, the secondary emission coefficient for most electrons will be larger than one, and an electron flux will appear in the dematron, which will transfer its energy to the wave. The process of the flux forming takes several seconds. When the input signal is turned off, the electron flux diffuses over several seconds and the output power becomes equal to zero. Thus the dematron can be controlled by the input signal, i.e., it can operate with a non-modulated power supply which significantly simplifies the transmitter circuit.

Dematrons are used in radar and electronic warfare systems in transmitters with a high pulse repetition frequency. In the decimeter band, the output power of dematrons reaches 1 MW, the gain is 10–15 dB, the bandwidth is 10%, and the efficiency is 40–60%.

8.3.2 The M-type Backward-Wave Oscillator

The M-type backward-wave oscillator was created by B. Epstein, in France, in 1952. Since the energy transfer in the SWS is carried out "backwards", i.e. from the

collector to the cathode, Epstein called his tube "the carsinotron" from the Greek word "karcinox"—the crawfish that moves backward. Later, this name was sometimes used also to denote O-type BWOs. According to its operating principles, an M-type BWO does not differ from an O-type BWO. It also uses the SWS spatial harmonic with negative dispersion. The linear theory of interaction described in Sect. 8.3 is also valid for the BWOM but the presence of internal feedback makes it possible to use this device as a self-oscillator. In contrast to a BWOO, backward-wave oscillators with crossed fields have a higher efficiency, which is due to the constancy of the drift velocity of electrons in the interaction space and the more efficient mechanism of energy extraction, which includes both bunching and sorting of electrons. Just like the TWTM, backward-wave oscillators are folded into a ring to reduce the weight and dimensions of the magnetic system. The band of electrical tuning of BWOM frequency is smaller than that for an analogous O-type device, since the SWSs used in the BWOM have stronger dispersion. M-type backward-wave oscillators have sufficiently large starting currents.

The BWOM scheme is shown in Fig. 8.10, and the parameters of some devices are given in Table 8.2.

In order that the width of the generated signal spectrum is small, the operating current of the device must be several times higher than the starting value, otherwise, parasitic oscillation modes having other frequencies are excited in the BWOM in

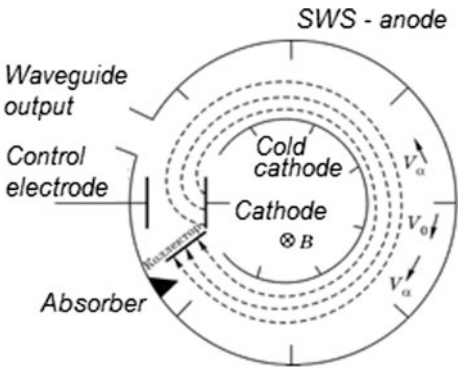


Fig. 8.10 Scheme of an M-type BWO

Table 8.2 Parameters of some BWOM

Device type	Waveband (GHz)	Output power (kW)	Anode voltage (kV)	Anode current (A)	Efficiency (%)	Weight (kg)
CMA 1241	0.8–1.06	0.9	6	0.6	25	14
L3724 A	2.5–3.55	0.23	5.2	0.35	43	7.3
CMA 1290	15.4–17.7	0.1	4.2	0.3	8	3.7

addition to the operating one. However, with any mode, the noise of these devices is much larger than for O-type generator devices.

8.4 M-type Devices with a Re-entrant Beam

8.4.1 The Multi-cavity Magnetron

M-type microwave electron devices have a long history. In 1921, A. Hall (USA) published an article in which he described the mechanism for controlling the anode current in the cylindrical diode with the aid of a magnetic field perpendicular to the electric field. The magnetic field changes the direction of the movement of electrons from the cathode to the anode and upon reaching the critical value of this field, the electrons stop falling on the anode and return back to the cathode. The anode current diminishes. Thus, Hall described the phenomenon of anode current cutoff under constant magnetic field action. However, Hall did not aim to generate high-frequency oscillations with the device.

Later, the split-anode magnetron was proposed which generated oscillations in the decimeter and centimeter bands. However, the output power of these devices and their efficiency remained low. The breakthrough appeared in 1939, when employees of the Leningrad Research Institute-9 N.F. Alekseev and E.D. Malyarov created a multi-cavity magnetron. In 1940, they published an article in which they described the multi-cavity magnetron with water-cooling and an output power of 300 W at a wavelength of 9 cm. Unfortunately, with the beginning of the war, this work was stopped.

Simultaneously, work on creating a powerful generator of centimeter waves was being conducted in the UK. These works took on special importance since to combat German night aviation attacks, it was necessary to create aircraft radars with greater power and accuracy than existed at that time. In late 1939, success came from H.A. Boot and J.T. Randall from Birmingham University, when they created a multi-cavity magnetron capable of generating oscillations with the power of 10 kW at a wavelength of 10 cm. This power was two orders of magnitude higher than that of all the generators then available. In 1940, the British gave the test model of the magnetron to the Bell company laboratory in the United States, where this device was updated and its mass-production began. The use of the magnetron in aircraft, ship and ground radars made a great contribution to the Allied victory in the Second World War. No wonder the commander of the US Pacific forces during the war General MacArthur said: “The atomic bomb ended the war but the radar won it.” After the war, the magnetron relinquished its place in radar to other devices that generated signals of better quality, but it continues to be used in simple radars, as well as in microwave ovens and industrial heating units.

The multi-cavity magnetron has a very simple design. In contrast to the M-type devices considered earlier, the multi-cavity magnetron (hereinafter simply “magnetron”) has a closed electrodynamics system and re-entrant electron flux, the cathode (emitter) being placed in the interaction space. Essentially, the magnetron is

a cylindrical diode placed in the magnetic field. The anode of this diode has a complex shape, which ensures the interaction of the high-frequency electromagnetic field with the electron flux.

The design of the multi-cavity magnetron is shown schematically in Fig. 8.11. It includes a thermo-cathode 5 and anode block 1 made in the form of a massive copper cylinder with cavities 2. The cavities play the role of resonators in which the electromagnetic field is concentrated. The inductive loop 4 located in one of the resonators and the coaxial line 3 serve to extract the high-frequency energy from the device. The figure also shows the cathode arms isolated from the anode block. The upper and lower covers that close the anode block ends and seal the magnetron internal chamber as well as the magnetic system that creates the magnetic field directed along the magnetron axis are not shown.

The static mode of the magnetron operation. Let's consider the motion of electrons in the magnetron in the absence of a high-frequency field. Usually, the ratio of the anode radius r_a to the cathode radius $\boldsymbol{r}_{\text{cat}}$ is not too large, and in the first approximation the cylindrical structure can be considered as flat with the distance between the anode and cathode $d = r_a - r_{\text{cat}}$. In this case, we can use all the results obtained in Sect. 8.1. We assume that electrons leave the cathode at zero velocities, hence in crossed fields they move along cycloids with radius $r_c = v_c/\omega_c$. If this radius is less than $d/2$, the electrons do not reach the anode and return to the cathode. At this, the anode current is equal to zero. The value of magnetic induction corresponding to $2r_c = d$ is called the cutoff inductance B_c and the anode current termination at $B > B_c$ is called the cutoff phenomenon. Since $r_c = v_c/\omega_c$ and $\omega_c = eB/m$ we obtain the formula for critical induction

$$B_c = \frac{1}{d}\sqrt{\frac{2m}{e}U_a}. \tag{8.47}$$

In a cylindrical magnetron, the electron trajectories have the form of a cardioid and the critical value of the magnetic induction is specified by the formula

$$B_c = \frac{\sqrt{8mU_a/e}}{r_a(1 - r_{\text{cat}}^2/r_a^2)}. \tag{8.48}$$

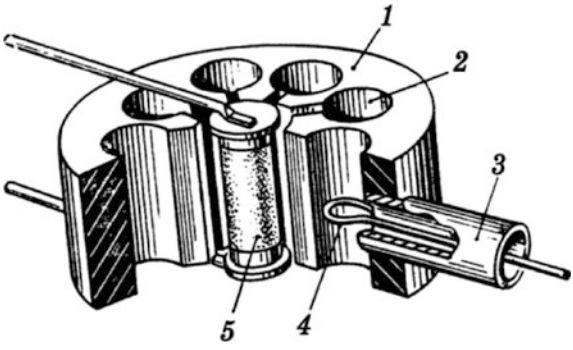


Fig. 8.11 Magnetron design

For $r_a/r_{cat} \approx 1$ this formula coincides with (8.47). It was proven that expressions (8.47) and (8.48) are strict, in spite of the fact that the space charge was not taken into account in their development.

Figure 8.12 shows the trajectories of electrons in the magnetron at various values of magnetic field induction. The trajectory 1 corresponds to the absence of a magnetic field. Electrons transit along straight lines from the cathode to the anode. When a magnetic field appears, the electron trajectories deviate but they still reach the anode (trajectories 2 and 3). Trajectory 4 corresponds to the critical value of magnetic field induction. The electron touches the anode at the top of the curve. Trajectory 5 corresponds to magnetic induction which is larger than the critical one. The electrons do not enter the anode; the anode current is equal to zero. This is the anode current cutoff mode.

Figure 8.13a shows the dependence of the magnetron anode current on magnetic field induction, and Fig. 8.13b shows the dependence of the critical anode voltage on critical induction, constructed from formula (8.48) (the critical-mode parabola). This parabola separates the area where all the electrons emitted by the cathode get to the anode from the cutoff area where the anode current is absent.

In the cutoff mode, in the absence of the high-frequency field, the electrons do not get to the anode and form an electronic "hub" around the cathode. The outer radius of this hub can be found from the balance of the electron's kinetic and potential energy at the hub boundary:

$$\frac{1}{2}mr^2\omega_e^2 = eU_1,$$

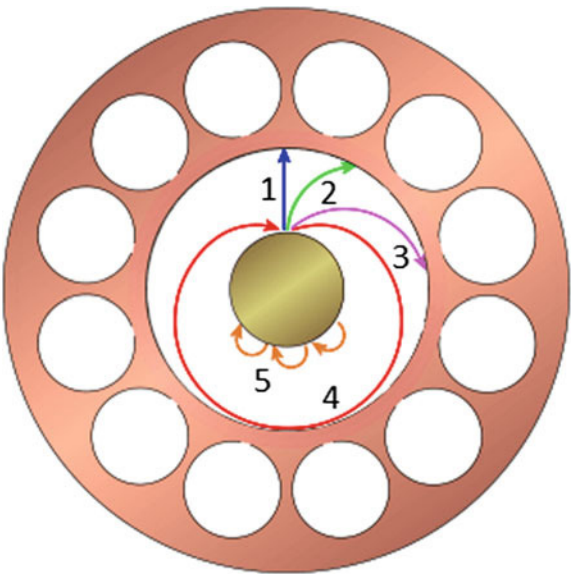


Fig. 8.12 Electron trajectories in magnetron in the static regime

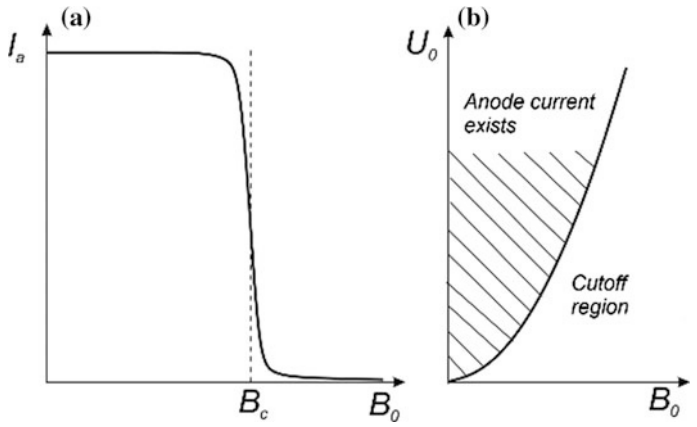


Fig. 8.13 Dependence of magnetron anode current on magnetic field induction (**a**). Cutoff parabola (**b**)

where ω_e is the electron angular velocity, and U_1 is the potential at the linear boundary. From here, after a series of intermediate calculations, we find the hub radius:

$$r_h = \frac{r_{\text{cat}}}{\sqrt{41 - 4m\omega_e/(enB)}}. \tag{8.49}$$

The electrodynamics system of the magnetron is the periodic structure in which slow waves can propagate (not taking into account the energy output). The dispersive characteristic of this structure is determined by its configuration. In magnetrons, a comb-type SWS with slots of various shapes is used. The most common are slots of the following types: hole-and-slot (Fig. 8.14a), slit-type (Fig. 8.14b), and vane-type (Fig. 8.14c). Figure 8.14d shows the rising-sun structure in which resonators of different sizes interchange. In such systems it is possible to use different types of slots, for example hole-and-slot and slit-type ones. At frequencies below 10 GHz, hole-and-slot or vane slots are most often used. At higher frequencies, slit-type slots are used.

Each slot can be considered as a resonator. These resonators are coupled by common magnetic and electric fields. Therefore, the traveling waves can propagate in the magnetron electrodynamics system. A typical dispersive characteristic of these waves is shown in Fig. 8.15.

The basic type of wave has direct dispersion, which is a characteristic of a comb-type SWS. Since the SWS is closed, it forms a ring resonator in which the resonances are observed when the integer number of wavelengths is laid on the ring length. Assuming the ring length as the mean radius of the interaction space, this condition is written in the form

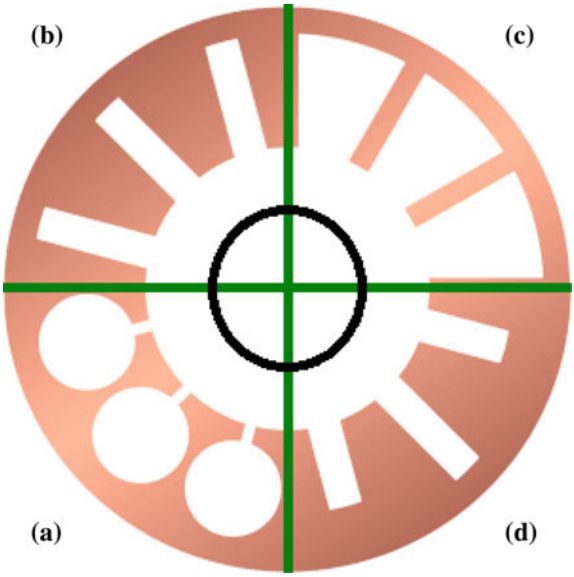


Fig. 8.14 Types of magnetron cavities

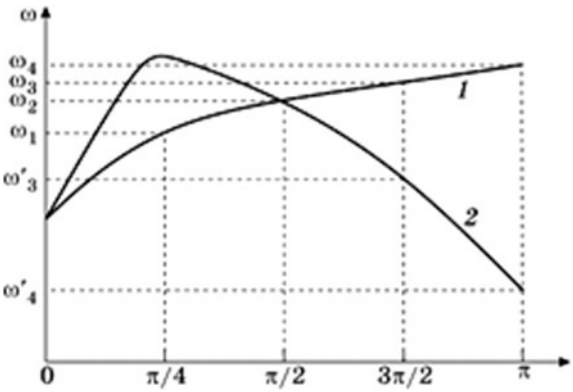


Fig. 8.15 Dispersion chart of the SWS of a magnetron

$$NL = n\lambda_g, \quad n = 0, \pm 1, \pm 2, \ldots,$$

where L is the SWS period, and N is the number of periods. Introducing the phase shift for the period $\varphi = 2\pi L/\lambda_g$ the resonance condition can be written in the form

$$\varphi_n = \frac{2\pi n}{N}, \quad n = 0, \pm 1, \pm 2, \ldots \tag{8.50}$$

Oscillations occur in the same phase in all periods for n=0. These oscillations are usually called "zero-type". Oscillations in neighboring periods occur in anti-phase for $n = N/2$, supposing even number of resonators. These oscillations are

called "π-type" and it is this type that is used as the operation mode of oscillations in the magnetron.

Therefore, in the magnetron oscillatory system there is always an even number of periods. The structure of the π-type oscillation high-frequency field is shown in Fig. 8.16. In the quasi-static approximation, we can assume that the given segment of the anode block is positively charged and the neighboring segments are negative.

Separation of the frequencies of the operation mode and the mode closest to it with respect to the oscillation frequency is of major importance for the stable operation of the magnetron and is determined by formula

$$\delta f = \left| \frac{f_{\text{work}} - f_{\text{par}}}{f_{\text{work}}} \right|.$$

for the magnetron stable operation it is necessary that the separation of frequencies between the operation mode of oscillation and the neighboring type is sufficiently large (usually not less than 10%). In a conventional comb-type or slot-and-hole resonators system, such separation cannot be obtained with the number of periods $N \geq 8$ (Fig. 8.15, curve 1). This, in particular, was the cause of the unstable operation of the first magnetrons.

In modern magnetrons, the necessary frequency separation is achieved in several ways. Historically, the first method proposed back in 1941 was the use of metal conductors (straps) connecting the segments of the anode unit to the next but one. The straps can be located on one side or on both sides of the anode unit, and be single or double. Figure 8.17 shows the anode unit with double bilateral straps. The straps form the capacitance with the segment above which they pass, and the inductance appears due to currents going along the strap.

These capacitances and inductances are connected in parallel with the equivalent capacitance and inductance of the slot (cavity).

In the zero mode, all segments of the anode unit have the same potential, so the current does not flow through the straps and the equivalent inductance of the straps is equal to zero. The difference of the potentials between the strap and the segment located below is equal to zero, so the equivalent capacity of the straps is equal to

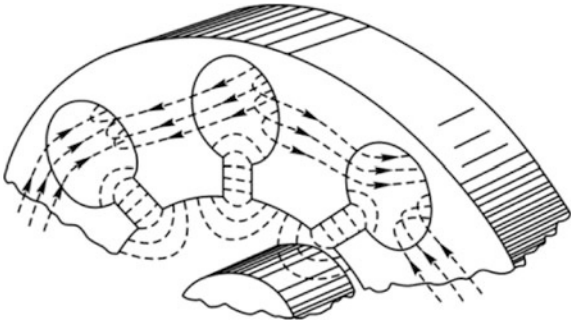


Fig. 8.16 Structure of the electro-magnetic field of π-type oscillations

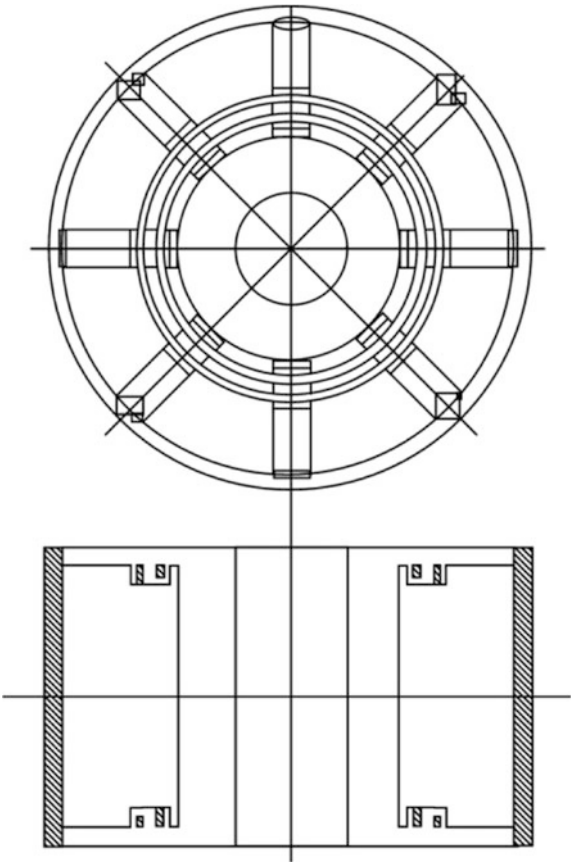


Fig. 8.17 The magnetron anode unit with double bilateral straps

zero. It follows that the frequency of zero-type oscillations with the introduction of the straps is practically unchanged.

In the π-mode, the straps connect segments with the same potential, the current does not flow through the straps, and their equivalent inductance is equal to zero. However, the strap passes over segments that have the opposite potential sign, and a capacitance arises between these segments and the strap, which is added to the capacity of the slots. As a result, the resonant frequency of the π-mode decreases. In other modes, both the capacitance and the inductance of the straps are included, therefore their resonant frequencies can both increase and decrease. As a result, the dispersive characteristic acquires the form shown in Fig. 8.15, curve 2. As can be seen, the dispersion curve with the straps is negative near the π-mode, and the separation of the π-mode frequency and the neighboring frequencies increases significantly in comparison with the system without straps. Frequency separation can reach up to 60% in short anode units with double two-sided straps. This value decreases as the height of the anode unit increases. In addition, the straps introduce additional losses into the electrodynamics system of the resonator, reducing its

Q-factor. The production of anode units with straps for frequencies of more than 10 GHz is associated with considerable technological difficulties due to the small size of the straps.

Another method of separating frequencies is based on the use of slots of different sizes or shapes—different-resonator anode units. The most widely used is the design with slit-type slots, shown in Fig. 8.14d, which is called "the rising sun". The sum of the large and small resonator lengths in such a unit should be approximately equal to half of the wavelength of the oscillations. In the dispersive characteristic of such an SWS, a hop is observed, the frequency of the π-type oscillations being located in the middle of this hop, thereby increasing frequency separation. Figure 8.18 shows the dispersive characteristic of a different-resonator unit containing 18 cavities.

The rising sun units are widely used in magnetrons operating at frequencies above 10 GHz. Their disadvantage is the decrease of the high-frequency field strength in the interaction space, since the maximum field shifts into the interior of the large resonator. In addition, in the different-resonator unit, along with the operation type, the zero-mode is excited, the intensity of which increases as the ratio of the lengths of the large and small resonators increases. The reason for this is that in the different-resonator unit in the π-mode, the current flows in the same direction along the surface of all resonator segments (Fig. 8.19). The field of this mode worsens the field structure of the operating π-mode, and reduces the efficiency of the magnetron. The ratio of the lengths of the large and small resonators is chosen within 1.8–2 and this ensures frequency separation up to 10%.

Figure 8.20 shows a magnetron anode unit with resonators of the hole-and-slot type and double straps (a) and with the different-resonator unit with slit-type resonators (b).

Self-excitation conditions. The axial construction of the magnetron and the cathode located in the interaction space complicate the mechanism of the interaction of electrons with the wave in comparison with the TWTM and the BWOM. In particular, since the constant electric field of the anode is non-uniform and decreases from the cathode to the anode, the synchronism conditions cannot be satisfied throughout the interaction space. Typically, the anode voltage is selected in

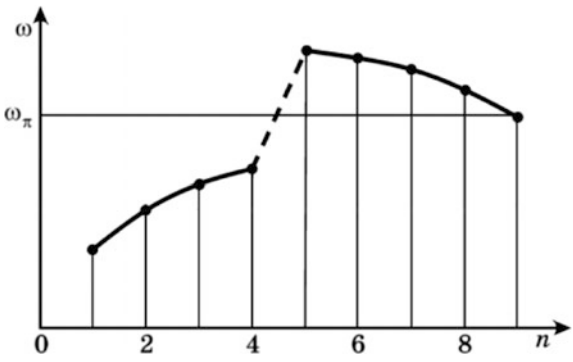


Fig. 8.18 Different-resonator unit dispersive characteristic

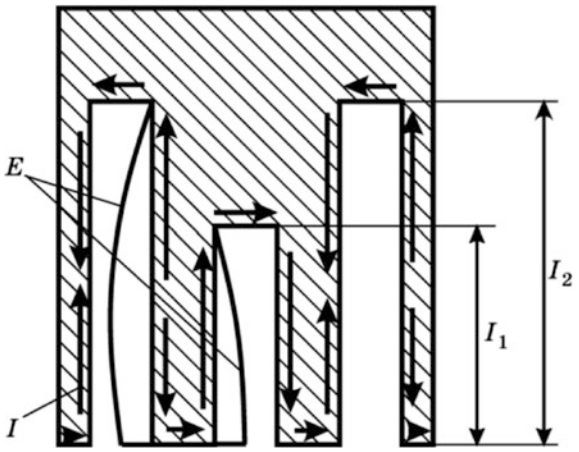


Fig. 8.19 Zero-mode excitation in the different-resonator unit

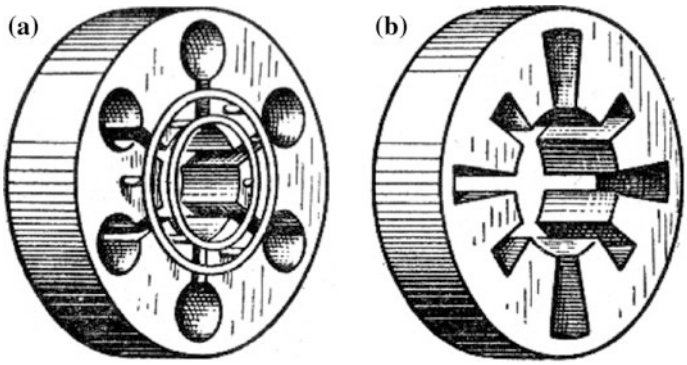


Fig. 8.20 Magnetron anode units with double straps (**a**) and rising sun type (**b**)

such a way that synchronism is ensured on the average radius of the interaction space. The average phase velocity of the fundamental (zero) spatial harmonic:

$$v_p = \frac{\boldsymbol{r}_a + \boldsymbol{r}_{\text{cat}}}{2}\frac{\omega}{n},$$

where n is the number of the oscillation mode. Equating this velocity to the guiding center velocity, we obtain

$$U_a = \frac{\omega n(\boldsymbol{r}_a^2 - \boldsymbol{r}_{\text{cat}}^2)}{2n}B. \tag{8.51}$$

This equation on the plane (B_0, U_0) defines the straight line on which the conditions of synchronism are satisfied for the given mode of oscillations and

magnetic induction. Below this straight line, the synchronism between the wave and the motion of electrons is absent and self-excitation is impossible. Above this straight line, the self-excitation of the device and the appearance of an anode current are possible. Therefore, it is said that (8.51) defines the threshold line. The slope of this straight line depends on the mode of oscillations. Figure 8.21a shows the threshold lines for different modes in the 8-cavity magnetron and the cutoff parabola. Generation in a given mode is possible above the threshold line but below the cutoff parabola. The smallest voltage for the given magnetic field is required to excite the π-type of oscillations, so it excites first when the anode voltage pulse is applied, and suppresses other types of oscillations.

A more accurate treatment of the threshold-line equation is based on an analysis of the forces acting on the electron in the interaction space. These are the electric force, the centrifugal force, and the magnetic force. The first two forces are directed from the cathode to the anode and the third is directed from the anode to the cathode. In order for the electron to reach the anode, it is necessary that the work done by the electric and centrifugal forces is no less than the work done by the magnetic force. Therefore, we obtain the Hartree equation for the threshold straight lines:

$$U_a = \frac{\omega_n\left(r_a^2 - r_{\text{cat}}^2\right)}{2n} B - \frac{m}{2e}\left(\frac{\omega_n \boldsymbol{r}_a}{n}\right)^2, \tag{8.52}$$

which are called Hartree lines. The Hartree line for the π-type of oscillations is shown in Fig. 8.21b. It touches the critical-mode parabola at the point

$$U_{\min} = \frac{mr_a^2}{e}\frac{\omega_{N/2}^2}{N}, \quad B_{\min} = \frac{2mr_a^2}{e}\frac{\omega_n}{n\left(r_a^2 - r_{\text{cat}}^2\right)}.$$

At the voltage U_{min} for electrons moving near the anode parallel to its surface, synchronism conditions with the wave field are satisfied.

As can be seen, the Hartree lines lie below the threshold lines determined by the approximate (8.51). The magnetron generation area lies between the Hartree line and the cutoff parabola (shaded in Fig. 8.22).

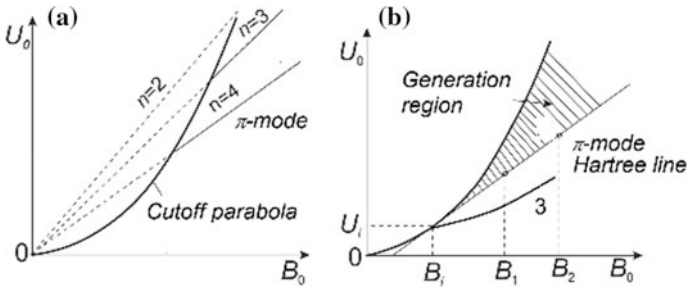


Fig. 8.21 Threshold straight lines (**a**) and the Hartree line for π-type oscillations (**b**)

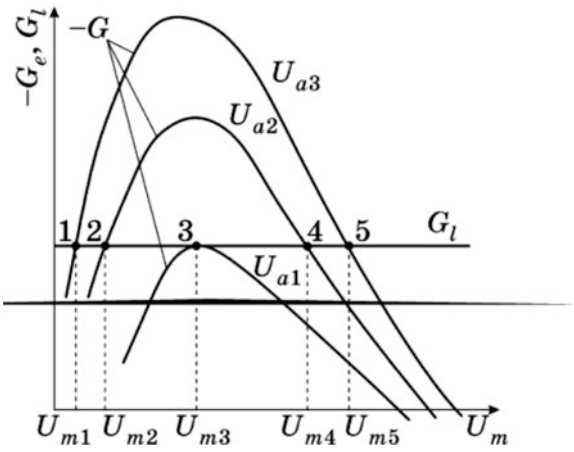


Fig. 8.22 Active component of electron conductivity versus anode voltage amplitude

To ensure synchronism in the entire interaction space, the constant electric field should increase linearly from the cathode to the anode: $E_0 = 2\pi f_n rB/n$. However, the actual field depends on the radius as $E_0 = A/r$, as in the coaxial capacitor and it does not satisfy the synchronism condition. The required distribution of the electric field can be realized only by taking into account the high-frequency field, the radial component of which increases exponentially while approaching the anode.

At the start of generation, the amplitude of the high-frequency field is equal to zero, so the minimum anode voltage ensuring synchronism with the electrons on the surface of the hub must be determined from the hub radius. Taking into account (8.49), we obtain the expression

$$U_{a\min} = \frac{\omega_n}{n} B r_h^2 \ln\frac{r_a}{r_h} + \frac{e}{8m} B^2 r_h^2 \left(1 - \frac{r_{cat}^2}{r_h^2}\right)^2. \tag{8.53}$$

The dependence $U_{a\min}(B_0)$ is also shown in Fig. 8.21b (the curve 3). This curve passes through the tangency point of parabola and the Hartree line and then lies below this line. It can be concluded that the conditions for self-excitation are facilitated by the hub space-charge action. In reality, self-excitation of a magnetron occurs at points located between the Hartree line and curve 3. Equations (8.52) or (8.53) determine the *phase conditions* for the self-excitation of the magnetron.

The amplitude conditions of self-excitation are determined by the equation

$$G_e + G_l = 0,$$

where G_e is the active component of the total electron conductivity $Y_e = G_e + jB_e = \dot{I}_i/\dot{U}_m, I_i$ is the complex amplitude of the induced current, and U_m is the voltage amplitude in the equivalent oscillatory circuit. The amplitude of the induced current can be calculated by the Shockley-Ramo theorem:

$$\dot{I}_i = nl_a \int_{r_h}^{r_a} \int_{\theta_l}^{\theta_r} \rho(\dot{E}_{1\varphi} v_\varphi + \dot{E}_{1r} v_r) r \mathrm{d}r \mathrm{d}\varphi, \tag{8.54}$$

where l_a is the length of the anode, and $\theta_l(r)$, $\theta_r(r)$ are the angles defining the left and right boundaries of the spoke. The shape of the electron spoke boundaries depends on the amplitude $\dot{U}_m$, the anode voltage U_0 and the induction of the magnetic field B_0.

Numerical simulation of interaction processes allows us to calculate the integral (8.54) for the specified U_m and find the electron conductivity. Typical dependences of magnetron electron conductivity on voltage amplitude at different anode voltages are shown in Fig. 8.22. Such non-monotonic dependences are typical for generators with rigid self-excitation. The same figure shows that the straight line G_l= const.

As the anode voltage increases, self-excitation conditions are first satisfied at point 3. However, the voltage amplitude U_{m3} necessary for this is too large for coherent oscillations to arise from the chaotic noise-like oscillations. With a further increase of the anode voltage to U_{a2} and U_{a3} values, the threshold amplitude decreases to U_{m2}, U_{m1} values respectively. If the U_{m2} value is sufficiently small in order for such an amplitude to appear in the magnetron due to noise or as a result of transient processes, generation occurs at point 2. However, at this point oscillations are unstable, since negative electronic conductivity increases at the accidental amplitude increase. Therefore, the amplitude of oscillations increases until it reaches the value corresponding to point 4. This point is stable and the oscillations continue in the stationary mode with amplitude U_{m4}. The larger amplitude of steady-state oscillations U_{m5} corresponds to an even larger value of the anode voltage U_{a3}. When the anode voltage decreases, the oscillations skip at the point with amplitude U_{m3}.

When the active part of electronic conductivity changes, its reactive part also changes. This results in the change of frequency of the oscillations in accordance with the equation

$$B_e + B_r(\omega) = 0,$$

where B_r is the reactive conductivity of the equivalent circuit. This phenomenon is named the electron frequency shift. The shift of frequency is smaller when the Q-factor of the magnetron resonator system is greater, but the increase of the Q-factor results in a decrease in the efficiency of the magnetron.

Magnetron operation in the generation mode. In the stationary generation mode, there are two main mechanisms that ensure the energy transfer from electrons to the high-frequency field. These are the sorting and bunching of electrons, which are described in Sect. 8.2. However, the cylindrical design of the magnetron and the cathode location in the interaction space give these processes a number of new features. Figure 8.23 shows the trajectories of harmful (A) and useful (B) electrons in the fixed coordinate system. As can be seen, the harmful electron is removed from the

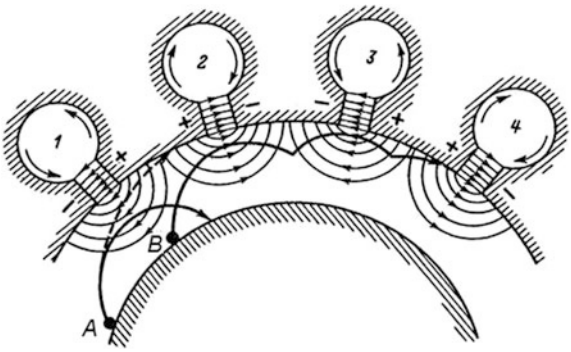


Fig. 8.23 Trajectories of electrons in the magnetron interaction space

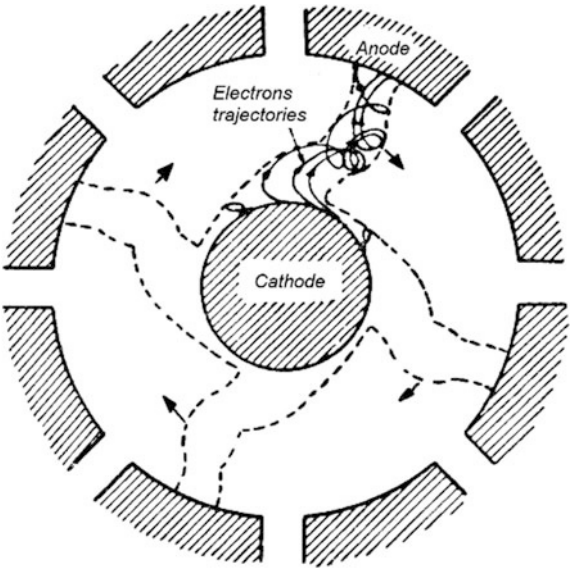


Fig. 8.24 Trajectories of electrons in the spoke

interaction space on the first turn, while the useful electron manages to make several revolutions around the guiding center before reaching the anode.

In the coordinate system rotating at angular velocity $\omega_s = v_p/r_a$ the shape of the spoke and the trajectories of electrons in it are shown in Fig. 8.24.

A magnetron still remains one of the most efficient sources of microwave energy. To estimate the electronic efficiency of a magnetron we take into account the potential energy which the electron possesses when leaving the cathode, $W_p = eU_a$ (thermal velocities are not taken into account). At the moment of arriving at the anode, its potential energy is equal to zero and the kinetic energy $W_c = mv^2/2$, where v is the electron velocity at the moment of contact with the anode. These velocities are different for different electrons. In the most unfavorable case, the electron hits the anode at the apex of the cycloid where its velocity

$$v = 2v_c = 2E_0/B.$$

Accordingly, its kinetic energy

$$W_c = 2m\left(\frac{E_0}{B}\right)^2 = 2m\left[\frac{U_a}{(r_a - r_{\text{cat}})B}\right]^2.$$

The difference between these energies is the energy that the electron transfers to the electromagnetic field. Hence the electronic efficiency

$$\eta_e = \frac{W_p - W_c}{W_p} = 1 - \frac{W_c}{W_p} = 1 - \frac{2m}{e}\frac{U_a}{r_{\text{cat}}^2 B^2 (1-\sigma)^2}, \tag{8.55}$$

where $\sigma = r_{\text{cat}}/r_a$ is the ratio of the cathode radius to the anode radius. Introducing the critical values of the anode voltage and magnetic induction using condition

$$r_a - r_{\text{cat}} = 2r_{\text{c}} = \frac{2m}{e}\frac{U_{\text{ac}}}{r_{cat}B_{\text{c}}^2(1-\sigma)},$$

the formula for electronic efficiency can be rewritten:

$$\eta_e = 1 - \frac{U_a/U_{\text{c}}}{(B/B_{\text{c}})^2}. \tag{8.56}$$

As can be seen, electronic efficiency is equal to zero for critical values of anode voltage and induction, and increases with increasing magnetic induction, if at this the anode voltage changes in accordance with the conditions for synchronism (8.52).

Substituting (8.51) into (8.55), we obtain

$$\eta_e = 1 - \frac{m}{e}\frac{\omega_n}{nB}\left(\frac{1+\sigma}{1-\sigma}\right). \tag{8.57}$$

From the obtained formula, it follows that electronic efficiency increases with increasing magnetic induction. In addition, under otherwise equal conditions, the magnetron has the highest efficiency with π-type oscillations ($n = N/2$). The efficiency also depends on the ratio of the cathode and the anode radii. It follows from (8.57) that this ratio should be reduced, but at too small a value of σ, it is impossible to provide synchronism in the greater part of the interaction space. Usually the optimum value of σ is chosen with the help of the "square rule": for the optimal value of σ, the cathode-anode distance should be equal to half the slow wavelength. From this rule we obtain

$$\sigma_{\text{opt}} = (2n - \pi)/(2n + \pi).$$

The recorded formulas only allow us to make a rather rough estimate of efficiency. They do not take into account that not all electrons hit the anode at

maximum velocity, and therefore they underestimate the result. On the other hand, these formulas do not take into account the presence of harmful electrons that overstates the result of the calculation. Since these two factors act in opposite directions, they partly compensate each other, reducing the error in the calculations.

The analysis does not take into account the effect of the space charge. Although in M-type devices this influence is not as important as in O-type devices, it is necessary to take it into account when analyzing the operation of the devices in detail. As shown above, a hub space-charge facilitates self-excitation conditions. On the other hand, the space charge makes it difficult to form the spokes and reduces the electronic efficiency of the magnetron

Figure 8.25 shows the dependence of electronic efficiency on magnetic induction for various types of oscillations, provided that the anode voltage varies in accordance with the Hartree line (8.52). As can be seen, with an increase in magnetic induction, the electron efficiency of the magnetron approaches 100%.

For the specified value of B_0 the highest efficiency is achieved in the π-mode. In real magnetrons, a "dip" in electron efficiency is observed (shown in dashed lines in Fig. 8.25) near the cyclotron resonance frequency $\omega_c = eB_0/m$. This dip is particularly noticeable in magnetrons with different-resonator anodes, and this is explained by the excitation of parasitic zero-mode oscillations. From the expression for cyclotron frequency, we obtain the induction value corresponding to the maximum "dip":

$$B_c = 1.2/\lambda.$$

The value of induction is obtained in Teslas, if the wavelength is substituted in centimeters. Magnetic induction in M-type devices is chosen, as a rule, to be less than B_c.

The straps also worsen electronic efficiency, since the zero-type oscillation field arises near the anode ends due to the straps. Therefore, the straps are usually screened by placing them in slots at the ends of the anode unit.

Electronic efficiency of modern centimeter wavelength magnetrons reaches 70–80% and can reach 90% in the decimeter wavelength. The efficiency of the oscillatory system (the circuit efficiency) depends on the ratio between its unloaded and external Q-factors. The higher this ratio is, the higher the efficiency is. However, it is difficult to obtain a high unloaded Q-factor, especially in the short-wave band, and too small an external Q-factor worsens the stability of the generation frequency. Typically, circuit efficiency is 70–90%.

Total magnetron efficiency is equal to the product of the electronic efficiency at the circuit one varies from 80% in the decimeter band to 30% in the millimeter wavelength band.

Magnetron characteristics. The magnetron volt-ampere characteristic (the dependence of the anode voltage on the anode current for the fixed value of the magnetic induction) is shown in Fig. 8.26. At an anode voltage less than the minimum, a very small current flows in the magnetron associated with noise oscillations. When the threshold voltage is exceeded, the anode current rises sharply

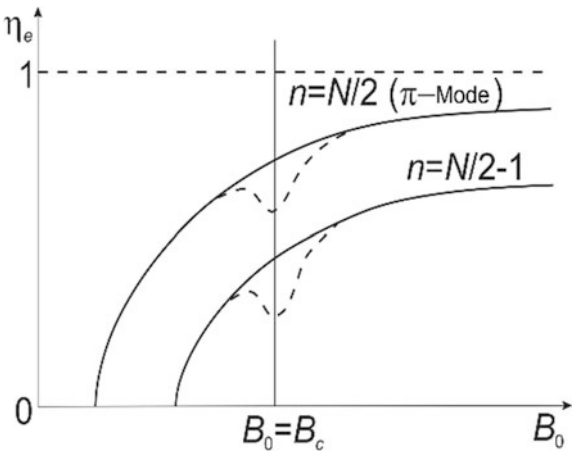


Fig. 8.25 Magnetron electronic efficiency versus magnetic induction

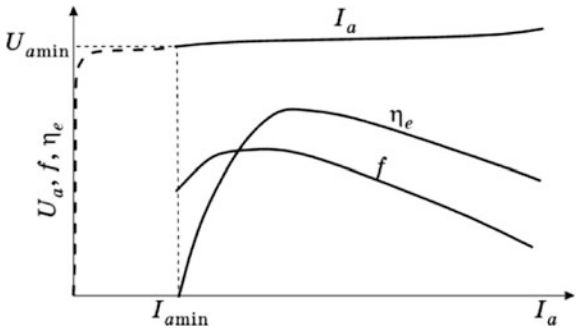


Fig. 8.26 Magnetron current-voltage characteristic

due to the appearance of generation and then changes very rapidly with small voltage changes. Therefore, to stabilize the operation mode, the magnetron should be fed by a current source with high intrinsic resistance.

The same figure shows the dependence of the electronic efficiency and the generation frequency on the anode current. As can be seen, there exists an optimum value of anode current at which efficiency is maximal. This current value is used as the operational current. The frequency of generation also depends on the anode current (Fig. 8.26). As already noted, this phenomenon is called the electronic frequency shift. At other values of magnetic induction, the current, efficiency, and generation frequency curves are similar to those shown.

Constructing the volt-ampere characteristics corresponding to different B values in the I_a, U_a plane and by drawing lines of constant output power and efficiency values, we obtain the *operation characteristics* of the magnetron shown in Fig. 8.27. These characteristics allow for the choosing of the operation mode of the magnetron. The strong coupling of the oscillatory system of the magnetron with the load results in the dependence of the frequency of oscillations and power on the

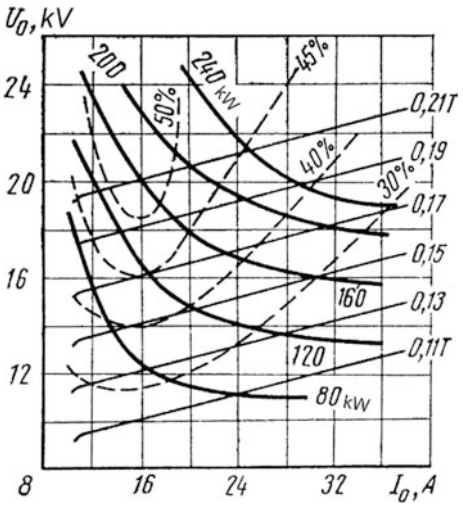


Fig. 8.27 Magnetron operation characteristics

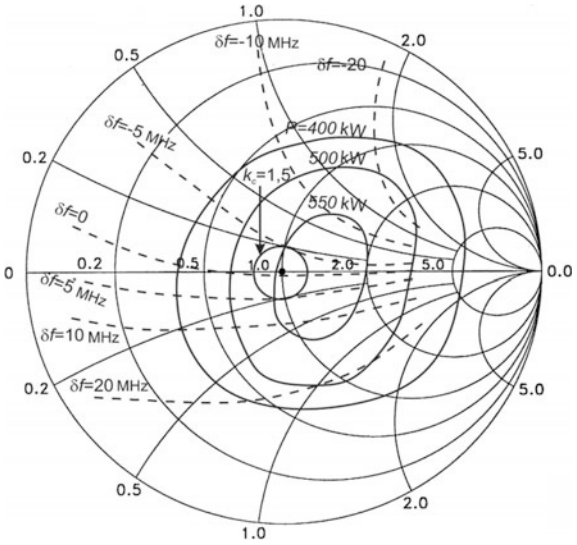


Fig. 8.28 Magnetron Rieke diagram

load parameters. These dependencies are customarily depicted on the circular chart where the active and reactive parts of the load total conductivity are used as the independent variables. The resulting set of graphs is shown in Fig. 8.28. It is called *the load characteristic* of the magnetron (Rieke diagram). The dotted line in this diagram shows lines of constant frequency and the bold solid curves show constant power lines. As can be seen, the magnetron transmits maximum power to the mismatched load.

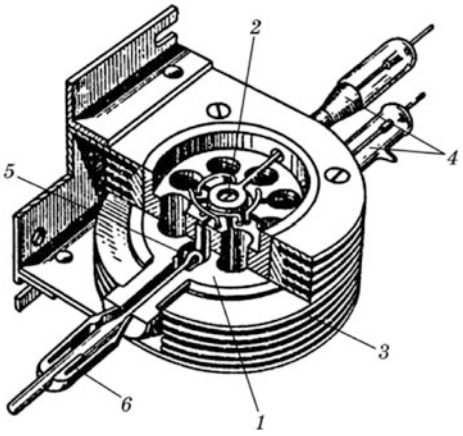


Fig. 8.29 Pulse magnetron design

When the reflection coefficient of the load changes, the generating frequency of the magnetron changes. This phenomenon is called frequency pulling. The quantitative pulling is determined by the maximum frequency change when the phase of the reflection coefficient changes by 360° with the module 0.2 (VSWR 1.5). The circle $|\Gamma| = 0.2$ ($k_c = 1.5$) and the tangent of the constant frequency lines to it are put on this characteristic. Frequency pulling in this case is 10 MHz.

Magnetron designs. The section of a typical pulsed multi-cavity magnetron is shown in Fig. 8.29. It consists of two main parts: the anode 1 and the cathode 2. The design of the anode resonators was considered above. The length of the anode is chosen with the following considerations: as length increases, the anode current increases, and hence the device's output power increases. However, with increasing anode length, the risk of axially inhomogeneous oscillation modes excitation occurs. In anodes with straps, frequency separation also deteriorates. Therefore, the length of the anode should not exceed half of the wavelength, and in magnetrons with straps it does not exceed (0.3–0.4)λ.

As a rule, in pulse magnetrons, a heating oxide cathode is used. In magnetrons with continuous operation (especially in magnetrons for consumer microwave

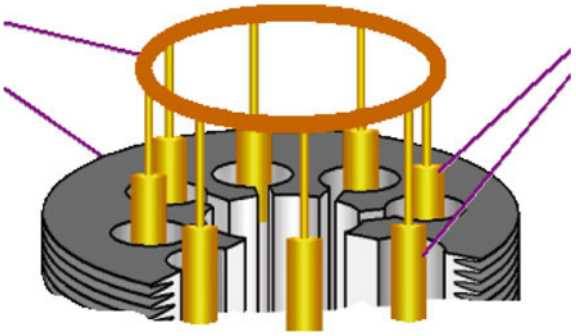


Fig. 8.30 Mechanism for mechanical tuning of a magnetron

ovens), the readiness time of the device is of great importance, which is determined mainly by the cathode heating rate. Therefore, in such magnetrons, a filamentary cathode is used. In modern oven magnetrons, the readiness time does not exceed 3 s. It should be noted that the harmful electrons bombard and additionally heat the cathode. To avoid the cathode overheating after the start of generation, the heating voltage can be reduced or even completely turned off.

The cathode is under high negative voltage, and the anode is under ground potential. The high-voltage inputs 4 serve to connect the source of the anode power supply and heater power. An air cooling system in the form of fins 3 is connected to the anode. In magnetrons with a large average power, a liquid cooling system is used. The inductive loop 5 with a vacuum seal 6 serves for the energy output from the magnetron.

The anode is closed above and below with metal covers (not shown in the figure). The distance from the anode to the covers should be chosen on a compromise basis. A small distance results in the violation of the field axial homogeneity in the interaction space and if the distance is too large, there is the probability of parasitic oscillations excitation in the space between the anode and the covers. Usually, the distance from the covers to the anode is 0.1–0.2 of the anode length.

For tuning generation frequency, the metal plungers can be introduced into resonator cavities and moved by the special mechanism through the flexible vacuum dense membrane are used (Fig. 8.30). In this way, it is possible to obtain a frequency tuning band of 5–15%.

The parameters of some magnetrons are given in Table 8.3. The MI-325 magnetron is designed to work in radar transmitters. MI-320 and MI-475 magnetrons are designed for linear accelerators and feeding plasmatrons. Quasi-continuous mode magnetrons, such as M-117 and M-170 are used for feeding industrial heating units. The Samsung om75s magnetron is designed for use in consumer microwave ovens.

Table 8.3 Parameters of some magnetrons

Type	Frequency (GHz)	P_{puls} (MW)	P_{av} (kW)	Pulse duration (μs)	Anode voltage (kV)	Magnetic system	Cooling	Mass (kg)
MI-325	15	0.2	–	0.18–0.3	15	Const. magn.	Air	8.5
MI-320	1.9	10	10	5	50	Solenoid	Liquid	70
MI-475	3	2	2.5	4	45	Const. magn.	Liquid	70
M-117	2.45	–	5.25	–	7.2	Const. magn.	Liquid/ air	5
M-170	0.915	–	100	–	12	Solenoid	Liquid	65
Samsung om75s	2.45	–	0.9	–	4.1	Const. magn.	Air	0.75

Fig. 8.31 Magnetron appearance

Figure 8.31 shows a pulse packed (i.e., mechanically connected to the magnetic system) magnetron for a 25 kW radar (*a*), a magnetron for linear accelerators feeding with a pulse power of 2 MW MI-475 (b) and a continuous mode magnetron for a microwave oven (c) with a power of 700 W.

8.4.2 Other Types of Magnetron

D. Wilbur and P. Peters invented a mitron or a magnetron tuned by voltage in the USA in 1949. The mitron design is shown in Fig. 8.32. Its interaction space is formed by the interdigital structure 1 and the cold cathode 2. The filamentary

cathode 3 and the conical control electrode 4 form the electron gun of the Kaino-Taylor type that is outside the interaction space. The ceramic vacuum-tight envelope 5 separates the vacuum part of the device from the atmosphere. The interdigital structure is connected to the external toroidal resonator 6.

The entire structure is in the axial magnetic field created by magnets 8 and the magnetic core 7. The coaxial line 9 and the coupling loop 10 serve the energy output of the oscillations.

The device operates as follows. The positive with respect to the cathode potential U_t is applied to the control electrode. Under the influence of the non-uniform electric field of the control electrode and the axial magnetic field, the electrons emitted by the cathode form a hollow electron beam, which is injected into the interaction space. The interdigital structure containing an even number of pins is under the potential U_a relative to the cold cathode, which is electrically connected to the thermionic cathode. The mitron resonator is excited by the basic axially symmetric type of oscillations. In the interaction space the pins connected the next but one to the upper and lower resonator covers create the high-frequency field corresponding to the π-type oscillations. The electron beam interacts with the field of the structure, forming spokes, as in the conventional magnetron, and transfers part of its potential energy to the field. This energy is transferred to the load with the help of the coupling loop and the external transmission line.

The mitron resonator loaded Q-factor is sufficiently small (less than 10) so the velocity of the rotation of electrons around the cold cathode mainly determines the generation frequency:

$$f_\pi = \frac{N}{2\pi(r_2 - r_{\text{cat}})}\frac{E_0}{B} = \frac{NU_a}{2\pi(r_a^2 - r_{\text{cat}}^2)B}.$$

As can be seen, the generation frequency for the specified induction B is proportional to the anode voltage. Since the electron gun is brought out of the interaction space, the change of anode voltage does not in practice effect the anode current, and electronic frequency tuning is carried out in the pure form without significant changing of other device parameters.

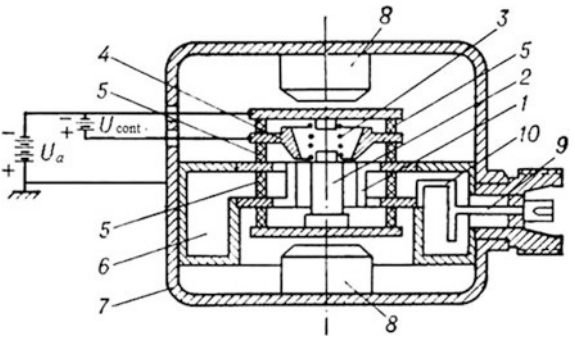


Fig. 8.32 Mitron scheme

The frequency tuning band is defined as the interval of the generated frequencies when the anode voltage varies from minimum to maximum. The width of the frequency band is characterized by the tuning ratio k—the ratio of the maximum generated frequency f_{max} to the minimum f_{min}:

$$k_f = f_{max}/f_{min}.$$

Figure 8.33 shows the operation (a) and control (b) characteristics of the mitron. It can be seen from the operating characteristics that the frequency is almost linearly dependent on anode voltage. The deviation of the frequency response from the linear pattern in broadband mitrons with frequency tuning within the octave band (k = 2) does not exceed ±1% of the average frequency f_{av} of the operation band. For narrowband high-power mitrons this deviation appears to be somewhat larger.

The steepness of the electronic tuning is determined by the expression

$$S_f = \Delta f/\Delta U_a.$$

The steepness value for mitrons of different power levels can lie in the band 0.2–10 MHz/V.

From the control characteristics (Fig. 8.33b), it can be seen that with an increasing control electrode voltage, the anode current and the output power increase. At very high U_t close to the anode voltage, the anode current falls, due to the deposition of a part of the electrons on the control electrode. Accordingly, the output power also decreases. With the control voltage increasing, the generation frequency decreases slightly which is due to the increasing influence of the space charge. Mitrons have a high efficiency value, which is determined by the "magnetron" mechanism for converting the energy of the power source into the energy of the alternating field. The efficiency of high-power mitrons can reach 30–60%.

The coaxial magnetron was proposed by J. Feinstein (USA) in 1964. The purpose of the invention was to increase the stability of the magnetron's frequency, which is subject to fluctuations due to load parameters changes (frequency pulling)

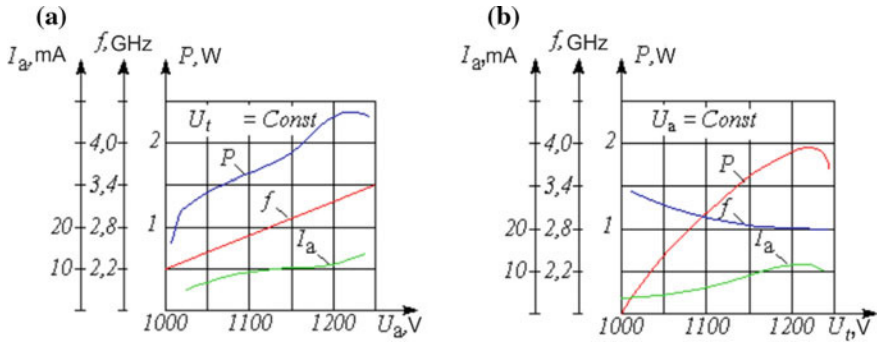


Fig. 8.33 Operation (**a**) and control (**b**) characteristics of the mitron

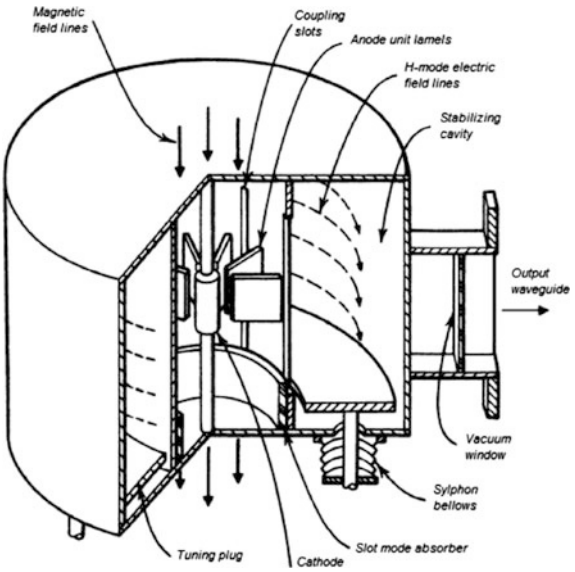


Fig. 8.34 Scheme of a coaxial magnetron

and cathode current fluctuations (electronic frequency shift). With an unfavorable combination of these factors, mode-skipping can be observed. The coaxial magnetron minimizes these phenomena due to the use of a stabilizing resonator. In addition, a magnetron of this type makes it possible to use a large number of cavities in the anode unit without worsening the separation of the mode of oscillations.

The coaxial magnetron design is shown in Fig. 8.34. The cathode and the anode with sector cavities form the interaction space. The resonators of the anode unit are connected by means of slots to the stabilizing coaxial resonator.

The magnetron operates with π-type oscillations and coupling slits located at the next-but-one ends of the sector resonators excite the azimuthally homogeneous oscillations of the H_{011} type in the stabilizing resonator. The resonance frequency of this device can be changed by means of a plug mechanically moved through the vacuum-tight membrane—sylphon. The absorber suppresses oscillations that arise in the anode unit and are not connected with the coaxial resonator. The energy output occurs through the waveguide connected with a coaxial resonator through the slot.

The high Q-factor of the coaxial resonator makes it possible to obtain good frequency separation even with a large number of anode unit resonators. In reality, the number of resonators can be increased by 6–8 times resulting in an increase of the cathode area, an increase of anode current and an increase of output power.

In coaxial magnetrons, the noise power is reduced by 10–20 dB and the probability of missing pulses decreases 100 times compared with conventional devices. The degree of frequency pulling decreases 3–5 times and the electronic frequency

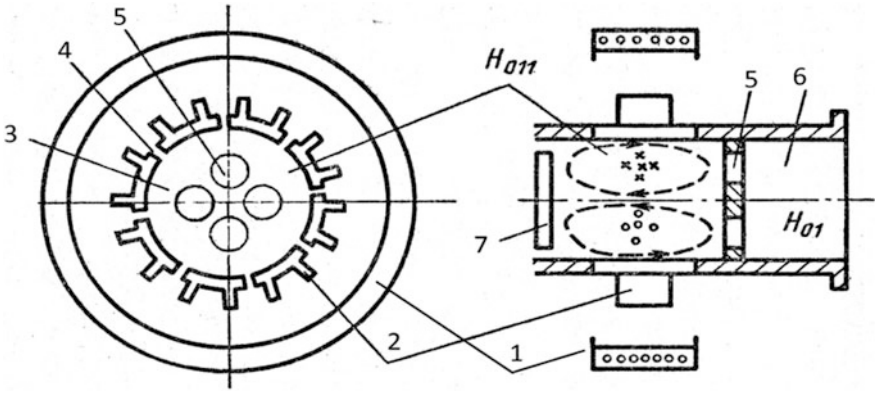


Fig. 8.35 Scheme of an inverted magnetron

shift decreases by an order of magnitude. Therefore, the coaxial magnetron can operate with a more mismatched load. Its efficiency is 40–50%.

The inverted magnetron is used mainly in the millimeter wavelength band. As its name indicates, the anode in this magnetron is located inside the cathode. The inverted magnetron circuit is shown in Fig. 8.35. The inner surface of the cathode 1 has the emitting coating. The electrons emerging from the cathode enter the high-frequency field created by the anode unit 2. Inside the anode unit, there is a cylindrical resonator 3 connected to the resonators of the anode unit by slots 4. In the anode, π-type oscillations are excited, so the coupling slots located in the resonators with the same field direction excite axially symmetric oscillations of the H_{011}-mode in the cylindrical resonator. This mode of oscillations has a large Q-factor and provides good frequency separation even with a large number of anode resonators.

The output of the energy of the oscillations is carried out through the holes 5 in the end wall of the cylindrical resonator, which extends into the circular waveguide 6. The holes located at the maximum of the radial component of the cylindrical resonator electric field excite the H_{01}-mode in the waveguide 6. Ceramic disks are usually sealed into the holes, so that they serve simultaneously as the vacuum window. The plunger 7 serves for the frequency tuning of the magnetron.

The high Q-factor of oscillations in the cylindrical resonator makes it possible to increase the number of resonators in the anode unit, thereby increasing the area of the cathode. As a result, it is possible to reduce the current density at the cathode and increase its lifetime.

The parameters of coaxial and inverted magnetrons are given in Table 8.4.

Table 8.4 Parameters of coaxial and inverted coaxial magnetrons

Magnetron type	Wave length (cm)	Number of resonators	Ppuls (MW)	U_a (kV)	I_a (A)	Efficiency (%)	$\Delta f/f_0$ (%)	Frequency shift (MHz/A)
Coaxial	3	40	1.5	33	83	50	10	0.08
Inverted	0.8	120	0.1	23	22	25	6	0.5

8.4.3 The Platinotron

In 1952, W. Brown from the American company Raytheon, developed a new broadband amplifier with crossed fields, called the platinotron (from the word *platy* —flat) in view of its flat amplitude-frequency response. The device was later called an amplitron or simply an amplifier with crossed fields. Amplitrons have high values of output power (up to 500 kW in pulse) and efficiency (70–80%) in combination with low weight and small dimensions. Since the amplitron is the amplifier, the signal at its output has much better characteristics than the signal of the magnetron. Therefore, the amplitron is used in many modern radars and communication systems. In particular, it was the amplitron that provided the transmission of the television image from the Moon of the astronauts landing on its surface in 1969–1972. Amplitrons have a relatively small gain (10–15 dB) and a highly nonlinear amplitude response, which limits their use.

The scheme of the platinotron is shown in Fig. 8.36. The cathode 1 together with the anode 2 form the interaction space. The lamellae of the anode unit are connected in a next-but-one pattern by the straps 3. The straps, in turn, are connected to the input and output of the tube.

Regarding the design, the amplitrons are close to cavity magnetrons (Fig. 8.36). The cathode 1 emits the electrons into the interaction space formed by the cathode and the anode 2. The anode contains slots-cavities, separated by lamellae. The lamellae are connected next-but-one with the straps 3 which in turn are connected to

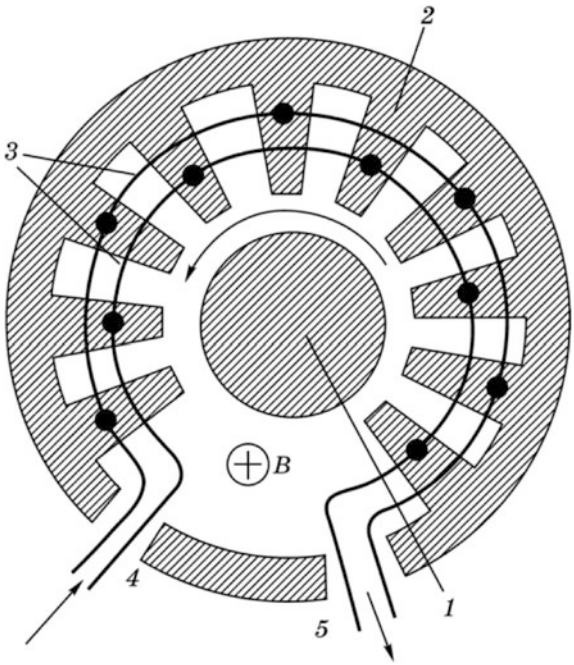


Fig. 8.36 Scheme of a platinotron

the input 4 and the output 5 of the device. A constant magnetic field is applied to the amplitron in the axial direction.

The slow-wave structure. The main difference between the magnetron and the amplitron is that the slow-wave structure in the amplitron is not closed. Therefore, separate types of oscillations cannot exist in it as they can in the magnetron.

The condition of phase constancy during the transition from one lamella to the neighboring lamella can be written in the form

$$\omega l/v_p = \varphi + \pi + 2\pi p, \quad p = 0, \pm 1, \pm 2, \ldots, \tag{8.58}$$

where $l = 2\pi r_a/N$ is the distance between the centers of the neighboring gaps; φ is the phase shift for the SWS period; and N is the number of SWS periods (number of lamellae). In contrast to the magnetron, in the platinotron the number of lamellae is odd, because one lamella falls on the SWS gap where the input and output couplers are located.

From (8.58), we can find the phase velocity of the spatial harmonic in the SWS:

$$v_p = \frac{\omega l}{\pi(2p+1) + \varphi}.$$

Obviously harmonic with the number $p = -1$ has the largest phase velocity in absolute magnitude:

$$(v_p)_{p=-1} = \frac{\omega l}{\pi - \varphi}.$$

The sign of the phase velocity indicates that this harmonic has negative dispersion. Therefore, the electron spokes must rotate towards to the motion of the wave energy. The dispersive characteristic of the basic spatial harmonic is shown in Fig. 8.37. In the magnetron, the direction of the spokes rotation is of no importance, since there is the standing wave in the SWS which can be represented as the sum of two waves traveling in opposite directions.

To transfer most of the energy to the high-frequency field, it is necessary that after a complete revolution, the spoke gets into the same phase as the wave. To do this, the turnover time of the spoke should be the integer number of the periods of the oscillations with the smallest permissible deviation Δt:

$$\tau = \gamma T \pm \Delta t, \quad \gamma = 1, 2, \ldots$$

The turnover time is determined by the velocity of spoke rotation:

$$\tau = \frac{2\pi r_a}{(v_p)_{p=-1}} = \frac{N}{\omega}(\pi - \varphi).$$

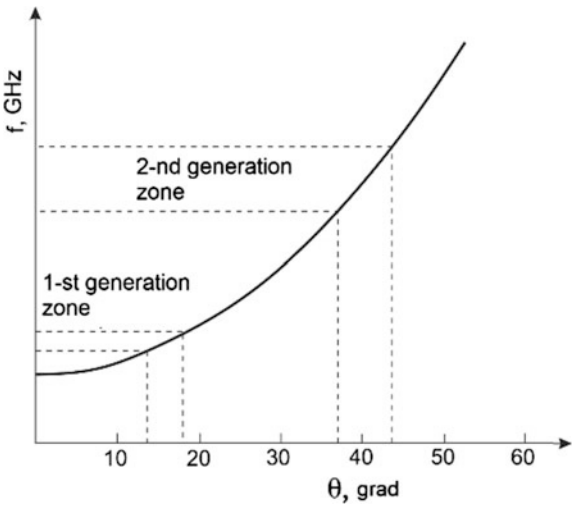


Fig. 8.37 Dispersion curve of the SWS of an amplitron

Equating the obtained expressions, we'll find

$$\varphi = \pi\left[1 - \frac{2}{N}\left(n \pm \frac{\Delta t}{T}\right)\right]. \tag{8.59}$$

At $\Delta t = 0$ we obtain a number of discrete values of the optimal phase shift:

$$\varphi_{\text{opt}\,n} = \pi(1 - 2n/N).$$

However, these phase shift values are not as rigidly fixed as in the magnetron. They allow deviations corresponding to the values of $\Delta t \leq \Delta t_{\max}$, where $\Delta t_{\max}$ is the maximum permissible deviation of the turnover time. It is natural to determine this deviation on the basis that after one revolution the spoke does not return to the field of another sign. Hence $\Delta t_{\max} < T/2$.

Substituting this value into formula (8.59), we obtain the admissible values. The admissible values correspond to permissible values of frequency, which are called amplification zones. These zones are shown in Fig. 8.37. In total, the platinotron has $(N - 1)/2$ amplification zones. The first zone is the zone corresponding to the maximum value $n = n_{\max} = (N - 1)/2$. In this zone, the rotation velocity of the spokes and the anode voltage are minimal. However, in the second zone $(n = (N - 3)/2)$ a wider frequency band is provided due to the larger slope of the dispersive characteristic. As a rule, this zone is used as the operation zone. Note that the number n determines the number of spokes in the interaction space, as it also does in the magnetron.

The amplitude response of the amplitron is shown in Fig. 8.38a. It is essentially nonlinear. In the shaded region, the input power is insufficient to form the spokes

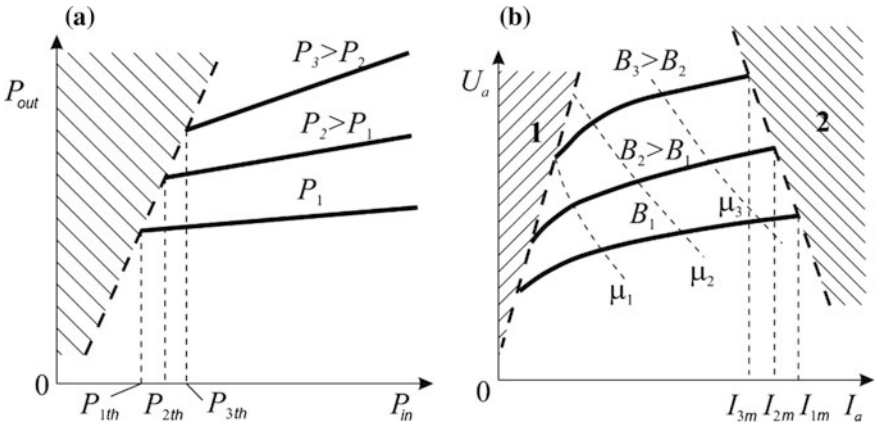


Fig. 8.38 Amplitude (**a**) and operational (**b**) characteristics of an amplitron

and the amplitron does not amplify the input signal, although significant noise power is observed at the output. The amplitron gain is relatively small and does not exceed 15 dB. It drops sharply as the power of the input signal increases. Since the input power is an appreciable part of the output power, the efficiency of the amplitron is determined by the so-called added power equal to the difference in the output and input power:

$$\eta = \frac{P_{\text{out}} - P_{\text{in}}}{P_o}$$

where P_0 is the supply power. The small gain is the amplitron's significant disadvantage and in some cases it is necessary to use an amplifying chain of several amplitrons. However, the high efficiency, reaching up to 60–80 in combination with the wide bandwidth compensates this disadvantage.

Figure 8.38b shows the performance of the amplitron. It is limited by both small and large currents. In the small currents region 1 the amplitron can self-excite on a parasitic oscillations type.

The disruption of the gain and the transition to the mode of noise-like oscillations are in the region of high currents 2. In the operation area, lying between the shaded sections, the anode current of the amplitron strongly depends on the voltage and on the magnetic induction. With increasing induction, the efficiency of the amplitron increases. The same graph shows lines of equal gain (dashed lines), which decrease as the anode current increases.

Figure 8.39 shows the construction of a high-power amplitron. It uses a bow-type SWS, consisting of copper tubes 1 connected to the anode through which the cooling liquid is pumped. From the side of the cathode on these tubes, the copper butt plates 2 are soldered to improve the structure of the field. Two rings of

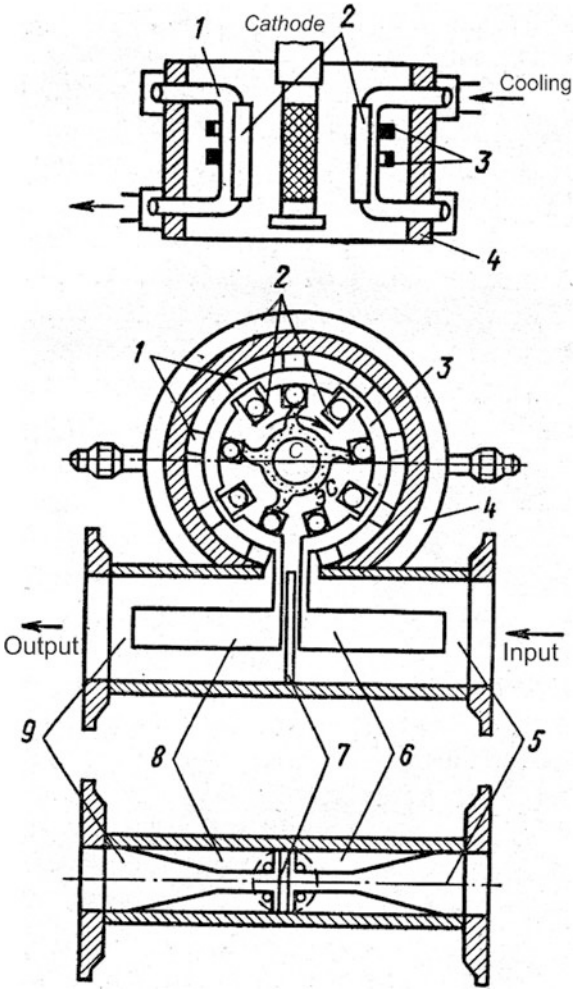


Fig. 8.39 Power amplitron construction

the straps 3 are located on the back side of the bows. The loops 6 and 8 are connected to the input and output waveguides 5 and 9 are linked with the ends of the straps. The waveguides are separated by the baffler strip 7.

Table 8.5 shows the parameters of some amplitrons.

If a high-Q resonator is connected to the input of the platinotron, feedback will arise in it due to the wave reflected from the resonator at the resonant frequency, and under the conditions of self-excitation it will turn into a generator. The frequency of this generator is stabilized by the external resonator which can be made

Table 8.5 Parameters of some amplitrons

Amplitron type	Frequency (GHz)	Pulse power (MW)	Av. power (kW)	Frequency band (%)	Gain (dB)	Efficiency (%)	Anode voltage (kV)	Anode current (A)
QKS 576	0.6	10	70	10	8	60	80	170
QKS 622	3	3	15	10	8	80	54	60
QKS 1243	10	0.5	0.5	5.3	13	50	37	26
QKS 1224	3	Cont. mode	400	5	9	74	–	–

with a high-Q and can be thermally stabilized. Such platinotron generators are called stabilotrons. Their output power and efficiency differ little from the parameters of the amplitrons on the basis of which these generators are designed.

Advancement Questions

1. Describe the interaction mechanism in M-type devices.
2. Describe the electron motion in the crossed fields. What determines the form of its trajectory?
3. Describe the mechanism of the interaction of electrons with the traveling wave field in M-type devices.
4. Write down the dispersion equation of M-type devices. What processes does it describe?
5. Describe the operating principle and parameters of an M-type TWT.
6. What determines the gain of an M-type TWT?
7. Describe the operating principle and parameters of an M-type BWO (the carcinotron).
8. Describe the circuits of an M-type TWT and BWO.
9. Describe the design and operation principle of a multi-cavity magnetron.
10. Describe the static operation mode of a magnetron. Explain what the critical-mode parabola is.
11. Describe the properties of the magnetron oscillatory system. Why is the π-type of oscillation chosen as the operating one?
12. What does it mean oscillation modes separation in the magnetron? In what ways can the separation of the modes be improved? What are the disadvantages of these methods?
13. Formulate the conditions for the self-excitation of a magnetron. What are Hartree lines?
14. What factors determine the efficiency of a magnetron?
15. Specify the magnetron operation area in an “anode current-anode voltage” graph. What is danger in going beyond the operation area?

16. Describe the design and operation of a mitron.
17. Describe the design and operation of coaxial and inverted coaxial magnetrons.
18. Describe the design and operation of a platinotron.
19. What factors determine the gain of a platinotron?
20. Describe the design and operating principle of a stabilotron.

Chapter 9
Gyro-resonant Devices

9.1 The Operating Principle of Gyro-resonant Devices

O-type and M-type electronic devices, as well as devices with quasi-static control are referred to as so-called classical microwave devices. The electron beam in them interacts either with the quasi-static field of the interelectrode gap, or with the slow electromagnetic wave. It follows that the characteristic dimensions of the electrodynamic system (the length of the gap, the SWS period) should be less than the wavelength. This condition limits the output power of the devices with high working frequency. In gyro-resonant devices, the electron beam interacts with fast electromagnetic waves. Since the field of the fast wave is not "retained" against the EDS surface, its characteristic dimensions can be much larger than the wavelength which makes it possible to substantially increase the output power of the device.

In 1958, Australian astrophysicist R.Q. Twiss was the first to point out the possibility of the induced emission of non-equilibrium electron-oscillators at the cyclotron frequency. In the following year, independently of Twiss, the work of A. V. Gaponov-Grekhov and J. Schneider appeared in which the possibility of induced radiation by the ensemble of electrons rotating in the magnetic field was proved. An idea stated by A.V. Gaponov on the necessity to take into account the relativistic effects when considering the interaction of rotating electrons with the electromagnetic field was of particular importance. Later, these ideas resulted in the creation of a whole class of devices capable of producing large powers at very high frequencies —*cyclotron resonance masers* (CRM). A great contribution to the creation of these devices was made by the staff of the Institute of Applied Physics at the Academy of Sciences in the USSR in Gorky (now Nizhny Novgorod) under the supervision of Academician A.V. Gaponov-Grekhov. In 1966, they were the first to create high-power CRM generators, called "gyrotrons".

A fundamental feature of the CRM, is the use of electron beams in which the electrons move along spiral trajectories, performing both translational (along the

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_9

device axis) and rotational motion. The generation and amplification of microwave radiation in these devices are based on the induced emission of the ensemble of nonlinear oscillators which are the rotating electrons. The name of the devices is related to this feature of the energy conversion mechanism (from the *Greek* giros—circular).

Since the rotation of electrons occurs in the plane almost perpendicular to the direction of translational motion, they effectively interact with the transverse components of the electric field. This makes it possible to use open resonators and smooth wavegudes in devises with standing and travelling electromagnetics waves, correspondingly. The linear dimensions of such systems can be of several wavelengths of the generated oscillations, and this provides a practical opportunity to create high-power devices in the short-wave region of the microwave band. As an example, Fig. 9.1 presents, at the same scale, the resonator of an O-type amplifying klystron (on the left) and the open resonator of the gyrotron (on the right) for the same wavelength $\lambda = 8$ mm.

At present, there are the following main types of gyro-resonant devices:

- The gyrotron—a generator device with an electrodynamic system in the form of a single open-type resonator[1];
- The gyroklystron—an amplifying device with an electrodynamic system containing two or more open-type resonators with drift spaces between them;
- Gyro-resonant traveling wave (gyro-TWT) and backward wave (gyro-BWO) tubes—devices in which the electromagnetic field has the form of a traveling wave (forward in the TWT, backward in the BWO);
- Gyro-resonant frequency multipliers.

9.2 Electron Beam Interaction with the High-Frequency Electrical Field

9.2.1 Cyclotron Resonance

The processes of the interaction of the electron beam with the transverse electromagnetic field are common to all types of gyro-resonant devices. Therefore, let's consider them using the example of the interaction of electrons with the field of an open resonator.

[1]In earlier publications concerning gyro-resonant devices, the device was called a "gyromonotron" and the term "gyrotron" ("gyrotrons") was used to designate the entire set of gyro-resonant devices. Currently, this term is used primarily to designate one type of gyro-resonant device, namely the single-resonator generator.

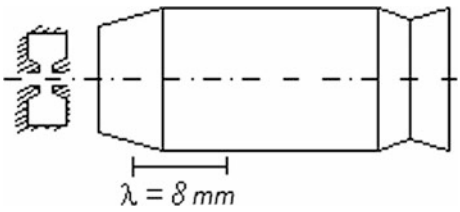


Fig. 9.1 Comparison of open and closed resonators

Let's consider an electron moving at a certain velocity in the homogeneous static magnetic field **B**. The electron trajectory in such a field is generally known to be a spiral (Fig. 9.2a). Motion along the spiral can be considered as the sum of the rotational motion with velocity $v_\perp$ and the translational motion with velocity $v_\parallel$ where $v_\parallel$ and $v_\perp$ are the velocity components parallel to vector **B** and perpendicular to it respectively. Radius R of the circle along which the rotation occurs and time of one revolution T_c and angular frequency ω_c of rotation are given by formulas

$$R = mv_\perp/(eB), \tag{9.1}$$

$$T_c = 2\pi R/v_\perp = 2\pi m/(eB), \tag{9.2}$$

$$\omega_c = 2\pi/T_c = eB/m. \tag{9.3}$$

An electron moves along magnetic field lines at distance $\lambda_c = T_c v_\parallel = 2\pi m v_\parallel/(eB)$ during one revolution, which is called the cyclotron wavelength. In terms of its geometric meaning, λ_c is the spiral trajectory step. If longitudinal velocity $v_\parallel = 0$, then the electron performs only rotational motion in the plane perpendicular to the magnetic field lines along the circumference of the radius R.

Let's suppose that in the area where the electron is moving, in addition to static magnetic field **B**, electric field $\mathbf{E} = \mathbf{E}_m \cos \omega t$ exists that varies with time according to the harmonic law, the force lines of which are perpendicular to the force lines of the magnetic field. Field **E** is assumed to be sufficiently weak, so for a time equal to several cyclotron periods T_c it does not appreciably perturb the trajectory of the electron.

For simplicity, we assume that longitudinal electron velocity $v_\parallel = 0$ and that it moves clockwise along the circular orbit. We assume that within the orbit of the electron, the electric field is homogeneous (Fig. 9.2b).

Let's consider the interaction of an electron with the alternating electric field. The electron at time $t_1 = 0$ is at point A on the trajectory. At this moment, the field reaches the amplitude value and is directed as shown in Fig. 9.2. Electric force $\mathbf{F} = -e\mathbf{E}$ is directed against the direction of motion and produces a decelerating effect on the electron. At time $t_2 = T_c/2$ this electron moves to point B on the trajectory. If the rotation frequency of the electron and the field variation frequency are equal over this time interval, then the field changes to the opposite direction, and at point B the electron again experiences the deceleration action of the electric field.

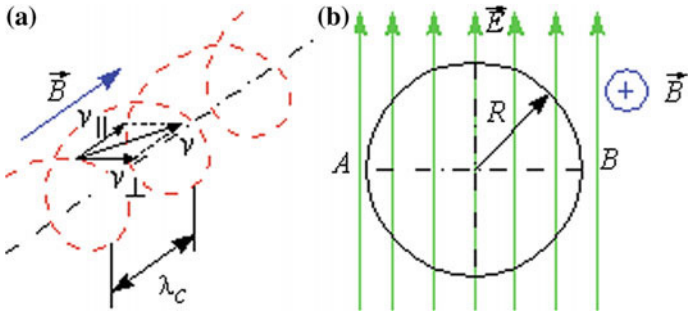


Fig. 9.2 Electron motion in the transverse electrical field

At time $t_3 = T_c$ the electron moves to point A on the trajectory and again meets the decelerating field, and so on.

Similarly, the electron being at initial time $t_1 = 0$ at point B on the trajectory, always "feels" the accelerating field and takes energy from it while moving along the orbit. Obviously, if the equality of frequencies is not fulfilled, then the electrons will either be accelerated or be decelerated by the electric field and the average result over several periods will be equal to zero. Thus, the interaction with the alternating electric field has resonant behavior and it is observed when the frequency of the field variation ω and the cyclotron frequency ω_c are equal. This phenomenon is called *cyclotron resonance*.

By repeating the preceding reasoning, it is not difficult to show that the effective interaction of rotating electrons with the transverse electric field is also observed in the harmonics of the cyclotron frequency. The general cyclotron resonance condition is written in the form $\omega = m\omega_c$, where $m = 1, 2, \ldots$ is the harmonic number of the cyclotron frequency.

9.2.2 Azimuthal Bunching

Figure 9.3a shows several trajectories of electrons moving in the transverse electric field E_θ and the dependence of this field on time (Fig. 9.3b). Figure 9.3c shows the elementary area—the "cutting" of the electron beam cross section containing the electron circular orbit and the force lines for time $t_1 = 0$ when the electric field component E_θ reaches its maximum value. We assume that the electric field in the cutting is almost homogeneous due to cutting's small size. Let's mark four characteristic electrons in orbit: 1, 2, 3, 4. For electron 2, force $\mathbf{F} = -eE_\theta \mathbf{e}_\theta$ is directed against its motion and, therefore, leads to its deceleration. Hence, the velocity of the electron 2 decreases. For electron 4, force $\mathbf{F}$ coincides in direction with the velocity vector and this electron accelerates at this moment of time. For electrons 1 and 3, the electric field force is directed perpendicular to their trajectories and, in practice, does not change their velocities.

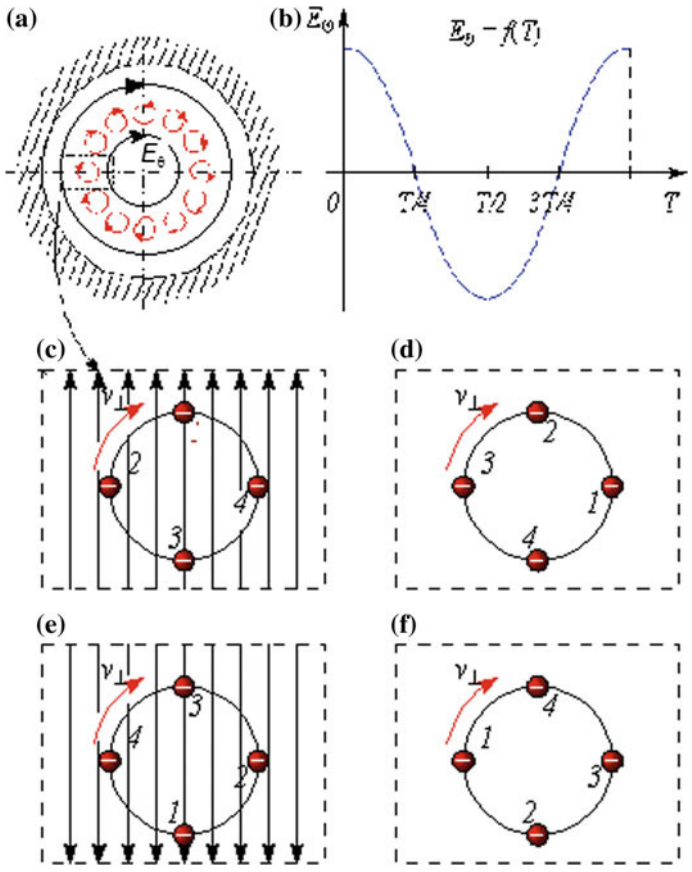


Fig. 9.3 Velocity modulation of electrons by the transverse electric field

Let's now consider the time $t_2 = T/4$. Since $T = T_c$ then the electrons under consideration at time $t_2 = T/4$ will move by a quarter of a turn and will occupy the positions shown in Fig. 9.3d. Since at this moment of time the field E_θ reduces to zero, the electrons do not experience the force action from this field.

At time $t_3 = T/2$ when the field E_θ again reaches its maximum value, electrons 1, 2, 3, and 4 occupy the positions shown in Fig. 9.3e. In this case, electron 2 is again subjected to deceleration and electron 4 is accelerated. At time $t_4 = 3T/4$ (Fig. 9.3f) the electric field reduces to zero and the electrons do not experience its force action. At time $t_5 = T$ the situation repeats.

Considering the interaction of electrons with the electric field during following and other periods, we obtain analogous results. Thus, in the process of multiple orbital rotation, some electrons are accelerated by the alternating electric field (electron 4 and electrons in its vicinity), some electrons are decelerated (electron 2 and electrons in its vicinity), while some electrons (1 and 3 and electrons located in their vicinity) do not change their velocity.

Consequently, during the interaction of electrons moving along circular trajectories in the high-frequency electric field in cyclotron resonance conditions, *velocity modulation* occurs.

The change in transverse electron velocities which occurs in the process of interaction with the alternating field results in the change of their rotational motion character. Since the circular orbit radius of the electron is proportional to velocity, its variation leads to a corresponding change of orbit radius. The effect of the cyclotron frequency ω_c changing due to the relativistic dependence of electron mass on its linear velocity is more delicate, but very important in the explanation of the interaction mechanism in gyrotrons. This effect, first described by A.V. Gaponov, becomes noticeable at accelerating voltages of 20–50 kV, typical for high-power gyro-resonant devices.

Such devices are often called weakly relativistic. From the expression of cyclotron frequency (9.3), it follows that in the framework of classical physics, the angular velocity of the rotation of the electron is determined only by the value of the magnetic field. A different result is obtained when the dependence of the relativistic mass on velocity is taken into account. In this case

$$\omega_c = \frac{eB}{\gamma m_0},$$

where m_0 is the electron rest mass; and $\gamma = \left(1 - v^2/c^2\right)^{-1/2}$ is the relativistic factor. As the linear velocity of the particle increases, its mass grows and the angular velocity of the rotational motion decreases.

As shown above, electrons 1 and 3 (and electrons in their vicinity) do not, in practice, change their velocities when interacting with the field, and therefore they move along trajectories close to the trajectories of the static case when the alternating electric field is absent. In contrast, for electron 4, for which the linear orbital velocity $v_\perp$ increases, orbit radius R increases and the angular velocity of the rotational motion decreases. Therefore, this electron, in the process of motion, shifts in the azimuthal direction towards electron 1 (Fig. 9.4).

The linear orbital velocity of electron 2 decreases and its angular velocity increases, and it also shifts in the azimuthal direction toward electron 1, decreasing the radius of its orbit at the same time.

Thus, near the azimuthal plane which rotates synchronously with electron 1, an electronic group is gradually formed. This process is called *azimuthal bunching*. The bunch forms around the electron, which does not change its velocity and gives no energy to the field.

In order to transfer its energy to the field, the bunch must be in the deceleration field. For this, it is necessary that the electron bunch shifts toward the azimuthal plane in which it will undergo deceleration, i.e., to the plane in which electron 2 was located initially. This can be achieved if we depart from the exact cyclotron resonance increasing slightly the angular velocity of electrons: $\omega = \omega_c\ (1 + \varepsilon)$, where ε is a small quantity in comparison with 1. In the self-oscillators to which the

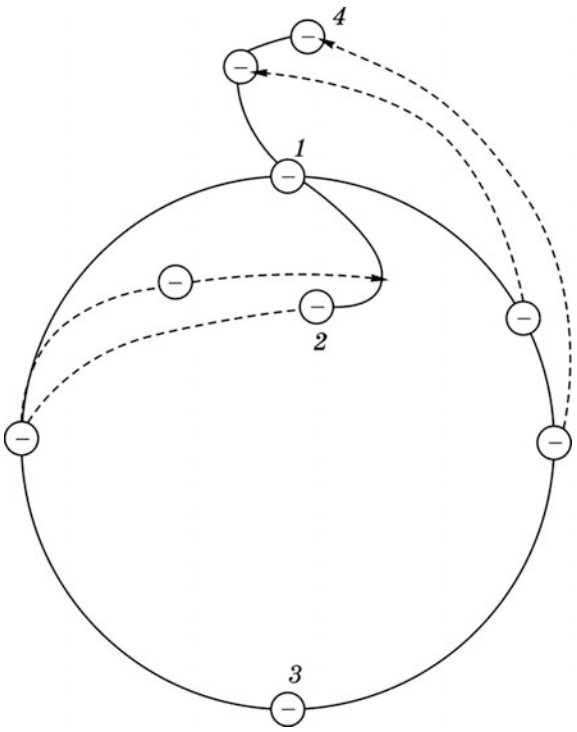


Fig. 9.4 Velocity modulation of electrons and azimuthal bunching

gyrotron belongs, such a generation frequency is set automatically, since it is this frequency at which self-excitation conditions are satisfied.

9.2.3 Equations of Electron Motion

In the derivation of the equations of motion, we confine ourselves to the case of the thin hollow electron beam interacting with the high frequency electric field of the H_{011} mode. The electromagnetic field of this mode has only one component of the electric field E_θ.

Let's introduce the local system of cylindrical coordinates r, φ, z the origin of which we align with the center of the circular orbit along which the electrons move before entering the zone of the alternating field. For the case under consideration, the relativistic equations of the motion of electrons can be written in the following form:

$$\frac{\mathrm{d}p_r}{\mathrm{d}t} = -e\tilde{E}_r - er\dot{\varphi}(B_0 + \tilde{B}_z) + m\frac{(r\dot{\varphi})^2}{r}, \tag{9.4}$$

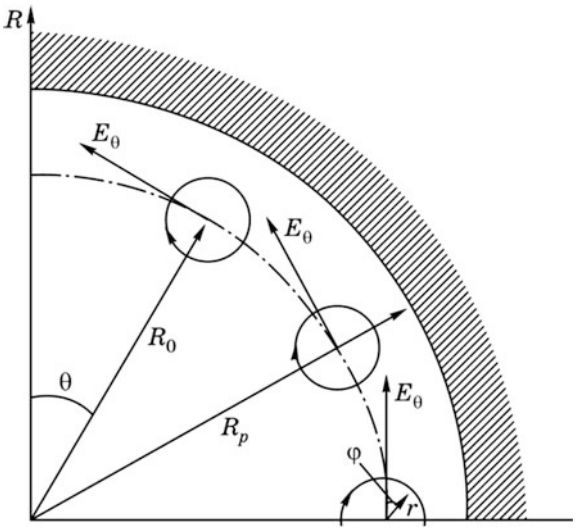


Fig. 9.5 For derivation of the equations of electron motion

$$\frac{\mathrm{d}p_\varphi}{\mathrm{d}t} = -e\tilde{E}_\varphi + e\dot{r}(B_0 + \tilde{B}_z) - m\dot{r}\dot{\varphi}, \tag{9.5}$$

$$\frac{dp_z}{dt} = 0, \tag{9.6}$$

where $p_r = m\dot{r}$, $p_\varphi = mr\dot{\varphi}$, $p_z = m\dot{z}$ are the components of the electron pulse $\mathbf{p} = m\mathbf{v}$ (the dots denote the derivatives of the corresponding quantities with respect to time); $m(r\dot{\varphi})^2/r$ is the centrifugal force; B_0 is the static magnetic field induction directed along the z axis; $\tilde{B}_z$ is the axial component of the alternating magnetic field; and $\tilde{E}_r, \tilde{E}_\varphi$ are the components of the alternating electric field.

Let's express these components of the electric field through the E_0 component of the resonator field. In accordance with the constructions in Fig. 9.5, we have

$$\tilde{E}_r = E_\theta \cos\varphi \quad \tilde{E}_\varphi = -E_\theta \sin\varphi$$

or

$\tilde{E}_r = E_m \cos\varphi \cdot \cos\omega t, \tilde{E}_\varphi = -E_m \sin\varphi \cdot \sin\omega t$. With the account of the earlier assumption about the smallness of the radius and the step of the electron orbit, the amplitude of the electric field E_m entering into these expressions can be regarded as having a constant value.

Substituting expressions for $\tilde{E}_r$ and $\tilde{E}_\varphi$ in the original system of equations and ignoring the z-component of the alternating magnetic field $\tilde{B}_z$ in it, then taking into account the smallness of the alternating magnetic field amplitude in comparison with the induction of the static field B_0, we obtain:

$$\frac{dp_r}{dt} = -eE_m \cos\varphi \cdot \cos\omega t - er\dot{\varphi}B_0 + mr\dot{\varphi}^2, \tag{9.7}$$

$$\frac{dp_\varphi}{dt} = eE_m \sin\varphi \cdot \cos\omega t + e\dot{r}B_0 + m\dot{r}\dot{\varphi}, \tag{9.8}$$

$$\frac{dp_z}{dt} = 0. \tag{9.9}$$

In the absence of an alternating field, this system has the particular solution based on the equality of the magnetic and centrifugal forces in (9.7):

$$er\dot{\varphi}B_0 = mr\dot{\varphi}^2$$

This solution corresponds to the motion along the spiral trajectory characterized by the following relations:

$$r = \text{const} = r_0, \quad er\dot{\varphi}B_0 = mr\dot{\varphi}^2, \quad m\dot{z} = \text{const} = m\dot{z}_0,$$

where index 0 denotes the initial values of the corresponding quantities.

The azimuthal velocity of the equilibrium motion is given by $\dot{\varphi} = eB_0/m = \omega_c$.

9.2.4 Abridged Motion Equations

The equations that are obtained from the initial equations by averaging the rapidly varying forces acting on the electrons from the side of the alternating field over time, are commonly referred to as abridged equations.

Assuming the electric field frequency is close to the cyclotron resonance frequency, i.e., $\omega \approx \omega_c = \dot{\varphi}$ we represent the time variation of the electron azimuthal coordinate φ in the following form:

$$\varphi = \omega_c t - \psi(t), \tag{9.10}$$

where $\psi(t)$ is a slowly varying function of time, such that its variation during the field oscillation's period or the cyclotron period is small.

Using (9.10), we can write

$$E_m \cos\varphi \cdot \cos\omega t = E_m \cos(\omega t + \psi)\cos\omega t = 0.5E_m \cos(2\omega t + \psi) + 0.5E_m \cos\psi,$$

$$E_m \sin\varphi \cdot \cos\omega t = E_m \sin(\omega t + \psi) \cdot \cos\omega t = 0.5E_m \sin(2\omega t + \psi) + 0.5\mathrm{E}_m \sin\psi.$$

Taking into account the expressions obtained, let's rewrite the system of (9.7)–(9.9) in the following form:

$$\frac{\mathrm{d}p_r}{\mathrm{d}t} = -\frac{1}{2}E_m \cos(2\omega t + \psi) - \frac{1}{2}E_m \cos\psi - er\dot{\varphi}B_0 + mr\dot{\varphi}^2,$$

$$\frac{\mathrm{d}p_\varphi}{\mathrm{d}t} = \frac{1}{2}eE_m \sin(2\omega t + \psi) + \frac{1}{2}eE_m \sin\psi + e\dot{r}B_0 - m\dot{r}\dot{\varphi},$$

$$\frac{\mathrm{d}p_z}{\mathrm{d}t} = 0.$$

We select from the right-hand sides of the equations, the rapidly oscillating components of forces with the argument $2\omega_c t + \psi$, the sign of which changes twice during the period of the alternating voltage. These components have little effect on the motion of electrons, and the action of the alternating electric field is mainly determined by terms containing the slowly varying function of time $\psi(t)$. This feature allows us to omit rapidly oscillating components from the equations of motion, equivalent to averaging the time-dependent terms over the period of the alternating field.

As a result, we obtain the following system of simplified (abridged) equations:

$$\frac{\mathrm{d}p_r}{\mathrm{d}t} = -\frac{1}{2}E_m \cos\psi - er\dot{\varphi}B_0 + mr\dot{\varphi}^2,$$

$$\frac{\mathrm{d}p_\varphi}{\mathrm{d}t} = \frac{1}{2}eE_m \sin\psi + e\dot{r}B_0 - m\dot{r}\dot{\psi},$$

$$\frac{\mathrm{d}p_z}{\mathrm{d}t} = 0.$$

The numerical integration of these motion equations makes it possible to analyze the bunching process of azimuthal electrons that arises because of the alternating field interaction, and to calculate the energy exchange between the electron beam and the field.

Figure 9.6 shows the results of calculating the motion of electrons for several different time moments reflecting the process of azimuthal bunching. The

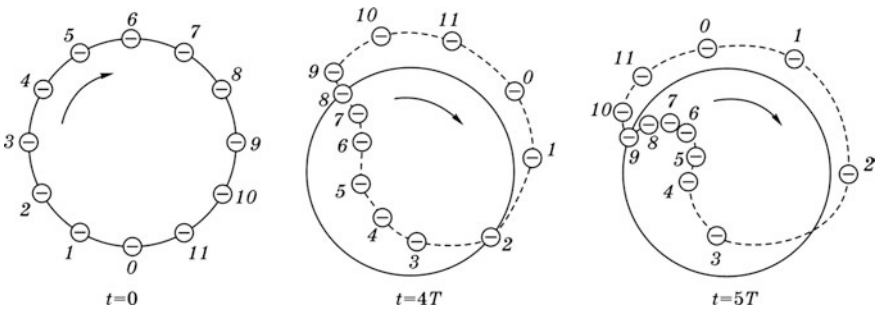


Fig. 9.6 Azimuthal bunching of electrons

calculation was carried out for $\omega_c/\omega = 0.93$ so the electron bunch formed around electron 6 gradually shifts to the decelerating phase of the electric field.

9.2.5 Field and Electrons Interaction on Cyclotron Frequency

The resonance interaction at the harmonics of cyclotron frequency is characterized by the condition $\omega \approx n\omega_c$, $n = 1, 2, \ldots$. The value $m = 1$ corresponds to the previously considered case of the main cyclotron resonance $\omega \approx \omega_c$.

Operation on the harmonics of cyclotron frequency allows us to reduce the value of magnetic field induction required for the operation of the device, which is the reason for the practical significance of such an operation mode. Indeed, from the cyclotron resonance condition, it follows that $\omega_c \approx \omega/\mathrm{n}$ and $B = m\omega/(en)$.

The operation modes at the second and third harmonics of the cyclotron frequency is of practical use, since at higher harmonics, the electronic efficiency of the device decreases.

9.3 The Gyrotron

9.3.1 The Design and Operating Principle of the Gyrotron

The design of the gyrotron is shown schematically in Fig. 9.7. It includes the following basic parts and assemblies: the electron gun of the magnetron type, consisting of the cathode electrode 1, the first and second anodes 2 and 3, the open cavity 4, the transition 5 to the output waveguide 6 serving simultaneously as the electron beam collector, the vacuum window 7 and the external waveguide 8. The magnetic field generated by the main 9 and auxiliary 10 solenoids is used to form the electron beam. The vacuum envelope of the gyrotron is formed by the metal case of the device, the vacuum window and the ceramic insulators 11. A system of liquid cooling 12 is provided for removing the heat generated in the resonator walls and the output waveguide. A typical axial distribution of magnetic induction is shown on the same figure.

The electron gun forms a hollow electron beam in which the electrons move along spiral trajectories.

In gyrotrons, an open waveguide cavity is used in the form of an irregular cylindrical waveguide segment narrowing at its ends (Fig. 9.8). The cavity mode H_{011} is often used as the operation mode. Its eigen frequency is close to the critical frequency of the corresponding wave mode of the cylindrical waveguide The figure shows the distribution of the electric field in this resonator for several modes of oscillations.

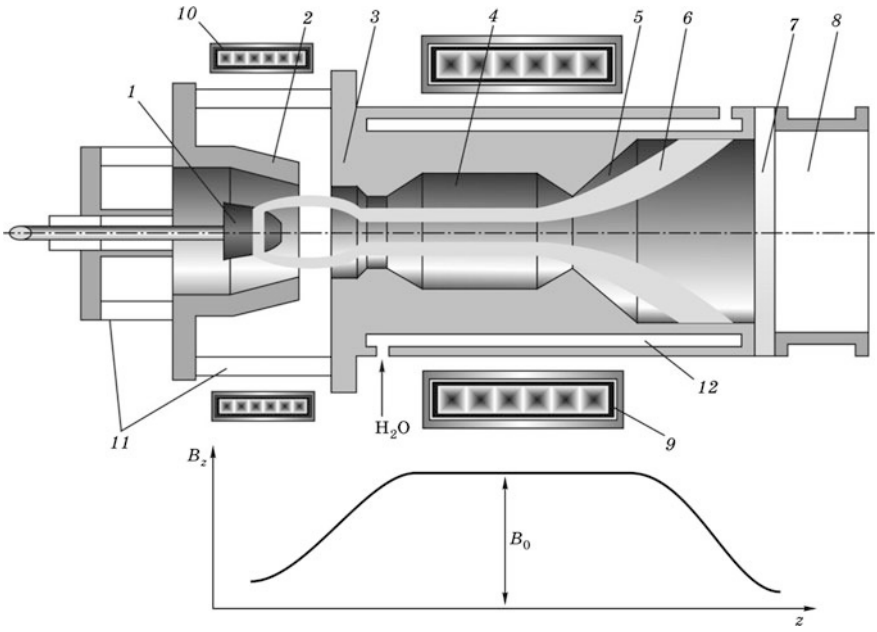


Fig. 9.7 Gyrotron design

When the cavity eigen frequency is close to the angular velocity of the rotational motion of the electrons, conditions arise for such an interaction of the beam and the field at which the generation of electromagnetic waves occurs and is maintained.

The electron beam that has passed through the resonator and has transmitted part of the kinetic energy associated with the rotational motion to the resonator field drops on the walls of the output waveguide. The energy of the electromagnetic field is transferred through the output window to the external waveguide connecting the device with the load.

In gyrotron-type devices, the electron-optical system should not only form the electron beam of the certain configuration, but also provide the rotational motion of the electrons. It is desirable that the energy of the rotational motion $W_\perp$ is substantially larger than the kinetic energy of the longitudinal motion $W_\parallel$ (usually $W_\perp/W_\parallel = (v_\perp/v_\parallel)^2 = 2{-}4$). For this purpose, special electron guns of the magnetron type are used (see Appendix B).

9.3.2 *Electronic Efficiency*

There are two types of efficiency: transverse efficiency and total electronic efficiency. We define transverse efficiency as the ratio of the energy $W_\sim$ transferred from the electron beam to the resonator field in the time interval equal to or a

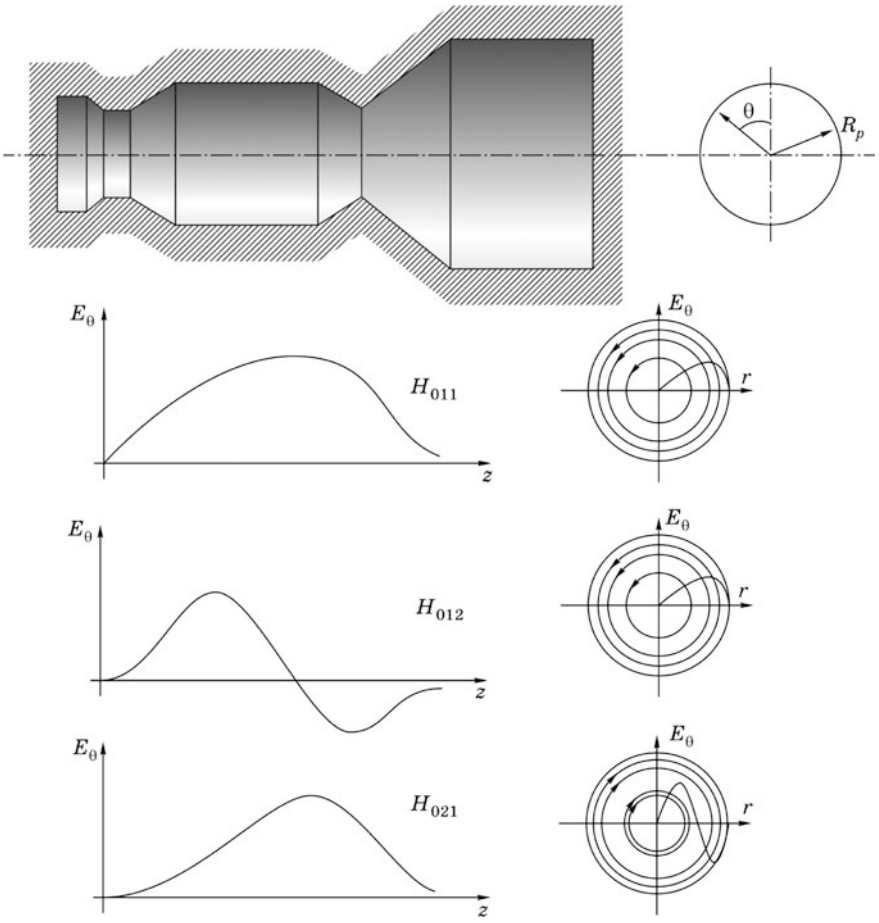


Fig. 9.8 Open resonator of a gyrotron

multiple of the cyclotron period T_c to the kinetic energy of the rotational motion $W_{\perp 0}$ of the electrons entering the resonator in the same time interval:

$$\eta_{e\perp} = \frac{W_\sim}{W_{\perp 0}}.$$

On the basis of the energy conservation law, the value of $W_\sim$ is equal to the difference of the kinetic energy at the cavity entrance $W_{\perp 0}$ and at its output $W_\perp$:

$$W_\sim = W_{\perp 0} - W_\perp.$$

Taking this formula into account, we obtain

$$\eta_{e\perp} = 1 - W_{\perp}/W_{\perp 0}.$$

The practical calculation of $\eta_{e\perp}$ is carried out by integrating the motion equations of N trial electrons which are initially uniformly distributed along the azimuthal coordinate.

The initial kinetic energy of these electrons is

$$W_{\perp 0} = N\frac{mv_{\perp 0}^2}{2},$$

or, since $v_{\perp 0n} = v_{\perp 0} = \text{const}$,

$$W_{\perp} = \sum_{n=1}^{N} \frac{mv_{\perp n}^2}{2},$$

where $v_{\perp 0}$ is the initial velocity of the electrons.

The kinetic energy of the rotational motion of the electrons at the resonator output is given by the formula

$$W_{\perp} = \sum_{n=1}^{N} \frac{mv_{\perp n}^2}{2},$$

where $v_{\perp n}$ are the velocities of electrons at the resonator output obtained as a result of integrating the motion equations. Taking into account the formulas given above, we obtain

$$\eta_{e\perp} = 1 - \frac{1}{v_{\perp 0}^2 N}\sum_{n=1}^{N} v_{\perp n}^2$$

The value $N^{-1}\sum_{n=1}^{N} v_{\perp n}^2$ in this formula is an average square of the velocities of electrons $\bar{v}_{\perp}^2$ at the output of the resonator. Consequently,

$$\eta_{e\perp} = 1 - \frac{\bar{v}_{\perp}^2}{v_{\perp 0}^2}.$$

Numerical calculations show that for uniform distribution of the high frequency electric field along the cavity axis, the maximum value of $\eta_{e\perp}$ is about 40%. By choosing the optimal law of the electric field distribution along the resonator axis, it is possible to achieve values exceeding 80%. The optimal distribution of the high-frequency field is characterized by the slow increase of the field amplitude at the beginning of the resonator, the faster increase of the field in the middle part of the resonator and the sharp decrease of the field at its end. With such a distribution, the optimal conditions for the interaction between the electron beam and the field

are ensured. In the weak, slowly increasing field, electron velocity modulation and dense azimuthal bunch formation occur. The large amplitude of the field in the output part of the resonator provides effective energy extraction from the bunched electron beam.

Total electronic efficiency is defined as the ratio of the energy $W_{\sim}$ transferred from the electron beam to the resonator field, to the total kinetic energy of the electrons at the resonator input W_0:

$$\eta_e = W_{\sim}/W_0.$$

The value of W_0 can be represented as the sum of the longitudinal and rotational motion kinetic energies:

$$W_0 = W_{\perp 0} + W_{\parallel 0}.$$

Taking this into account, we obtain

$$\eta_e = \frac{\eta_{e\perp}}{1 + W_{\parallel 0}/W_{\perp 0}}.$$

Since we assume that at the cavity input all electrons have the same values of longitudinal and transverse velocities, then

$$\frac{W_{\parallel 0}}{W_{\perp 0}} = \left(\frac{v_{\parallel 0}}{v_{\perp 0}}\right)^2.$$

Consequently,

$$\eta_e = \eta_{e\perp}\frac{p^2}{1+p^2},$$

where $p = v_{\perp 0}/v_{\parallel 0}$ is the ratio of transverse velocity to longitudinal velocity (*pitch factor*).

From this formula, it follows that to obtain a high total electronic efficiency it is necessary to use electron beams with a large pitch factor. In real devices, its value is 1.4–1.7. Further increase of the pitch factor is limited by the difficulty of the design of the corresponding electron optical systems.

9.3.3 Total Efficiency and Output Power

The total efficiency of the gyrotron is defined as the ratio of the HF field power delivered by the device to the output waveguide, to the supply power. As in other types of electronic device, the total efficiency can be represented as the product of electronic efficiency η_e and circuit efficiency η_r:

$$\eta = \eta_e \eta_r.$$

Circuit efficiency can be expressed in terms of loaded Q_l, external Q_e and intrinsic Q_0 Q-factors of the cavity:

$$\eta_r = Q_l/Q_{\text{out}} = 1 - Q_l/Q_0.$$

Loaded Q-factor can, in turn, be expressed in terms of unloaded and external, Q-factors:

$$Q_l = \frac{Q_e Q_0}{Q_e + Q_0}.$$

For an open waveguide-type resonator, the unloaded Q-factor is determined as the ohmic Q-factor Q_m which characterizes the energy losses in the resonator walls. The external Q-factor is identified by the diffraction Q-factor of the cavity Q_d, which characterizes the energy loss due to radiation. Since the radiated energy in this case is directed to the output waveguide and represents the output energy of the device, the diffraction Q-factor plays the role of the external Q-factor.

Thus, for a loaded Q-factor and resonator efficiency, we obtain the following formulas:

$$Q_l = \frac{Q_d Q_m}{Q_d + Q_m},$$

$$\eta_r = Q_l/Q_d = 1 - Q_l/Q_m.$$

Consider the open cavity formed by the cylindrical waveguide segment with H_{01} mode. Its ohmic Q-factor is calculated from the formula $Q_m = R_r/\delta$ where R_r is the cavity radius, and δ is the skin depth defined by the formula $\delta = (2/\omega\mu_w\mu_0\sigma_w)^{1/2}$ in which σ_w and μ_w are the conductivity and the relative permeability of the cavity walls.

The diffraction Q-factor of H_{011}-type oscillations is expressed by formula $Q_d = 4\pi(L_r/\lambda)^2\alpha$ where λ is the free space wavelength. Since $\lambda_{cf} \approx \lambda$,

$$Q_d = 4\pi(L_r/\lambda_{cf})^2\alpha,$$

where L_r is the length of the operating part of the cavity, λ_{cr} is the critical wavelength of the H_{01}-type in the cylindrical waveguide forming the operating part of the resonator; and α is the coefficient depending on the resonator configuration and the geometry of the output window. In those cases when Q_m significantly (in several times) exceeds the diffraction Q-factor Q_d, η_r is close to one.

The output power is determined by the power of the electron beam and the value of the total efficiency of the device:

$$P_{out} = P_0\eta = I_0 U_{a2}\eta,$$

where P_0 and I_0 are the power and the current of the electron beam; and U_{a2} is the potential of the second anode.

9.3.4 Gyrotron Starting Current

The starting or inrush current refers to the minimum value of the electron beam current I_0 at which the self-excitation of oscillations begins. At a current equal to the starting current, the power given by the beam is equal to the power of the losses. The analysis shows that the value of the starting current I_{st} expressed in amperes for a gyrotron with an operation mode H_{011} is determined by the formula

$$I_{\text{start}} = \frac{2.3\lambda\beta_{\parallel}^2 \mathrm{e}^{2\chi^2}}{(Q_l \cdot 10^{-3}) L_r G(\mu\chi - 1)},$$

where

$$\chi = \frac{1}{2}\left(\frac{1}{\mu} + \sqrt{\left(\frac{1}{\mu}\right)^2 + 1}\right), \quad \beta_{\parallel 0} = \frac{v_{\parallel 0}}{c}, \quad \mu = \pi p^2 \frac{L_r}{\lambda},$$

$$G = 0.42 J_1^2\left(3.83\frac{R_0}{R_r}\right),$$

λ is the wavelength of electromagnetic oscillations in the free space; L_r, R_r, Q_l are the length, radius and loaded Q-factor of the resonator; R_0 is the average radius of the electron beam; and $J_1(x)$ is the first-order Bessel function of the first kind.

The value of ratio R_0/R_r is chosen so that the beam passes through the zone of maximum intensity of the high-frequency field. For H_{011} oscillations, the maximum field occurs at a distance from the resonator axis equal to $0.48R_r$. Assuming that $R_0 = 0.48R_r$, we obtain $G = 0.142$.

9.3.5 Influence of the Spread of Electron Velocities on Gyrotron Operation

Until now, when analyzing the processes in a gyrotron, it was assumed that the electron optical system forms the electron beam with identical orbital and longitudinal velocities for all electrons. However, the real electron beams have a spread of velocities due to the following reasons:

- The spread of initial (thermal) velocities;
- The roughness of the emitting cathode surfaces;
- The violation of the axial symmetry of the electron-optical system (for example, due to the radial displacement of the cathode);
- Position spread due to the difference between the electric and magnetic fields for electrons emerging from different points of the cathode;
- The spread occurring under the influence of the electron beam space-charge field.

It is customary to characterize the total velocity spread by the relative magnitude of the orbital (oscillator) velocity spread:

$$\delta v_{\perp} = \frac{\Delta v_{\perp}}{v_{\perp 0}},$$

where $\Delta v_{\perp}$ is the absolute value of the orbital velocity spread. In different beams, the value $\delta v_{\perp}$ can reach several tens of percent. As the analysis shows, the main contribution to the velocities spread gives beam space charge.

Velocity spread results in the deterioration of the electron beam and efficiency of the HF field interaction, and in the decrease of the efficiency of the device. For example, with a velocity spread equal to $\delta v_{\perp} = 20\%$, the electronic efficiency value decreases by about 30%. Velocity spread also results in the change of the starting conditions and the starting current force.

9.3.6 Large-Orbit Gyrotrons

In devices called large-orbit gyrotrons, electrons move along spiral trajectories that span the symmetry axis of the device. The radii of their orbits can be comparable with the resonator radius and this explains their name.

There are two systems for the formation of such beams. The first of them includes an electron gun of the Pierce type, and a magnetic system that creates the homogeneous magnetic field B to accompany the electron beam and the magnetic system called "kicker" (from English *kick*). The electron gun forms a thin electron beam propagating along the axis of symmetry. The magnetic kicker system creates two local transverse magnetic fields directed perpendicular to the system's axis of symmetry, located at a distance equal to half the cyclotron wave length. The first transverse field deflects (and, more precisely, sharply rejects) the electron beam from the symmetry axis. The beam begins to rotate along the circle whose center is shifted relative to the symmetry axis by the radius of rotation. The second local transverse field imparts an azimuthal impulse to the electrons and completes the formation of the spiral electron beam rotating relative to the symmetry axis with the cyclotron frequency ω_c.

The second system includes an electron gun that forms a hollow thin electron beam and and a magnetic system with a magnetic field reversed in the region of the gun. The electron beam passing through the reverse zone receives an azimuthal

impulse and transforms into a rotating spiral beam, which rotates with respect to the symmetry axis of.

The interaction of such a beam with the resonator field, has the following important feature. The electron beam excites rotating electromagnetic field H_{mn1} and interacts effectively with it. The azimuthal index m is equal to the number of the operation harmonic of the cyclotron frequency: $m = n$ $(\omega = n\omega_c, n = 1, 2, \ldots)$.

This condition ensures the electronic selection of the oscillation types and makes it possible to use harmonics with high numbers $n = 4, 5$ as operation harmonics. At the specified frequency of the generated electromagnetic waves ω, the operation on the cyclotron frequency harmonics makes it possible to reduce the value of the cyclotron frequency $\omega_c = \omega/n$ and, hence, to reduce the required value of the magnetic field $B = (m/e)\omega_c = (m/e)(\omega/n)$. The latter becomes especially necessary in the development of gyrotrons at frequencies of f = 0.3–1 THz and higher.

Indeed, to generate electromagnetic waves with a frequency f = 1 THz (the wavelength λ = 0.3 mm) when operating at the basic harmonic of the cyclotron frequency, the required value of magnetic induction appears to be equal to B = 1.07/λ = 1.07/0.03 = 35.6 T. To obtain a magnetic field with such a value of magnetic induction in a sufficiently large volume, it is necessary to use solenoids with pulsed power.

When the gyrotron operates at the 4th harmonic, the value of the required magnetic induction is reduced to the acceptable figure of 9 T, and such a field can already be realized with the help of superconducting magnets. However, the efficiency of a gyrotron operating at the frequency of cyclotron harmonics decreases.

Gyrotrons with large orbits operating at the frequency of cyclotron harmonics are considered promising for the generation of electromagnetic waves of the T-band (f = 0.3–3 THz, λ = 1–0.1 mm).

9.3.7 Parameters and Applications of Gyrotrons

For all their advantages, gyrotrons have a rather wide spectrum of oscillations and a high noise level. Therefore, they are not used in practice in communication and radiolocation systems where a high coherence of oscillations is required.

High-power gyrotrons are widely used in units for controlled thermonuclear fusion such as the “tokamak”, and “stellarator” for plasma heating. The parameters of such devices are: a band of operation frequencies of 40–170 GHz, output power of up to 1 MW in an impulse lasting up to hundreds of microseconds with an efficiency value of 40–50%.

Most high-power gyrotrons are equipped with converters of the operating resonator mode into a paraxial wave beam. The latter is output with the aid of the mirror system through the diamond output window in a direction perpendicular to the axis of the device.

Another area of gyrotron application is in technological processes, among which are ceramics sintering, thin film modification, welding and coating. The typical

Fig. 9.9 The gyrotron developed at the Research Center of the University of Fuku (Japan)

parameters of this group of gyrotrons are a frequency of 28 GHz and an output power of 10–30 kW in continuous mode.

Devices intended for the generation of electromagnetic waves with frequencies of 0.3–1 THz and higher (wavelengths of 1–0.3 mm) with a relatively small output power level constitute the third group of gyrotrons. These waves are used as a tool for scientific research in physics, chemistry, biology, and medicine. Figure 9.9 shows a gyrotron developed at the Research Center of the University of Fuku (Japan) as an example. It has the following parameters: an output power in the continuous mode of 250 W at a frequency of 1 THz, an electron beam current of 200 mA, and an accelerating voltage of 30 kV.

9.4 Gyroklystrons

9.4.1 Gyroklystron Design

Gyroklystrons are amplifying gyro-resonant devices. The electrodynamics systems of a gyroklystron include two or more resonators (multi-cavity gyroklystrons) connected by drift tubes. Just as in conventional klystrons, the first resonator from

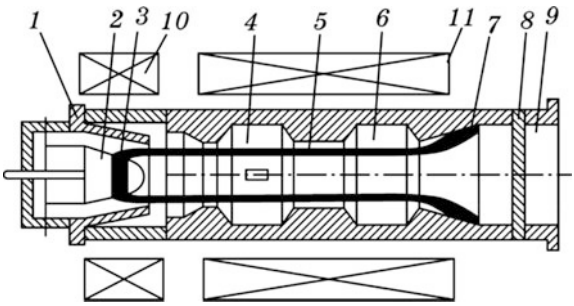


Fig. 9.10 Scheme of a gyroklystron

the cathode is the input. Then, intermediate (idle) resonators follow with the last one, being the output resonator. The purpose of the resonators is the same as in a conventional klystron. The resonators, with the exception of the last, form the cascade buncher. The last resonator takes power from the electron beam and transfers it to the load. The simplest version of the gyroklystron is the double-cavity device.

The construction of a gyroklystron differs from that of a gyrotron by the design of the electrodynamic system. Figure 9.10 shows the design of a double-cavity gyroklystron, which includes the following main components and assemblies: a cathode 2 with an emitting belt 3, the first anode 1, the input resonator 4, the drift tube 5, the output resonator 6, the collector 7, the output window 8, the section of the external waveguide 9, the solenoid of the auxiliary magnetic field 10, and the solenoid of the main magnetic field 11.

Both resonators and the drift tube are placed in the strong uniform magnetic field, the induction vector of which is parallel to the axis of the device. Therefore, along the entire electrodynamic system the motion of the electrons has the same character as in the gyrotron. An external source excites the operational mode in the input cavity. The electrons transiting the resonator interact with the electrical component of its electromagnetic field. Usually, as in gyrotrons, H_{mn1}-modes are used as operation ones. As a result of interacting with such a field, the transverse components of the electron velocities change depending on their phase of entry into the resonator. The total transverse velocity at the output from the resonator $v_\perp$ is equal to the sum of the constant and the variable components $v_\perp = v_{\perp 0} + \tilde{v}_\perp$. This is the process of velocity modulation which is the basic physical process in the input resonator.

A velocity-modulated electron beam enters the drift tube where azimuthal bunching of electrons occurs. The length of the tube is optimized to obtain the maximum values of such device parameters as gain and efficiency.

The bunched electron beam enters the output resonator and excites the H-type electromagnetic field in it. The interaction of this field with the electrons ensures the transformation of the rotational kinetic energy of the electrons into the energy of the electromagnetic field.

To increase the gain and electronic efficiency of the device, multi-cavity gyroklystrons containing three or more resonators connected by drift tubes are used.

9.4.2 Azimuthal Bunching in Gyroklystrons

In a gyroklystron drift tube, there are no alternating fields and the motion of electrons is determined by the external static magnetic field and by those velocities that electrons possess at the entrance to the drift tube.

Let's consider the process of azimuthal bunching in the kinematic approximation neglecting the action of the space-charge field. At the entrance to the drift tube, the total transverse velocity is equal to the sum of the constant and the variable components $v_\perp = v_{\perp 0} + \tilde{v}_\perp$. In the small-signal mode ($\tilde{v}_\perp \ll v_{\perp 0}$) the variable velocity component is the harmonic function of time and the azimuthal coordinate of the electrons entering the drift tube θ_1. We assume that the transverse velocity variable component depends on the azimuthal coordinate of the electrons entering the drift tube θ_1 by the law

$$\tilde{v}_\perp = v_{\perp m} \sin \theta_1, \tag{9.11}$$

where $v_{\perp m}$ is the variable velocity amplitude. The azimuthal velocities of the electrons in the drift tube are given by

$$\dot{\theta} = \omega_c = \frac{e}{m} B_0 \sqrt{1 - \frac{v_\perp^2}{c^2} - \frac{v_{z0}^2}{c^2}}, \tag{9.12}$$

where v_{z0} is the longitudinal velocity component, which is the same for all electrons if the action of the space-charge field is not taken into account. $v_\perp$ is the total transverse velocity: $v_\perp = v_{\perp 0} + \tilde{v}_\perp$.

Substituting (9.11) into (9.12), we obtain

$$\dot{\theta} = \frac{e}{m} \sqrt{1 - \frac{v_{\perp 0}^2 + 2 v_{\perp 0} \tilde{v}_\perp + \tilde{v}_\perp^2}{c^2} - \frac{v_{z0}^2}{c^2}}.$$

Assuming that $\tilde{v}_\perp \ll v_{\perp 0}$ and $\tilde{v}_\perp \ll v_{z0}$ and ignoring the square of the transverse velocity variable component, then after the transformations, we obtain the following expression for the azimuthal velocity:

$$\dot{\theta} = \omega_{c0} \left(1 - \frac{v_{\perp 0} \tilde{v}_\perp}{(1 - v_0^2/c^2) c^2} \right),$$

where

$$\omega_{c0} = \frac{e}{m} B_0 \sqrt{1 - v_0^2/c^2} \quad v_0^2 = v_{\perp 0}^2 + v_{z0}^2$$

Using this formula we can calculate the angle θ_2 which determines the azimuthal position of the electrons in some plane located at distance l from the entrance to the drift tube:

$$\theta_2 = \theta_1 + \dot{\theta}\tau, \tag{9.13}$$

where $\tau = l/v_{z0}$ is the electrons transit time from the entrance to the drift tube to the considered plane. Consequently,

$$\theta_2 = \theta_1 + \left(1 - \frac{v_{\perp 0}\tilde{v}_\perp}{1 - v_0^2/c^2}\right)\frac{l}{v_{z0}},$$

or taking into account (9.13)

$$\theta_2 = \theta_1 + \theta_0 - \frac{\omega_{c0} l}{c^2\sqrt{1 - v_0^2/c^2}}\frac{v_{\perp 0}}{v_{z0}}\tilde{v}_\perp \sin\theta_1,$$

where $\theta_0 = \omega_{c0} l/v_{z0}$ is the increment of the azimuth angle for unperturbed motion. Introducing the notation

$$X = \frac{\omega_{c0} l}{c^2\sqrt{1 - v_0^2/c^2}}\frac{v_{\perp 0}}{v_{z0}}\tilde{v}_\perp,$$

we obtain

$$\theta_2 = \theta_1 + \theta_0 - X \sin\theta_1.$$

This formula establishes the relationship between the azimuthal coordinates of the electrons at the entrance to the drift tube and in some current plane of the drift tube.

It can be seen that this relationship coincides in form with the expression determining the process of longitudinal grouping in a conventional klystron:

$$\omega t_2 = \omega t_1 + \theta_0 - X \sin\omega t_1,$$

where t_1 is the time of an electron entering the klystron drift tube; t_2 is the time of some current plane passing; $\theta_0 = \omega l/v_0$, is the unperturbed transit angle; and X is the bunching parameter.

Analogous to the way it is done in theory with a conventional klystron, the degree of azimuthal grouping in a gyroklystron can be analyzed by plotting the graphical dependencies $\theta_2 = f(\theta_1)$ at $X = \text{Const}$ as a parameter. We will consider the electrons which are uniformly distributed at the beginning of the drift tube in the interval of phases $-\pi \le \theta_1 \le \pi$.

Figure 9.11 shows the dependences of the azimuthal phases $\theta_2 = f(\theta_1)$ for four values of the bunching parameter $X = 0, 1, 1.5, 1.84$. In the absence of grouping (X = 0), parameters θ_2 and θ_1 are proportional and the electrons which are initially distributed evenly the phase intervals of $-\pi \le \theta_1 \le \pi$ are also uniformly distributed in the phase intervals of $\theta_0 - \pi \le \theta_2 \le \pi$.

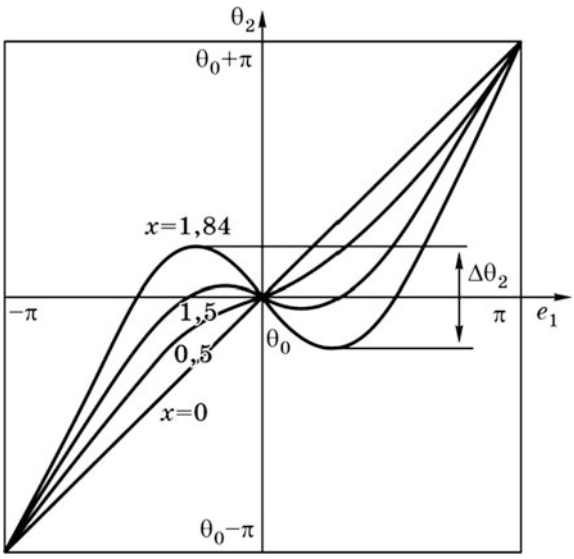


Fig. 9.11 Gyroklystron phase bunching diagram

At $X \neq 0$ the redistribution of electrons arriving at plane z take place. What is fundamentally important, is the increase that occurs in the number of electrons in a relatively narrow phase interval near the phase angle θ_0. Thus, during the drift, an azimuthal bunching of electrons occurs. For example, the bunch has an azimuthal extension of $\Delta\theta_2 = 1.3$ radians and contains about 2/3 of the charges initially uniformly distributed in the angular interval equal to 2π if the value of the bunching parameter $X = 1.84$.

To improve the bunching quality, cascade bunching can be used in multi-cavity gyroklystrons, making it possible to obtain a more compact electronic bunch with a large percentage of charge content.

The gyroklystron's output resonator can be tuned to the basic frequency $\omega = \omega_c$ or to the harmonic of this frequency $\omega = \omega_c$. In the latter case, the klystron operates in the frequency multiplication mode (the klystron—"gyro multiplier").

The use of the frequency multiplication mode makes it possible to reduce the required value of magnetic field induction by n times. The second and third harmonics are usually used as operation harmonics, as the electronic efficiency of the device decreases with increasing harmonic number.

9.4.3 Parameters and Applications of Gyroklystrons

Gyroklystrons have found application in high-precision radars, and at frequencies of the order of 90–100 GHz in particular, where there is a region with small absorption of electromagnetic waves by the Earth's atmosphere (the so-called

Table 9.1 The gyroklystron parameters

Parameters	3-cavity gyroklystron	4-cavity gyroklystron
Output power (MW)	7	1
Output signal frequency (GHz)	95	35
Input signal frequency (GHz)	45.5	17.5
Efficiency (%)	34	40
Gain (dB)	50	40
Accelerating voltage (kV)	500	82
Electron beam current (A)	45	67
Input resonator oscillation type	H_{011}	H_{011}
Output and intermediate resonators oscillation type	H_{021}	H_{021}
Input resonator operation harmonic	1	1
Output and intermediate resonators operation harmonic	2	2
Magnetic field induction in the operation area (T)	2.9	0.6

"transparency window"). High-power pulsed gyroklystrons are also used to feed charged particle accelerators.

In Table 9.1, the parameters of high-power pulsed gyroklystrons operating in the frequency multiplication mode are given. In such devices, the input resonator is excited by a signal with a frequency equal to the cyclotron frequency determined by the value of magnetic induction. The output and intermediate cascades are tuned to one of the cyclotron frequency harmonics (usually the second or the third). This mode allows a reduction in the required value of magnetic field induction.

9.5 The Gyro-TWT

9.5.1 Gyro-TWT Design

The design of a gyro-TWT is shown schematically in Fig. 9.12. In contrast to a gyroklystron, the electrodynamic system of a gyro-TWT is the segment of the regular cylindrical waveguide 4 in which a fast H-type wave is excited from the external source through the energy input 3.

Interaction of the wave field and the beam in the cyclotron resonance mode results in the modulation of the transverse velocity of the electrons, their azimuthal bunching and eventually in the conversion of part of the kinetic energy of the beam into field energy. The wave amplitude increases while moving toward the electron collector 5. The energy output to the external waveguide is performed through the vacuum window 6. The magnetic field is created by solenoids 1 and 2.

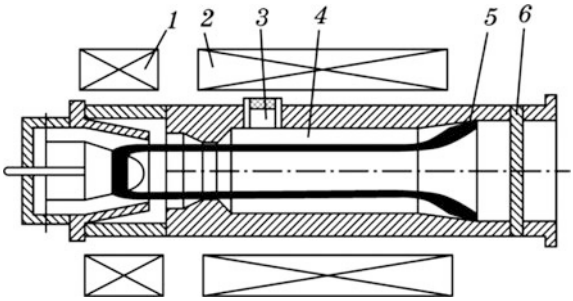


Fig. 9.12 Scheme of a gyro-TWT

9.5.2 *Features of Beam and Field Interaction*

The fundamental feature of the gyrotron's mechanism of the interaction of electrons with the transverse fields of the electrodynamic systems, is that this interaction occurs equally effectively with both standing waves of the resonator systems, and with the traveling fast wave field of waveguide systems.

Let's consider the interaction of electrons with the traveling wave field of the H_{01} mode in a cylindrical waveguide. This wave has only one transverse component of the electric field—E_θ, which alternates in time and space according to the law

$$E_\theta = E_m(r)\cos\left(\omega t - \frac{\omega z}{v_p}\right),$$

where $E_m(r)$ is the field amplitude depending on the radius; z is the longitudinal coordinate along the waveguide axis; and v_p is the phase velocity of the wave. The transverse structure of the electric field coincides with the field structure of the resonator system shown in Fig. 9.8. In the fixed plane $z = \text{Const}$, for example, $z = 0$, the high-frequency field alternates in time according to the harmonic law

$$E_\theta = E_m(r)\cos(\omega t).$$

Let's first consider the electron beam in which the electrons have only transverse (rotational) velocities ($v = v_\perp, v_\parallel = 0$) as in Sect. 9.2.2. The electric field of the traveling wave, which acts on electrons rotating in the fixed plane, $z = \text{Const}$, alternates in time according to the harmonic law and for them it is indistinguishable from the field of the standing wave. Consequently, the mechanism of the field interaction with the electrons analyzed earlier in Sect. 9.2.2 is completely valid for the case under consideration. When the cyclotron resonance condition $\omega = \omega_c$ is satisfied, the processes of velocity modulation and azimuthal bunching proceed in a similar manner. The condition $T = T_c$ ensures phase synchronism, which consists of the fact that the electron which performed the complete revolution "meets" the same phase of the field again that it "saw" at the beginning of the revolution.

Suppose now that the electron has a longitudinal velocity $v_{\parallel}$ in addition to transverse velocity $v_{\perp}$. At the beginning of the revolution, the electron is in the $z = 0$ plane and "sees" the field in the phase

$$E_{\theta} = E_m(r)\cos(\omega t).$$

For the time interval $\Delta t = T_c$, the electron will perform a complete revolution and will move along the waveguide axis by distance $z = v_{\parallel} T_c$. The phase of the wave which it will meet at the end of the revolution is determined by the formula:

$$E_{\theta} = E_m(r)\cos\left(\omega t_0 + \omega T_c - \omega T_c \frac{v_{\parallel}}{v_p}\right).$$

In order for this phase to coincide with the initial phase and for the phase synchronism condition to be satisfied, it is necessary that the equality

$$\omega T_c - \omega T_c \frac{v_{\parallel}}{v_p} = 2\pi n, \quad n = 1, 2, \ldots,$$

is ensured. From this, we find

$$T_c = \frac{nT}{1 - v_{\parallel}/v_p} \tag{9.14}$$

and

$$\omega = \frac{n\omega_c}{1 - v_{\parallel}/v_p}. \tag{9.15}$$

Conditions (9.14), and (9.15) ensure synchronism of electrons moving along the spiral phase with the high-frequency field and can be considered as the cyclotron resonance condition in the case of the interaction of electrons with the traveling wave field.

The case $n = 1$, $\omega = \omega_c/(1 - v_{\parallel}/v_p)$ corresponds to the fundamental cyclotron resonance, i.e., the resonance at the cyclotron frequency first harmonic.

In a gyro-TWT with an electrodynamic system in the form of a smooth circular waveguide, the waves of H_{11}, H_{01}, H_{21} modes are used as operating waves. The parameters in the centimeter wavelength band are as follows: the output power in the mode of microsecond pulses is tens of kilowatts, the efficiency is 20%, the gain is 20–25 dB, and the operation frequency band is about 10%.

However, a gyro-TWT with a smooth waveguide has not been practically applied due to the number of significant disadvantages: the inclination to self-excitation, criticality to the spread of electron velocities, relatively low values of gain, and the operation frequencies band.

A significant improvement in the parameters and performance is expected from gyro-TWTs with an electrodynamic system in the form of a cylindrical waveguide

Table 9.2 Parameters of a gyro-TWT with a helical waveguide

Parameters	Gyro-TWT of the cm-band	Gyro-TWT of the mm-band
Output power (MW)	1.1	0.18
Output signal frequency (GHz)	9.4	36
Efficiency (%)	29	26
Gain (dB)	50	40
Accelerating voltage (kV)	200	80
Electron beam current (A)	25	9
Frequency band (%)	21	9
Pulse width (μs)	0.12	12
Magnetic field induction in the operation area (T)	0.34	1

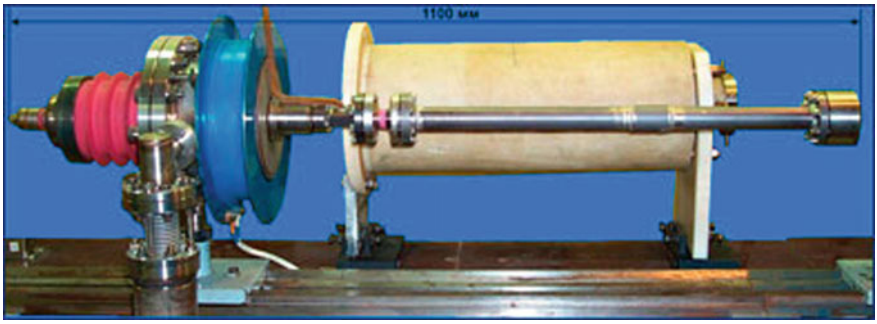


Fig. 9.13 A Gyro TWT with a helical waveguide

with a helical (screw) corrugation of the inner surface. Such a system provides a wider band of amplified frequencies, resistance to self-excitation, and less criticality to the spread of electron velocities. Such a device is termed a gyro-TWT with a helical waveguide. Table 9.2 gives the parameters of an experimental pulsed gyro-TWT with a helical waveguide developed at the Institute of Applied Physics at the Russian Academy of Sciences (RAS), under the guidance of S.V. Samsonov (Fig. 9.13).

9.6 The Gyro-BWO

The experiments performed at the Institute of Applied Physics (IAP) of the RAS (Nizhny Novgorod) show that waveguides with a helical inner corrugation make it possible to create a gyro-BWO with an output power of hundreds of megawatts per

Table 9.3 Parameters of a gyro-BWO

Parameters	Output characteristics					
Center frequency, GHz	35	34.9	35.1	93.2	93.2	91.6
Gyro-frequency harmonic	1	1	2	1	1	1
Output power, MW	750	200	260	300	100	2.5
Gain, dB	20	22	17	25	30	32
Instantaneous bandwidth, GHz	0.24	0.6	0.035	0.4	0.92	0.33
Efficiency, %	24	33	18	26	30	25
Pulse width, μs	100	10	100	100	100	Cont. mode
Manufacturer	IAP RAS	NRL	IAP RAS	IAP RAS	NRL/ CPI	GYCOM (Russia)

Fig. 9.14 A gyro-BWO

pulse (Table 9.3). These devices have a high efficiency and a sufficiently wide band of smooth electronic frequency tuning. For example, at an accelerating voltage of 600 kV and a current of 4 kA, an output power of 600 MW is obtained. At this, a change of generation frequency was registered, both within the pulse (the pulse was not perfectly rectangular, the voltage in the impulse decreased from the beginning of the impulse to its end) and from pulse to pulse. When the voltage changed from 600 kV at the beginning of the impulse, to 200 kV at its end, the frequency varied smoothly from 10 to 9.5 GHz. One of these tubes is shown in Fig. 9.14.

Advancement Questions

1. Describe the history of the creation of gyro-resonant devices.
2. What relativistic effect is used in gyro-resonant devices?
3. Why do gyro-resonant devices have larger electrodynamic system dimensions compared with O-type and M-type devices?
4. Describe the phenomenon of cyclotron resonance.
5. Describe the phenomenon of electron beam azimuthal bunching.

6. Describe the design and operating principle of a gyrotron.
7. What are the abridged motion equations? What processes do they describe?
8. Explain the mechanism of electron beam interaction with the high-frequency field on cyclotron frequency harmonics.
9. What factors determine a gyrotron's efficiency?
10. What factors determine a gyrotron's starting current?
11. Explain the design and operating principle of a large-orbit gyrotron.
12. Describe the design and operating principle of s gyroklystron.
13. Describe the design and operating principle of a gyro-TWT.
14. Describe the design and operating principle of a gyro-BWO.

Chapter 10
Relativistic Microwave Devices

10.1 General Characteristics of Relativistic Microwave Devices

Relativistic microwave devices are a large group of generators and amplifiers of microwave radiation in which the relativistic effects are noticeably manifested: the dependence of the electron mass on energy as well as the normal and anomalous Doppler effects. Conditionally, the devices in which the kinetic energy of electrons is commensurable with their rest energy, i.e., in which the accelerating voltage is 200 kV or more (let's recall that the rest energy of an electron is 511 keV) are called relativistic.

Relativistic microwave devices are divided into two groups. The first consists of classical devices (klystrons, klystrodes, TWTs, BWOs, and magnetrons) operating at accelerating voltages of over 200 kV. The relativistic effects influence the mechanism of interaction in these devices, which makes it necessary to modify the interaction theory and their design. The second group consists of devices using specific interaction mechanisms based on the relativistic effects. These devices include free-electron lasers, gyrocons, magnicons, vircators and their modifications.

The article by Ginzburg (1947), in which the possibility of generating coherent oscillations of microwave band wavelengths by rectilinearly moving relativistic electron fluxes and relativistic Doppler oscillator fluxes was shown, heralded the origins of relativistic microwave electronics. However, the real development of relativistic electronics began in the late 1970s, when high-current electron accelerators were created.

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_10

10.2 Classical Relativistic Devices

10.2.1 Relativistic Klystrons

The desire to increase the output power of klystrons resulted in the increase of the accelerating voltage up to several hundred kilovolts. With such voltages, the efficiency of velocity modulation decreases, since under the action of the high-frequency voltage in the gap, the electron mass changes rather than its velocity.

This effect requires the use of appropriate bunching designs, an electron gun, a focusing system and a collector. In particular, with an unchanged number of resonators, and with increasing accelerating voltage, the gain of the device decreases. It is difficult to realize guns with a hot cathode for accelerating voltages above 500 kV, therefore a high-current linear induction accelerator (LIA) is used as the electron source. As an example, let's show the parameters of the SL4 relativistic klystron (Table 10.1) developed in the SLAC laboratory (USA).

As can be seen, this device is capable of producing power of 200 MW in short impulses with a duration of 40 ns. The electron beam was injected into this klystron from the LIA. In the SL4 klystron, a single-gap output resonator is used, which results in a decrease of the output power pulse duration at power levels of more than 100 MW. This phenomenon is apparently connected with plasma formation in the gap of the output resonator, and the use of multi-gap resonators is thought to allow the elimination of this undesirable phenomenon. Information exists concerning a relativistic klystron with an output power of 1 GW and a pulse duration of 15 ns. The device operated with a high-current accelerator with an electron energy of 2 meV, and a current of 1 kA.

Further increases of the accelerating voltage make the classical bunching mechanism ineffective, since all electrons move with practically identical velocities. As an alternative, bunching in the rotating magnetic field (RMF) can be used (Fig. 10.1). The bunching section consists of the resonator 1 modulating the electron beam energy, the annular drift tube 2 located in the magnetic field, and the output resonator 3.

The radius of the rotation of an electron in the magnetic field is determined by its mass:

$$R = \frac{m_0}{eB}\sqrt{\gamma^2 - 1}$$

Table 10.1 Parameters of the SL4 klystron

Frequency (GHz)	Output power (MW)	Gain (dB)	Efficiency (%)	Accelerating voltage (kV)	Beam current (A)	Number of cavities	Device length (cm)
8.57	200	65	50	1000	750	6	98

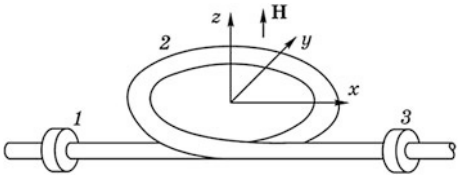


Fig. 10.1 Rotating magnetic field buncher

where γ is the relativistic factor. Electrons with higher energy move along the larger radius, and since the velocities of all electrons are approximately the same, this electron spends more time per revolution, and, vice versa, an electron passing through the modulating resonator at the moment of the decelerating field has less energy and a smaller turning radius. As a result, it takes less time for a revolution. Thus, electrons are grouped around the electron that has passed through the modulating resonator at the moment of the transition of the voltage on it from accelerating to decelerating.

The analysis shows that, as in the classical klystron, the amplitude of the n-th harmonic of the convection current at the exit from the ring can be calculated from the formula

$$I_i = 2I_0 J_1(nX),$$

where

$$X = v\theta$$

is the bunching parameter,

$$v = M\frac{U_m}{2U_0}\left(\frac{\gamma+1}{2\gamma^2}\right)^{1/2}$$

is the parameter referring to the energy modulation coefficient, U_m is the amplitude of the voltage on the modulating resonator, U_0 is the accelerating voltage, M is the interaction coefficient, $\theta = 2\pi R\beta_e$ is the transit angle in the rotating field, and $\beta_e = \omega/v_e$ is the electron propagation constant. The comparison of this bunching mechanism with the classical mechanism shows that the ratio of the gap voltages providing the same bunching factor for the two types of bunching is determined by the formula

$$\frac{U_m^{rmf}}{U_m^{kl}} = \gamma^2 - 1,$$

i.e. the efficiency of the bunching in the RMF is increasing as the energy of the electrons increases.

10.2.2 Relativistic TWTs and BWOs

As the accelerating voltage increases, the electron velocity approaches the velocity of light and a slight slowing is required for electromagnetic wave for syncronism with the electron beam. Such a slowing can be provided by the corrugated waveguides with small corrugation depth or by waveguides with diaphragms. The scheme of such a device developed at the Lebedev Physical Institute of the Russian Academy of Sciences (LPI RAS) is shown in Fig. 10.2. The magnetron gun 1 forms the electron beam 2 propagating in the magnetic field produced by the solenoid 3. The output power is brought out through the vacuum window 4. The length of the SWS section is 12 cm, the inner radius is 1.6 cm, and the depth of the corrugation is 0.4 cm. The dispersion characteristic of this SWS is shown in Fig. 10.3 (curve 1).

The same graph shows the electron line $2\omega = \beta_e v_e$ where v_e is the velocity of the electrons. The intersection of this line with the dispersive characteristic determines the frequency at which synchronism is observed. Usually, the intersection point (point of synchronism) in a relativistic TWT is selected near to the upper cutoff frequency (frequency of the π-mode) so that the electron beam interacts with both the forward and backward space harmonics. Therefore, such devices are called the relativistic TWT-BWO. The interaction with the backward wave provides the feedback necessary for self-excitation of the device. Therefore, such tubes are used as generators. The parameters of some relativistic TWT-BWOs are given in Table 10.2.

The basic relationships for determining the parameters of the device can be written as follows:

$$\frac{2}{\pi}\frac{e}{m_0 c^2}\gamma^3 R_c I_0 = 1,$$
$$\frac{\pi L n_e}{\lambda \gamma^2} = 1.2,$$
$$\gamma^2\left(\frac{v_p}{c} - n_e\right) = 2.25 n_e.$$

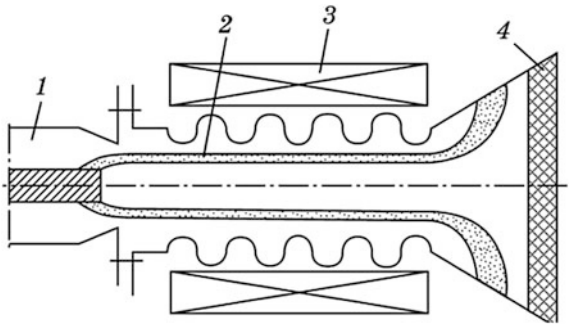


Fig. 10.2 Scheme of a relativistic TWT

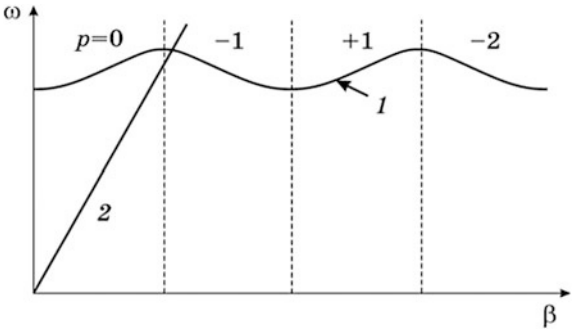


Fig. 10.3 Dispersive characteristic of the corrugated waveguide

Table 10.2 Parameters of a relativistic TWT-BWO

Designer	F (GHz)	Wave mode	U_0 (Kv)	I_0 (kA)	τ_u (ns)	P (MW)	η (%)
LPI	9.7	E_{01}	600	4	15	350	15
USA	10.3	E_{11}	500	3	20	500	17
USA	10	E_{02}	1400	10	1400	200	10

In these expressions, $n_e = c/v_e$ is the electron velocity deceleration, L is the SWS length, and λ is the radiation wavelength.

At present, large-scale studies on the creation of high-power millimeter wave generators based on the Cherenkov effect are being carried out at the Institute of Applied Physics at the Russian Academy of Sciences (IAP RAS) and at the Aragon National Laboratory (USA). New technical solutions for the formation of short electron packets (impulses) with durations in the region of several picoseconds and a charge content of 10–50 nC are being used in these works.

The initial batching is carried out in a high-frequency photoinjector (Fig. 10.4).

The source of the electrons in the injector is the photocathode 1 which is irradiated by pulsed laser radiation of high intensity. The cathode is built in the cavity 2 excited from the external source. The laser impulse 3 irradiates the photocathode at the time when the alternating electric field E near the cathode surface is accelerating and reaching its amplitude value. The electrons emitted by the photocathode are accelerated by this field and a short-time convective current impulse (the electronic bunch) is formed in the photoinjector. The duration of the convection current impulse can be controlled by varying the laser pulse duration. To limit the transverse dimension of the electron bunch, the magnetic field which is created by the solenoids located on both sides of the photoinjector is used.

Under the action of the accelerating alternating field and the magnetic field, the electrons move in the direction of the output channel 5 inside the region bounded by the line 4. The generated electron bunch is also accelerated in the linear accelerator, and then passes through the system of quadrupole magnetic lenses which correct its

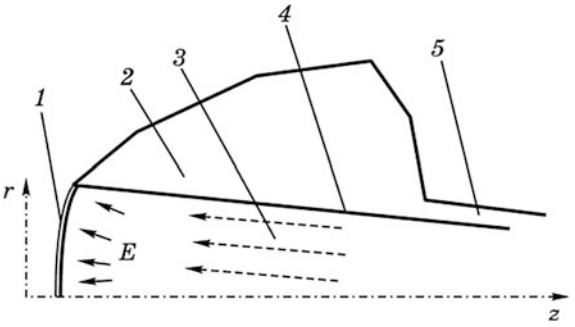


Fig. 10.4 High-frequency photoinjector of the Cherenkov generator

transverse dimensions and phase characteristic before it is directed to the electrodynamic system. The latter is a cylindrical waveguide partially filled with the dielectric. Passing through the channel in the dielectric, the convection current impulse excites the Cherenkov radiation. The intensity of the radiation depends on the magnitude of the charge that this impulse carries.

The Cherenkov generator developed at the Aragon National Laboratory has an output power of 3.5 GW at a frequency of 7.8 GHz. The beam energy is 16 meV, and the pulse width is 4 ps. The radius of the cylindrical waveguide is 11 mm, the radius of the channel in the dielectric is 6 mm; the length of the channel is 100–200 mm, and the dielectric materials are steatite with dielectric permittivity $\varepsilon = 5.8$ and borosilicate glass $\varepsilon = 4$.

IAP RAS (Russia) developed Cherenkov generators and amplifiers with output power up to 2 GW at a frequency of 10 GHz, and 600 MW at a frequency of 30 GHz.

10.2.3 Relativistic Magnetrons

It is also possible to create relativistic magnetrons on the basis of high-current accelerators. Since accelerators produce a current impulse of short duration (tens of nanoseconds), it is required to ensure that the duration of the transient period is in units of nanoseconds. In addition, when a high voltage is applied to the anode, plasma is formed near the cathode and it is necessary to minimize its effect on the operation mode of the magnetron. The setting time of the oscillations can be reduced by lowering the Q-factor of the magnetron resonators. For this purpose, the anode unit is not made of copper, but of materials with lower conductivity, for example, stainless steel or invar.

Experiments have shown that the cathode plasma does not disrupt the magnetron's operation until it completely covers the cathode–anode gap. The overlap time is more than 100 ns.

According to the synchronism conditions, the limiting electronic efficiency of a relativistic magnetron can be estimated by the formula

$$\eta_e = 1 - \frac{2m_0c^2}{eU_0}\left(\gamma^2 - 1\right).$$

It follows from this formula that at U_0 = 1 MV, an electron efficiency of about 70% can be obtained. The real efficiency of relativistic magnetrons is lower, since they use a magnetic field with induction below the optimal value.

Figure 10.5 shows the design of a relativistic magnetron developed at the Institute of Applied Physics at the USSR Academy of Sciences in the late 1970s. It contains a graphite cold cathode 1, an anode 2 with 8 slot resonators, an irregular waveguide segment 3, a vacuum window 4 and a solenoid 5. A feature of this magnetron is the diffraction energy output which, in addition to the increasing electrical strength, allows excitation of the operation wave mode in the output waveguide (in this case, H_{41}). For this purpose, the depth of the slots in the anode unit decreases smoothly to zero toward the transition to the regular waveguide.

The magnetic system of relativistic magnetrons must provide the necessary value of magnetic field induction which increases with increasing anode voltage. To obtain the required induction, solenoids or superconducting magnets are usually used.

At the present time, the relativistic magnetron is the simplest and most reliable device allowing the use of pulses at the gigawatt power level (Table 10.3).

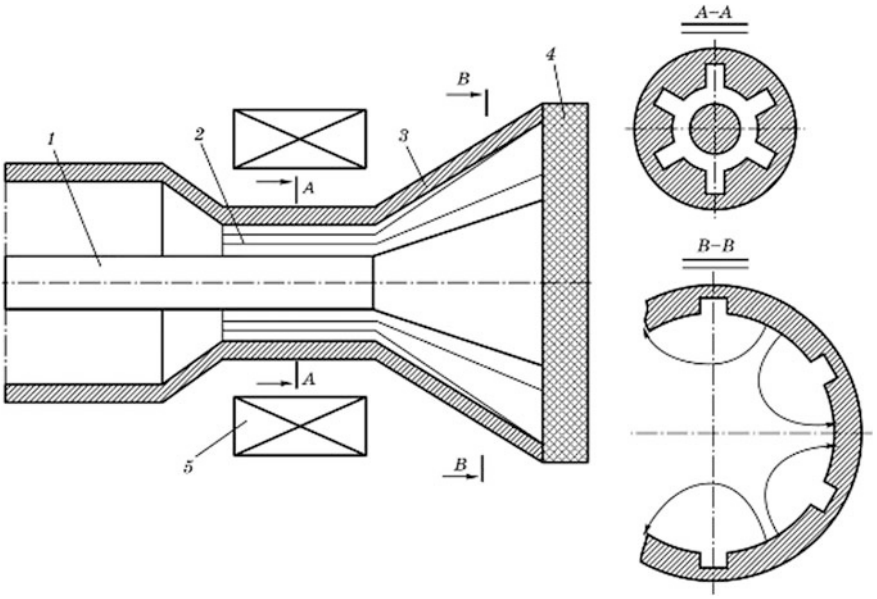


Fig. 10.5 Relativistic magnetron

Table 10.3 Parameters of relativistic magnetrons

Designer	Number of resonators	Wavelength (cm)	Anode voltage (MV)	Output power (GW)	Pulse width (ns)	Efficiency (%)
IAP RAS	8	3.3	0.57	0.5	20	15
Tomsk Polytechnic University (TPU)	6	10	1	10	30	43
CPI (Canada)	6	10	0.9	4.5	–	25
SLAC (USA)	6	10	0.36	1.7	20	36

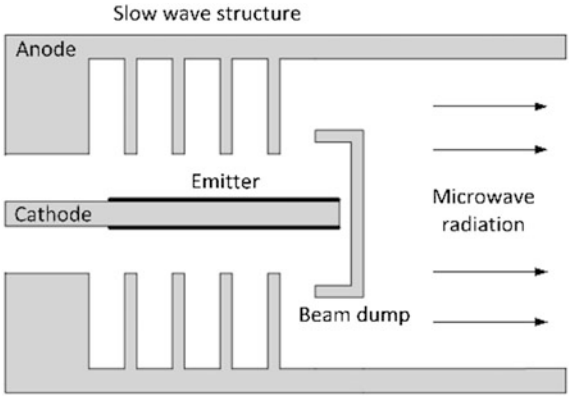


Fig. 10.6 Scheme of the MILO

Unfortunately, these devices cannot effectively operate at frequencies above 10 GHz because the dimensions of the anode unit become too small, causing breakdowns and difficulties with heat removal.

Another type of relativistic magnetron is the MILO (Magnetically Insulated Line Oscillator). It is basically a linear magnetron, see Fig. 10.6. The big difference compared to an ordinary cylindrical magnetron is that the MILO does not require an external magnetic field. The current itself generates an axial magnetic field that prevents the electrons from short-circuiting the anode–cathode gap. To generate a self-insulating magnetic field, an applied voltage of >500 kV is needed.

10.3 Free-Electron Lasers

10.3.1 Working Principle of Free-Electron Lasers

A free-electron laser (FEL) is an electronic device designed to amplify and generate electromagnetic waves, mainly in the submillimeter and X-ray regions of the

spectrum. A FEL working in the microwave band is often called a Free-Electron Maser (FEM). The working principle of the FEL is based on the interaction of a relativistic electron beam with a spatially periodic electric or magnetic field. In such a field, electrons move along complex curvilinear trajectories, performing translational and periodic transverse motions.

In these devices, the kinetic energy of electrons is transferred to electromagnetic wave energy due to their interaction with the transverse electric field component of the wave, which propagates in the waveguide structure or in free space. This process can be regarded as the forced oscillator radiation of an ensemble of electronic oscillators. Since the oscillators move in the longitudinal direction at near-light speed, a strong Doppler effect is observed in the FEL, which greatly increases the frequency of the radiation relative to the frequency of the electron oscillations.

Periodic magnetic field action causes transverse periodic electron motion. Such motion can also be achieved by means of counter propagating an electromagnetic wave tagged as a "pump wave". The frequency of this wave can be much lower than the frequency of the generated electromagnetic wave.

FELs with a periodic magnetic field are sometimes called relativistic ubitrons, since they use the principle of electron interaction with a wave field, first achieved in a centimeter-band electronic device, called the ubitron.

10.3.2 The Ubitron—The Predecessor of the FEL

An ubitron (ubitron—Undulator Beam Interaction) is an amplifying electronic device in which an electron beam interacts with a fast electromagnetic wave propagating in a waveguide. The ubitron was first described by Phillips (USA) in 1960. Its design is represented schematically in Fig. 10.7. An electron gun 1, creates and accelerates an electron beam. A waveguide system 3, with input 2 and output waveguides 5, ensures the propagation of a fast wave. The magnetic system, including permanent magnets 4, creates a periodic magnetic field distribution, and is called an undulator or a wiggler. The electron collector 6, receives the electron beam.

Figure 10.8 shows the undulator magnetic system formed by a number of prismatic magnets. The arrows point in the magnetization direction of the prismatic magnets. The function $B_x(x)$ in such an undulator is of a periodic nature and can be approximated using the formula

$$B_x = B_m \sin \frac{2\pi}{b} z,$$

where b is the magnetic field period of the undulator; and B_m is the amplitude of the magnetic field.

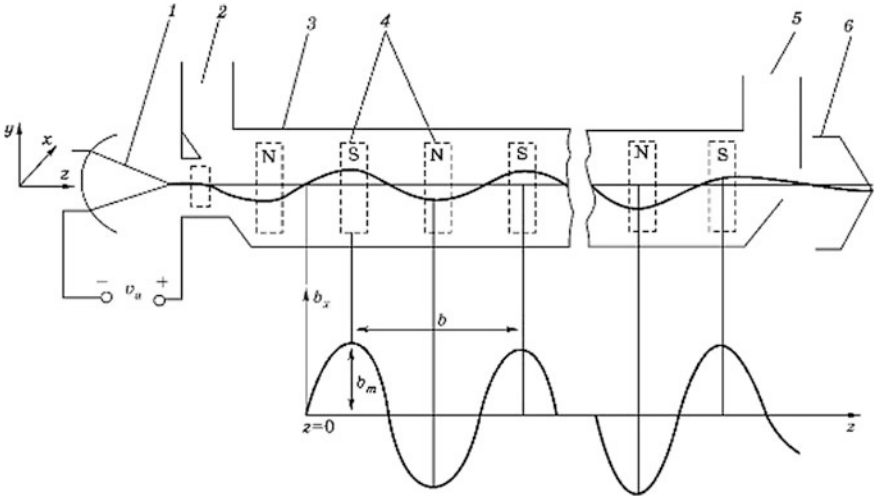
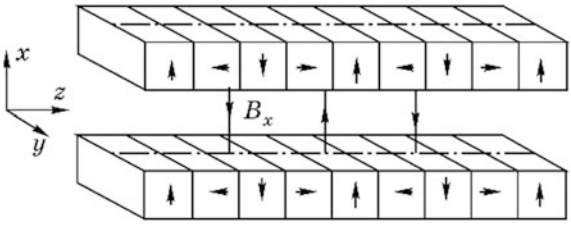


Fig. 10.7 Scheme of Ubitron design

Fig. 10.8 An undulator



The motion of electrons in the undulator field occurs along periodic trajectories, as shown in Fig. 10.7. If we assume that the electron velocity component v_y is much smaller than the longitudinal velocity v_z, then the following formulas can be obtained for the velocity and displacement of electrons in the transverse direction:

$$v_y = \frac{e}{m}\frac{B_m b^2}{2\pi\gamma} v_z \cos\frac{2\pi}{b} z, \tag{10.1}$$

$$y = \frac{e}{m}\frac{B_m b^2}{4\pi^2\gamma v_z} \sin\frac{2\pi}{b} z, \tag{10.2}$$

where v_z - avarage longitutional electron velocity, γ is the relativistic factor.

The electromagnetic field with which the electron beam interacts has the form of a traveling wave that propagates in a smooth waveguide with a phase velocity $v_p > c$. The given ubitron uses a rectangular waveguide and H_{01} mode as the working wave having a transverse component of electric field E_y. In this case, the interaction of the beam electrons with the field occurs only because of the

transverse motion of the electrons. Thus, the ubitron belongs to the class of devices with transverse interaction.

Transfer of the kinetic energy of an electron to field energy occurs only when the interaction of the electrons and the field is ordered. In O-type travelling-wave tubes, ordered interaction occurs due to the synchronous motion of electrons and the field at approximate equality of electron velocity v_{ez} and phase wave velocity v_p. (In gyro-resonant devices, such an interaction occurs in the cyclotron resonance mode).

Since the electromagnetic wave propagates in the smooth waveguide with phase velocity $v_p > c$, and velocity of the beam electrons $v_z < c$, in the ubitron the synchronism between the wave and the beam cannot occur in the exact meaning of this term. At a phase velocity exceeding the beam velocity, $v_p > v_{ez}$, the wave slides along the beam, ahead of its motion along the axis of the device. However, even in this case, it is possible to find a wave and beam motion mode, where the interaction between the field and the beam is ordered, and part of the beam's electrons is accelerated and part is decelerated.

This kind of interaction is ensured if in time T_b, while the electrons move a distance equal to the period of undulator field b, the wave moves along the device's axis for a distance equal to

$$\Delta z = v_p T_b = b + \lambda_g, \tag{10.3}$$

where λ_g is the wavelength (spatial period) of the HF electromagnetic field in the waveguide.

We shall follow the motion of an electron that enters the regular part of the undulator (plane $z = 0$, Fig. 10.9) at time t_0.

In accordance with formulas (10.1) and (10.2), the transverse coordinate of the electron is $y = 0$, and its transverse velocity v_y is maximum, $v_y = v_m$. We assume that at this moment, the strength of the transverse field of the electromagnetic wave reaches its amplitude value $E_y = E_m$ (Fig. 10.9). The electron moves along the modulator following a zigzag curve $y(z)$, the arrows on it show the direction and relative values of the transverse velocity of the electron. At the point where the electron enters the undulator, the transverse electric field has maximum value and its direction coincides with the direction of the transverse velocity.

Hence, an electron experiences maximum deceleration in the electric field. The electromagnetic wave shifts to the right at a speed greater than the speed of the

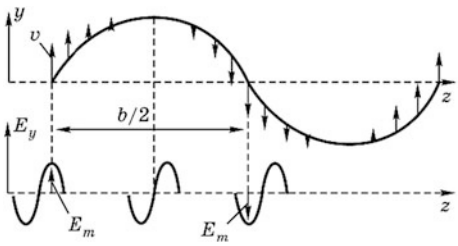


Fig. 10.9 Analysis of electron beam motion in the undulator

moving electron, so the value of the transverse field acting on the electron gradually decreases, but everywhere its direction coincides with the direction of the transverse velocity. When the electron is at a distance from the entry point equal to a quarter of the period of the undulator field ($z = b/4$), the electromagnetic wave precedes it by a distance equal to a quarter of the wavelength. At this moment, the field acting on the electron is nullified and then changes direction.

Consequently, on the entire ascending part of the trajectory, up to plane $z = b/4$, the electron experiences deceleration. On the descending section of the trajectory, the absolute value of the field increases, reaching a maximum on the $z = b/2$ plane and its direction coincides with the direction of the transverse velocity. Thus, on the descending section of the trajectory, the electron again experiences deceleration. Tracing the electron and field motion in section $z = b/2 \ldots b$, it can be seen that here again, the considered electron experiences deceleration and its transverse velocity decreases.

We shall now consider the interaction between the electromagnetic field and the electron entering the undulator later, when the field upon entrance assumes a maximum negative value. We shall trace the electron and field motion during one undulator period. Considering that the field moves ahead of the electron, it can be established that throughout the entire undulator field period, the directions of the field vectors and the transverse velocity are opposite. Consequently, the electron is accelerated by the field, and its velocity increases.

Considering the motion of the electron entering the undulator in the intermediate (zero) phase of the field, we find that the electron accelerates in certain parts of its trajectory and decelerates in other parts. On average, for the period of the undulator field, its velocity remains unchanged.

Thus, the analysis performed shows that the if condition (10.3) is satisfied, the interaction between the beam and the field is of an ordered nature, in which, upon moving along the undulator, part of the electrons increases its transverse velocity (is accelerated), and part reduces it (is decelerated). Consequently, the condition expressed by this formula can be regarded as a generalized condition of synchronism.

The transverse velocity modulation of the electron beam, arising because of the action of an alternating electric field, further leads to electron bunching. In the observed case, the electron bunch is formed around the electron, which on average does not change its velocity during the period of the undulator field. Gradually, both electrons moving in front of it with a decreased velocity, and electrons moving behind with an increased velocity, shift towards it.

The formation of such an electron bunch, however, does not lead to the effective conversion of electron kinetic energy to field energy, since the electrons composing the bunch, just like electron 2, will periodically accelerate and decelerate, eventually leading to a zero energy balance.

In order to effectively transfer energy from the electrons making up the bunch to the alternating field, it is necessary for the bunch to gradually shift to the zone of decelerating field. This is achieved by refusal of the exact synchronism mode expressed by the formula (10.3).

Assuming that $T_b = b/v_z$, we rewrite the synchronism condition as follows:

$$v_z = v_p(1 + \lambda_g/b)^{-1}.$$

For effective energy exchange between the bunched beam and the field, it is necessary for the beam velocity to be somewhat higher than the synchronous velocity determined by this formula, i.e. $v_z > v_p(1 + \lambda_g/b)^{-1}$.

Ubitron prototypes had a pulse output power of 1.6 MW at a frequency of 16 GHz, and 150 kW at a frequency of 54 GHz, with an accelerating voltage of 150 and 70 kV respectively, and a beam current of 400 A. However, these devices were abandoned since they had no advantages when compared to gyrotrons.

10.3.3 The FEL—Relativistic Ubitron-Self-Oscillator

The idea of using a relativistic electron beam and the Doppler effect to increase radiation frequency was expressed in 1968 by Pantell and Soncini. The first device using these phenomena and called the *free-electron laser* (FEL) was created in 1976 in Stanford. Later, the design and parameters of the FEL were quickly improved.

A scheme of the FEL device is shown in Fig. 10.10. It includes: an undulator formed by a system of permanent magnets 3, an open Fabry-Perot resonator formed by mirrors 2 and 4. Mirror 4 is translucent and part of the electromagnetic energy 5 stored in the resonator is radiated through it.

Electron beam 1 enters the undulator and exits it by special deflecting systems. While moving through the undulator it interacts with the wave field of the open resonator.

The resonator field forms a standing wave. Its boundaries are indicated in Fig. 10.10 by dashed lines. The standing wave can be represented as the sum of two traveling waves: direct (incident) and inverse, moving towards the electron beam. The velocities of these waves are equal to the free space speed of light. At certain values of electron beam velocity (energy), the distance between the mirrors and other parameters of the FEL it will generate coherent radiation at a frequency that

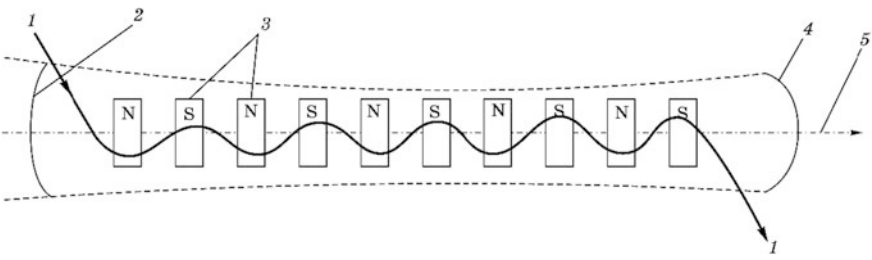


Fig. 10.10 Scheme of the free-electron laser

can be found from the synchronism condition of electron beam motion and a direct electromagnetic wave determined by the above-mentioned formula (10.3).

We transform this formula, bearing in mind that in this case the direct wave moves along the undulator axis at the speed of light $v_p = c$, and its spatial period is equal to the wavelength in free space $\Lambda = \lambda = cT$:

$$c(T_b - T) = b,$$

or, bearing in mind that $b = T_b v_z$, we find

$$\frac{2\pi}{T} = \frac{2\pi}{T_b}\frac{c}{(c - v_p)}.$$

The value $2\pi/T_b$ is the angular frequency of electron beam oscillations in the magnetic field of the undulator. We shall denote this frequency by symbol ω_b and call it, in accordance with the adopted terminology, the "bounce frequency" $\omega_b = 2\pi/T_b$. Considering that $2\pi/T = \omega$, the preceding formula can be rewritten as follows

$$\omega = \omega_b \frac{1}{1 - v_z/c}$$

or

$$\omega = \omega_b \frac{1 + v_z/c}{1 - (v_z/c)^2} = \omega_b \gamma^2 (1 + v_z/c). \tag{10.4}$$

This formula expresses the condition for electron and wave motion synchronism and determines the frequency of the oscillations generated. At ultra-relativistic electron velocities, it takes the following form $\omega = 2\gamma^2 \omega_b$.

Two important practical consequences derive from the formula (10.4). First, it implies that the frequency of the generated oscillations can be changed by changing the velocity (energy) of the electrons, that is, electronic frequency tuning is possible. Secondly, for a given undulator period b, the frequency of the generated oscillations will increase sharply when shifting to relativistic and ultra-relativistic electron velocities.

Table 10.4 illustrates the last point. It shows frequency ω, calculated by (10.4) for b = 12 cm, which is equal to the ubitron undulator period described in Sect. 10.3.2. The calculations show that the transition to relativistic electron

Table 10.4 Dependence of FEL parameters on electron energy

W (MeV)	f_b (GHz)	γ	F (GHz)	Λ (mm)
0.1	1.33	1.2	3	100
50	2.5	100	5×10^4	0.006
100	2.5	200	2×10^5	0.0015

energies provides the possibility of generating electromagnetic waves from infrared and visible portions of the spectrum without any reduction in the linear dimensions of the main FEL parts.

10.3.4 Analysis of Radiation Processes in the FEL

The above expressions for the generated frequencies can be obtained directly, based on the physical concepts of a radiating electron.

To analyze the radiation of electrons passing through the undulator, it is reasonable to use a coordinate system moving along the undulator axis z together with the electron at velocity v_{ez}. For brevity, we shall henceforth call this coordinate system, the *e*-system. An observer in this coordinate system perceives the undulator field as an incident electromagnetic wave field.

The components of this wave can be found by means of Lorentz transformations. In the laboratory coordinate system (*L*-system), in which the undulator is at rest, the field has only one component

$$B_x = B_m \sin(2\pi z/b).$$

Applying the Lorentz transformation, we find that in a moving coordinate system (*e*-system) the field will have both magnetic and electrical components:

$B'_x = \gamma B_x$ и $E'_y = \gamma v_z B_x$.

In the *e*-system, these fields vary in time at frequency ω'_b defined by the following formula

$$\omega'_b = 2\pi v_z/b',$$

where b' is the spatial period of the undulator field, calculated using Lorentz length contraction: $b' = b/\gamma$.

Considering that $\omega_b = 2\pi v_z/b$, we obtain

$$\omega'_b = 2\pi\gamma v_{ez}/b' = \gamma\omega_b.$$

Field E'_y, acting on the electrons that were initially at rest in the *e*-system, leads to the buildup of their oscillations at frequency $\omega'_b = \gamma\omega_b$. The oscillating electrons create electromagnetic radiation at frequency $\omega' = \omega'_b$ in the *e*-system. An observer (receiver) in the *L*-system perceives its radiation as the radiation of a source moving in relation to him. If the observer is located such that the radiation source is approaching him at velocity v_{ez}, then according to the relativistic Doppler effect, radiation frequency amounts to

$$\omega = \omega'\gamma(1 + v_z/c).$$

Table 10.5 Parameters of an FEL-self-oscillator

Electron beam	
Beam energy (MeV)	43
Relativistic factor (γ)	85
Pulse current (A)	1.3
Pulse duration (ps)	3
Spatial pulse length (mm)	1
Spatial pulse repetition period (mm)	25.4
Beam radius (mm)	0.25
Undulator	
Undulator length (m)	5.3
Magnetic field period (cm)	3.3
Field amplitude (T)	0.23
Resonator and radiation parameters	
Distance between mirrors (m)	12.7
Mirror curve radius (m)	7.5
Radiation wavelength (μm)	3.3
Radiation spot size (mm)	1.6

Since $\omega' = \gamma\omega_b$, the expression for radiation frequency ω is as follows

$$\omega = \omega_b \gamma^2 (1 + v_z/c), \tag{10.5}$$

that corresponds to formula (10.4).

Thus, the motion of an electron in the undulator field results in its individual (spontaneous) radiation at frequency ω, the magnitude of which depends on the velocity of the electron and the period of the undulator field.

If the FEL open resonator is tuned to frequency ω, then radiation will accumulate in it, and as a result a wave field is formed in the beam propagation zone. By acting on the electrons in the beam, this field leads to their bunching (auto phasing) and, as a result, to stimulated collective radiation at a frequency close to ω.

Table 10.5 shows parameters of one of the FEL-self-oscillators.

10.3.5 FEL-Scattertron

The aforementioned electron radiation analysis shows that the effect of the undulator field on electrons can be replaced by an equivalent electromagnetic pump wave with relatively low frequency ω_p propagating towards the electron beam. If the pump wave is linearly polarized and the electric field vector is parallel to the y-axis, the effect of such a field on the electrons will be analogous to the effect of the undulator field. Since the beam and the pump wave move towards each other, due to the Doppler effect the frequency of the field oscillations in the e-system will be

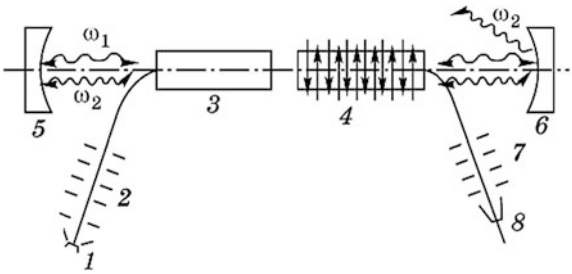


Fig. 10.11 Free-electron laser with electromagnetic wave pumping

$$\omega' = \omega_p \gamma (1 + v_z/c).$$

The action of the pump field leads to electron oscillation and radiation at the same frequency ω'. When this radiation is received in the laboratory system (L-system) it will have the frequency

$$\omega = \omega_p \gamma^2 (1 + v_z/c).$$

It should be noted that another explanation can be given for the process of electron emission arising under the action of the pump wave. The emission can be regarded as Compton scattering of the pump wave on moving electrons. Thus, FELs, in which the pump wave is used instead of an undulator, are called scattertrons (from "scatter").

Usually FELs with a pumping electromagnetic wave are used in a two-section laser, as shown in Fig. 10.11.

The laser includes the following elements and assemblies: an electron gun 1, an acceleration system 2, two laser sections 3 and 4, mirrors of the open resonator 5 and 6, a deceleration system 7 and an electron collector 8.

The laser section 4, containing the undulator, is used to generate primary electromagnetic radiation at frequency $\omega_p = \omega_b \gamma^2 (1 + v_z/c)$. This radiation is used as a pump wave in laser section 3, which ensures the generation of output radiation (signal radiation) at frequency

$$\omega_2 = \omega_1 \gamma^2 (1 + v_z/c)^2 = \omega_b \gamma^4 (1 + v_z/c)^3. \tag{10.6}$$

The deceleration system and the electron collector ensure energy recovery and electron capture from the used electron beam.

It can be deduced from formula (10.6) that a two-section FEL actually enables generation of radiation ranging from infrared to visible portions of the spectrum at lower electron beam energies (velocities) than a single-section FEL.

10.3.6 High-Current FEL

High-current FELs are designed to generate high-power electromagnetic waves in the millimeter and submillimeter wavelength bands. One type of FEL represents a

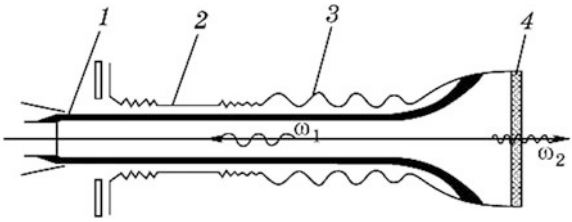


Fig. 10.12 High-current FEL

combination of a backward-wave tube and a FEL using a common electron beam. The FEL device, schematically shown in Fig. 10.12, includes the following units and elements: an electron gun 1, a FEL cavity section 2, an electrodynamic backward wave system 3 and an output window 4.

The Center for the Study of terahertz emissions and technologies at the University of California in Santa Barbara (USA) has three FELs with different parameters. The parameters of a millimeter band FEL are given as an example:

- electron energy—up to 6 meV;
- beam current—2 A;
- type of accelerator—electrostatic;
- radiation wavelength—2.5 mm–338 μm;
- output power—1–15 kW;
- pulse duration—1–6 μs.

The appearance of this FEL is shown in Fig. 10.13.

10.3.7 X-Ray Free-Electron Laser

X-rays are electromagnetic radiation with a wavelength of $\lambda = 10–10^{-3}$ nm. The creation of high-power coherent X-ray radiation sources enables the study of complex biomolecular and material structures at the atomic level, the dynamics of molecule formation during chemical reactions, as well as the creation of new materials and nanostructures. Ultrashort pulses of coherent X-ray radiation allow us to literally see the stages of molecular and atomic processes in materials and biomolecules.

Free-electron lasers can be the source of such radiation. Initially, the development of lasers for this wavelength range faced two specific problems. It turned out that the classical version of a laser—a generator with an open resonator formed by mirrors—is not suitable in this case, since the radiation is of a penetrating type and it is impossible to create a resonator—an effective energy store. The laser amplifier requires a coherent input signal to operate. Since there is no source of coherent radiation in this wavelength range, this option turned out to be impractical.

Modern lasers developed for the generation of coherent X-ray radiation (X-ray lasers) use an idea expressed 30 years ago by Russian scientists from the Budker

Fig. 10.13 The University of California FEL

Institute of Nuclear Physics in the Siberian Branch of the Russian Academy of Sciences (BINP SB RAS). According to this idea, free-electron lasers with long undulators can be excited and switched to coherent radiation mode due to primary spontaneous emission of electrons. This idea found experimental and theoretical confirmation.

In accordance with modern concepts, self-excitation of the FEL occurs as follows. It is assumed that the FEL is tuned to generate X-rays of a certain wavelength. Spontaneous emission of electrons occurs at the initial portion of the undulator, caused by their motion along curvilinear trajectories. The radiation is directed along the undulator axis and is concentrated in a small solid angle. Since the length of the packet is much larger than the wavelength of the radiation, and the charge is uniformly distributed along its length, the total radiation of the electrons forming the packet is initially incoherent and has low intensity. Radiation propagates along the undulator at a speed exceeding the longitudinal velocity of the electrons and outruns electrons moving along curvilinear trajectories. As a result of this effect,

further motion of electrons occurs both under the influence of the magnetic field of the undulator and the field of spontaneous emission.

The key moment determining the nature of the interaction of electrons with the field is the coincidence of the spontaneous emission frequency determined by (10.5) and the frequency ensuring the synchronism of electron motion and the field (10.4). With this type of motion, part of the electrons forming the bunch accelerates and another part decelerates, whereby the electrons become velocity modulated (or mass modulated for an electron velocity close to the speed of light).

Such a modulation ultimately leads to the bunching of electrons and the formation of electron groups (micro-packets) with an axial extension that is substantially smaller than the wavelength of spontaneous emission. Therefore, the radiation of electrons forming a micro-packet occurs synchronously, coherently, and has high intensity.

The micro-packet formation process takes a finite time and is realized over a large number of undulator periods. Therefore, FELs using the principle of self-excitation, have long undulators with a large number of periods.

One of the first X-ray FELs was built at the BINP. This complex setup consists of an electron gun, an accelerator-recuperator, an undulator, an optical system including an open resonator, a radiation output channel, a radiation distribution system for user stations and a FEL parameter monitoring system. The open resonator circuit of this device is shown in Fig. 10.14. Some parameters of the first stage FEL are listed in Table 10.6 and a general view of this installation is shown in Fig. 10.15.

The introduction of the second and third stages allowed an increase in the average radiation power to 10 kW and enabled the supply of three FELs with different parameters from one accelerator.

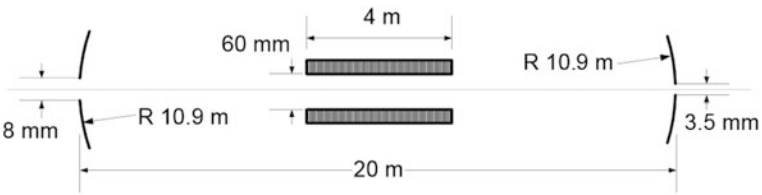


Fig. 10.14 Scheme of the Novosibirsk FEL open resonator

Table 10.6 FEL parameters

Beam energy (MeV)	12
Repetition rate (MHz)	11.75
Average current (mA)	20
Average laser emission power (W)	400
Wavelength range (μm)	120–235
Relative spectrum width (%)	0.3

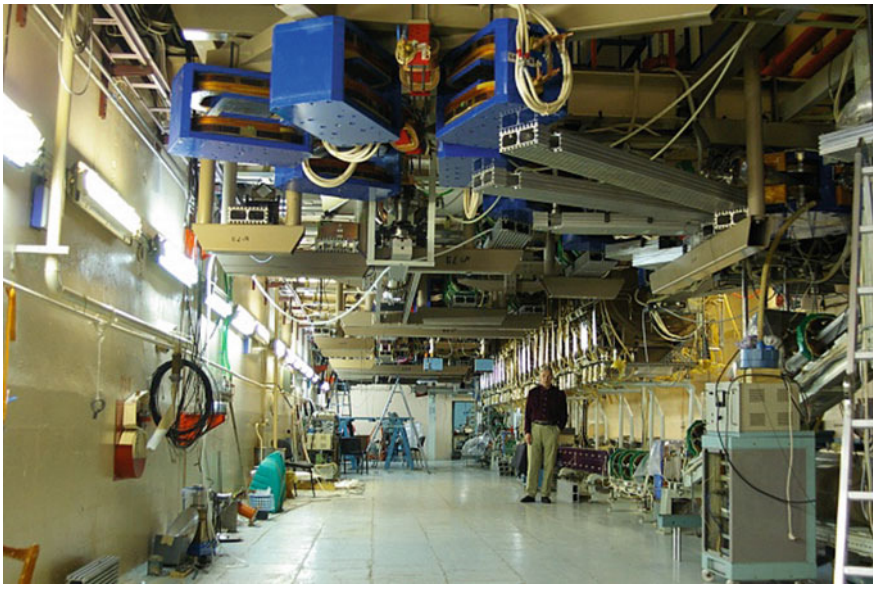


Fig. 10.15 General view of the Novosibirsk FEL

Another FEL example is the European X-ray laser XFEL, being built in Germany with the participation of several European countries. The laser is a complex engineering structure, located in an underground tunnel 3.5 km long. It includes the following main assemblies: a photoinjector, a linear electron accelerator, an undulator, an open resonator and an energy recuperation system for the spent electron beam. It began to work in September 2017, and the cost of the project amounts to 1.23 billion euros. Figure 10.16 shows the device in the tunnel where it is located.

Photoemission occurs by irradiating the photocathode with short pulses of an optical laser. The pulse duration is a few dozen femtoseconds (1 femtosecond = 10^{-15} s). The pulse is synchronized with the alternating field of the resonator and is applied when the electric field near the photocathode surface is accelerating and reaches its maximum value. Electrons accelerated by the field form a pulse (packet) of electron current, which enters the accelerator.

The photo-injector is a three-dimensional microwave resonator in which a photocathode is embedded. The resonator is excited from an external microwave source and an alternating electric field, used to accelerate the electrons emitted by the photocathode, forms within.

A Linear accelerator is formed by a system of resonators separated by transit tubes. The resonators are excited by external microwave radiation sources—powerful pulsed klystrons.

The European X-ray laser XFEL uses superconducting resonators made of niobium and cooled to a temperature of 4.2 K. The use of superconducting resonators allows a sharp reduction in the resonators' power supply. The length of the accelerator section of the XFEL is about 1.7 km.

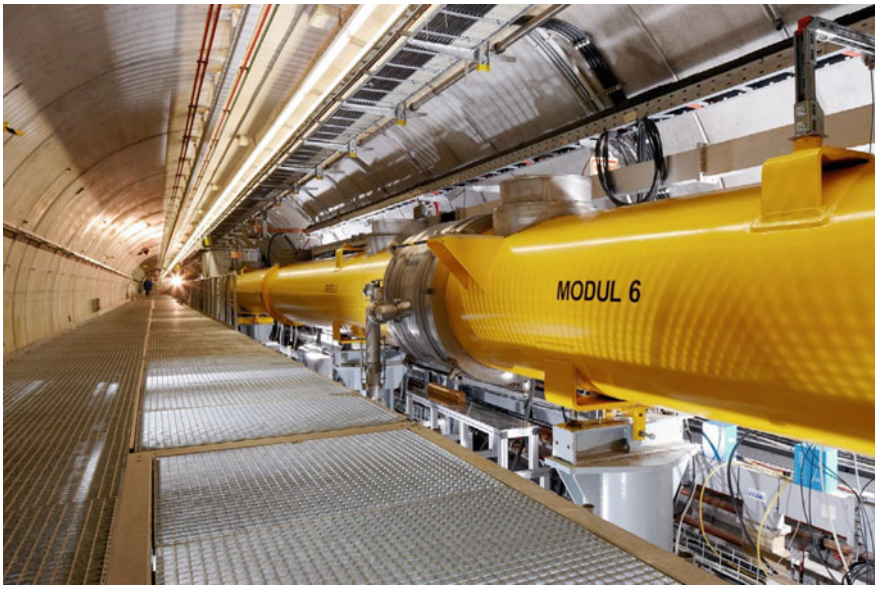


Fig. 10.16 The XFEL

The electron pulse (packet) accelerates while passing through the gaps of the resonators under the influence of an alternating field and its velocity and energy increase to the values necessary for the generation of X-rays.

The XFEL **undulator** consists of permanent magnets and works in the self-excitation mode due to the initial spontaneous emission of electrons. The length of the undulator is about 200 m. The undulator is placed between the open resonator mirrors. The undulator parameters make it possible to obtain powerful coherent X-ray radiation with a wavelength of up to 0.1 nm. The main parameters of the XFEL are given in Table 10.7. It is assumed that the brightness of the X-ray pulse will be 10^9 times greater than that of existing sources. This will allow us to

Table 10.7 XFEL Parameters

Parameters	Value
Electron energy (GeV)	Up to 17.5
Radiation wavelength (nm)	0.06–6
Pulse duration (fs)	100
Pulsed radiation power (GW)	120
Average radiation power (W)	300–800
Pulsed radiation brightness[a]	5×10^{33}
Average radiation brightness	1.6×10^{25}

[a]Brightness in terms of the number of photons per second on 1 mm^2 in solid angle 1 mrad^2 in the frequency band 0.1% from the central frequency

study new effects of the interaction of hard X-rays with matter, the structure of organic and inorganic molecules and to conduct other experimental studies.

The Free-electron laser is perhaps today's ideal tool for nanotechnology. It is able to record the processes that occur in the nano-world, at the atomic level.

10.4 Vircators

10.4.1 Virtual Cathode Effect

Let us consider a virtual cathode when an electron beam is injected into a cylindrical metal channel. We assume that length of the channel l greatly exceeds its radius a, $l \gg a$. The ends of the channel are considered to be closed by electron-permeable grids. We assume that transverse motion of electrons in the channel is frozen by a strong longitudinal magnetic field. As the channel is filled with the electron beam, the potential inside the channel decreases due to the action of the beam space charge. Gradually, over a time interval approximately equal to twice transit time of the electrons, a stationary potential distribution with a minimum in the middle channel plane is established. The electron beam passes through the channel and is deposited on the collector. The decrease in potential in the channel depends on the value of the injected current, increasing proportionally with current value. When an injection current becomes equal to some *space charge limiting current*, the minimum of the potential on the channel axis abruptly drops to zero and its position shifts toward the inlet end of the channel. The transit of the electron beam through the channel is broken, part of the electrons are reflected and moves to the inlet end of the channel. Such a state is known as a "virtual cathode" in the channel with the electron beam.

In the case when the injection current exceeds the limiting (critical) value from the beginning, the virtual cathode formation pattern is substantially different. When the channel is filled with electrons, the potential in the near-axis region of the channel falls to zero and become negative even before that part of the electrons reaches the outlet end of the channel. The potential minimum is located near the inlet end of the channel and its value varies periodically in time. A dynamic virtual cathode mode appears in the channel. Normal electron beam flow through the channel is broken, part of the beam electrons is reflected from the potential minimum and moves in the opposite direction towards the inlet end of the channel.

To estimate the value of the limiting (critical) current I_{cr} in a cylindrical channel (under condition $l \gg a$) the following approximate formulas can be used:

For beams with relativistic velocities $v \approx c$ the Bogdankevich-Rukhadze interpolation formula is valid:

$$I_{cr} = \frac{17.10^3(\gamma_a^{2/3} - 1)^{3/2}}{G},$$

Table 10.8 Value of coefficient K

b/a	0.8	0.6	0.4	0.2
K	0.5	0.33	0.22	0.15

where γ_a is the relativistic factor calculated for the accelerating voltage U_a, $\gamma_a = 1 + 1.96 \times 10^{-6} U_a$; $G = 1 + 2 \ln (b/a)$, a is the channel radius, and b is the electron beam radius.

For beams with low velocities $v \ll c$:

$$I_{cr} = 32.4K(b/a)U_a^{3/2},$$

where $K(b/a)$ is the coefficient depending on the ratio of the radii of the electron beam b and the channel a (Table 10.8).

It is important to note that the virtual cathode mode can arise when an electron beam is injected not only into a cylindrical channel, but also into a metal cavity of different configuration.

The virtual cathode phenomenon is used in devices designed to generate electromagnetic waves in the microwave band, called vircators.

10.4.2 Types and Parameters of Vircators

There are two groups of vircators. Powerful impulse relativistic vircators with an electron beam power in pulses reaching tens and hundreds of megawatts, and vircators with a moderate electron beam power, using low accelerating voltages and comparatively small currents.

Relativistic vircators.

There are experimental and measured data to indicate the existence of two mechanisms for the generation of microwave radiation. One of them is related to the oscillations of electrons reflected from the virtual cathode in the virtual cathode–anode–cathode space of the electron gun. This type of oscillations is analogous to the oscillations of electrons in a triode with a large positive potential of control grid (Barkhausen-Kurz oscillations). Another mechanism of oscillations is due to the vibrations of the virtual cathode.

Vircators, using the first generation mechanism, are sometimes called “triodes with a virtual cathode”, and those based on the second mechanism are called “reditrons”.

In *triode with a virtual cathode type of vircators* (Fig. 10.17), the electrons emitted by the cathode 1 accelerate in the cathode–anode electrode space of the electron gun 2 and enter the cavity 4 through the hole into anode 3 covered with a grid.

If the electron beam current exceeds the space charge limiting current, a dynamic virtual cathode 5 is formed in the cavity of the resonator. Part of the beam electrons

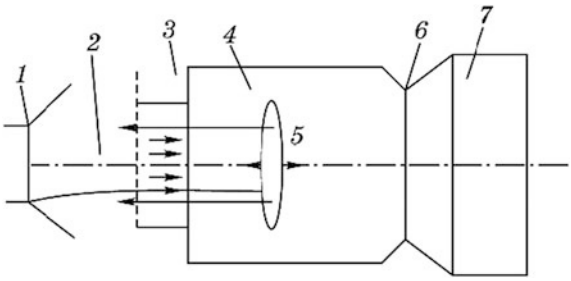


Fig. 10.17 A triode with a virtual cathode

is reflected from the virtual cathode region and while passing through the anode electrode hole, it enters the anode–cathode gap. Moving in the deceleration field, the electrons lose velocity, turn back and enter the cavity of the resonator through the anode hole once more, where they can be reflected again from the virtual cathode. Thus, the motion of the reflected electrons is of an oscillatory nature.

When the *E*-mode oscillation frequency of the resonator is close to the frequency of electron oscillations, oscillations are excited in the resonator.

The energy effect of electron interaction with the alternating resonator field depends on the transient phase. If the electrons entering the resonator meet the field in the decelerating phase, they transfer part of their energy to the field, decelerate and turn back. When returning, depending on the transient phase, transient time and the resonator oscillation period, they can again transfer their energy to the resonator field.

The electrons that first enter the resonator in the accelerating phase of the alternating field increase their kinetic energy. They can overcome the potential barrier created by the virtual cathode and drop out of the interaction process, leaving the resonator.

Among the electrons reflected from the virtual cathode, there are electrons accelerated by the alternating field the kinetic energy of which is sufficient to overcome the difference in cathode–anode potentials of the gun and to land on the cathode. Such electrons also drop out of the interaction with the alternating field. Thus, in a triode with a virtual cathode, there is a mechanism of selection (sorting) of electrons, that ensures the removal of electrons taking energy from the alternating field.

Analysis shows that along with the selection mechanism, there is an electron bunching process around favorable phase electrons, that transfer their energy to the field. These two processes ensure the conversion of the electron beam energy into electromagnetic oscillations energy and the generation of microwave radiation. The latter exits the resonator through diffraction window 6 into outlet waveguide 7. The parameters of the Tomsk Polytechnic University vircator (Table 10.9) can be used as an example to illustrate the possibilities of this type of vircator.

In *reditron type of vircators* (Fig. 10.18, references on this figure have the same values as in Fig. 10.17), the electrons reflected from the virtual cathode do not enter the cathode–anode gap of the device. For this purpose, a long-distance channel is

Table 10.9 Parameters of the triode type of vircators

Accelerating voltage (kV)	600
Pulse current (A)	20
Pulse duration (ns)	
Emission power (MW)	500
Duration of emission pulse (ns)	80
Emission frequency (GHz)	3
Energy conversion efficiency (ECE) (%)	5

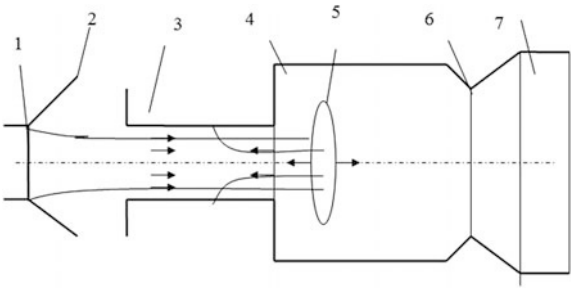


Fig. 10.18 Scheme of a reditron

used. The reflected electrons are deposited on the walls of this channel. Upon energy exchange with the alternating field of the resonator, only electrons coming through the channel from the electron gun are involved. An Output power of 250 MW was obtained for such a vircator with the following electrical parameter values: accelerating voltage 800 kV, pulse current 30 kA, pulse duration 80 ns, and energy conversion efficiency 1%.

Vircators generate powerful microwave radiation in a wide frequency range and make electronic frequency tuning possible. The disadvantages of the devices are low efficiency, instability of frequency and output power level within the pulse, as well as from one pulse to another.

In order to overcome these shortcomings, several devices were proposed, with a more complex electrodynamic system. These include, in particular, a vircator with a pre-modulated electron beam, called a vircator-klystron (Fig. 10.19).

In addition to a main resonator 4 with virtual cathode 5, the electrodynamic system of this device includes an additional resonator 3, which is connected to the main resonator via the coupling gap. Frequency tuning in the electrodynamic system occurs with the help of plungers 8, one of them serving to change the resonator coupling degrees.

When resonators are tuned to frequencies close to the oscillation frequency of the virtual cathode, an alternating voltage appears at resonator gap 3 to pre-modulate the electron beam. Calculations and experiments show that the efficiency of pre-modulated vircators increases to 10%. Also, frequency stability along the pulse, and from one pulse to another is improving. With a pulse duration of 25 ns, an output power of 1 GW in the decimeter wavelength band was obtained.

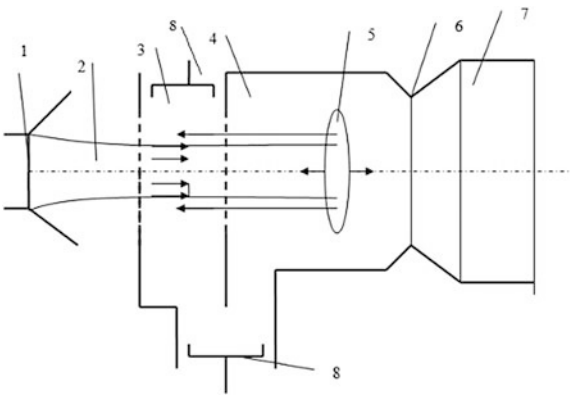


Fig. 10.19 Vircator with a premodulated electron beam

10.4.3 Low-Voltage Vircators

Low-voltage vircators can serve as noise-like signal generators with different power levels in a wide band of centimeter and millimeter wavelengths. Such generators can be used in electronic warfare and radar systems, communication systems based on chaotic signals, industrial process units and medicine.

Unlike relativistic vircators, these devices operate at low current values and accelerating voltages, both in pulsed and continuous modes.

To reduce the space charge limiting current value, in the cavity where the virtual cathode is formed, an additional electrode (for example, a grid or an annular electrode) is introduced, and its adjustable potential is lower than the potential of the cavity shell. Such an electrode reduces the potential in the cavity space and leads to an additional electron deceleration, thus contributing to the formation of a virtual cathode for a lower electron beam current value. In addition, by regulating the potential of this electrode, it is possible to change the amplitude and frequency band of chaotic oscillations, from narrowband, close to single-frequency, to chaotic bands with a bandwidth of the order of an octave. The low-power vircators of this group operate with accelerating voltages not exceeding several kilovolts and currents not exceeding one ampere.

10.5 Gyrocons and Magnicons

A gyrocon is a relativistic electronic device with circular electron beam scanning. This device was proposed by Academician Budker in 1967. Figure 10.20 shows the schemes of two gyrocone types.

An electron gun 1 creates an electron flow that is injected into a cylindrical resonator 2 excited by external source 3. TM_{110} mode with rotating electromagnetic field is excited in resonator.

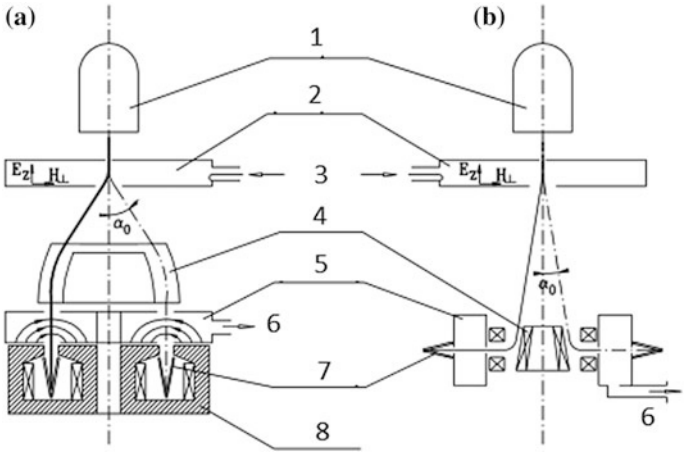


Fig. 10.20 Gyrocones schemes **a**—with an axial beam, **b**—with an radial beam

This field has a transverse magnetic field maximum and a longitudinal electric field node on the resonator axis. Electrons passing through the resonator get an impulse directed from the axis of the resonator under the action of the magnetic field. Impulse direction varies in accordance with the change in azimuthal orientation of the magnetic induction lines. Electrons deflected at angle α_0 pass through the electrostatic deflection system 4, where their trajectories are corrected and become parallel to the axis of the device. The electron beam formed in this way rotates continuously in the azimuthal direction.

In the first device (Fig. 10.20a), electrons transfer through the outlet resonator parallel to the axis, exciting a rotating field of TM_{110} type and transfer their energy, which is output from the resonator into load 6. A ring gap is made for the flow of electrons in the resonator cavities. After interaction, electrons drop into the collector 7. Compensating magnets 8 serve to preserve the rectilinear motion of electrons in the resonator.

In the second gyrocone type (Fig. 10.20b), before entering the resonator, the electron beam is deflected by 90°. The output resonator 6 is formed from a waveguide folded into a ring. Ring gaps are cut in the wide walls for the electron beam to pass through.

The point of entrance of the beam into the resonator moves along the annular gap with angular velocity $\dot{\theta}$, numerically equal to the circular oscillation field frequency of the inlet resonator: $\dot{\theta} = \omega$. The electron beam excites a traveling wave of TE_{10} type in the resonator. Interacting with the electric field of this wave, it transmits part of its kinetic energy to the resonator field and enters the collector 7. Energy exits the resonator through waveguide 6.

A fundamental feature of gyrocones is the absence of electron velocity modulation by absolute value and subsequent bunching. The electron current entering the

annular gap of the output resonator remains constant and does not depend on the azimuthal position of the beam. In this case, energy transmission from the electron beam to the resonator field is possible because the electron entry point into the resonator moves synchronously with the displacement of its maximum decelerating phase and all the electrons crossing the resonator experience only the braking effect of this field. Such a mechanism for converting the kinetic energy of the electron beam into the energy of an alternating field makes it possible to expect the electronic efficiency to tend to 100%. The parameters of gyrocones developed at the BINP and at the Los Alamos National Laboratory in the USA (LANL) are listed in Table 10.10.

Theoretical and experimental studies of gyrocones have revealed a number of inherent limitations that prevent the advancement of these devices into the shortwave region. These include overheating and breakdown of resonators due to a reduction in their dimensions, the complexity of passing a long beam through narrow slots without magnetic support and the associated limitations of perveance and beam current.

These restrictions were mostly reduced in a device invented under the leadership of Academician Budker at the INP in 1982 and called a *magnicon*.

Figure 10.21 shows a scheme of the first magnicone. The device works as follows: an electron gun creates an electron beam that enters the input cavity excited by an input signal. Oscillations are induced in the resonator in the form of a wave traveling along the azimuth. Transverse electric field deflects the electron beam from the device axis. The passive resonator coupled to the input cavity enhances this deflection. The input and passive resonators are in a longitudinal magnetic field, which causes the electrons to move along spiral trajectories. The induction value of the magnetic field is chosen such that the frequency of electron rotation (gyromagnetic frequency) coincides with the input signal frequency.

The "twisted" electron flow enters the output resonator and induces an oscillation with a field that rotates in azimuth with signal frequency ω. The oscillation phase is automatically set in such a way that the electrons give up their rotational kinetic energy to the resonator field. Resonator energy is transferred to the load and the spent electrons hit the collector.

Table 10.10 Parameters of gyrocones

Parameter	Gyrocone				
	First	Cont.	Imp.	Imp.	Imp. LANL
Developer	INP	INP	INP	INP	LANL
Operational frequency (MHz)	430	182	430	7000	450
Power (MW)	0.6	0.4	65	60	0.15
Pulse duration (μs)	20	–	10	0.7	50
Voltage (kW')	320	240	1600	1200	82
Efficiency (%)	65	60	75	26	23
Gain (dB)	7	17	26	60	–

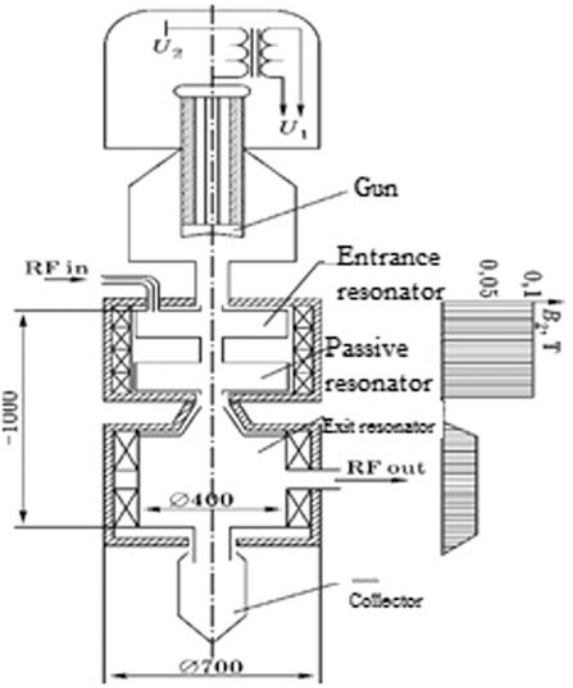


Fig. 10.21 Scheme of a magnicone

Due to the magnetic field, the electron beam does not spread out sideways and retains a small radius. The output resonator can be long, since electron and resonator field interaction remains synchronous for a long time. As a result, the electric field strength and enegy losses in the resonator decrease. These magnicone features make it possible to significantly increase their operating frequency while maintaining high output power and efficiency. As an example, Table 10.11 shows the parameters of a 3-cm band magnicone developed at the National Naval Research Laboratory (USA), to supply the accelerator. This device is shown in Fig. 10.22.

This magnicone operates on the second harmonic of the input frequency.

Table 10.11 Parameters of the NRL magnicon

Operational frequency (GHz)	11.424
Output power (MW)	66
Efficiency (%)	63
Pulse duration (μs)	1
Repetition frequency (Hz)	10
Gain coefficient (dB)	62
Accelerating voltage (kV)	500
Beam current (A)	210
Beam diameter (mm)	1.5

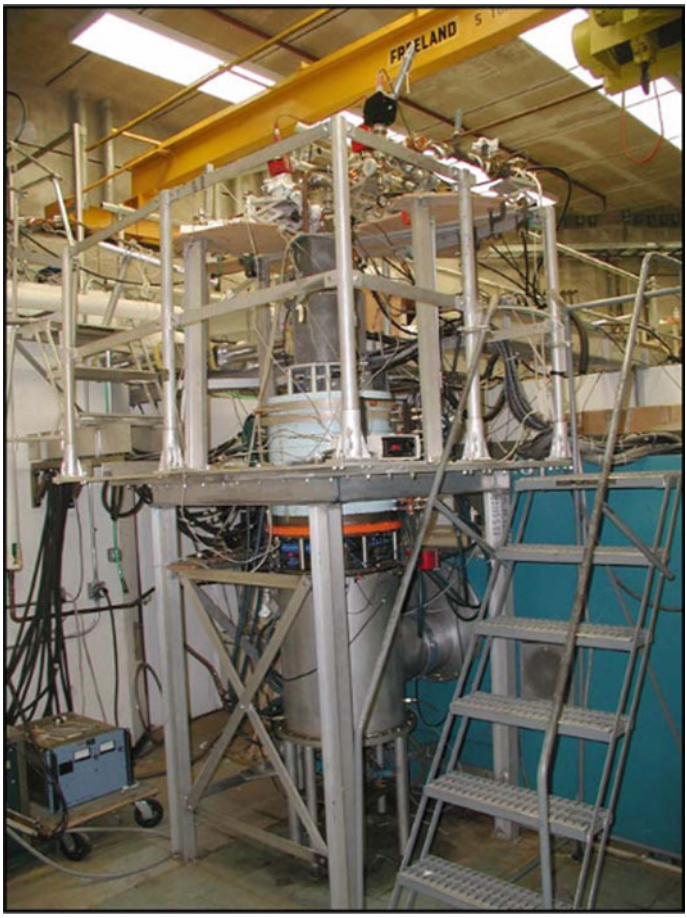


Fig. 10.22 Magnicone at the US Naval Research

Advancement Questions

1. Describe the design features of relativistic klystrons.
2. Describe the electron grouping process in a rotary magnetic field.
3. Describe the structure and working principle of a relativistic BWT-TWT.
4. Describe a relativistic Cherenkov generator, its working principle and parameters.
5. Describe the structure and working principle of a relativistic magnetron.
6. Describe the working principle of a free-electron laser.
7. Describe the structure and working principle of a ubitron.
8. Formulate the synchronism conditions in the ubitron.

9. Describe the structure and working principle of the FEL.
10. What factors determine the frequency of FEL radiation?
11. Describe the structure and working principle of the scattertron FEL.
12. Describe the design features of a high-current FEL.
13. Describe the design and working principle of an X-ray FEL.
14. Describe the phenomenon of a virtual cathode. Under what conditions does it arise?
15. Describe the structure and working principle of the triode type relativistic vircator.
16. Describe the structure and working principle of the reditrone vircator.
17. Describe the structure and working principle of the gyrocone. What kind of oscillation is used in the scanning resonators of these devices?
18. Why do gyrocones not work at frequencies above 10 GHz?
19. Describe the structure and working principle of the magnicone. What role does the longitudinal magnetic field play in these devices?

Part III
Semiconductor Microwave Devices

Chapter 11
Key Functional Elements of Semiconductor Microwave Devices

11.1 Elements of the Electronic Band Structure

Before considering the processes in semiconductor devices, we shall remind the reader some solid-state physics concepts. Electrons in a solid have different energies. Since electrons are attracted by positively charged lattice ions, their energy in a solid is less than the energy of electrons at rest in vacuum, which is generally assumed to be zero. With such a choice of reference point, the energy of electrons in a solid is negative.

The probability that a state with energy W is occupied by an electron is determined by the quantum function of the Fermi-Dirac distribution

$$f(W,T) = \frac{1}{e^{(W-W_F)/(kT)}+1}, \tag{11.1}$$

where T is the lattice temperature; and W_F is the Fermi energy (level). It is also called the *chemical potential*.

It follows from expression (11.1) that at any temperature, the probability that the Fermi energy level to be occupied is $1/2$ and the maximum energy of an electron in a metal at absolute zero temperature is equal to the Fermi energy. Quantum systems described by the Fermi-Dirac function are called *degenerate*.

At sufficiently high energies $W - W_F \gg kT$ the unit in the denominator (11.1) can be ignored. Then

$$f(W,T) \approx e^{W_F/(kT)} e^{-W/(kT)}. \tag{11.2}$$

Function (11.2) is known as a *classical distribution function*, or the *Boltzmann function*. The graphs of the quantum and classical distribution functions for the Fermi energy $W = -2\,\text{eV}$ and temperature 1000 K are shown in Fig. 11.1. It can be noted that at an energy equal to the Fermi energy, these functions differ by a

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_11

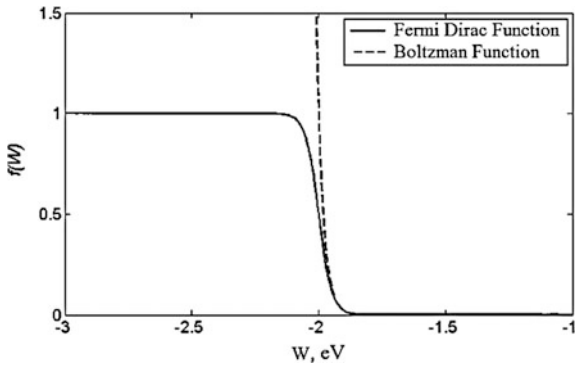


Fig. 11.1 Fermi-Dirac and Boltzmann distribution functions

factor of two, however, at an energy of $0.75\,W_F$, there is practically no difference between them.

The thermodynamic electron *work function* Φ from the solid-state is equal to the work required to transfer the electron from the Fermi level to level $W_0 = 0$ (in vacuum), therefore

$$\Phi = -W_F. \tag{11.3}$$

According to the band theory, electrons in a solid-state can only have certain energy values located inside the so-called energy bands. Each electron pair in such a band occupies a separate energy level in accordance with the Pauli principle. The number of these levels in each band is finite. If all levels in a given band are occupied, the electrons in that band cannot move on macroscopic distances, since they cannot change their velocity and energy. Therefore, only the electrons in the partially filled bands contribute to the conduction current. In this case, if the band is almost filled, the displacement of the vacant position in the electron sheath is equivalent to the displacement of a positively charged particle (hole) in the opposite direction. Therefore, in solid bodies, there are two types of charge carriers: electrons and holes.

A rough approximation shows two energy bands (or zones) in the solid—the lower band (valence) and the upper - conduction band (Fig. 11.2). The electron energy in this diagram increases from the bottom up and the energy of the holes is from the top down. At absolute zero, the valence band is completely filled and the conduction band is not filled (empty). As temperature rises, free electrons appear in the conduction band, and holes appear in the valence band. As a result, an electric current can flow in the material. In a number of materials, the valence band and the conduction band intersect. In such a unified band, there are always free energy levels and the material conducts electric current at any temperature. Such materials are called conductors. These include, for example, all metals.

In other materials, the valence band and the conduction band are seperated a gap—the forbidden band (bandgap). In pure (intrinsic) semiconductors and dielectrics, the

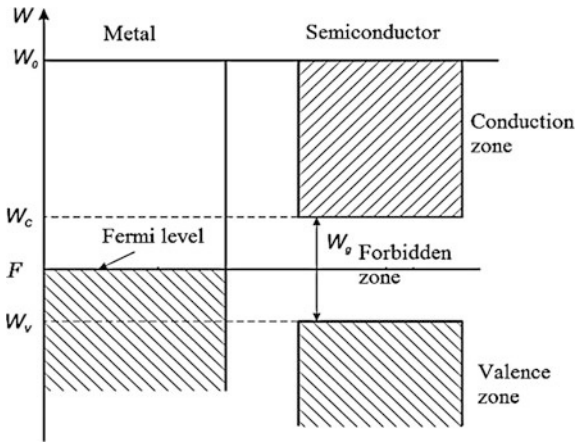


Fig. 11.2 Band diagrams of a metal and a semiconductor

Fermi level is in the middle of the forbidden band (Fig. 11.2). In order to pass into the conduction band, the electrons need to have energy not less than the width of the forbidden band. Therefore, at low temperatures, such materials have a high resistivity. Depending on the width of the band gap, they are called dielectrics or semiconductors. The difference between these materials is purely quantitative. If the width of the forbidden band is less than 2–3 eV, the material is called a semiconductor, otherwise it is a dielectric. Recently, semiconductor electronics has been using materials with a wider band gap (up to 5.5 eV), which are called *wide-band semiconductors.*

The concentration of free charge carriers in the valence band, and in the conduction band of the semiconductor are given by the following formulas:

$$n = 2\left(\frac{2\pi m_e^* kT}{h^2}\right)^{3/2} \exp\left(-\frac{W_c - W_F}{kT}\right); \tag{11.4}$$

$$p = 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{3/2} \exp\left(-\frac{W_F - W_V}{kT}\right). \tag{11.5}$$

In these formulas, m_e^*, m_p^* are the effective mass of electrons and holes, respectively, h is the Planck constant, W_c, W_V are the energies of the conduction band bottom and the valence band top respectively (Fig. 11.1).

Since in an intrinsic semiconductor, each electron passing from the valence to the conduction band due to thermal excitation (generation), leaves a hole behind itself, the electron and hole concentrations in the intrinsic semiconductor are equal. It follows that the Fermi level of an intrinsic semiconductor lies approximately in the middle between the bottom of the conduction band and the top of the valence band.

The concentrations of electrons and holes satisfy the following condition

$$np = n_i^2 = N_c N_V \mathrm{e}^{-W_g/(kT)},$$

where N_c, N_V are the effective densities of states in the conduction band and the valence band respectively. The greater the width of the forbidden band and the lower the temperature, the lower the carrier concentration will be. At room temperature, electron and hole concentrations are small, hence, an intrinsic semiconductor has a low conductivity at this temperature.

Semiconductor doping with various impurities changes their band diagram and the position of the Fermi level. There are donor impurities in which atoms have an "extra" electron that is easily detached from the neutral atom (ionizing it) and acceptor impurities having a vacancy (absence of an electron). The electrons of the impurity atoms are located at additional energy levels within the forbidden band. The donor levels are located closer to the bottom of the conduction band, while the acceptor levels are closer to the top of the valence band.

If the distance from the impurity level to the bottom (top) of the nearest band ΔW is greater than kT, the level is called deep. At working temperature T, it is weakly ionized and almost does not change the carrier concentration in the corresponding band. If $\Delta W < kT$, the impurity level is called shallow. At working temperature, it is almost completely ionized, and as a result, the carrier concentration in the semiconductor increases and becomes weakly temperature dependent. Impurities that form shallow levels are actually the ones used in semiconductor electronics. The volume of semiconductor device production with this material largely depends on the possibility of creating shallow impurity levels in it and the degree of technological development for their implementation.

In a semiconductor doped with a donor impurity (n-type semiconductor), the concentration of electrons in the conduction band is approximately equal to the concentration of donor atoms: $n_{n_0} = N_d + p_{n_0}$ and much higher than the concentration of holes in the valence band p_{n_0}. Electrons in this case are called the majority carriers and the holes are minority. The Fermi level in this semiconductor is shifted to the bottom of the conduction band.

In a semiconductor with an acceptor impurity (p-type semiconductor), the majority charge carriers are holes and their concentration $p_{p_0} = N_a + n_{p_0}$, where N_a is the concentration of acceptor impurity. The concentration of electrons n_{p_0} is much less than the concentration of holes. These carriers in the p-semiconductor are minority carriers. The Fermi level is shifted to the top of the valence band.

Along with the processes of carrier generation (formation of electron-hole pairs), the inverse process, that is the recombination of carriers, also occurs in the semiconductor, and as a result, the free electron and hole disappear. Recombination causes the excess concentration of minority carriers created at a certain instant of time to decrease and tend to equilibrium as the generation and recombination processes counterbalance each other. The recombination process is defined by the lifetime of excess minority carriers $\tau_{e,h}$, for which their concentration decreases by a factor of e.

The characteristic length of the region, where recombination occurs, is determined by diffusion and is defined by the following expression

$$L_{0e,h} \approx \sqrt{D_{e,h}\tau_{e,h}},$$

where D is the diffusion coefficient. Value $L_{0e,h}$ is called the diffusion length.

11.2 Semiconductor Materials for Microwave Electronics

11.2.1 Common Semiconductor Materials

The parameters and application field of a semiconductor device are largely determined by the semiconductor material used for the device's structure. A wide variety of semiconductor materials is used in modern semiconductor electronics, but only some of them are used in microwave devices (Table 11.1). Figure 11.3 compares field-velocity characteristics of commonly used materials.

Silicon (Si) is traditionally used in semiconductor diodes and transistors of relatively low power and with relatively low limiting frequencies (up to 10 GHz). Group V elements are used as donor impurities to silicon: phosphorus (P), arsenic (As), and antimony (Sb).

Group III elements are used as acceptor impurities: gallium (Ga), indium (In), aluminum (Al), and boron (B). The most commonly used are phosphorus and

Table 11.1 The main parameters of some semiconductor materials

Item No.	Material parameter	Unit of measurement	Material					
			Si	C (Diamond)	4H-SiC	GaAs	GaN	InP
1.	Width of the bandgap, W_g	eV	1.12	5.45	3.26	1.42	3.4	1.34
2.	Critical field strength, E_c	kV/cm	300	10,000	2200	400	3000	350
3.	Electron mobility at 300 K, μ_e	$cm^2/(V\,s)$	1500	480	900	8500	1000	5000
4.	Hole mobility at 300 K, μ_p	$cm^2/(V\,s)$	600	1600	–	400	200	200
5.	Saturation velocity, $v_e \cdot 10^{-5}$	m/s	1.0	1.5	1.5	1	2	2.2
6.	Heat conductivity, k	W/(m K)	1.5	22	4.9	0.5	1.5	0.68
7.	Relative dielectric permeability	-	11.9	5.5	10.1	12.5	9.5	12.4
8.	Maximal work temperature, T_{max}	°C	100	500	300	150	400	100

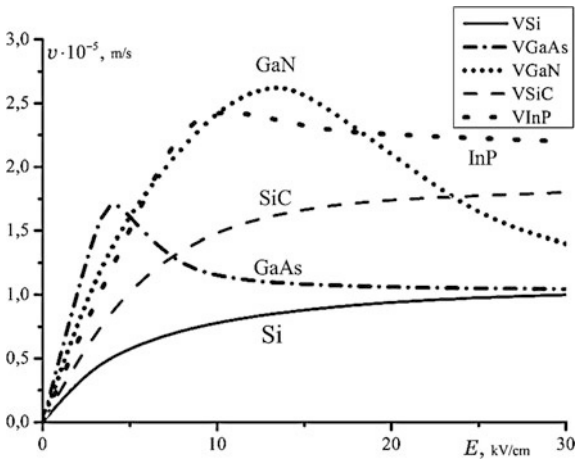


Fig. 11.3 Field-velocity characteristics of semiconductor materials

boron. Production technology of these devices is well-established and silicon itself is the cheapest semiconductor. Therefore, the use of silicon devices and microcircuits is more appropriate, if technically possible.

Gallium arsenide (GaAs) is today the most widely used material for manufacturing microwave diodes and transistors, although such devices cost more than silicon ones. Silicon (Si) is often used as a donor impurity, while zinc (Zn) and tellurium (Te) are used as acceptors. However, the relatively small width of the forbidden band does not allow the creation of high power devices made of GaAs.

Wide-band GaN and SiC semiconductors are rapidly gaining a leading position in microwave semiconductor electronics. The high saturation velocity in these materials allows the creation of transistors with a limiting frequency of up to 100 GHz. High heat conduction combined with high breakdown strength of the electric field, especially of silicon carbide, opens the way to the creation of devices with an output power of up to several hundred watts in a frequency range of up to 10 GHz.

Silicon (Si) as a donor impurity and magnesium (Mg) as an acceptor impurity are used for the doping of gallium nitride. Nitrogen and phosphorus are used as donor impurities for silicon carbide, while boron, aluminum, gallium and vanadium are used as acceptors. We note that electrons in GaN have a negative differential mobility, so this material can be used to create Gunn diodes. However, the technology of producing devices based on gallium nitride and silicon carbide is not sufficiently developed yet.

The unique properties of diamond allow the creation of powerful high-frequency transistors based on this material. Boron is used as an acceptor impurity for diamond. Conductivity of *n*-type is obtained by introducing boron, but an impurity level lies deep and carrier mobility is very small because of a large number of defects in the crystal lattice. The best results have been obtained recently with the doping of diamond by sulfur. The technology of growing single-crystal diamond substrates and creating necessary structures on them is still not well established and is very expensive.

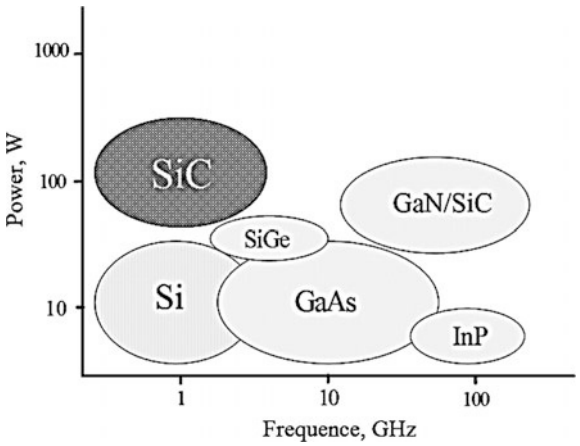


Fig. 11.4 Application fields for various semiconductor materials

Indium phosphide shows the highest saturation velocity (Fig. 11.3), therefore it is considered as a promising material for the creation of devices in the millimeter and submillimeter bands. To form *n*-type conductivity, the material is doped with selenium, tellurium, silicon, germanium or tin The best results were obtained with silicon doping. Acceptor impurities are zinc and cadmium. The relatively small forbidden band width of indium phosphide prevents the creation of powerful microwave devices based on it.

Figure 11.4 shows a diagram in terms of output power and frequency: (*P–f*-diagram), demonstrating which materials are used to create devices of the appropriate band and power. Specific types of devices based on these materials are discussed in Chap. 14.

11.2.2 Graphene as a Semiconductor for the Microwave Band

Graphene—an allotropic form of carbon has good potential to be used in microwave electron devices. Graphene is a mono-atomic layer of carbon atoms, forming a hexagonal structure (Fig. 11.5). It was first observed in an electron microscope in 1962, but intense study of its properties only began after its rediscovery in 2004 by A. Geim and K. Novoselov at the University of Manchester. As a 2D structure, graphene has very unusual properties. It is 100 times stronger than steel and is a perfect thermal conductor. A single graphene sheet is not stable and has a tendency to wrap into a sphere. So graphene sheets have to be placed on a substrate, which changes graphene's properties. The most suitable substrate is SiC.

Graphene is a semiconductor with zero forbidden zone. Graphene's energy diagram has conduction and valence zones, which touch at six points, where the

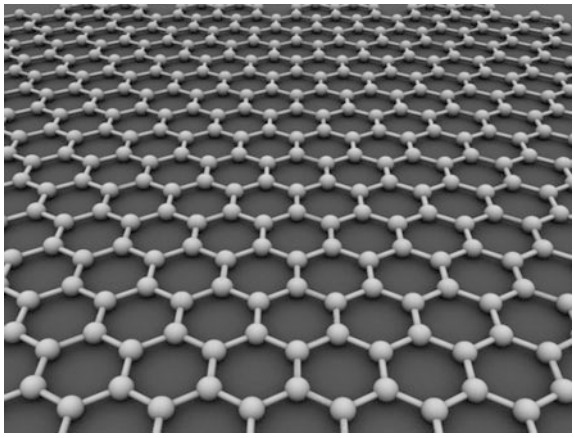


Fig. 11.5 Atomic structure of graphene

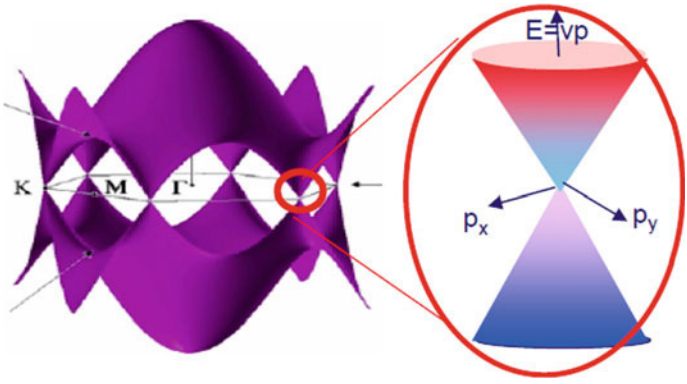


Fig. 11.6 Energy diagram of graphene

density of states is zero, but there is no band gap (see Fig. 11.6). The graphene sheet thus displays semi-metallic properties.

Figure 11.6 illustrates the structure of graphene energy zones. At low energy, the motion of electrons is described by the massless Dirac equation. Hence, the electron dispersion relation has the form

$$W = \hbar v_F \sqrt{k_x^2 + k_y^2}.$$

This equation is similar to the equation for a photon, but instead of light velocity c here stands Fermi velocity ($v_F \approx 10^6 = c/300$ in graphene).

As a result, graphene has a high electron mobility of more than 15,000 cm^2/(V s) at room temperature and high saturation velocity compared with commonly

used semiconductors (Fig. 11.7). Mobility of the holes has the same order of value. Notice the value of threshold electric field, which is of the same order as in GaAs and smaller than in GaN and SiC. Graphene electrons can travel micrometer distances without scattering, even at room temperature. These features make the use of graphene in terahertz electron devices very perspective. However, the absence of the band-gap makes it impossible to build typical FETs on graphene (low on/off ratio). Moreover, it is difficult to dope graphene and make contacts. Nevertheless, these difficulties were overcome and graphene transistors were constructed.

There are several methods to obtain a band gap in graphene. They include the use of narrow graphene ribbons (nanoribbons, GNR), substrates and bilayer graphene structures. Energy zones of GNR depend on their edge structure. Zig-zag structure (Fig. 11.8a) has no band-gap, while armchair orientation can be metallic or semiconducting depending on the GNR width (Fig. 11.8b). In order to have a

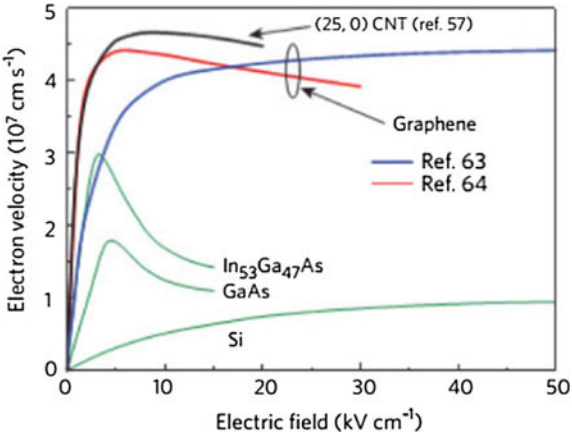


Fig. 11.7 Field-velocity curves for graphene compared with other materials

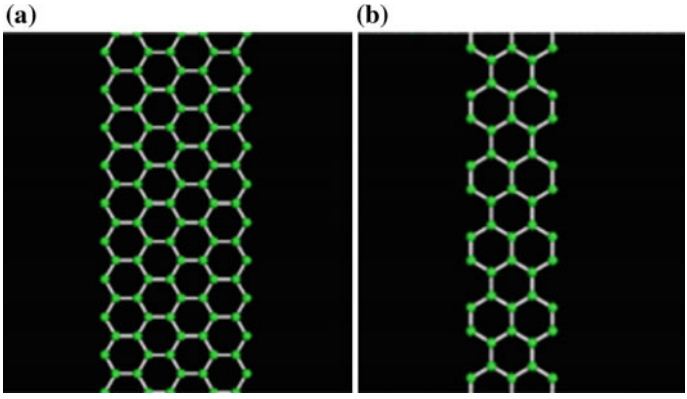


Fig. 11.8 GNR edge structures. **a** Zig-zag, **b** armchair

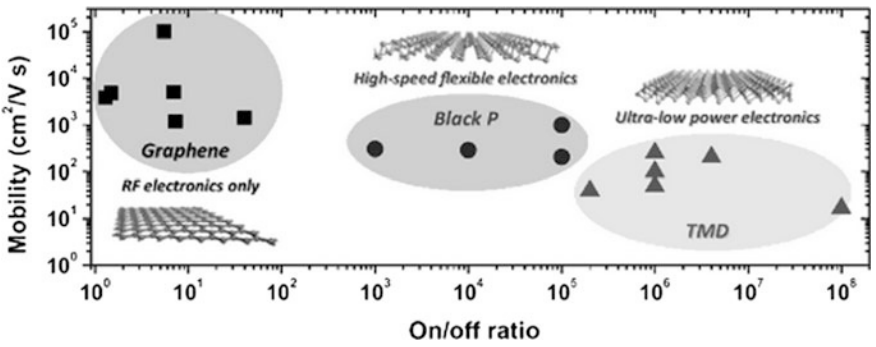


Fig. 11.9 Fields of 2D material applications

sufficiently wide band gap, very thin GNR should be used (about several lattice constants).

All methods can provide a band gap of up to 250 meV, which is sufficient for designing MOSFET transistors. However, all of them lower electron mobility and need very complicated technology for implementation. Nanoribbons, for example, suffer from edge lattice disorder and bilayer structures need a strong electric bias field between layers.

Recently other 2D-materials were discovered, for example, phosphorene (the 2D form of black phosphorus) and transient metal dihalcogenides, such as MoS_2, WS_2, $MoSe_2$ and others. Their physical properties show great potential for high-performance transistors. Fields of preferable applications of these materials are shown on Fig. 11.9. We can see that graphene devices can be used in analog RF frequency circuits only, due to their small on/off ratio, but the mobility of high charge carriers make them optimal for microwave and terahertz applications.

Properties of graphene microwave transistors are briefly described in Chap. 14.

11.3 Functional Elements of Microwave Semiconductor Devices (MSD)

11.3.1 Features of the MSD Functional Scheme

It is noted in Sect. 5.3 that any electronic device contain certain functional elements: a charged particle emitter, a flow control device, an energy extracting system from the CP flux, a collector of "spent" particles, an electron-optical system, a cooling system and some others. In most vacuum devices, each of these functional elements is an independent structural assembly.

Many of these systems are absent or combined in semiconductor devices. In particular, there is no electron-optical system, since the charge carrier electric field is largely compensated by the fixed ion field of the crystal lattice and the distance to

which the charge carriers move is usually much smaller than the transverse dimensions of the charged particle flux. In addition, it is difficult to distinguish the elements of the control system and the elements of the power extraction system in semiconductor devices. As a rule, they are integrated in one space, which is called the *gap*, or the *interaction space* of the charge flow and the field.

The main functional element of the device is a *semiconductor layer*, or a *boundary (barrier)* between layers. This can be a homogeneous semiconductor layer, a metal-semiconductor contact, a *p-n* junction (homo- or heterostructural) or an ohmic contact. It is important to note that essentially the properties of these elements depend not only on the type of material used, the level of doping and the thickness of the layer, but also to a large extent on the applied electric field. This property is actually used to obtain a useful effect.

11.3.2 Uniformly Doped Semiconductors

Under the action of an electric field, electrons and holes move, thus creating a conduction current. The density of this current is given by the following

$$J = -J_e + J_h = env_e + epv_p = e(n\mu_e + p\mu_e)E, \tag{11.6}$$

where μ_e, μ_p are the mobility of electrons and holes. Formula (11.6) expresses Ohm's law, which is obviously valid in semiconductors, if we assume that mobility is independent of the value of the electric field. In fact, formula (11.6) is valid only in comparatively weak fields and in samples, which dimensions are large in comparison with the Debye length and the mean free path of the charge carriers. In addition, a homogeneous distribution of the electric field strength is assumed along the length of the sample.

In strong fields, the linear dependence of the carrier drift velocity on the electric field strength is disrupted (see Chap. 2). In most semiconductors, as the electric field strength increases, the carrier velocity tends to the limit, called the saturation velocity. Accordingly, the resistivity of the sample, defined as the ratio of the voltage on the field sample to the current, increases as field strength rises. The differential resistance of the sample is often used $R_d = \mathrm{d}U/\mathrm{d}I$, which tends to zero as the voltage of the sample increases.

In many semiconducting materials of the $A^{III}B^{V}$ group, electron *negative differential mobility* (NDM) is observed for a certain range of electric field strengths: electron velocity decreases with increasing field strength. If a uniform distribution of field and carrier density is maintained along the sample length, the volt-ampere characteristic will follow the velocity-field one and the sample will exhibit negative dynamic resistance. Note that materials with holes NDDM are not known.

However, a homogeneous distribution of field and density is unstable with NDM. The Gunn effect is observed in such samples and is considered in more detail in Chap. 13.

If the length of the sample is less than the mean free path of the carriers, the latter move in such a sample without collisions and their velocity is determined by Newton's laws. In a strong field, carriers can acquire a velocity several times larger than the saturation velocity. This effect is called overshooting (super-speed). It is used in modern microwave transistors (see Fig. 2.8 and Chap. 14).

In thin samples (less than 100 nm), carriers can move freely only in two directions. So-called two-dimensional electron gas (2DEG) occurs. The probability of carrier scattering on phonons and structural defects decreases, leading to an increase in mobility. Actually, an electron mobility of $35 \cdot 10^6\,\mathrm{cm}^2/(\mathrm{V\,s})$ was obtained from the 2DEG heterostructure formed in AlGaAs/GaAs that exceeds the mobility in a bulk sample by 4000 times.

11.3.3 Metal-Semiconductor Contact Properties

Figure 11.10a shows band diagrams of metal and *n*-type semiconductor before their contact. The Fermi levels in the metal and semiconductor are denoted by F_M and F_S, and work function by Φ_M and Φ_S respectively. The energies corresponding to the bottom of the conduction band and the top of the valence band in the semiconductor are denoted by W_c and W_v, and the energy corresponding to the top of the conduction band in the metal is $W_m = F_m$. When the materials are in electrical contact, electrons from the material with a lower work function start diffusing to the material with a larger work function. Diffusion occurs because the energy levels in the material with the lower work function lying near the Fermi level are mainly occupied, while in the other material, the levels with the same energy are basically free. As a result, one material becomes positively charged and the other - negatively charged. A potential difference and a corresponding electric field strength are formed between them. This electric field causes the charge carriers to move (drift) in a direction opposite to the direction of their motion under the action of diffusion.

A stationary state arises when dynamic equilibrium is established, as the drift and diffusion currents are equal in absolute value.

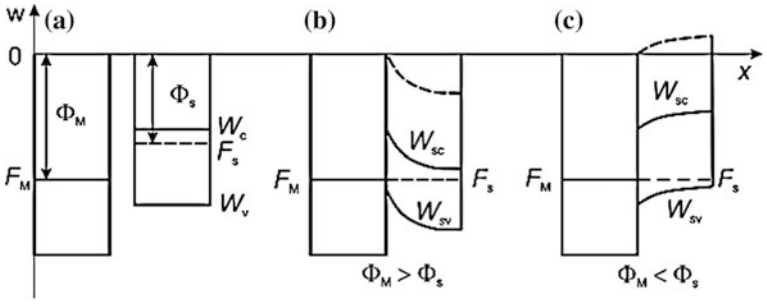


Fig. 11.10 Band diagrams of a metal-semiconductor contact

As a result, a space charge ρ forms in the semiconductor near the contact with the metal. The voltage on this contact layer in the steady state is called the *contact potential difference* (CPD) $\varphi_k = (W_M - W_s)/e$. The term *built-in potential* is often used in English sources and it is denoted as U_{bin}.

In a metal, due to its high conductivity, the electric field is concentrated in a very narrow contact layer with a width less than the lattice period. In the remaining volume of the metal, there is no electric field and the potential of this volume is constant. Therefore, the energy bands in the metal are not deformed and the entire electric field is concentrated in a semiconductor.

Unlike a metal, a semiconductor has a significantly lower conductivity and any electric field can penetrate it to a greater depth. Near the metal-semiconductor contact point, the energy bands of the semiconductor are bent, since $-\partial\varphi/\partial x = \rho/\varepsilon$ and $\rho \neq 0$. The energy bands are curved downwards, if $\Phi_s > \Phi_M$, and upwards, if $\Phi_s > \Phi_M$ (Fig. 11.10b, c).

The electric field in the contact layer changes the concentration of the charge carriers, which in turn leads to a change in the conductivity of the layer. The layer enriched in main charge carriers is called *enriched*. It has an increased conductivity, so it is also called anti-blocking. A layer depleted of main charge carriers is called depleted. It has a lower conductivity, for which it is called *blocking*. *Inversion layers* can also appear, in which the type of conductivity changes as a result of strong curvature of the bands. Such layers are formed through a contact with heavily doped regions when the electric field in the depletion layer proves to be strong enough to inject the carriers of the opposite sign near the depletion layer. The conditions for the formation of enriched, depletion and inversion layers are indicated in Table 11.2.

Early studies of metal-semiconductor contacts have shown that the CPD is only slightly dependent on the choice of metal. This fact is explained by the surface states near the contact point. The density of these states can be so great that the curvature of the bands in the semiconductor caused by them is much greater than the curvature of the bands due to the difference in work functions. Modern technologies for creating metal-semiconductor contacts can significantly reduce the density of surface states.

In this case, the theory presented above turns out to be valid.

In practice, the depletion layer is most often used. It was called the *Schottky barrier* in honor of the German scientist Walter Schottky, who investigated it in 1939.

The law of variation of potential inside an n-type semiconductor is determined by the concentration of free electrons $n(x)$ and the concentration of donors $N_d(x)$. This is true for an n-type semiconductor, in which the influence of the acceptor

Table 11.2 Occurrence conditions of various types of pre-contact layers

Type of semiconductor	Enriched layer	Depletion layer	Inversion layer
Electron	$F_M > F_s, \Phi_M < \Phi_s$	$F_M < F_s, \Phi_M > \Phi_s$	$F_M < F_s, \Phi_M \geq \Phi_s + \Delta W$
Hole	$F_M < F_s, \Phi_M > \Phi_s$	$F_M > F_s, \Phi_M < \Phi_s$	$F_M < F_s, \Phi_M \geq \Phi_s - \Delta W$

Note ΔW—Some additional energy

impurity and the charge of the holes can be neglected. The one-dimensional Poisson equation for homogeneous doping with donors $N_d = \text{const}$ has the form

$$\frac{dE_x}{dx} = -\frac{e[N_d - n(x)]}{\varepsilon},$$

where $n(x) = n_0 \exp[-e\varphi(x)/kT] \approx N_d \exp[-e\varphi(x)/(kT)]$ is the concentration of free carriers, provided that their energy is subject to the Boltzmann distribution, $\varphi(x)$ is the potential in x cross section. Thus, to calculate the field, the potential and the charge, it is necessary to solve the following equation

$$\frac{dE_x}{dx} = -\frac{eN_d[1 - \exp(-e\varphi(x)/(kT))]}{\varepsilon},$$

for the following boundary conditions: $\varphi(0) = \varphi_k - U, \varphi(\infty) = 0$. Here U is the external voltage against the barrier.

When the voltage on the barrier significantly overcomes thermal potential $U > kT/e$, the expressions for the field and the potential take the following form:

$$\begin{aligned} |E(x)| &= -\frac{eN_d}{\varepsilon}(\omega - x) = E_{\max} + \frac{eN_d}{\varepsilon}x; \\ \varphi(x) &= -\frac{eN_d}{\varepsilon}\left(\omega x - \frac{1}{2}x^2\right) - \varphi_\kappa. \end{aligned} \tag{11.7}$$

In these expressions $w = \sqrt{2\varepsilon(\varphi_\kappa - U_m)/(eN_d)}$ is the width of the depletion layer.

Formula (11.7) makes it possible to determine the plane in which the field is zero, that is, the surface charge of electrons on the metal is completely compensated by the positive charge of the donors. A more rigorous solution takes into account the change in the width of the depletion layer due to the thermal energy of the mobile carriers. Using the concept of thermal potential, the expression for the width of the depletion layer takes the form

$$w = \sqrt{\frac{2\varepsilon(\varphi_k - U - 3kT/(2e)}{eN_d}}. \tag{11.8}$$

In this expression, term $3kT/(2e)$ determines the equivalent thermal blurring of the channel boundary. In practice, expression (11.8) simplifies, taking $3kT/(2e) \approx kT/e$.

When the contact potential difference is compensated by an applied voltage, this formula corresponds to the expression for the Debye length (see Appendix A)

$$w = L_D = \sqrt{\frac{3\varepsilon kT}{e^2 N_d}}.$$

We define the current through the barrier, using the expressions obtained. Under thermodynamic equilibrium and $U = 0$, the current from the metal to the semiconductor is equal to the current from the semiconductor to the metal and both are determined by the Richardson-Deshman formula, namely

$$J_s = J_{ms} = -J_{sm} = AT^2 \exp(-e\varphi_k/(kT)).$$

If external voltage $U > 0$ is applied to the barrier, the current density from the semiconductor exceeds the current density from the metal. For $U < 0$, on the contrary, the current density from the semiconductor is less than the current density from the metal. The resultant current density is given by

$$J = J_{ms} - J_{sm} = AT_e^2 \mathrm{e}^{-\frac{e\varphi_k}{kTe}}\left(1 - \mathrm{e}^{\frac{eU}{kTe}}\right) = J_s\left(1 - \mathrm{e}^{\frac{eU}{kTe}}\right), \tag{11.9}$$

where $A = emk^2/(2\pi h^3) = 120.4\,\mathrm{A}/\left(\mathrm{cm}^2\,\mathrm{K}^2\right)$ is the Richardson constant. Value J_s determines the thermionic current through the barrier, also called the saturation current.

The positive (direct) branch of the volt-ampere characteristic of such a contact on a linear scale (for a contact potential difference $\varphi_k = 0.7\,\mathrm{V}$) is shown in Fig. 11.8a. A sharp exponential increase in the current near φ_k, described by a function of the form $\sim e^{40(U-\varphi_k)}$ [see formula (11.9)], prevents the analysis of the volt-ampere characteristic over a wide range of voltage variations. Therefore, in practice, the volt–ampere characteristics are plotted on a logarithmic scale, as shown in Fig. 11.11b.

The current through the Schottky barrier with reverse bias, according to (11.9), is practically independent of the applied voltage. However, with significant reverse voltages, a sharp increase in current is possible due to avalanche breakdown. The breakdown voltage value depends on many factors, in particular, the width of the forbidden band, the level of doping and the temperature. Naturally, structures based on wide-band semiconductors of SiC, GaN, and diamond (C) demonstrate the best resistance to breakdown (Table 11.1). The breakdown limits the amplitude of the operating voltages used and the permissible power of the device respectively. In the microwave band, avalanche breakdown is used to inject charged bunches in an Avalanche Transit-Time diode (IMPATT diode). Chapter 13 examines the features of such a breakdown.

When AC voltage is applied to the Schottky barrier, due to the passage of electrons through the barrier, a capacitive current is added due to the capacity of the depletion layer. As the frequency rises, the role of this capacity increases. We note that in the considered metal-semiconductor junction, only charges of one type (electrons) participate in the current flow.

To calculate depletion layer capacity C_b we shall use definition $C_b = \mathrm{d}Q/\mathrm{d}U$ and the formula for layer depletion charge:

$$Q = eN_d wS = S\sqrt{2e\varepsilon N_d(\varphi_k - U - kT/e)}.$$

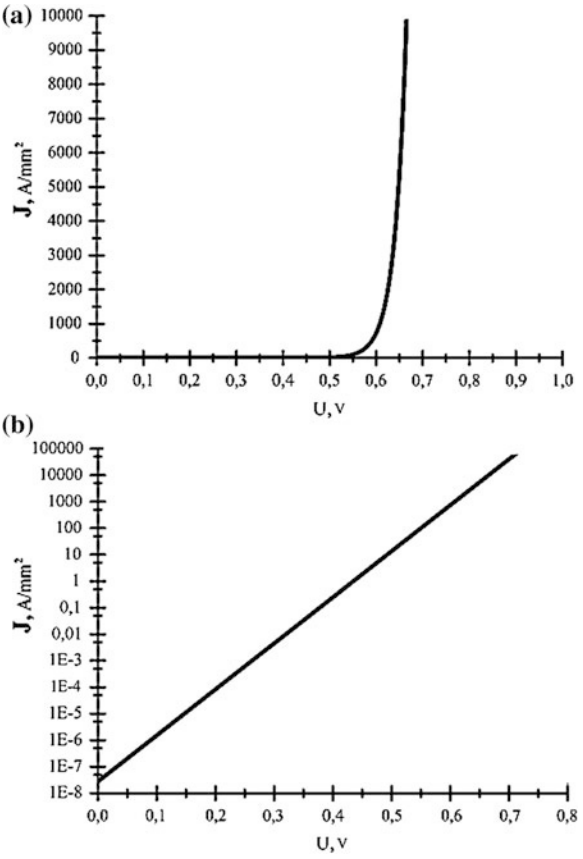


Fig. 11.11 Volt-ampere characteristic for direct bias voltage: **a** linear scale; **b** logarithmic scale

Differentiating this expression, we obtain

$$C_b = S\sqrt{\frac{e\varepsilon N_d}{2(\varphi_k - U - kT/e)}} = \frac{\varepsilon S}{w}. \tag{11.10}$$

The voltage-capacitance characteristic of the Schottky barrier in coordinates $C_b/C_0 = f(U/\varphi_k)$ is shown in Fig. 11.12, where C_0 is the capacity at zero bias voltage. This characteristic is discussed in more detail in Chap. 12.

Using expression (11.10), we find

$$\frac{1}{C_b^2} = \frac{2(\varphi_k - U - kT/e)}{e\varepsilon \mathrm{N}_d}.$$

and define the derivative of this expression for the voltage applied to the barrier:

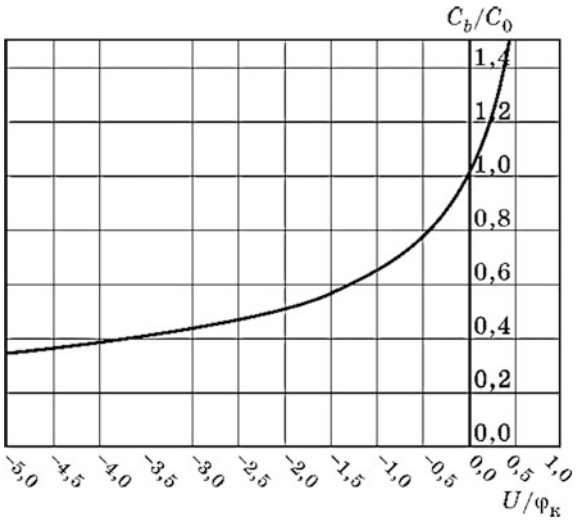


Fig. 11.12 Voltage-capacitance characteristic

$$\frac{d\left(1/C_b^2\right)}{dU} = -\frac{2}{\varepsilon e N_d},$$

from which it is easy to obtain an expression for the semiconductor doping level:

$$N_d = -\frac{2}{\varepsilon e}\frac{1}{d\left(1/C_b^2\right)/dU}.$$

This formula is widely used for the experimental determination of semiconductor structure parameters.

In practice, the presented ideal characteristics are influenced by many factors: imperfection of metal- semiconductor contact, surface states, and resistance of the semiconductor neutral region, etc. However, parametric optimization of the experimentally measured volt-ampere and voltage-capacitance characteristics allows determination of the equivalent parameters of the semiconductor structure, such as N_d, φ_k and «parasite» resistances, which are very important in improving the accuracy of the physical, topological and circuit simulation of devices.

11.3.4 Properties of the p-n *Junction*

An *electron-hole*, or *p-n junction* is the contact of two semiconductors with different types of conductivity. If both parts have the same forbidden band width, then the electron-hole junction is called a *homogeneous p-n junction*, or simply a *p-n*

junction. If, on the other hand, contact is made between semiconductors with different bandgaps, then it is called a *heterojunction*. It is important to note that the boundary between semiconductors is realized in the volume of the material, where there is no effect of surface states and surface defects (yields of dislocations, surface vacancies, etc.) on the properties of the resulting transition. Technically, this contact is easier fabricate than a metal-semiconductor contact. This circumstance caused the historically earlier development of bipolar devices in comparison with field devices. Even the first field-effect transistor was created with a control *p-n junction*, and not with a Schottky barrier.

Let us consider the basic properties of *p-n* junctions and the features of their application in the microwave band (it is assumed that the theory of *p-n* junctions is known to the reader). Figure 11.13 shows the distribution of the charge, the electric field strength, and the band diagram for a homogeneous *p-n* junction. When a junction occurs, electrons move from the *n*-region to the *p*-region where they are

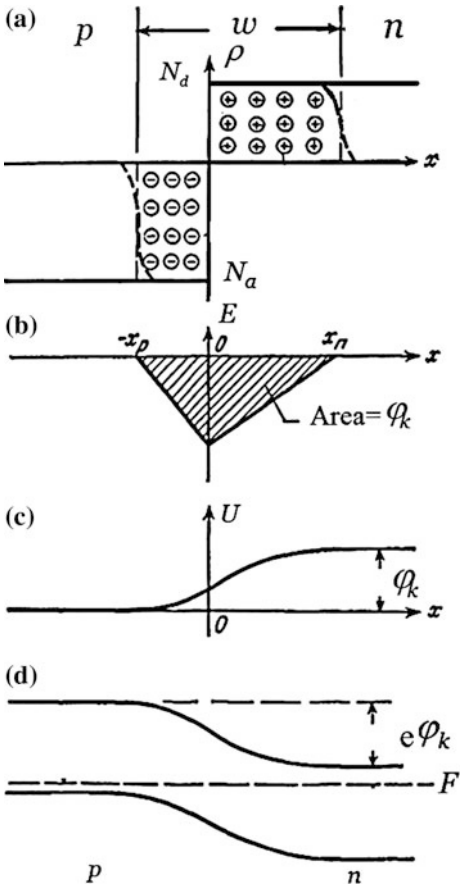


Fig. 11.13 Distribution of the charge, the electric field, and the band diagram of a *p-n* junction

less concentrated, under the action of diffusion, and the holes, on the contrary, move from the *p*-region to the *n*-region. As a result, an uncompensated positive charge of donor ions appears in the *n*-region and a negative charge of the acceptors is in the *p*-region (Fig. 11.13a). An electric field arises between these charges (Fig. 11.13b), that prevents further diffusion of charge carriers. Thermodynamic equilibrium occurs when the drift current under the action of the electric field is equal in magnitude to the diffusion current. As a result, a depletion layer of mobile carriers and a potential barrier appear near the physical (metallurgical) *p-n* junction. The considered potential barrier formation process is analogous to the barrier formation process between a metal and a semiconductor. However, we note that in the processes of current transfer in the *p-n* junction, charges of two signs and two types are involved: majority and minority. If the potential barrier prevents transfer of the majority charge carriers in the *p*- and *n*-regions from one region to another, then it facilitates the transfer of minority carriers.

Therefore, there are minority carriers (both electrons and holes) in the depletion layer.

If an external voltage is applied to the junction, a current flows through it. The voltage is of blocking or inverse type if a plus sign is on the *n*-region and a minus sign is on the *p*-region, and of opening or direct type if the polarity of the voltage source is reversed. Under the influence of external voltage U, the width of the depletion layer changes:

$$w = \sqrt{\frac{2\varepsilon(\varphi_k - U)(N_d + N_a)}{eN_dN_a}}, \tag{11.11}$$

where N_a, N_d are the concentrations of acceptor and donor impurities (it is assumed that they are completely ionized); and φ_k is the contact potential difference between *p*- and *n*-regions. At a sufficiently high reverse voltage, only the minority carrier current flows through the junction with a density determined by carrier diffusion to the boundary of the barrier:

$$J_s = e\left(\frac{p_{n_0}}{\tau_p}L_p + \frac{n_{p_0}}{\tau_n}L_n\right), \tag{11.12}$$

where p_{n_0}, n_{p_0} are the concentrations of minority carriers (holes in the *n*-region and electrons in the *p*-region); τ_n, τ_p are the lifetime of non-basic carriers; and L_p, L_n are the diffusion lengths of charge carriers (holes in the *n*-region and electrons in the *p*-region). Value J_s is called *density of diffusion heat current*, or *saturation current density*. It follows from (11.12) that the reverse current through the barrier does not depend on the applied voltage.

Upon direct bias, the current of the majority carriers is added to the saturation current, with opposite directions. The resulting volt-ampere characteristic of an ideal *p-n* junction is as follows

$$J = J_s\left(e^{\frac{e(U-\varphi_k)}{kT}} - 1\right). \tag{11.13}$$

Figure 11.14a shows the initial section of the volt-ampere characteristic, constructed according to formula (11.13). This dependence in logarithmic scale for current and direct bias (upper curve) and reverse bias (lower curve) is shown on Fig. 11.14b. The contact difference in potentials is considered to be 0.5 V, the saturation current is 0.1 mA. For clarity of the current value on the negative branch the characteristics are multiplied by 10. Real volt-ampere characteristics differ by a smaller slope of the positive branch due to the influence of p- and n-region resistance. This resistance leads to the fact that at large positive voltages upon the junction, the characteristic changes from exponential to linear. With even greater voltages, heating of the junction by the flowing current is developed, as a result of which the active resistance of the junction decreases and the current begins to grow faster. Then a thermal breakdown occurs and the junction is destroyed. At large reverse voltages, charge carriers in the depletion layer acquire sufficient energy to ionize the atoms of the crystal lattice. The electrons knocked out of these atoms in turn ionize another atoms, and an avalanche break occurs, characterized by a sharp increase in current.

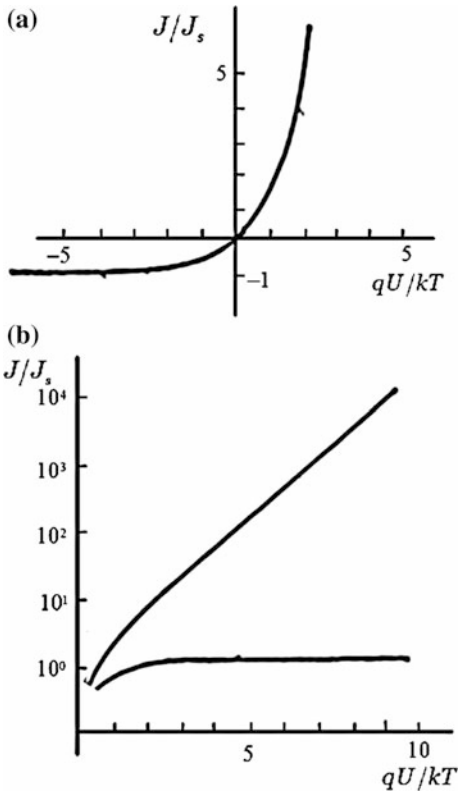


Fig. 11.14 Volt-ampere characteristics of a p-n junction

If we do not limit the growth of this current by external resistance, then a thermal breakdown occurs. Details of these processes are discussed in Chap. 13.

Since the depletion layer has a high resistance, it can be regarded as an insulator located between two plates: *p*- and *n*-layers, on which there are charges created by stationary ions, as well as mobile charges that have passed the barrier. The capacitance of such a capacitor, caused by stationary charges, is called the *barrier capacitance*.

It depends on the thickness of the insulator (depletion layer), which, in turn, depends on the applied voltage:

$$C_b(U) = \frac{\varepsilon S}{w(U)}.$$

Depending on the doping profile of the semiconductor near the junction, function $w(U)$ may have different forms. Respectively, the voltage-capacitance characteristic has different forms.

The so-called *diffusion capacity*, caused by the accumulation of free carriers that have crossed the barrier, is added to *p-n* junction to the barrier capacitance. These charges are concentrated on the boundaries of the depletion layer: positive charge of the holes is in the *n*-region and negative charge of the electrons is in the *p*-region. The spatial localization of these charges is determined by the corresponding diffusion length and the charge density—by the flowing current. When the current is increased (at a forward bias voltage), the concentration of these charges increases. The ratio of the change in the charge to voltage increment of the mobile carriers is called the *diffusion capacitance*. It is added to the barrier capacitance, forming the full capacity of the *p-n* junction. Assuming that the distribution of free charge carriers with respect to energies is determined by the Boltzmann formula, for the diffusion capacitance we obtain the following expression

$$C_d = \frac{e^2}{2kT}\left(L_n n_{p0} + L_p p_{n0}\right)\mathrm{e}^{eU/(kT)}. \tag{11.14}$$

The dependences of the barrier and diffusion capacitances of a *p-n* junction on the voltage across it are shown in Fig. 11.15. It can be observed that only the barrier capacitance exists with reverse voltages at the junction, while for a forward bias the contribution of the diffusion capacitance to the total capacitance of the junction can be significant.

In modern microwave devices, *p-n* junctions have different functional roles. In particular, in bipolar *n-p-n* transistors, the emitter-base junction injects electrons into the base, and holes from the base to the emitter. To characterize the properties of such a junction, emitter efficiency factor γ, defined as the ratio of the electron current to the total current through the junction, is introduced:

$$\gamma = \mathrm{I}_n/\left(I_n + I_p\right). \tag{11.15}$$

To ensure a predominant fraction of electrons in the total current, it is necessary to use an asymmetric *n-p* junction in which $n_n \gg p_p$ and, respectively, $N_d \gg N_a$. In this case, the emitter efficiency factor is close to one.

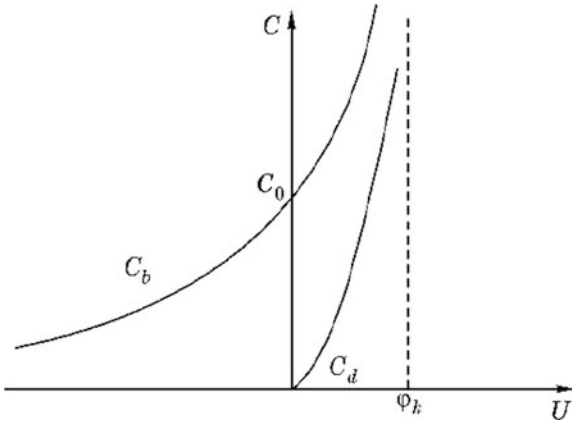


Fig. 11.15 Voltage-capacitance characteristics of a *p-n* junction

When the concentration of doping impurity increases both in the *p*- and *n*-regions to the level $10^{19}-10^{20}\,\mathrm{cm}^{-3}$, the thickness of the depletion layer, according to (11.11), decreases to values of the order of 10 nm. This gives rise to a non-zero probability of electron tunneling through the potential barrier. The inertia of this process is very small (less than 1 ps), and that makes it possible to apply this effect even in the terahertz range (see Chap. 13).

Heterojunctions are formed between semiconductors with different electrical properties: dielectric constant ε, width of the forbidden band ΔW and work function Φ Φ_2. To obtain a heterojunction close to ideal, it is necessary to select semiconductor pairs with the same type of crystal lattice, and the lattice constants should differ by not more than 0.5%. In addition, semiconductors should have similar coefficients of temperature expansion, so that while the temperature changes there is no strong mechanical stress.

Heterojunctions can be *isotype* (both semiconductors have the same type of conductivity) and *anisotype* (*n-p* and *p-n* junctions). Depending on the ratio of the thermodynamic work functions and the width of the forbidden band, these junctions can be either rectifying or ohmic.

Figure 11.16a shows band diagrams of two semiconductors before contact is formed between them. Semiconductor 1 is narrow-bandgap and, as the position of Fermi level F_1 indicates, it has *p*-type conductivity.

Semiconductor 2 of the *n*-type has a wider forbidden band. The electron affinity for the first semiconductor is greater than for the second semiconductor $(\chi_1 > \chi_2)$. When contact is formed, the Fermi levels in both semiconductors are equalized (Fig. 11.16b). As a result, diffusion and carrier drift begin, a layer of space charge of length L is formed, on which contact potential difference φ_k arises. There arises also discontinuity of the conduction band $\Delta W_c = \chi_1 - \chi_2$ and discontinuity of the valence band $\Delta W_v = W_{v2} - W_{v1} - \Delta W_c$.

Materials forming a heterojunction can be chosen such that the potential barrier in the valence band is greater than in the conduction band. This limits the current of

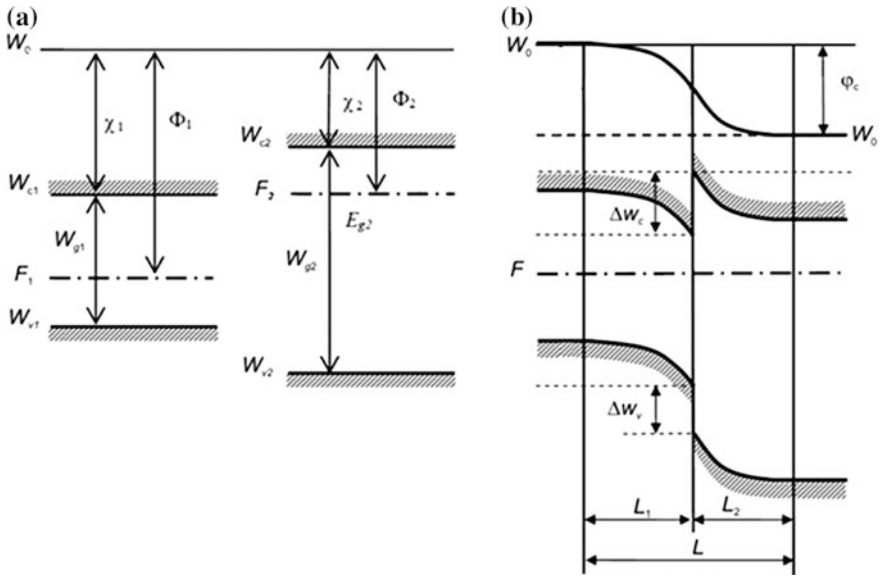


Fig. 11.16 The band diagram of a heterojunction

minority carriers from the second material to the first, thereby allowing bipolar transistor parameters to be improved. Details of hetero-structural bipolar transistors are discussed in Chap. 14.

11.3.5 Ohmic Contact

Ohmic contact means a metal-semiconductor contact which has the properties of ordinary resistance, that is, obeying Ohm's law. The volt-ampere characteristic of such a contact is linear. In order not to affect the basic functions of the device, its resistance should be much less than the resistance of other elements of the semiconductor structure. The ohmic contact is an integral part of any semiconductor device. In order for a metal-semiconductor contact to have ohmic contact properties, any potential barrier between a metal and a semiconductor should be absent.

To ensure that there is no barrier between the semiconductor and the metal, it is possible to select pairs with a small difference in the work function. Unfortunately, it is difficult to find such materials. In addition, they must meet the stringent technological requirements for possible alloying and the formation of stable compounds. To solve this problem, in practice, we use a heavily doped N_d^+ (degenerate) layer. It provides an additional component to the thermionic current-tunneling. With increased doping, the energy barrier for electrons does not in practice change and barrier thickness w, according to (11.8), decreases, creating conditions for tunneling. The density of the tunneling current is given by

$$J_t = A \exp\left(-2L\sqrt{2m^* e(\varphi_k - U)/\hbar^2}\right),$$

where m^* is the effective mass of electron carriers; and A is the constant. Length of the depletion layer w is determined by formula (11.8) and decreases with increasing donor concentration (doping level). Respectively, the density of the tunnel current increases. To obtain an appreciable tunneling current density, the donor concentration should be $10^{19}\,\mathrm{cm}^{-3}$ and over.

Thus, to create a good ohmic contact, it is necessary to select a pair with minimum contact potential difference φ_k and increase the doping level of the additional layer N_d^+. A band diagram of this contact is shown in Fig. 11.17.

The figure shows the distribution of electron energy in two modes: without bias (Fig. 11.17a) and with a direct bias (Fig. 11.17b). The tunneling effect is observed for both forward and reverse bias voltages, since the barrier length varies slightly. This causes a linear current dependence (Ohm's law) through the contact even at low voltages.

The greatest difficulty is the creation of such contacts for wide-band semiconductors. As an example, we shall mention some metal-semiconductor pairs used in practice. Au–Zn (99:1) or Au–Si (94:6) is used for GaAs, and a combination of Al–In is used for gallium nitride, GaN. In this case, the doping level of the additional semiconductor layer is $(1-3) \times 10^{19}\,\mathrm{cm}^{-3}$. Technologically, the process of obtaining the ohmic contact is to create a highly-doped layer by diffusion (or ion implantation), to which the desired metal is then deposited. The choice of this metal is based on the required mechanical and chemical properties of the contact.

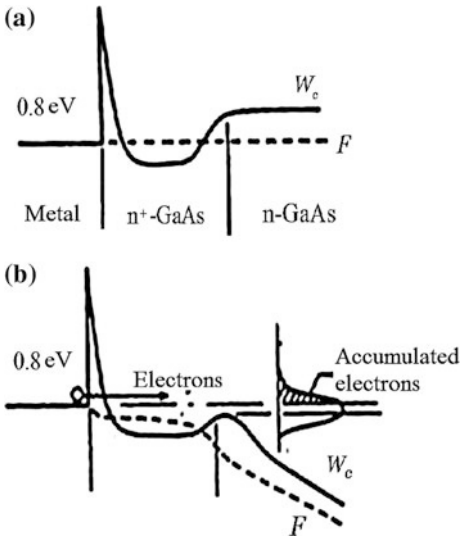


Fig. 11.17 Band diagrams of ohmic contact

11.4 Classification of Microwave Semiconductor Devices

The physical processes and functional elements described in this chapter are used to create various microwave semiconductor devices (MSD). These devices constitute a significant part of a large family of modern microwave electronic devices. Not a single microwave radio-electronic system is used without them nowadays.

Currently, the MSD class includes devices with very different purposes and structures. A simplified scheme for classifying these devices is shown in Fig. 11.18. As observed, MSDs are divided into three large families: diodes, transistors and functional electronic devices.

In turn, microwave semiconductor diodes are divided into diodes with positive and negative dynamic resistance. The first group serves to convert signals and control their passage in the microwave circuit and the second group generate and amplify microwave oscillations. Devices with positive dynamic resistance are discussed in Chap. 12 and those with negative dynamic resistance are considered in Chap. 13.

Microwave transistors are also divided into two large groups—field and bipolar. The devices included in these groups are discussed in detail in Chap. 14.

Functional electronic (FE) devices are the most diverse in terms of structure, principle of operation and the materials used. They include delay lines, phase shifters, attenuators, modulators and other devices on acoustic and spin waves, and

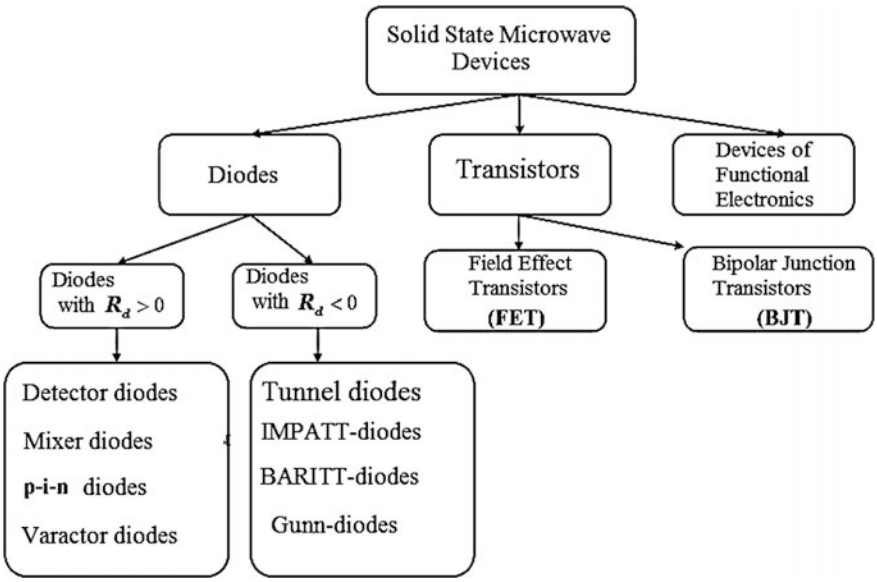


Fig. 11.18 Classification of microwave semiconductor devices

electro-acoustic effects, etc. Along with semiconductors, FE devices use ferroelectrics, ferrimagnets, superconductors and artificial media such as ferroids, superconductors, metamaterials, and photonic crystals. Radiophotonics has recently been developing rapidly and is implemented in devices that use electro-optical phenomena in a solid. Functional electronic devices are not considered in this textbook.

Advancement Questions

1. List the main types of semiconductor devices with positive dynamic resistance.
2. Draw a band diagram of the Schottky barrier at different bias voltages. What charge carriers are involved in the current transfer across the barrier?
3. What charges are in the depletion layer? Which current components pass through the depletion layer when the microwave voltage is applied to the barrier?
4. Draw the volt-ampere curve of the Schottky barrier with logarithmic and linear scales for the current. What is the complexity of representing the volt-ampere characteristic on a linear scale?
5. What are the parameters connected by the voltage-capacitance characteristic? Can it describe the doping profile? With what accuracy?
6. What is the difference between the Schottky barrier and the barrier formed through the contact of *n*- and *p*-type semiconductors?
7. What is the difference between barrier and diffusion capacitances?
8. What is the ohmic contact? What is its function in the device?

Chapter 12
Diodes with Positive Dynamic Resistance

12.1 Detector Diodes

12.1.1 Designation and Design of Detector Diodes

The functional purpose of detector diodes (DD) is a conversion of weak microwave signals 10^{-9}–10^{-2} W in to a direct current or a low-frequency envelope of signal. The detector diode performs the function of an AC rectifier operating in the microwave band.

The design and technology of detector diodes have undergone great changes, ranging from simple pin contacts with a semiconductor crystal (1935, "crystal detectors") to modern planar structures based on gallium arsenide. However, these devices still work based on the properties of the Schottky barrier (see Chap. 11).

Figure 12.1 shows designs of modern detector devices. The point diode (Fig. 12.1a) consists of a polycrystalline *p*-type silicon piece 1 contacting with a tungsten spring wire sharpened at the end 2. The silicon sample and the spring are soldered to electrodes 3 and 4, which in turn are welded to a ceramic case 5.

The sharp end of the spring is welded to the semiconductor by current pulse passing. The Schottky barrier with a small area (unit μm^2) is formed at the welding point that results in a small junction capacity and relatively high resistance R_s. The non-ideality coefficient of the volt-ampere characteristic of a point diode is $n \simeq (3–4)$, this indicates a significant number of defects in the crystal structure, and impurities in it.

These shortcomings are absent in an epitaxial structures (Fig. 12.1b). An example of such Schottky Barrier Diodes (SBD) is a gallium arsenide mesa-structure. Metallization 1, deposited on a single crystal of gallium arsenide by vacuum evaporation, forms a Schottky barrier close to ideal, with an *n*-type epitaxial film 2. A highly doped n^+ substrate 3 has low resistance R_s. Therefore, despite a greater capacitance compared to a point diode, this SBD has higher critical frequency values and a lower noise level. Figure 12.1c shows a diode with a planar

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_12

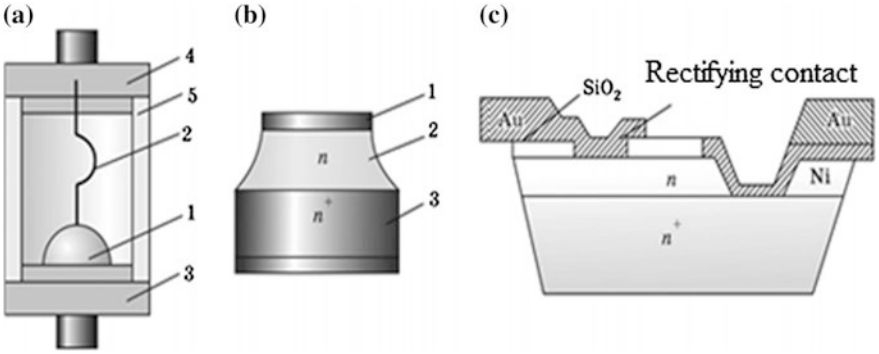


Fig. 12.1 Main types of diode design

design used in integrated circuits. The barrier is formed at the contact point of the metal electrode with *n*-type semiconductor. Covering contacts by gold ensures the working stability of such devices in time.

The diode structures need to be protected against mechanical damage due to their micron sizes. Therefore, they are placed in special housing (case). The main element of such cases is high-quality radio transparent ceramic that is soldered with metal electrodes. The case shape is usually chosen with the convenience of being included in a microwave transmission line or a resonant system in mind. For convenience of use, the shape and dimensions of the cases are specified. Figure 12.2 shows some types of cases used. In hybrid and monolithic integrated circuits, chip diodes with smaller parasitic parameter values are used.

The inertia of the diode structures is determined by two factors: the time of the transition of an electron through the active region (depletion layer) and the time for recharging of barrier capacitance, taking into account "parasitic" resistances of the structure and of the case. To calculate the limiting frequency, we shall consider the equivalent circuit of the device shown in Fig. 12.3, reflecting the role of individual elements of the device structure.

Fig. 12.2 Types of detector diode cases

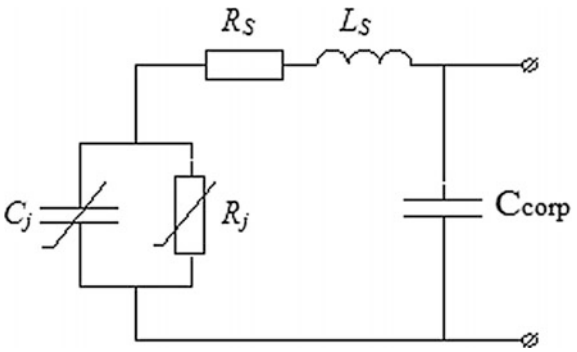


Fig. 12.3 The equivalent circuit of the device

The properties of the barrier are described by nonlinear elements $C_j(U_b)$ and $R_j(U_b)$. They correspond to the capacitance and resistance of the Schottky barrier, determined by the volt-ampere and voltage-capacitance characteristics of the diode (see Sect. 11.3). Resistance R_s is equal to the sum of the neutral part of semiconductor resistance and the contact resistance. Parameters L_s and C_{corp}—«cold» case reactivity—are the contact inductance and the case capacitance.

Let us determine the limiting frequency at which the conductivitys of such a device is equal for opposite polarities of applied voltage. Let's idealize the situation, neglecting the inductance of the contacts and the capacitance of the case. With a positive voltage polarity, there is no potential barrier on the junction and the current through the diode is limited only by resistance R_s. When the polarity is reversed, the properties of the diode are mainly determined by reactive conductivity $i\omega C_j$. From the condition that the current amplitudes are equal in both half-periods, we obtain

$$R_s = 1/(\omega C_j)$$

The frequency at which this equality is satisfied is called the *cutoff* frequency and is calculated using the formula

$$\omega_c = 1/(R_s C_j). \tag{12.1}$$

It should be noted that barrier capacitance $C_j(U_b)$ depends on the applied bias voltage, and causes the dependency of the diode cutoff frequency on its operating mode.

Parasitic reactivities L_s and C_{corp} significantly reduce the diode's frequency range and change the parameters of the devices. The housing capacitance C_{corp} shunts the diode structure and inductance reduces the useful signal (voltage) applied to the actual barrier.

From the expression for the cutoff frequency, it follows that in order to obtain a high-frequency device, R_s and C_j should be reduced. We can determine how to choose the topological parameters of the structure to fulfill this requirement. Resistance R_s for a neutral region of the semiconductor structure is given by: $R_s = \rho l/S = l/(eN_d\mu_0 S)$. In this formula, S is the contact area, μ_0 is the low-field

mobility, and l is the thickness of the neutral region (base) of the semiconductor. To determine the capacitance, we use the formula for flat capacitors $C_j = \varepsilon S/w$. As a result, we obtain an expression for the cutoff frequency:

$$\omega_c = \frac{eN_d\mu_0 w}{\varepsilon l} = \frac{e\mu_0\sqrt{2\varepsilon e U N_d}}{\varepsilon l} = \frac{\mu_0}{l}\sqrt{\frac{2e^3 N_d}{\varepsilon}U}. \tag{12.2}$$

It follows from the expression obtained, that under the assumed conditions, the cutoff frequency does not depend on the area of contact. Preference in choosing materials for DD should be given to semiconductors with high electron mobility μ_0 in weak fields and with a high level of doping N_d. The materials most used in DD are gallium arsenide with $\mu_0 \sim 8500$ cm^2/(V s) and indium phosphide with $\mu_0 \sim 5000$ cm^2/(V s)

We note that mobility decreases with an increasing doping level. According to the empirical formula, for gallium arsenide

$$\mu_0 = \frac{0.85}{\sqrt{1 + N_d 10^{-23}}}.$$

Substituting this formula in (12.2), we obtain

$$\omega_c = \frac{0.85}{l\sqrt{1 + N_d 10^{-23}}}\sqrt{\frac{2e^3 N_d}{\varepsilon}U}.$$

The expression obtained shows that the increase in the cutoff frequency due to an increase in the level of doping is limited by decreasing of mobility. This effect determines optimum value of the doping. In practice, the level of doping for different materials varies within 1–5 $\times$ 10^{17} cm^{-3}. The typical contact area is of the order of 20–100 μm^2. The cutoff frequency for such devices is 100–500 GHz.

The presented estimates of the cutoff frequency are incomplete if we do not take into account the transit time τ (the transit angle $\theta = \omega\tau$) of the electrons through the depletion region. The effect of finite transit time can be analyzed using the simple expression $\tau \approx w/v_s$. The validity of this calculation is based on the fact that in detector diodes, even at voltages of a fraction of a volt at the depletion zone, the electric field strength is above critical, i.e. more than 10 kV/cm for GaAs. So, electrons velocity is about $v_s \sim 10^5$ m/s. It follows from (11.6) that the value of w is 0.1 μm for a reverse voltage of 0.5 V. Then $\tau = 10^{-12}$ s. Naturally, for diodes operating at frequencies in the order of 1000 GHz, transition time significantly affects the operation of such a device.

When analyzing the operating mode of DD, it is important to take into account the probability of avalanche breakdown at reverse voltages. From the previous example, it follows that the field strength at a reverse bias of 0.5 V is 50 kV/cm. The breakdown voltage of the semiconductors used lies within the range 150–300 kV/cm (see Table 11.1). Thus, the detector diode is a structure sensitive to both

direct and reverse voltages. To avoid breakdown with an irreversible change in structure due to static electricity during installation, storage and operation, the diode is placed in a lead capsule or packed in aluminum foil after manufacture. The breakthrough can also be caused by a high power signal. Therefore, it is necessary to protect the diodes from their action.

12.1.2 Static and Dynamic Characteristics

To analyze the operation of the detector diode in the microwave band, it is not enough to know the parameters of the equivalent circuit and the cutoff frequency. In particular, a static volt-ampere characteristic is used to determine the dynamic parameters of the device. This is explained by the fact that the inertia of the current flow is very small and current follows voltage, according to the volt-ampere characteristics.

The base of the DD is the Schottky barrier. Its volt-ampere characteristic is described by (11.9). However, this formula should include the dependence of current on parasitic resistance R_s and take into account the imperfection of a real metal-semiconductor contact. Then expression (11.9) takes the form

$$I_d(U_d) = AT^2 S \mathrm{e}^{-\frac{e\varphi_k}{kT_e}} \left[1 - \mathrm{e}^{-\frac{e(U_d - I_d R_s)}{nkT_e}}\right] \tag{12.3}$$

According to this expression, the current through the diode at external voltage U_d is equal to the current through the Schottky barrier with applied voltage $U_d - I_d R_s$. In addition, the imperfection of the barrier is taken into account by including the *coefficient of non-ideality* n into the formula. The coefficient of non-ideality n = 1.1–1.5 for Schottky diodes and n = 1.5–2 for point diodes. Usually, an approximate formula is used to describe the volt-ampere characteristic

$$I = I_s \left(\mathrm{e}^{\frac{eU}{nkT}} - 1\right) \tag{12.4}$$

where $I_s = AT^2 S \mathrm{e}^{e\varphi_k/(kT)}$ is the thermal current of the diode (saturation current). This formula has sufficient accuracy if $R_s \ll R_j$, i.e. for rather small direct voltages on the diode. At T = 300 K, expression (12.4) is transformed to

$$I \approx I_s \left(\mathrm{e}^{\frac{40U}{n}} - 1\right) \tag{12.5}$$

The volt-ampere characteristic of real diode structures is similar to the theoretical dependence for SB, especially in the region of voltages less than the contact potential difference, but there are also several important differences. Figure 12.4 shows the volt-ampere characteristic of a gallium arsenide diode in the range of direct voltages of 0–1 V. Scale for current is logarithmic. The curve can be divided

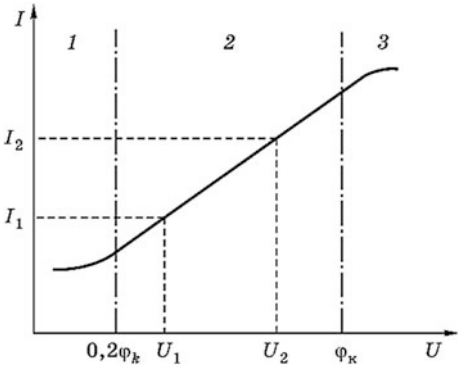


Fig. 12.4 Determination of diode parameters from volt-ampere characteristic

into a number of characteristic areas: linear section 2, which maps the exponential growth of current (as in an ideal barrier), as well as nonlinear sections 1 and 3. The deviation from the ideal characteristic in the region of low voltages (of order of $0.2\varphi_k$) is due to leakage currents through the surface, therefore, it is called the *leakage current region*. In region 3, an increase in current through the diode leads to an increase in the fraction of the voltage on resistance R_s in comparison with the fraction on the barrier. As a result, the characteristic deviates from exponential law, and at sufficiently high voltages, becomes linear.

Approximation of real characteristics by dependence (12.3) makes it possible to determine parameters such as contact potential, the non-ideality coefficient, and parasitic resistance.

Naturally, in order to determine these parameters, experimental points that are most sensitive to the desired parameter are chosen. So, φ_k and n are defined in the voltage range of section 2, determined by coordinates U_1, I_1 and U_2, I_2, while R_s is in section 3.

12.1.3 Dynamic Parameters

The dynamic parameters of DD include: current and tangential sensitivities, noise ratio, maximum dissipated power, and diode impedance.

A typical equivalent scheme of a detector is shown on Fig. 12.5a. A signal source throw the mathing transformer deliver signal to the detector diode (DD). High frequency current component goes throw the capicitor C_L while constant (rectified) comonent flowes throw the resistor R_I. Voltage formed on this resistor by the current makes up a useful signal.

Figure 12.5b shows implementation of this circuit in microwave range. The wire pin serves as a signal sours, and DD is connected in series with it. In order to close circuit for rectified current an inductor connects central wire and screen of the coaxial line. There is no need in such inductor if loop serves as a signal soursce (see Fig.12.5c). A loop connects inner and outer wires of the coaxial line.

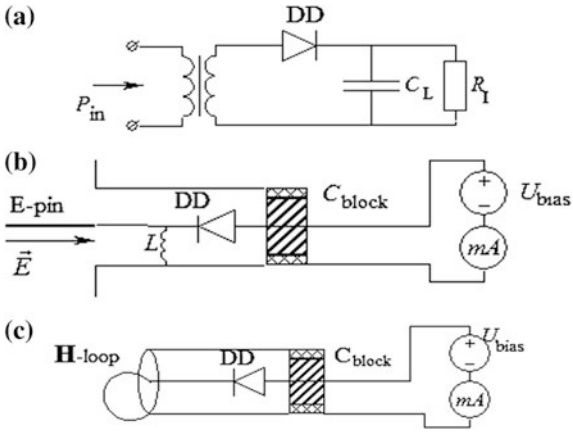


Fig. 12.5 Connection circuits of the detector diode in a microwave line: **a** detection circuit; **b**, **c** coaxial detector heads design

Usually, detection circuits in the microwave band are supplemented by matching circuits that reduce the power reflection from the diode. This ensures that sensitivity remains constant within the operating frequency range.

Current Sensitivity

Figure 12.6 graphically represents one of the possible modes of operation of the diode. Output direct current I_0, due to direct bias $U_{\text{bias}} > 0$, changes by ΔI_0 when applying microwave voltage $U_m \sin(\omega t)$, due to the nonlinearity of the volt-ampere characteristics of the diode, as shown in Fig. 12.6.

The efficiency of converting microwave power into rectified current is estimated by *current sensitivity* β. Its value is equal to the ratio of the rectified current increment ΔI_0 to the value of supplied microwave power P_f:

$$\beta = \frac{\Delta I}{P_f} = \frac{I_f - I_0}{P_f},$$

where $I_0 = I(U_0)$ is the current determined by a constant bias; and I_f is the diode current when a high-frequency signal and bias are applied to it.

Suppose that the voltage across the diode has constant and alternating components, the latter changes in harmonic order and its amplitude is much smaller than the constant component:

$$U = U_0 + U_m \cos(\omega t), \quad \text{and} \quad U_m \ll U_0.$$

The current through the diode is determined by its volt-ampere characteristic (12.4):

$$I = I_s(\mathrm{e}^{eU/(nkT)} - 1) = I_s(\mathrm{e}^{bU} - 1),$$

where $b = e/(nkT)$.

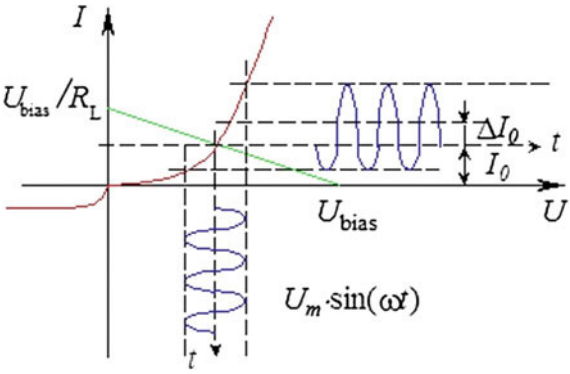


Fig. 12.6 Determination of current sensitivity

Expanding the exponent in the Taylor series around U_0 and truncating the series to the first three terms of the expansion, we obtain

$$I \approx I_0 + I' U_m \cos \omega t + \frac{1}{2} I'' U_m^2 \cos^2 \omega t$$
$$= I_0 + I' U_m \cos \omega t + \frac{1}{4} I'' U_m^2 + \frac{1}{4} I'' U_m^2 \cos 2\omega t.$$

Resistance of the diode to an alternating current with signal frequency ω is given by

$$R_f = \frac{U_m}{I_m} = \frac{U_m}{I' U_m} = \frac{1}{I'}.$$

The change in the constant component of the current under the action of microwave power is given by

$$\Delta I = \frac{1}{4} I'' U_m^2.$$

As can be seen, for weak signals, any detector (with an arbitrary form of volt-ampere characteristic) is quadratic, that is, the change in the constant component of the detector current is proportional to the square of the alternating voltage.

High-frequency signal power dissipated by a diode:

$$P_f = \frac{1}{2} U_m^2 / R_f.$$

Hence the current sensitivity

$$\beta = \frac{\Delta I}{P_f} = \frac{1}{2} \frac{I''}{I'}.$$

For volt-ampere characteristic of form (12.4)

$$\beta = \frac{1}{2}b = \frac{e}{2nkT} = \frac{20}{n}\ (\text{mA/mW}). \tag{12.6}$$

Thus, the maximum current sensitivity of the diode is limited to 20 A/W.

Taking into account the parasitic elements of the diode circuit in accordance with the equivalent circuit in Fig. 12.3 leads to the following formula:

$$\beta = \frac{20}{n\left(1 + R_s/R_j\right)^2\left(1 + (\omega/\omega_c)^2\right)},$$

which assumes an ideal match between the detector and the signal source.

The essential difference between the resistance of the diode and the wave resistance of the transmission lines necessitates the use of special matching circuits at the input of the detector sections. The sensitivity measured upon mismatch differs from the maximum for a given diode, in accordance with the formula

$$\beta_m = \beta_{\max}(1 - |\Gamma_d|^2),$$

where $|\Gamma_d|$ is the modulus of the diode reflection factor.

Figure 12.7a represents a typical Smith diagram for one example of a diode. An impedance hodograph of the diode in the frequency range 1–6 GHz is shown. Knowledge of the diode's resistance makes it possible to calculate the parameters of the matching circuits to obtain maximum sensitivity. As an example, Fig. 12.7b shows the manufacturer's proposed matching circuit, performed on the microstrip line (MSL) segments. Such a circuit makes it possible to provide a standing wave ratio VSWR of 1.2–1.4 in the frequency range under consideration. The parameters of the matching elements: substrate—thickness is 0.5 mm, dielectric permeability ε_r = 4; MSL1—the width of the strip is 0.45 mm, and the length is 4.1 mm; MSL2—the width of the strip is 0.2 mm, and the length is 4.1 mm.

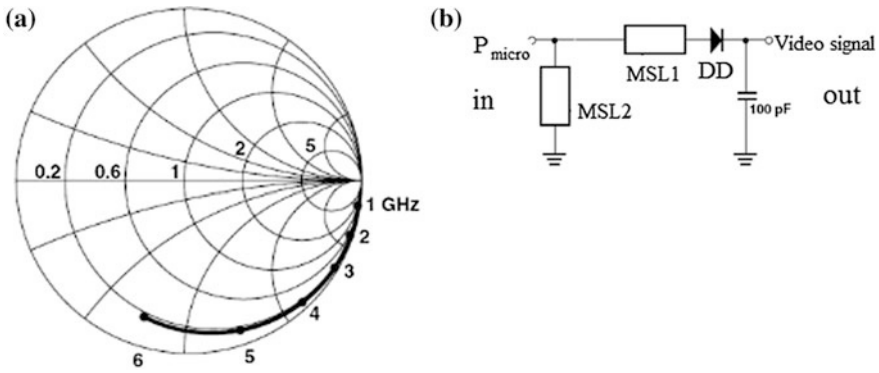


Fig. 12.7 Detector diode hodograph and a version of a matching circuit

Tangential Sensitivity, Noise Ratio

The ability to indicate a weak signal is limited by the self-noise of the detector diode used. The spectral density of the diode noise is subject to the distribution law presented in Sect. 2.5. In the low frequency region up to 10^5 Hz, flicker noise prevails in the device. Thus, the closer the modulation frequency is to the flicker noise cutoff frequency, the greater the sensitivity that can be obtained from the diode. In any case, an embedded and applied external electric field causes the difference of its noise characteristics from the noise of an equivalent diode resistance at the operating point. Usually this difference is characterized by the *noise ratio*. It determines how many times the average square of diode noise voltage $\langle U_{nd}^2 \rangle$ exceeds the average square of noise voltage $\left\langle U_{nR_d}^2 \right\rangle$ of resistance $R_d = 1/(\partial I_d/\partial U_d)$ i.e.

$$n_{\text{noise}} = \langle U_{nd}^2 \rangle / \left\langle U_{nR_d}^2 \right\rangle.$$

To observe the microwave signal envelope against the background of self-noise, it is necessary to send a signal from the detector to a narrow-band amplifier with a low noise factor. Figure 12.8 shows such a circuit: a low-noise amplifier LFA is attached to the detector section. LFA output signal is observed on an oscilloscope. This allows a visual observation of the modulating signal on the screen (the so-called "video signal"). If there is no signal on the screen, diode noise is observed (Fig. 12.8, left oscillogram). A signal with a carrier frequency modulated by rectangular pulses causes the appearance of the image shown in Fig. 12.8 (middle oscillogram).

At certain signal power it is possible to achieve an image in which the upper level of noise pulsations without a microwave signal coincides with the lower level

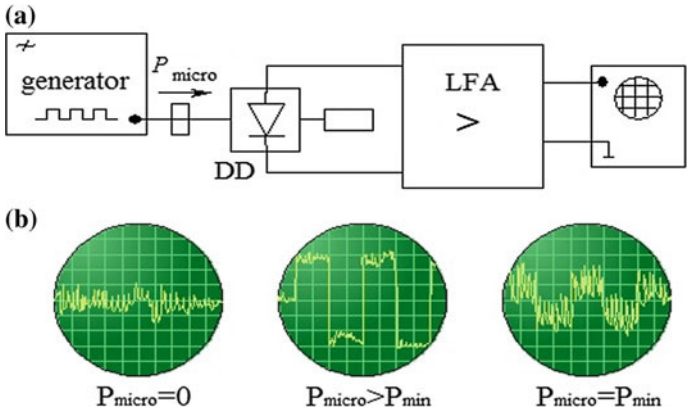


Fig. 12.8 Determining tangential sensitivity: **a** measurement circuit; **b** oscillograms at different power levels

in the presence of a signal (Fig. 12.8, right oscillogram). The power at which such a picture is observed is known as the tangential sensitivity P_{tg} (TSS).

It is customary to measure this power relative to 1 mW. In this case, tangential sensitivity is expressed in dBm:

$$\mathrm{TSS} = 10\lg\left(\frac{P_{tg}}{1\ \mathrm{mW}}\right),\ (\mathrm{dBm}).$$

It should be noted that TSS is not a strictly objective characteristic of the diode, since it is influenced by amplifier noise and subjectivity in estimating the coincidence of noise levels on the oscilloscope screen. However, in practice, this parameter has found wide application, since it reflects with sufficient accuracy the main functional purpose of the detector—detection of low-power signals. The typical value of tangential sensitivity in the decimeter band can be −60 to −80 dBm.

The heating of detector diodes limits the permissible maximum power supplied to the diode. It is called the *burnout power*. Usually, on the detector sections or device inputs, where the diode is used, the maximum power value is indicated. Typical values are 0.5–5 mW.

12.1.4 Circuit Application

In microwave devices, detector diodes are mounted in so-called detector heads, where the matching elements and connectors are integrated for connection to the typical microwave path: coaxial, microstrip, waveguide, etc. In hybrid IC diode is mounted on a printed circuit board on which the matching and connecting elements are located.

Figure 12.9a shows an example of the equivalent circuit of such an IC and the topology (Fig. 12.9b) of the device. The circuit parameters are given for the SMS7630 diode. The circuit provides a standing wave ratio (SWR) of the order of

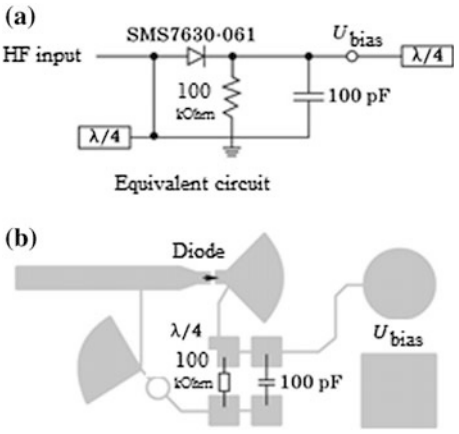


Fig. 12.9 **a** Equivalent circuit; **b** device topology

1.2 in the range 1–18 GHz. In comparison with the circuits shown in Figs. 12.6 and 12.7, this circuit contains additional elements that ensure isolation of the bias supply circuits, the microwave part and the output LPF chain. To do this, filters are applied made from distributed and lumped elements. A quarter-wave line segment at the input is used to transform the short circuit into an open circuit. This ensures the matching of input TL and the diode. The same segment connected to the other contact of the diode provides an open circuit condition at the output contact of the diode and the absence of the RF signal passage into the bias supply circuits.

Currently, the application field of detector diodes is expanding due to their use in wireless transmission systems. They receive microwave energy and convert it into direct current energy, similar to the work of solar cells. A large number of diodes forms a receiving antenna, called a rectifying antenna (or a rectenna). In such an application, small noises are secondary and energy conversions efficiency (ECE) and resistance to high-power are of primary concerns.

12.2 Mixer Diodes

12.2.1 Functional Designation and Usage Principle of the Mixer Diode

The nonlinearity of the diode volt-ampere characteristic can be used not only for obtaining DC signals or low-frequency envelopes, but also for transforming the signal to another carrier frequency. To do this, an unmodulated signal with frequency ω_{lo} is added on the diode, in addition to the useful signal with frequency ω_s. An additional generator with frequency ω_{lo} is called a heterodyne (from Greek "hetero"—other, and "dyn"—force) or *local oscillator* (LO).

Diodes converting signals from one frequency to another are called *mixer diodes* (MD).

Let us consider the "mixing" of signals of different frequencies on a nonlinear element and determine the specific requirements for diodes used in this circuits. For the analysis, we use the power series expansion of the volt-ampere characteristic of the diode, truncating it to the first three terms of the expansion:

$$\begin{aligned} i_d &= a_0 + a_1 U_d + a_2 U_d^2 + a_3 U_d^3 + \cdots \\ &= a_0 + a_1 (U_s \sin(\omega_S t) + U_{lo} \sin(\omega_{lo} t)) \\ &\quad + a_2 (U_s \sin(\omega_S t) + U_{lo} \sin(\omega_{lo} t))^2 + \cdots, \end{aligned} \tag{12.7}$$

where $U_s \sin\sin(\omega_S t)$ is the "useful" signal; $U_{lo} \sin\sin(\omega_S t)$ is the harmonic local oscillator signal; and a_0, a_1, …, a_n are the coefficients of expansion, dependent on the form of the volt-ampere characteristics.

We confine ourselves to a small-signal analysis, omitting all the terms of the series, except for the first three. As seen from (12.7), the first two terms do not

change the spectrum of the signals. New *combinational frequencies* occur due to the quadratic term (and to higher order terms):

$$\begin{aligned}
&[a_2(U_s\sin(\omega_s t)+U_{lo}\sin(\omega_{lo}t)]^2\\
&=a_2U_s^2\sin^2(\omega_s t)+2a_2U_sU_{lo}\sin(\omega_s t)\sin(\omega_{lo}t)+a_2U_{lo}^2\sin^2(\omega_{lo}t)\\
&=\frac{1}{2}a_2U_s^2+\frac{1}{2}a_2U_{lo}^2-\frac{1}{2}a_2U_s^2\cos(2\omega_s t)-\frac{1}{2}a_2U_{lo}^2\cos(2\omega_{lo}t)+\\
&\quad+a_2U_sU_{lo}\cos(\omega_s-\omega_{lo})t+a_2U_sU_{lo}\cos(\omega_s+\omega_{lo})t
\end{aligned}\tag{12.8}$$

This expression shows that the quadratic term not only determines the harmonic components with frequencies $2\omega_s$ and $2\omega_{lo}$, but also components with frequencies $\omega_s \pm \omega_{lo}$. The amplitude of these combination components $a_2U_sU_{lo}$ depends on the expansion coefficient a_2 and is proportional to the product of the amplitude of the "useful" signal and the local oscillator signal amplitude. Thus, the "useful" signal can be converted to one of the combination frequencies. We talk of *intermodulation*, i.e. the transfer of modulation from one frequency to another. In the approximation considered, only the quadratic term in the expansion of the volt-ampere characteristic was taken into account. So it is *second order intermodulation*. It is not difficult to show that preserving in the series higher order terms leads to appearence in the output signal spectrum combination frequencies $n\omega_{lo} \pm m\omega_s$, where n and m are the whole numbers $0, \pm1, \pm2, \ldots$. The sum of the numbers n and m characterizes the order of intermodulation. Figure 12.10 shows the second-order intermodulation frequency spectrum. In addition to the original frequencies, there are signals of difference and total frequencies $\omega_\pm = \omega_{lo} \pm \omega_s$, as well as second harmonics of the signal and the local oscillator $2\omega_s$ and $2\omega_{lo}$. The appearance of the second harmonic of the signal is interpreted as an increase in the nonlinear distortion of the signal. To reduce this undesirable effect, the amplitude of the signal should be small in comparison with the bias voltage.

Usually, the local oscillator frequency is chosen to be higher than the signal frequency, and the difference frequency, called the *intermediate* frequency $\omega_i = \omega_{lo} - \omega_s$ is chosen for further use.

Since the amplitude of the intermediate frequency signal is directly proportional to the amplitude of the local oscillator signal, in order to increase the conversion efficiency it is necessary to increase the power of the local oscillator.

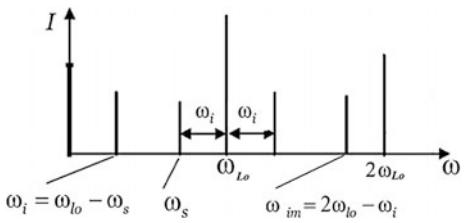


Fig. 12.10 The simplest spectrum of mixers

Analysis of expression (12.8) shows that the combinational frequencies arise due to the component containing the multiplication of signals $2a_2U_sU_{lo}\sin(\omega_s t)\sin(\omega_{lo}t)$. This explains the use of the multiplication sign $\otimes$ to denote mixers on block diagrams.

An important role in the mixer operation is played by the so-called image frequency $\omega_{im} = \omega_{lo} + \omega_i$ (Fig. 12.10). If an interference signal at the image frequency is present at the input of the device, then its combination with the local oscillator signal also generates an intermediate frequency signal. The second LO harmonics can also serve as a source of the image frequency, since

$$2\omega_{lo} - \omega_i = 2\omega_{lo} - \omega_{lo} + \omega_s = \omega_{lo} + \omega_s = \omega_{im}$$

An even more complicated picture arises when there is an interference with a large amplitude in the input signal. In Fig. 12.11, this interference is represented by additional signal $U_n\sin(\omega_n t)$. It can be represented as a signal from an additional heterodyne. In this mode, additional combinational frequencies arise that complicate the spectral composition. The increase in signal power also significantly changes the spectral composition at the output. This is explained by the faster growth of the amplitude of higher order terms than the quadratic term. Intermodulation components appear of a higher order than the second. To defuse the spectrum, filters and balanced compensation circuits to dump the local oscillator harmonics and its noise are used.

The main difference between the operation modes of the mixer diode and the detector diode is the large forward and reverse voltages due to the large amplitude of the LO signal. Thus, for reliable operation of diodes in mixer circuits, it is necessary to increase the breakdown voltage of the diode and increase its tolerance to heating by a high current. Nevertheless, the requirements of low inertia should be preserved. The fulfillment of these requirements is possible when using semiconductor structures with a lower level of doping than in DD.

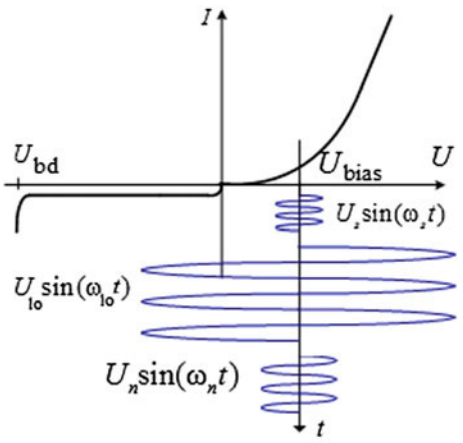


Fig. 12.11 Analysis of mixer operation

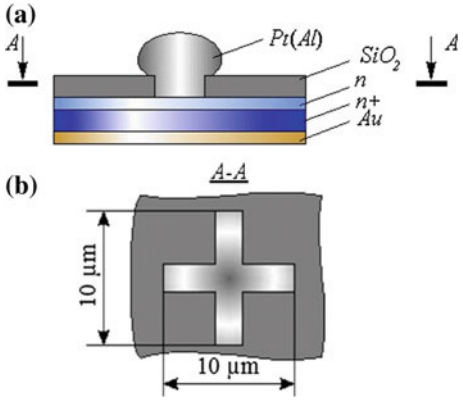


Fig. 12.12 Form of a mixer diode

Typically, gallium arsenide is used as the structure material as having higher breakdown field strength (250 kV/cm) than silicon (100 kV/cm) and higher low-field mobility.

Figure 12.12 shows a structure of the mixer diode. The cross-shaped structure of the metal electrode forming the Schottky barrier provides a reduction in the junction capacity in comparison with the circle and improves the mechanical properties of the connection with the semiconductor structure. The concentration of the impurity in the low-doped n-region does not exceed 10^{17} cm^{-3}. The concentration of the impurity in the n^+-region is usually of the order of 10^{19}–10^{20} cm^{-3}.

12.2.2 Mixer Diode Schemes

Figure 12.13 shows the circuit of the simplest mixer. In it, signals from antenna A and from the local oscillator arrive at the mixer diode. A constant bias from source U_{bias} is also applied to the diode. A bandpass filter BPF (preselector) allows tuning to the signal with the desired frequency. The local oscillator signal is fed to the MD through a directional coupler. The bias to the diode is fed through a low-pass filter

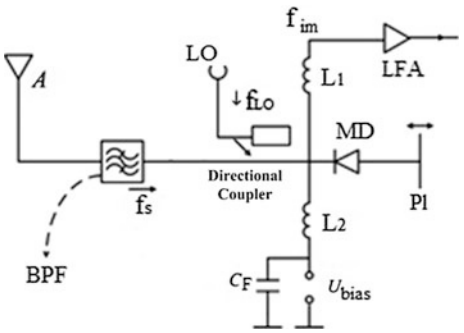


Fig. 12.13 The simplest microwave band mixer circuit

formed by components C_F, L_2. The intermediate frequency signal is fed to an intermediate frequency amplifier (IF amplifier) via inductance L_1 to prevent the passage of the high-frequency components of the spectrum to the amplifier input. In order to match the diode to the microwave path, a moving plunger Pl is used.

To compensate the amplitude noise of the local oscillator, different balanced circuits are used. Figure 12.14 shows a widely used two-diode balanced mixer (BM) circuit with a 3 dB waveguide directional coupler (Riblet coupler). It consists of two diodes, D_1 and D_2, included in the output coupler ports, to which a signal at frequency ω_s with power P_s and a local oscillator signal at frequency ω_{lo} with power P_{lo} are fed through a Riblet coupler. Let the initial phases of these oscillations be equal to zero. Then due to the phase shift of the signals in the output arms of the Riblet coupler at $\pi/2$ we get:

- on diode D_1, the voltage is composed of the components

$$u_{1s} = U_s \cos(\omega_s t) + U_{lo} \cos(\omega_{lo} t + \pi/2),$$

- on diode D_2:

$$u_{2s} = U_s \cos(\omega_s t + \pi/2) + U_{lo} \cos(\omega_{lo} t).$$

The diodes are connected in parallel with opposite polarities, so the current difference flows through the load R_0 with intermediate frequency $\omega_i = \omega_{lo} - \omega_s$. At full circuit symmetry, the resulting current is

$$\begin{aligned} i_i &= i_{1i} - i_{2i} = I_i\{(\cos[(\omega_s t - \pi/2 - \omega_{lo} t] - \cos[\omega_s t - (\omega_{lo} t - \pi/2)]\} \\ &= 2I_i \sin(\omega_s - \omega_{lo})t. \end{aligned}$$

Useful signal currents are combined in the load in-phase.

Let us represent the amplitude noise of the local oscillator acting in the signal band ω_s and in the image band channel ω_{im} as the side bands of the amplitude modulation oscillations with carrier ω_{lo}, i.e. $u_{nlo} = U_n \cos((\omega_{lo} + \omega_{im})t - \varphi_n)$, where φ_n is the phase change of the noise signal. Reception of local oscillator noise in the signal band on the counter-switched diodes gives noise current: $i_{ns} = i_{1ns} - i_{2ns}$. Following the previous analysis, we obtain mutual compensation of the local oscillator noise in the load.

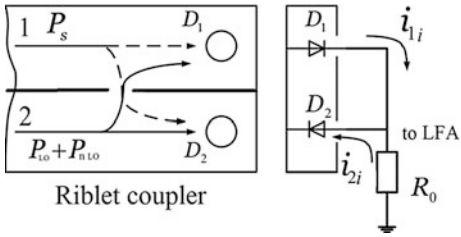


Fig. 12.14 Balance mixer circuit

Due to the inevitable difference in the diode parameters of the real BM, the local oscillator noise is suppressed by 15–30 dB. In practice, pairs of diodes are selected for BM.

12.3 *p-i-n* Diodes

12.3.1 *Structure, Principle of Operation and Equivalent Circuit of the* p-i-n *Diode*

Resistance of detector and mixer diodes strongly depends on the applied voltage. The low power levels of such diodes severely restrict their use.

For switching powerful signals, special structures have been developed, called *p-i-n* diodes. They are also known as *control diodes*.

It follows from the name that the structure of these diodes contains *p* and *n* layers of a semiconductor with an undoped *i*-region placed between them. Naturally, the inertia of such a device sharply increases due to the influence of a number of factors, in particular, the transit time of carriers, diffusion capacity, and carrier recombination time in the *i*-region. Such a device can be used as a rectifier at low frequencies, but cannot be used as a detector of weak microwave signals. The volt-ampere characteristic of such a diode in comparison with a typical *p-n* diode or SBD has a positive branch shifted towards higher voltages. The presence of the *i*-region reduces the electric field intensity in the base and provides an increase in the breakdown voltages.

For effective switching of microwave signals, the diode must have minimum impedance under positive control voltage (forward direction) and maximum impedance negative control voltage (reverse direction). The second condition is equivalent to reducing the capacity of the device, which can be provided not only by reducing the area of the electrodes (as in DD and DM), but also by increasing the distance between them. An extended (from 10 to 400 μm) *i*-region reduces the barrier capacity of the device.

Consider a typical *p-i-n* diode, with a structure as shown in Fig. 12.15. The structure is a semiconductor (most often Si), which has two heavily doped regions of p^+ and n^+ types (10^{18} cm^{-3}), and the *i*-region (base) between them. The resistivity of this region is 10^3–10^4 Ωcm. On the outer sides of the p^+ and n^+ regions, ohmic contacts are deposited. The area of the active diode part is in the range 10^{-1}–1 mm^2.

Let us analyze the operation of the device when it is exposed to a small microwave signal and a rectangular pulse of the control voltage. Suppose that the diode is connected to the microwave transmission line in parallel. Figure 12.16a shows the variation of control voltage U, the shape of the current through the diode I and the microwave voltage U_{micro} on the diode as functions of time. At the initial moment, the diode is biased in the reverse direction with voltage U_{rev}, the current through the diode is small, and is equal to the saturation current I_s. Alternating voltage across the diode U_m is determined by the level of incident power and the

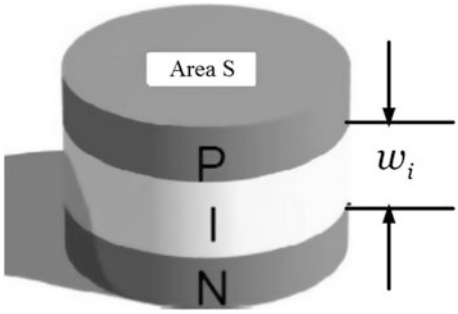


Fig. 12.15 Structure of the *p-i-n* diode

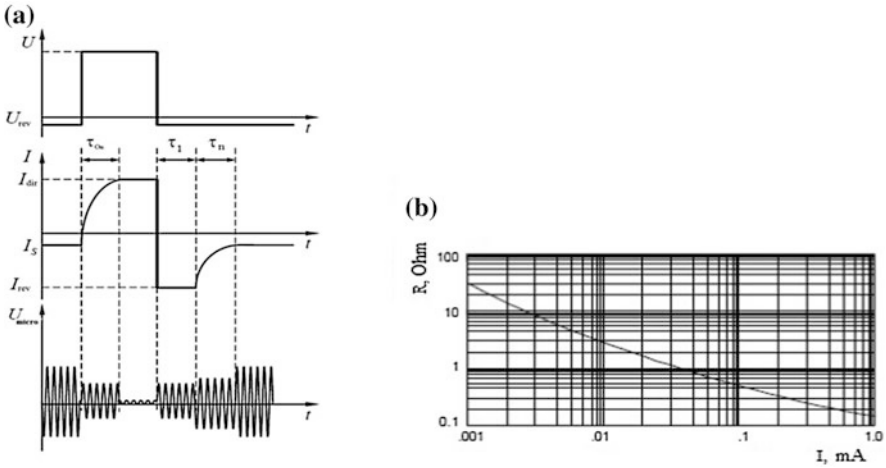


Fig. 12.16 Analysis of the switching process in the diode: **a** transient processes in the diode; **b** typical resistance change

shunting action of the diode capacitance. In this mode, it is maximum, and the resistance of the diode is large. When the polarity of the control voltage is changed to forward direction, charge carriers are injected into the high-resistance *i*-region from p^+ and n^+ regions, forming electron-hole plasma.

The filling time of the *i*-region with mobile carriers $\tau_d \approx w_i/2v_s$. As the *i*-region is filled, the current through the diode increases and the resistance drops. The time interval of this process is determined by the value of τ_d and shunt capacitance (Fig. 12.16a). This time interval is called *turn-on time* τ_{on}. The current reaches a maximum value I_{dir}, and resistance becomes minimal. The microwave voltage on the diode U_m drops almost to zero (the electromagnetic wave is reflected from the diode). When the polarity of the control voltage is reversed, the number of mobile charge carriers does not change instantly. Therefore, just after switching, reverse current I_{rev} is equal in amplitude to direct polarity current I_{dir}. Further, the process of extracting charge carriers from the base (τ_1) and their recombination $(\tau_{n,p})$ starts.

The current through the diode decreases slowly. When no free carrier is left in the base, the resistance of the diode increases sharply and the current through it will decrease to the value I_s. Time $\tau_{rec} \approx \tau_{on} + \tau_{n,p}$ from the moment of switching to the moment when the reverse current reaches the steady-state value I_s is called *recovery time*. Microwave voltage U_{micro} reaches its maximum value again. The diode stops shunting the transmission line, and a signal with minimal attenuation passes through it again. Thus, the switching of the microwave signal occurs with a delay relative to the moment of switching the bias voltage.

Let us estimate the resistance of the diode quantitatively under forward bias, assuming that the main contribution is provided by the resistance of the *i*-region. Figure 12.17 shows the distribution of the mobile charge carrier concentration with a positive control voltage polarity at a high injection level from the p^+ and n^+ regions.

The lifetime of electrons and holes is much longer than the transite time of the *i*-region. The middle part of the structure, filled with electrons and holes, has low resistance and has no large voltage drop. This corresponds to the condition when injected carrier concentrations are equal, i.e. $n_i = p_i$, as shown in Fig. 12.17. The concentration of minority carriers in the p^+ and n^+ regions p_n and n_p is small, due to the short time of charge recombination in the volume of heavily doped regions. Under these assumptions, the total charge of the mobile carriers in the *i*-region is equal to the current multiplied by the recombination time of these carriers τ in this region, i.e. $Q = I \cdot \tau_{n,p}$. Then the resistance of the *i*-region is calculated by the formula for the resistance of the sample at the specified conductivity $\sigma = 1/q(n_i\mu_n + p_i\mu_p)$, base length w_i and its area S:

$$R_{i+} = \frac{w_i}{e(n_i\mu_n + p_i\mu_p)S} \approx \frac{w_i}{(Q/V_i)(\mu_n + \mu_p)S} = \frac{w_i^2}{(\mu_n + \mu_p)Q} = \frac{w_i^2}{(\mu_n + \mu_p)I\tau_{n,p}} \tag{12.9}$$

where V_i is the volume of the *i*-region; S is the contact area; n_i, p_i is the concentration of injected electrons and holes; and $\mu_{n,p}$ is their mobilities.

It follows from (12.9), that the resistance of the *i*-region does not depend on the area of contact. In practice, recombination on the lateral surface of the *i*-region reduces the

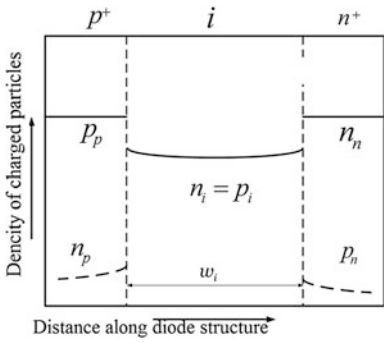


Fig. 12.17 Distribution of concentration of mobile carriers upon forward bias

effective lifetime of carriers $\tau_{n,p}$, complicating the simple dependence obtained, but the general trend remains. As an example, Fig. 12.16b shows the experimental dependence of the resistance on the flowing current for one of the *p-i-n* diodes.

Formula (12.9) does not fully describe the impedance of the diode when exposed to a microwave signal. The equivalent circuit of the *p-i-n* diode also includes barrier capacitance $C_i \approx \varepsilon S/w_i$, parasitic series contact resistance R_s, and case capacitance C_p, as well as contact inductance L_s (Fig. 12.18), where equivalent circuits of the diode biased in forward and backward directions are shown.

A sharp change in the resistance, depending on the control voltage, led to the use of *p-i-n* diodes in switchers, modulators, controlled attenuators, and phase shifters. In these circuits, both sequential and parallel diode connections in the microwave line are used. Some circuits are shown in Fig. 12.19. In these circuits: C_b is the blocking capacity, L_1, L_2 are the inductors playing the role of a low-pass filter, R_{lim} is the limiting resistance, and K_1 is the key in control circuit.

When control voltage U is applied with forward polarity in the series connection circuit (Fig. 12.19a), diode D is equivalent to a small resistance (short circuit) connecting two ends of one of the transmission line wires. This ensures that the U_m signal is transmitted without significant reflection loss. At reverse bias, the diode has a small capacitance C_i, equivalent to a large resistance (idle stroke) and leads to reflection of the signal from the diode.

For a parallel connection circuit of the diode, the picture is reversed: with forward bias, the diode shorts the transmission line, causing practically complete reflection of the signal. With reverse bias, the diode is equivalent to a large resistance, which does not practically affect the propagation of the signal in the transmission line.

The quantitative assessment of switching quality is determined by two main parameters: *transmission loss and isolation*. Transmission loss L is the ratio of the power transmitted through switch P_{out} to incident power P_{in}, in the transmission mode, expressed in decibels. In such a way, $L = -10lgP_{\text{out}}/P_{\text{in}}$.

Isolation A is characterized by the ratio logarithm of the power passing through the switch to the incident power in reflection mode, i.e. $A = -10lgP_{\text{out}}/P_{\text{in}}$. In

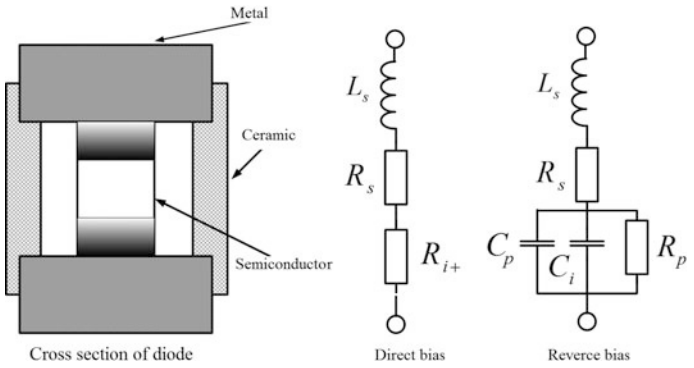


Fig. 12.18 Design and equivalent circuits of the *p-i-n* diode

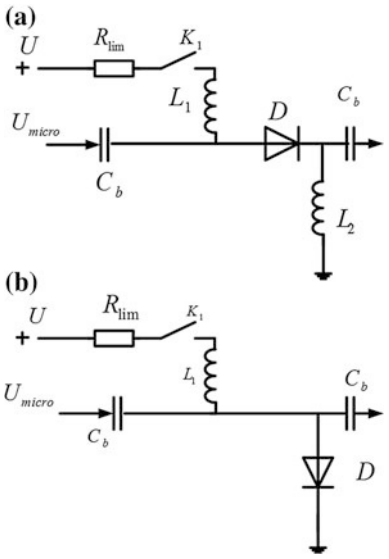


Fig. 12.19 Diode installed in microwave line

modern switch circuits, it is possible to obtain a transmission loss of less than 0.5 dB and isolation of more than 30 dB at a frequency of 2,400 MHz.

Carriers injection in the i-region increases during positive phase of high amplitude microwave signal. These carriers do not have time to fully recombine during the negative phase of the signal. Hence the number of charge carriers in the i-region increases from period to period while the region is completely filled. Naturally, the resistance of the *i*-region decreases in this case. The greater the incident power, the greater the microwave voltage at the diode and the higher the injection level, and the greater the charge of mobile carriers in the *i*-region, leading to a decrease in the resistance of the diode.

This property of *p-i-n* diodes is used when creating limiters and controllable attenuators. Figure 12.20 shows diode connection circuits in a waveguide and typical amplitude characteristic of the limiter. It is important to note that the reduction in signal power at the output of the device occurs due to reflection from the diode and only a small part is scattered inside diode. This allows switching

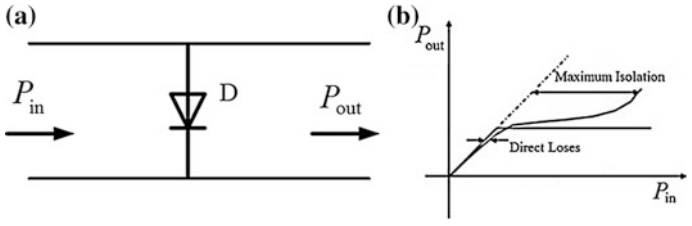


Fig. 12.20 Limiter circuit operation: **a** diode in waveguide; **b** amplitude characteristic

signals with a power exceeding 1 kW in continuous mode in the decimeter range. However, the nonlinearity of the diode leads to undesirable intermodulation distortions in the switched signal with increasing power.

12.3.2 *Peculiarities of the Use of* p-i-n *Diodes in Circuits*

Switchers

Consider the simplest circuits of switchers, phase shifters and limiters. Switchers are the most important elements of the transmitter-receiver paths of any radio electronic system. Low power losses in transmission mode, high isolation with locking mode, and low nonlinear distortions are required from switchers.

As an example, Fig. 12.21 shows the topology of a two-channel switcher SPDT (Single Pole Double Throw) using four diodes D_1–D_4. In this circuit, the microwave signal coming from port 1 falls into port 2 or 3, depending on the polarity of the control voltage applied to the diodes from sources U_A and U_B. When forward voltages are applied to D_1 and D_2 and reverse voltages to D_3 и D_4, the signal moves to port 3. With reverse polarity it moves to port 2. The length of the microstrip lines, which determines the diode places with respect to the input T-branch, is selected from the condition of identical phases of the signals arriving at the diodes and of ensuring small losses. Blocking capacitors exclude the closure of the control circuits through the conductors of the microstrip path.

It should be noted that with the development of MEMS (Micro Electro Mechanic Systems), switches with *p-i-n* diodes have now got decent competition in the form of MEMS switches. Such devices function as ordinary mechanical keys in the low-frequency range. Their parameters in the low-frequency portion of the microwave band exceed the parameters of *p-i-n* diodes, especially in terms of dynamic range of use, distortion and introduced loss. However, the lifetime and stability of MEMS are worse than in *p-i-n* diodes.

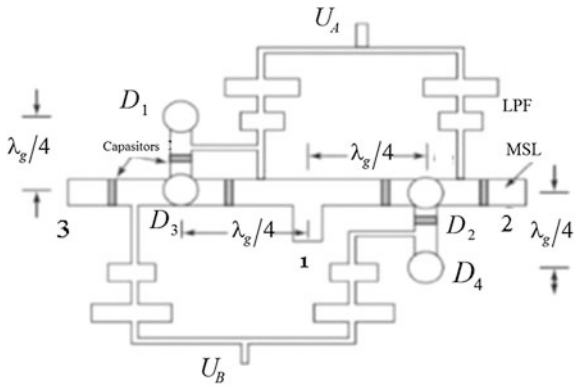


Fig. 12.21 Two-channel switch circuit

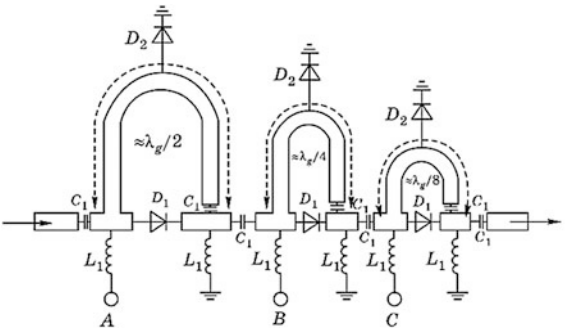
Phase-Shifters

With the development of the phased array antenna technique (PAA) and the improvement of the measurement technique, it became necessary to create fast-controlled phase shifters. For this purpose, *p-i-n* diodes in particular are used, providing a discrete change in the phase of the transmitted signal (digital phase shifters). Usually such a change is achieved by changing the distance traveled by the wave.

An example of a phase shifter using *p-i-n* diodes is shown in Fig. 12.22. The device consists of sections of transmission lines and stubs of various electrical lengths. Six diodes play the role of commutating keys, determining the way the wave travels. The state of these keys is determined by control signals supplied from sources *A*, *B* and *C*. A table of control voltages providing a phase change with a certain step ("discrete") is shown in Fig. 12.22. At level 1 the diode is open, at level 0 it is closed. Inductances L_1 exclude the passage of the microwave signal into the control circuits. Blocking capacitors C_1 provide isolation between the elements by direct current, without introducing losses into the transmitted microwave signal. If a single level exists on all control inputs (*A*, *B*, and *C*) the signal will pass along the shortest path. The resulting phase incursion of the wave is considered to be zero, since the phase difference, and not its absolute value, is important for the operation of the device. In the case of a zero level at all control inputs, the signal travels the longest path along all stubs, providing the maximum phase incursion in comparison with the smallest path. With different signals on the diodes, the path difference and the phase incursion will be between these limit values. The number and dimensions of the arc-wise elements are selected so as to provide a phase incursion multiple to the phase incursion in a smaller element. This phase incursion is called *discrete*. In this circuit, it amounts to 22.5°.

Limiters and Attenuators

In the operation of these devices, the nonlinearity of the diode at large values of the supplied microwave power is used. The simplest circuit of a limiter with auto-bias is shown in Fig. 12.23. In addition, a Schottky barrier diode is used along with a *p-i-n* diode, and a SBD detects part of the transmitted signal power and provides a



A	B	C	Δφ
0	0	0	157,5°
0	0	1	135°
0	1	0	112,5°
0	1	1	90°
1	0	0	67,5°
1	0	1	45°
1	1	0	22,5°
1	1	1	0

Fig. 12.22 Circuit of the discrete phase shifter and table of control signals with resulting phase incursion

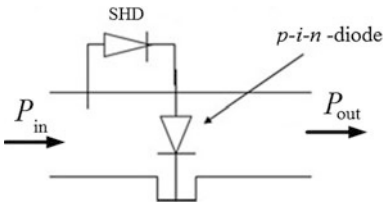


Fig. 12.23 Limiter circuit with auto-bias

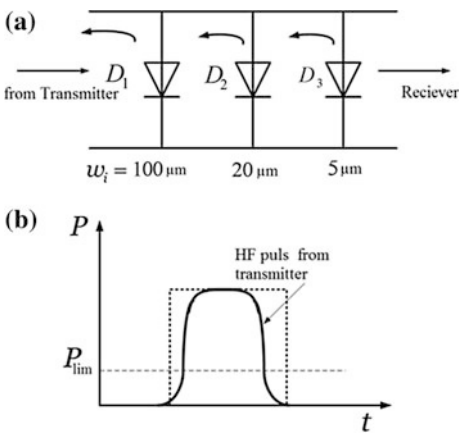


Fig. 12.24 High power level limiter: **a** connection circuit; **b** time diagram

forward bias to the limiting diode. This connection circuit allows a reduction of the threshold power of operation of such a device and its inertia. In it, the SBD acts as a device that reacts quickly to the supplied microwave signal of the device, providing a forward bias to the limiting diode.

For operation at a high power level, a block of diodes with different *i*-region widths are used. They are connected one after other, as shown in Fig. 12.24a. In general, the duration of the front of the input signal (Fig. 12.24b) should not be less than the response time of the most powerful diode in the circuit. In such a limiter, at the initial moment, a diode with a smallest thickness in the *i*-region of 5 μm, having a small inertia, is triggered. After that, the diode with average thickness of 20 μm is switched, and finally, at an increased power level, a "slow" high-power diode with a thickness of 100 μm are triggered.

12.4 Varactor Diodes

12.4.1 Structure, Equivalent Circuit and Applications of Varactor Diodes

A *varactor diode* (VD) is a diode, specially designed to use the dependence of its reactance on an applied voltage. Diodes with a Schottky barrier (see Sect. 12.1) or

p–n junctions (see Sect. 12.2), in which the capacitance depends on the applied voltage, can be used as a structure that realizes this nonlinearity. Varactor diodes are used for electrical frequency tuning, in frequency multipliers, parametric amplifiers and other facilities. In transceivers, these diodes are used, first of all, as frequency-tunable oscillators (VCO, a voltage-controlled oscillator) and tunable filters.

12.4.2 Varactor Structures

One of the most common structures used for varactor diodes is shown in Fig. 12.25. This is a metha-structure (from the English "metha," meaning a flat-top mountain) consisting of n^+ substrate 5, with an *n*-type epitaxial film 4 growing on it. By diffusion, a *p*-layer 2 is formed in this film. Between the *n*- and *p*-regions, a depletion layer 3 is found. Ohmic contacts 1 are deposited on the *p*-layer and substrate. A Schottky barrier can also be used in varactor diodes.

The form of voltage–capacity characteristic $C = f(U)$ of these structures depends on the doping profile (see Chap. 11). In the general case, the distribution of dopants near the metallurgical boundary is described by the function$N_d(x) - N_a(x) \sim Bx^m$, where m is the coefficient determining the doping profile. Coordinate x is directed from the donor region to the acceptor region starting from the metallurgical junction plane.

For $m = 0$ the level of doping on different sides of the metallurgical boundary does not depend on the coordinate. Such a junction is called "abrupt." At $m \approx 1$, the level of doping varies linearly. In this case, the junction is called "gradient." For $m < 0$, the degree of doping increases when approaching the metallurgical junction. Such a junction is called "hyper abrupt." The doping profiles for these three cases are shown in Fig. 12.26.

The difference in the doping profiles causes a difference in dependencies $C = f(U)$ Solving the Poisson equation for the depletion layer in the general case, we obtain:

$$C = C_0 \left(1 - \frac{U_{bias}}{\varphi_k}\right)^{-\frac{1}{m+2}}, \tag{12.10}$$

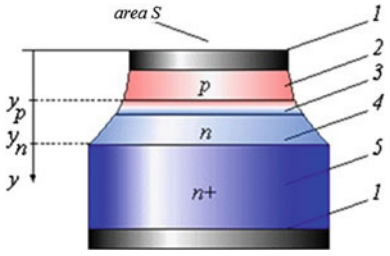


Fig. 12.25 Structure of the varactor diode

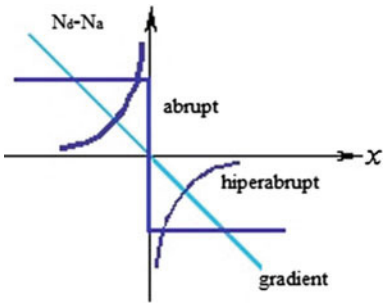


Fig. 12.26 Varieties of doping profiles of p-n junctions

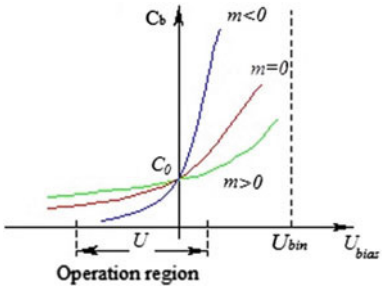


Fig. 12.27 Voltage-capacitance characteristics at different doping

where $C_0 = S\sqrt{\frac{e\varepsilon N_A N_D}{2(N_A+N_D)\varphi_k}}$ is the diode capacity at zero bias voltage. For $m = 0$, the formula coincides with the previously obtained expression for SBD.

Figure 12.27 shows the dependence of the capacitance on the applied voltage for the three typical barriers under consideration. The graphs are normalized to the same capacitance C_0. It follows from the graphs that diodes with a gradient junction have the weakest dependence of capacitance on voltage, the hyper abrupt junction, in contrast, causes the highest value of derivative $\partial C_b/\partial U$.

Varactor diodes are characterized by the following parameters: capacity at zero voltage C_0; overlap factor $K = C_o/C_{\min}$, where $C_{\min}$ is the capacitance of the diode with maximum allowable reverse voltage; sensitivity $(\Delta C/C)/(\Delta U/U_{bias})$; and steepness $\Delta C/\Delta U$. A decrease $C_{\min}$due to the increase (in absolute value) of the bias voltage is limited by the avalanche breakdown of the depletion layer. In this connection, the value of permissible reverse voltage $U_{\text{rev max}}$ is specified for diodes. For different types of diodes, this value varies from 5 to 50 V. Overlap factors vary from 1.5 for an abrupt junction to 3 for a hyper abrupt junction.

Figure 12.28 shows real doping profiles, the use of which are most appropriate. Donor profile 2, in the accepted terminology, corresponds to an abrupt barrier, and profile 1 corresponds to a hyper abrupt barrier. For better understanding, Fig. 12.28(b)

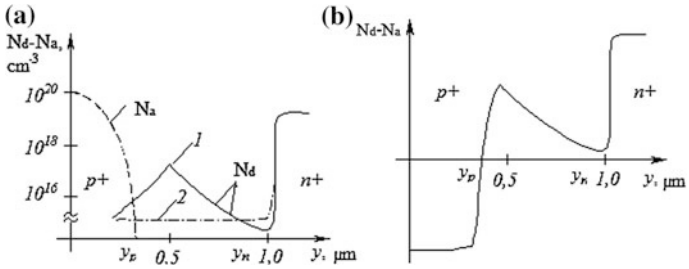


Fig. 12.28 Examples of real doping profiles for varactor diode

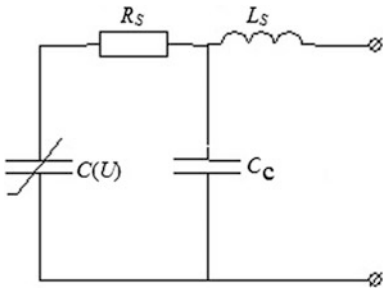


Fig. 12.29 Equivalent circuit of the varactor diode

shows characteristic $N_D(x) - N_A(x)$ for a hyper abrupt junction. To the right of the junction boundary ($y \approx 0.4\mu$m), the donor concentration decreases sharply.

An equivalent circuit is used to describe the RF properties of the varactor diode with reverse bias voltage (Fig. 12.29). This circuit is similar to the circuits for detector, mixer, and *p-i-n* diodes (see, for example, Fig. 12.4). The nonlinear properties of the device are represented by capacitance *C(U)*.

Parasitic resistance R_s determines the noise produced by the diode in parametric amplification circuits. In addition, this resistance reduces the efficiency of frequency multipliers. To quantify the effect of these parameters, *the Q-factor and the cut-off frequency* f_c are used. The Q-factor is determined by the ratio of the reactive energy stored in capacitance *C(U)*, to the power dissipated in resistance R_s.

Analysis of the circuit in Fig. 12.29 shows that the Q-factor calculation can be reduced to the calculation of the ratio of resistances under consideration, i.e.:

$$Q = \left|\frac{1/\mathrm{i}\omega C_0}{R_s}\right| = \frac{1}{2\pi f R_s C_0} = \frac{f_c}{f} \tag{12.11}$$

where $f_c = 1/(2\pi R_s C_0)$ is the cut-off frequency, analogous to the critical frequency of the DD. According to (12.11), an increase in the quality factor is achieved by reducing R_s and C. One of the ways to increase Q and critical frequency is to use structures based on gallium arsenide. They have R_s values much lower than silicon structures due to a higher mobility of electrons.

12.4.3 Heterostructure Barrier Varactor (HBV diode)

Apart from the varactor structure shown in Fig. 12.15 there also exists heterostructure barrier diodes (HBVs). The HBV is a semiconductor device which shows a variable capacitance with voltage bias, similar to a varactor diode. Unlike the usual varactor, it has an anti-symmetric current–voltage characteristic and symmetric capacitance–voltage dependence, as shown in Fig. 12.30. The device was invented by Erik Kollberg together with Anders Rydberg in 1989 at Chalmers University of Technology.

The inset of the figure shows the circuit schematic symbol for an HBV. From the symbol, one can conclude that the HBV consists of two, back to back connected diodes (such as Schottky diodes for instance) with opposite forward directions. The gap in the middle of the diode symbol represents the inherent capacitance of the device.

The electrical characteristics of an HBV are realized by separating two layers of a semiconductor material (1) with a layer of another semiconductor material (2) as shown in Fig. 12.31. The band gap of material (2) should be larger than for material

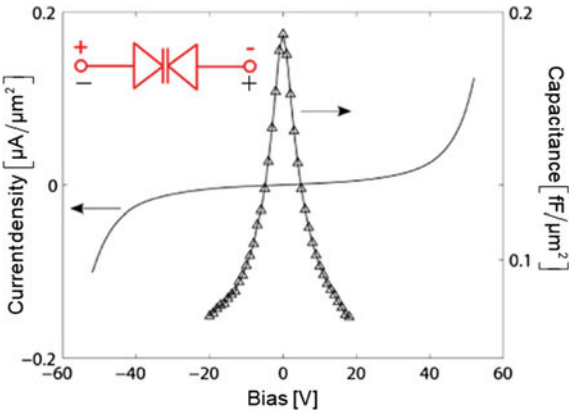


Fig. 12.30 Typical IĐV and CĐV characteristics of the HBV diode

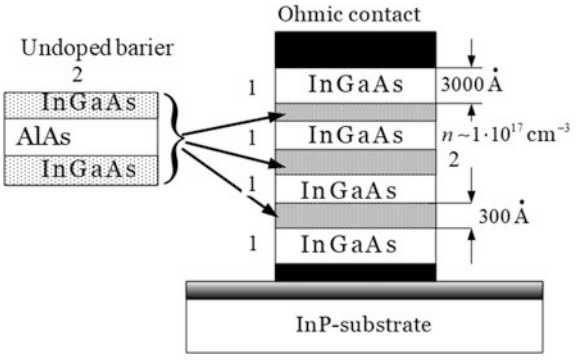


Fig. 12.31 Typical structure of the HBV diode with 3 barriers

(1), resulting in a series of barrier to carriers travelling through layers (1)-(2)-(1). Layer (1) is usually n-doped which means that the majority of carriers in this device are electrons. At different bias voltages the carriers are redistributed and the distance between the carriers on each side of the barrier (2) is different. As a consequence the HBV has electrical properties resembling a parallel plate capacitor, with a voltage dependent gap d.

12.4.4 Applications of Varactor Diodes

Electrical Tuning of Resonance Frequency

A change of capacitance due to an applied bias voltage is used in schemes for the electrical tuning of resonance frequency, as well as in special switchers with low losses. Fig. 12.32 shows an equivalent scheme of an oscillatory circuit with a VD. The resonance frequency of the oscillatory circuit $\omega_0 = 1/\sqrt{L_0(C_0 + C)}$ changes due to a change in the varactor capacitance $C(U_{\text{bias}})$, which in turn depends on the control voltage. A large capacitor C_b divides bias and microwave circuits.

A generator circuit varactor frequency tuning is shown in Fig. 12.33. The generator uses two coupled cavities made from waveguide sections. One of the cavities includes a generation diode, and the other—a VD. Resonators are coupled by a waveguide window O_1. The balance of phases and amplitudes of the entire

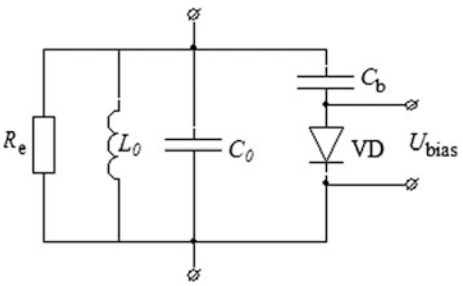


Fig. 12.32 Inclusion of a varactor diode into the oscillatory circuit

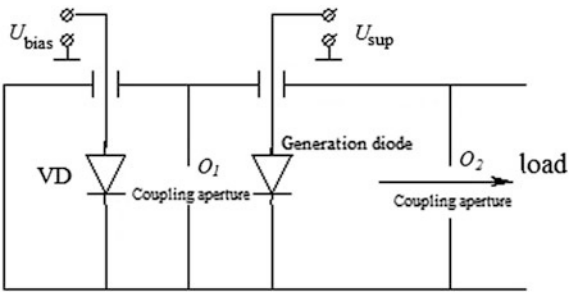


Fig. 12.33 Generator circuit with varactor frequency tuning

device determines the generator frequency. When the control voltage on the VD changes, the generation frequency also changes.

The use of a VD as a switch is based on the significant difference in the reactive conductivity of the diode under negative and small positive bias voltages. Such switchers differ from the switchers on *p-i-n* diodes with small active losses, and hence small noise insertions into the switched signal.

In addition, they have higher speeds. The drawbacks of these switches include difficulty matching the varactor to the transmission line over a wide frequency range, and the small values of the switched power.

Frequency Multiplication

Unlike varistors (DD, MD), in which the nonlinearity of ***the active nonlinear resistance*** is used to convert the signal spectrum, in varactors a ***nonlinear reactance*** is used. The essential difference between these nonlinear devices lies in the level of dissipation (absorption) of power of the acting signal. Varactors have power losses significantly lower than varistors, since varactors operate at a reverse bias voltage with small active currents.

Consider the simplest frequency multiplier circuit (a harmonic generator) on a VD. Consider an element that provides the third harmonic (frequency tripler). The equivalent circuit of this device is shown in Fig. 12.34. Signal with frequency ω_1—$U_{\omega 1} = U_{m1}Sin(\omega_1 t)$ is fed through a VD, through band-pass filter BPF1. In addition, a bias voltage is applied to the diode (not shown in Fig. 12.34). The filter BPF3 is connected in parallel to the diode. This filter is tuned on the third harmonic $3\omega_1$of the input signal and selects the third harmonic on the load Z_3.

However, the operation of the multiplier cannot be performed without connecting an "idle" circuit tuned to the second harmonic (elements Z_2 and BPF2). Then necessity of inclusion of an "idle" circuit follows from the law of energy conservation. This effect is considered in more detail in Sects. 12.4.5 and 12.4.6.

Let us analyze the operation of the circuit with the algorithm used to study varistor properties. We expand the voltage–capacitance characteristic in a power series in the vicinity of constant bias U_0 and find the response of the diode to the action of a microwave signal $U_{\omega 1} = U_{m1}Sin(\omega_1 t)$. We limit ourselves to a low-signal approximation, i.e., $U_{m1} \ll |U_0|$.

Using (12.10) and expanding it in a power series to a second-order term, we obtain:

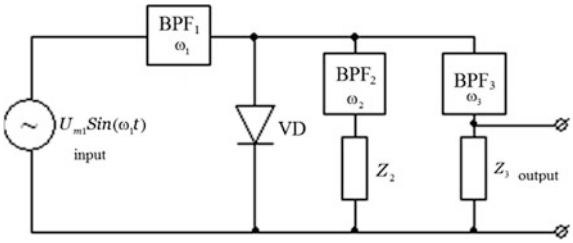


Fig. 12.34 Equivalent circuit of the frequency multiplier

$$C(U) = C(U_0)\left(1 + \alpha_1 U_{\omega 1} + \frac{1}{2}\alpha_2 U_{\omega 1}^2\right)$$

where

$$C(U_0) = C_0\left(1 - \frac{U_0}{\varphi_k}\right)^{-\frac{1}{\gamma+2}},\ \alpha_1 = \frac{\frac{1}{\gamma+2}}{\varphi_k - U_0},\ \alpha_2 = \frac{\frac{1}{\gamma+2}\left(\frac{1}{\gamma+2}+1\right)}{(\varphi_k - U_0)^2}$$

Let us determine the spectrum of the current flowing through the diode for an abrupt junction $m = 0$:

$$i(t) \approx \omega_1 C(U_0) U_{m1}\left(\left(1 + 1/8\alpha_2 U_{m1}^2\right)\sin(\omega_1 t) + \alpha_1/2U_{m1}^2 \sin(2\omega_1 t) + \alpha_2/8U_{m1}^2 \sin(3\omega_1 t)\right) \tag{12.12}$$

The presence of harmonics in the current spectrum leads to a rising voltage on the elements of the circuit both on the second and third harmonics. Moreover, the presence of a branch tuned to the second harmonic ensures the occurrence of an appropriate voltage on the diode, which is added to the voltage of the first harmonic and generates an additional signal at a combinational frequency (particularly, on the third harmonic). The conversion coefficient for such a device can be quite large, since all the power supplied by the first harmonic (assuming ideal capacitance of the diode and the filters, as well as the absence of losses in the idle circuit) is converted to higher harmonic powers:

$$P_1(\omega) = P_2 + P_3,$$

where P_i is the power delivered on i-th harmonic.

Naturally, in real circuits there are always losses, which reduce the conversion factor$\eta = P_3(3\omega)/P_1(\omega)$. Therefore, even with an optimally tuned idle circuit, in practice, it does not exceed 60–80%.

Since capacitance $C(U_0)$ depends on the voltage, the conversion coefficient depend on voltage also. This is manifested in the dependence of the conversion factor on both the bias voltage (with constant input power) and the input power (at constant bias). Both dependencies have a maximum, observed when the bias voltage and the amplitude of the microwave signal are equal.

At present, it is possible to obtain power up to 100 μW at a frequency of 1 THz based on varactor multipliers.

Currently, circuits with oppositely connected varactors are used to construct multipliers in the millimeter and sub-millimeter bands, which do not require the supply of a constant bias. The voltage–capacitance characteristic of such a compound is symmetrical with respect to zero bias voltage. In the output spectrum of a such a multiplier there are only odd harmonics of the input signal: $3f_s$, $5f_s$, $7f_s$. These features make it possible to eliminate the idle circuit and the bias circuits, something which greatly simplifies the construction of multipliers. For this purpose,

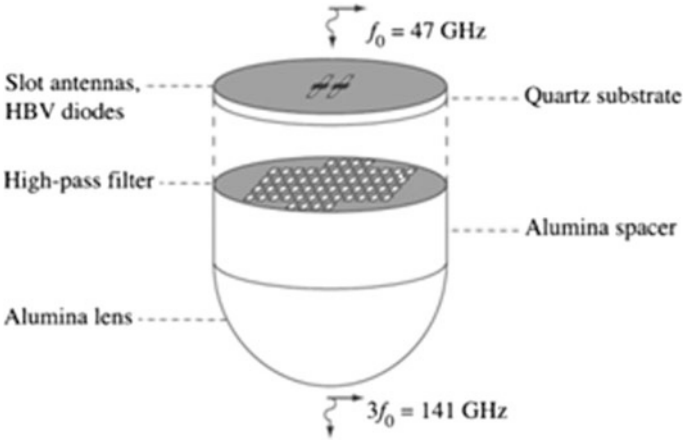


Fig. 12.35 A 141 GHz integrated quasi-optical slot antenna tripler

special hetero-barrier varactors (referred to as HBVs) are created (see Sect. 12.3). The main application for HBV diodes is generating extremely high-frequency signals from lower frequency inputs.

Only odd harmonics are generated, since even harmonics are cancelled due to the symmetric nature of the nonlinearity. Also, by using this inherent symmetry of the device, it can operate without DC biasing. This represents an advantage over the Schottky diode which has to be biased.

Figure 12.35 shows an example of quasi-optical tripler which has a simple mechanical construction— a quartz substrate with printed slotted antenna with mounted arrays of HBVs. A high-pass filter with a cut-off frequency of about$3f_1$ prevents passing the input signal into the antenna. The signal with the $3f_1$is radiated by an alumina lens antenna.

This type of frequency multiplication is demonstrated by triplers at 100 GHz through 282 GHz, and up to 450 GHz, and also as quintuplets (5× multiplication) at 175 GHz. Signals generated at these frequencies (100 GHz to 3 THz) have applications in diverse areas such as radioastronomy, security imaging, biological and medical imaging and high-speed wireless communications.

12.4.5 Manley-Rowe Relations

Let us analyze the operation of the varactor from the general energy positions. To do this, consider the power balance in a circuit with an ideal nonlinear reactivity subjected by two harmonic voltages with frequencies f_s and f_p. In addition, we design an electrodynamics system such that it has an active resistance component at some combinational frequency f_c, for example, $f_c = f_s + f_p$. The functional diagram of such a device is shown in Fig. 12.36

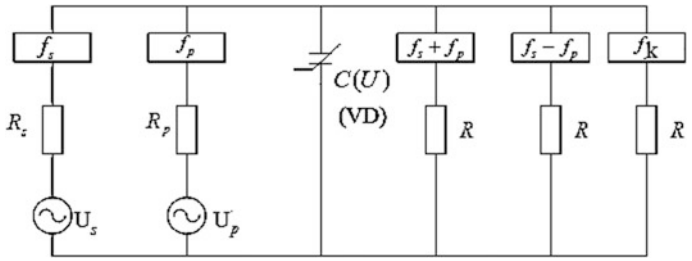


Fig. 12.36 For the analysis of nonlinear capacitance operation

We write the power balance for such a circuit as:

$$P_s + P_p + P_c = 0, \tag{12.13}$$

where $P_{(s,p,c)}$ is the average power for the period:

$$P = \frac{1}{T}\int_0^T u \cdot i\, dt = f \cdot W,$$

where W is the energy accumulate over the period. Applying these expressions from (12.13), we obtain

$$f_s W_s + f_p W_p + f_c W_c = 0 \tag{12.14}$$

We rewrite (12.14) representing the combinational frequency as $f_c = mf_s + nf_p$, where m and n are the whole numbers:

$$f_c W_s + f_p W_p + \left(mf_s + nf_p\right) W_c = 0;$$

or

$$f_s(W_s + mW_c) + f_p(W_p + nW_c) = 0. \tag{12.15}$$

Taking into account the arbitrariness of the frequency choice in (12.15), the equation is satisfied under the condition that each of the terms is equal to zero, i.e.:

$$\left\{ \begin{array}{l} W_s + mW_c = 0 \\ W_p + nW_c = 0 \end{array} \right\}. \tag{12.16}$$

Equation (12.16), when considering (12.14), is rewritten:

$$\begin{aligned} \frac{P_s}{f_s} + \frac{mP_c}{mf_s + nf_p} = 0; \\ \frac{P_p}{f_p} + \frac{nP_c}{mf_s + nf_p} = 0. \end{aligned} \tag{12.17}$$

Given the arbitrary of choice m and n, we rewrite (12.17) as:

$$\begin{aligned} \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{mf_s + nf_p} = 0; \\ \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \frac{nP_{m,n}}{mf_s + nf_p} = 0. \end{aligned} \tag{12.18}$$

Equation (12.18) represents the Manley–Rowe relations for circuits with nonlinear ideal reactivity, relating reflect the law of conservation of energy in circuits. These two mathematical relationships have the following important properties:

- They are independent of the particular shape of the voltage–capacitance or current–inductance curves for a nonlinear capacitance or nonlinear inductance, respectively.
- The power levels of the various sources are irrelevant.
- The external circuitry connected to the nonlinear reactance does not affect how the power is distributed to the harmonic frequencies.

Let us apply (12.18) to analyze the energy relations in the multiplier and the parametric amplifier. In the multiplier there is one signal with frequency f_s, and combinational frequency $f_c = mf_s$. Setting $n = 0$, we obtain:

$$\frac{P_s}{f_s} + \frac{mP_c}{mf_s} = 0$$

from which $P_c = -P_s$, i.e., the efficiency of transformation equals 100%.

12.4.6 Parametric Amplifier

Let us consider a more general case, when the diode is fed by two harmonic signals instead of one. Such a scheme is called a *parametric amplifier* (PA). We can see an analogy arises here with the operation of a mixer using the nonlinearity of the volt–ampere characteristics arises.

L.I. Mandelstam and N.D. Papaleksi discovered the principle of parametric amplification. Signal amplification in PA is carried out due to the energy of an external source (a so-called **pump oscillator**), which periodically changes the capacitance (or inductance) of the nonlinear reactive element that is included in its composition. Thus, a PA converts the power of a pump oscillator to the power of the output signal, i.e. it realizes an RF–RF energy conversion instead of a DC to RF energy conversion which happens in other devices.

The main advantage of such amplifiers is their low noise due to their small dissipation energy in the parasitic resistance R_s of the varactor. Since the 1950s,

PAs, along with quantum amplifiers, have become one of the main components of radio telescopes. The noise temperature of which, cooled by parametric amplifiers is 10–20 K, with gains of the gain is 20–80 dB.

These PAs were traditionally grouped into two types: *phase-incoherent* ***upconverting*** and *negative-resistance parametric amplifier* (in the case of a ***downconverting*** frequency***).*** Nowadays they are categorized as: *phase-coherent degenerate negative-resistance PAs* (when $f_s = f_p - f_s$ *or* $f_p = 2f_s$) and *nondegenerate negative-resistance PAs* (when $f_s \neq f_p - f_s$).

In phase-incoherent upconverting PAs a signal with a fixed frequency f_p from a pump generator mixes with an RF small-signal source at a frequency f_s, to produce an upconverted output $f_p + f_s$ with a gain that can be predicted by the Manley–Rowe relations.

In this variant $m = 1$and $n = 1$, so (12.18) can be rewritten as:

$$\left\{ \begin{array}{l} \frac{P_s}{f_s} + \frac{P_c}{f_s + f_p} = 0 \\ \frac{P_p}{f_p} + \frac{P_c}{f_s + f_p} = 0 \end{array} \right\},$$

where $P_c = -P_s - P_p$,

with the gain coefficient being:

$$K = \frac{P_c}{(-P_s)} = \frac{f_c}{f_s} = \frac{f_s + f_p}{f_s} > 1.$$

Using such kinds of amplifiers is limited by the impossibility of achieving large gain factors: $(f_p + f_s)/f_s$. It is difficult to design and manufacture wide-band circuits with only three resonances on frequencies f_s, f_p and $(f_p + f_s)$. Parametric amplifiers of this type are called *stable boost converters*.

If we assume $m = 1$and $n = -1$ then (12.17) can be rewritten as:

$$\frac{P_s}{f_s} + \frac{P_c}{f_s - f_p} = 0\,, \quad \frac{P_p}{f_p} + \frac{-P_c}{f_s - f_p} = 0$$

or

$$\frac{P_s}{f_s} = \frac{P_c}{f_p - f_s}\,, \quad \frac{P_p}{f_p} = -\frac{P_c}{f_p - f_s}\,,$$

As can be seen from the first equality, powers P_s and $P_c$are positive (>0), since the power consumed by the load $P_c > 0$. This means that a portion of the output of the pump generator is supplied to the signal circuit and compensates the loss of power, i.e., in the amplifier exists the regeneration frequency of the signal. From these equations it is impossible to find the gain, because P_s includes not only the power consumed by the load, but part of the power generated by regeneration. However, writing the first equation in the form $P_c = P_s(f_c/f_s)$, it can be argued that the increase will be larger when the ratio f_c/f_s becomes larger. Amplifiers of this type are unstable in operation, since the signal circuit is supplied power even in the absence of a signal, which under certain conditions can lead to self-excitation.

When $f_s = f_p - f_s$ or $f_p = 2f_s$ and $f_c = f_p - f_s = f_s$ we have *phase-coherent degenerate negative-resistance PAs.* If the condition is satisfied strictly, the circuit is provided with one reinforced oscillation, the amplitude is equal to the sum of the oscillations at the signal frequency and pump frequency. This mode of operation is called synchronous operation. As we have shown, this operation depends on the phase relations of the oscillations of the pump and signal.

In real conditions it is impossible to perform such a synchronization condition. Therefore, the single-circuit regenerative amplifier always operates in an asynchronous mode, when: $f_c - f_s = \delta f \neq 0$. The value of phase misbalance $\delta\omega t$ becomes a function of time, since it gains a random additive. Insertion resistance also becomes a random function of time, and as a result there are random changes in amplification. This represents a serious drawback to these kinds of amplifiers.

Parametric amplifiers are used in frequency ranges from hundreds of megahertz to tens of gigahertz. They have a relatively narrow bandwidth of 1–3% and a low noise factor (NF) due to the low level of dissipation energy in the parasitic series resistors R_s in a varactor. To minimize NF the first stage of an amplifier is usually put in a Dewar vessel containing liquid nitrogen, N_2.

Consider, as an example, the typical functional scheme of a two-stage PA in Fig. 12.37. The input of the PA signal is fed through circulator C1, then passed through the resonator and the low-pass filter that filters the signal, with the pumping frequency of the signal being transmitted to the resonator containing VD1. Operating

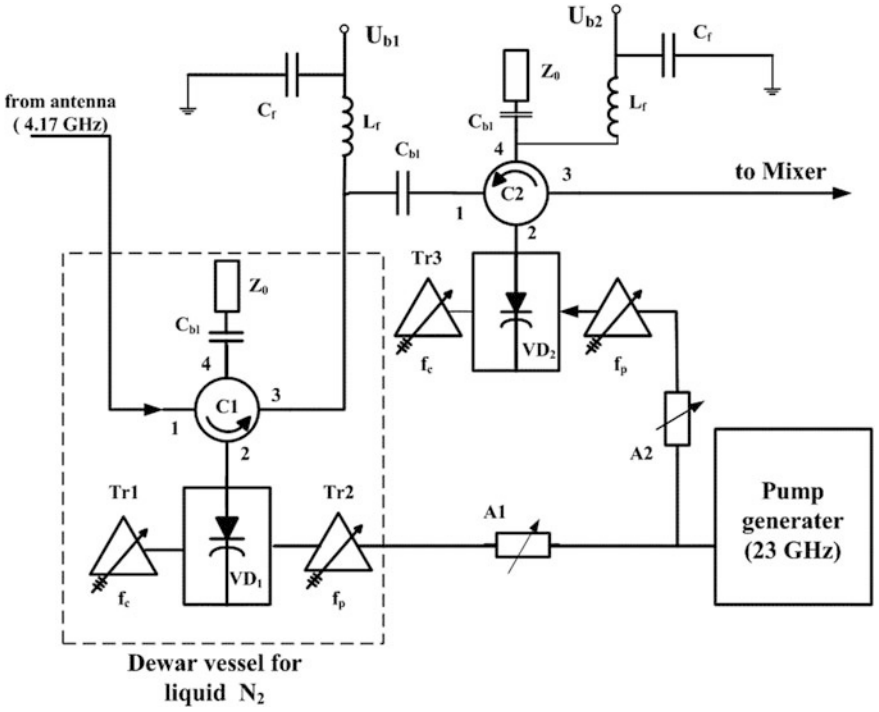


Fig. 12.37 Parametric amplifier

point VD1 is set using a DC voltage source (Ub1). On the other hand in a resonant system the voltage supplied from the generator pump is passed through a regulating attenuator A1. Transformers of impedance Tr1 and Tr2 are connected to allow adjustments of the PA in terms of its frequency band. All input components are placed in a Dewar vessel. The second stage is organized in the same manner, but without cryogenic cooling: the noise factor NF of the system determined mainly by the first stage.

Another simple example of a double-balanced phase-coherent degenerate PA is illustrated in Fig. 12.38 (printed board) and its electrical scheme given in Fig. 12.39.

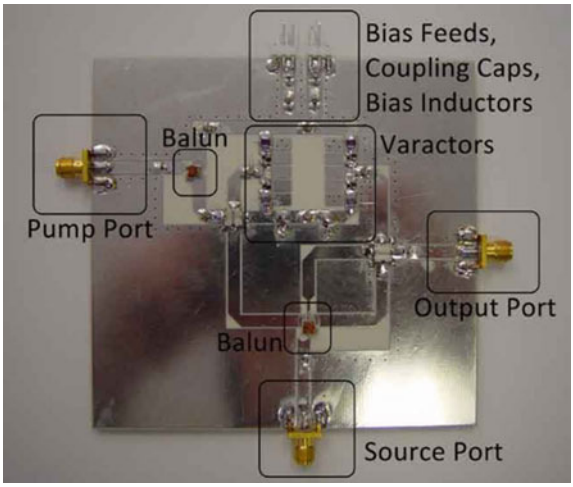


Fig. 12.38 Printed board of a double-balanced phase-coherent degenerate parametric amplifier

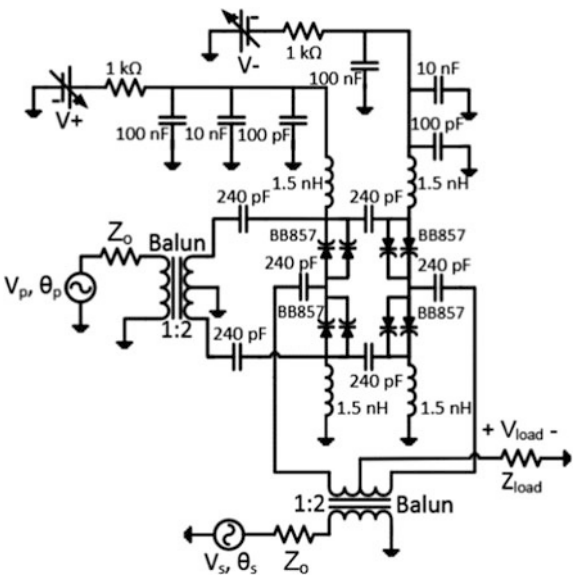


Fig. 12.39 The electrical scheme of the PA shown in Fig. 12.38

The component "Balun" in there figures means balanced–unbalanced transformer, which converts the electrical signal from a balanced one to an unbalanced one and vice versa. The working (source) frequency of such an amplifier is 1 GHz.

Advancement Questions

1. List the main types of semiconductor devices with positive dynamic resistance.
2. Draw a band diagram of the Schottky barrier for different bias voltages. Which charge carriers participate in current transport through the barrier?
3. What charges are in the depletion layer? Which current components pass through the depletion layer when the microwave voltage is applied to the barrier?
4. Draw the VAC of the Schottky barrier with logarithmic and linear current scales.
5. What dependence does the voltage-capacitance characteristic depict? Can it be used to define the doping profile?
6. With what accuracy?
7. What is the difference between the Schottky barrier and the barrier formed through semiconductor contact of *n*-type and p-type?
8. What is the difference between barrier and diffusion capacity?
9. What are the parameters of the detector diode VAC? What is its difference from the VAC of an ideal Schottky barrier?
10. What dynamic parameters characterize DD? What is the cutoff frequency?
11. How does the DD turn on in the microwave transmission line? What is the detector section?
12. What is the difference between a mixer diode and a detector diode?
13. What is the reason for the spectrum change in a mixer diode?
14. What is a heterodyne and why is it needed in a mixer?
15. Draw the spectrum of the mixer with second order intermodulation. What combinational frequencies can occur in this case?
16. Draw the simplest mixer circuit.
17. What is the structure of a *p-i-n* diode? Compare the speed of DD and a *p-i-n* diode.
18. How do the lifetime of the charge carriers and the transient time of the *i*-region affect the inertial properties of the device? Can a *p-i-n* diode be used as a rectifier?
19. What parameters are used to describe the RF properties of switches?
20. How do the limiter and attenuator work in a *p-i-n* diode?
21. Draw a phase shifter circuit with a discrete phase change level of 90°.
22. How does the doping profile affect the shape of the VCC of the varactor diode?

23. Why are varactor diodes not used to convert signal frequency in the input circuits of receivers?
24. Explain the peculiarities of a VD operating in frequency-tuning circuits.
25. Analyze how a frequency multiplier operates using Manley-Rowe relations.
26. Analyze the operation of the parametric amplifier using Manley-Rowe relations.

Chapter 13
Diodes with Negative Dynamic Resistance

13.1 General Characteristics of Diodes with Negative Dynamic Resistance

Microwave radio electronic systems use as devices that convert the signal spectrum and switch direction of microwave signals propagation (see Chap. 12), as well as devices that provide amplification and generation of microwave oscillations (*active devices*). In low-frequency technology, the conversion and switching of signals are provided by diodes, and amplification and generation are provided by transistors. However, in the 1960s, there were no transistors capable of operating in the microwave band due to technological limitations. Practice, however, required the creation of active devices meeting the requirements of microminiaturization, high cutoff frequencies, and small supply voltages, etc. The cumulated experience of creating vacuum devices with dynamic control was not useful, since bunching methods used in vacuum devices cannot be used in solids. During the period of microwave oscillations, charge carriers undergo a large number of scattering events, «forgetting» about the initial pulse. The efforts of researchers and developers were aimed at finding new physical mechanisms for controlling convection current and for energy extraction in the microwave band. A number of such physical effects were discovered to the middle of the 1960s. The tunnel effect, avalanche breakdown and inter valley carrier transport were the most important of them. These physical effects formed the basis for the creation of diode structures able to amplify and generate microwave oscillations: tunnel diodes (TD), avalanche transit-time diodes (IMPATT DIODE), Gunn diodes, and injection transit time diodes. In most cases, such devices have a positive differential resistance (with the exception of TD), determined by the static volt-ampere characteristic and a negative dynamic resistance in the microwave band. Therefore, this class of devices was called *devices with negative dynamic resistance* (NDR).

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_13

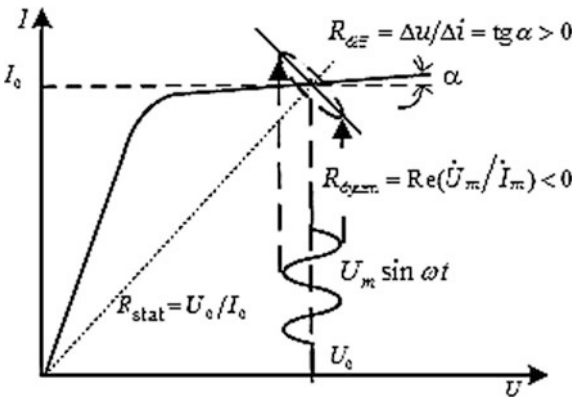


Fig. 13.1 Static and dynamic characteristics of the diode

Figure 13.1 shows a typical static volt-ampere characteristic of a diode and a graphical definition of static $R_{\text{stat}} = U_0/I_0$ and differential resistance $R_{\text{diff}} = tg\alpha = u/i$ at the selected operating point. The differential resistance together with the "cold" capacitance adequately describe the behavior of the device in the steady mode.

When an alternating signal $u = U_m \sin \omega t$ is applied to a diode, it is necessary to take into account the ratio of the acting voltage period and the time parameters characterizing the motion of the carriers in the sample. These include: the charge carrier lifetime, the carrier transit time through the diode structure, and the Maxwellian relaxation time, etc. (see Appendix A). If the signal period is comparable with the carrier transit time in the device, the phase of the induced current shifts substantially relative to the phase of the applied voltage (see Sect. 4.2). Under certain conditions, this phase can be greater than $\pi/2$, being equivalent to the appearance of a negative real resistance component in the total complex resistance of the diode. This component is called the *negative dynamic resistance*. In Fig. 13.1, this is indicated by a straight line with a negative slope. Such a linear representation is valid only for a small-signal approximation, i.e. for $U_m \ll U_0$. If the amplitude of the alternating voltage increases, then the waveform of the current becomes substantially more complicated.

Diodes with negative resistance can be used in amplifiers and generators. These devices unite flow control function and the function of energy extraction in the same spatial gap. Because such a device has only one port, it is impossible to separate the input and output circuits. As a result, in the microwave band, such devices can be used as reflective amplifiers or generators.

It should be noted that, despite significant progress in improving the parameters of such diodes, the development of microwave transistors has substantially limited the field of application of diodes with NDR.

The detailed description of the operation features of these devices in this book is based on an analysis of the general principles used to obtain negative resistance in semiconductor diode structures with further application of theoretical results to

specific devices. First, we shall analyze the general principles to obtain negative dynamic resistance in *small-signal approximation*.

Furthermore, when considering specific devices, the above principles and design formulas are used. Nonlinear analysis, as an example, is considered only for an avalanche transit diode, most often used in powerful generators.

13.2 Analysis of Semiconductor Sample Dynamic Resistance

To analyze the energy aspects of the flow interaction of charged particles with a field in semiconductors, we use the results of one-dimensional small-signal wave carrier analysis in these materials (see Sect. 3.3). Consider a sample (Fig. 13.2) of length L in which the left edge is called the cathode, and the right edge is called the anode. We assume that voltage

$$U = U_0 + U_m \sin \omega t$$

is applied between the cathode and the anode.

Let's assume that a solution was found for the static components of field E_0, charge density qn_0 and drift velocity v_0 in the structure under consideration. We also assume that their values do not depend on the coordinate and do not take into account diffusion. According to the calculations of Sect. 3.3 for $D = 0$, one wave of charge inhomogeneity propagates in the semiconductor with a velocity equal to the drift velocity v_0. The attenuation constant of this wave depends on the sign of the differential frequency of Maxwellian relaxation ω_m.

$$\widetilde{n} = \dot{n}_m \mathrm{e}^{-j\Gamma} = \dot{n}_m(0)\mathrm{e}^{\alpha z}\mathrm{e}^{-j\beta z}, \tag{13.1}$$

where $\Gamma = \beta - j\alpha = \omega/v_0 - j\omega_m/v_0, \alpha\omega_m/v_0, \beta = \omega/v_0$.

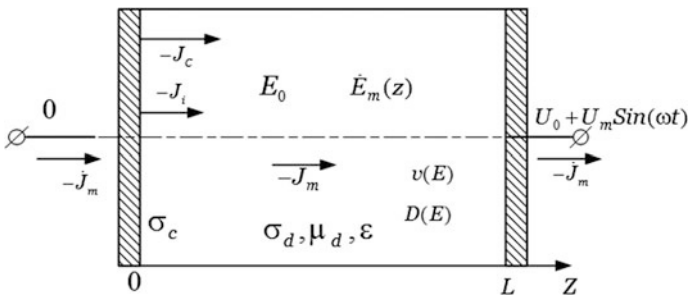


Fig. 13.2 Longitudinal cross section of a semiconductor sample

We determine equivalent resistance Z and specific equivalent resistance $Z_s = \dot{U}_m/\dot{J}_m$ for the sample, linking variable density of the current flow ω_m and sample voltage $U = \int_0^L \dot{E}_m(z)dz$. We shall use index s for specific values.

$$Z = \dot{U}_m/\left(\dot{J}_m S\right) = Z_s/S = R_s/S + jX_s/S, \tag{13.2}$$

where S is the sample area.

Let us define the couple between total current density $\dot{J}_m$ and alternating electric field in the structure. We find the field by substituting (13.1) in the right-hand side of the Poisson equation for the variable charge component:

$$\frac{\partial \widetilde{E}}{\partial z} = \frac{e}{\varepsilon}\dot{n}_m \mathrm{e}^{-j\Gamma}.$$

On the left boundary of the sample alternating electric field, $\dot{E}_m(0)$ may exist. Then

$$\dot{E}_m(z) = -\frac{e\dot{n}_m(0)v_0}{\varepsilon(\omega_m + i\omega)}\left(\mathrm{e}^{-j\Gamma z} - 1\right) + \dot{E}_m(0), \tag{13.3}$$

where $\dot{n}_m(0)$ is the complex charge amplitude on the cathode.

Let us write the expression for the variable component of the total current density, using (3.3):

$$\dot{J}_m = e\dot{n}_m v_0 + en_0\mu_d\dot{E}_m(z) + j\omega\varepsilon\dot{E}_m(z), \tag{13.4}$$

where $\mu_d\dot{E}_m(z)$ is the charge variable velocity component.

We transform (13.3) by expressing the components $\dot{n}_m(0)$ and $\dot{E}_m(0)$ in terms of the density of the total current flowing through the sample. We denote the conductivity of the cathode contact σ_c. Current density j_m and field $\dot{E}_m(0)$ at the contact are related by Ohm's law

$$\dot{J}_m = (\sigma_c + j\omega\varepsilon)\dot{E}_m(0). \tag{13.5}$$

According to (13.4), the current density at $z = 0$ is written as follows:

$$\dot{J}_m = e\dot{n}_m(0)v_0 + en_0\mu_d\dot{E}_m(0) + j\omega\varepsilon\dot{E}_m(0). \tag{13.6}$$

The current injected from the cathode with density j_i is defined through alternating charge density $\dot{n}_m(0)$ or contact conductivity σ_c, i.e.

$$\dot{J}_i = e\dot{n}_m(0)v_0 = \sigma_c\dot{E}_m(0).$$

Using (13.5) and (13.6), we form (13.3) as follows

$$\dot{E}_m(z) = \frac{J_m \exp(-j\Gamma_z)}{\sigma_c + j\omega\varepsilon} + \frac{J_m}{\varepsilon(\omega_m + j\omega)} - \frac{J_m \exp(-j\Gamma_z)}{\varepsilon(\omega_m + j\omega)}. \tag{13.7}$$

Using (13.7) and the expression for voltage $\dot{U}_m = \int_0^L \dot{E}_m(z)\mathrm{d}z$, we obtain

$$Z_s = \frac{(-j\Gamma)^{-1}[1 - \exp(-j\Gamma L)]}{\sigma_c + j\omega\varepsilon} + \frac{L}{\varepsilon(\omega_m + j\omega)} + \frac{(-j\Gamma)^{-1}[1 - \exp(-j\Gamma L)]}{\varepsilon(\omega_m + j\omega)}. \tag{13.8}$$

As can be seen, the resistance has reactive and active components, depending on the properties of contact σ_c, and properties of semiconductor bulk ω_m, as well as the length of the sample L (transit angle $\Theta = \omega z / v_0$).

Before using (13.8) to calculate the resistance of real semiconductor structures, we will analyze the energy balance of the interaction of the field with charged particles using the relations.

According to (4.5), the real part of the specific power in the interaction P of charge carriers with a time-varying field is generally given by

$$P = \frac{1}{2}\mathrm{Re}\left(\int_0^L \dot{J}_m \overset{*}{E}_m \,\mathrm{d}z\right), \tag{13.9}$$

where P is the real part of the power of interaction; and df is the conjugate value of the complex amplitude of the alternating electric field. Let us set current phase $\dot{J}_m$ to zero, i.e. $\dot{J}_m = J_m$.

We rewrite (13.9) considering the assumptions:

$$P = \frac{1}{2}J_m \int_0^L \mathrm{Re}\left(\overset{*}{E}_m(z)\right)\mathrm{d}z = \frac{1}{2}J_m \int_0^L \mathrm{Re}\left(\dot{E}_m(z)\right)\mathrm{d}z. \tag{13.10}$$

Formula (13.10) shows that the sign of the specific interaction power P is determined by the integral of the real part of field amplitude $\dot{E}_m(z)$ along coordinate z. Possible dependences of $\mathrm{Re}(\dot{E}_m)$ on the transit angle obtained using (13.3) are shown in Fig. 13.3. From a physical point of view, the change in the sign of $\mathrm{Re}(\dot{E}_m(z))$ reflects the fact that charges moving through the sample fall into the accelerating or decelerating phase of the field. The acceleration and deceleration phases alternate with period $T = 2\pi/\omega$. During this time, the charges travel distance $l = v_0 T$. The initial phase shift between the current density and the field is determined by the boundary condition at the left edge of the sample. Integral $\int_0^L \mathrm{Re}(\dot{E}_m(z))\mathrm{d}z$ is proportional to the area bounded by the abscissa and the curve

$\mathrm{Re}(\dot{E}_m(z))$. The region above the abscissa axis characterizes the positive power of the interaction, when the energy from the field is transferred to the electron, and the area under the axis corresponds to the deceleration of the charge carrier, i.e., the transfer of the kinetic energy of the charge to the alternating field. Figure 13.3 shows four distributions of $\mathrm{Re}(\dot{E}_m(z))$ for various differential frequencies of Maxwellian relaxation and various boundary conditions at the cathode. Curve 1 corresponds to $\omega_m = 0$, when the charge inhomogeneity is constant, field $\dot{E}_m(0) = 0$. Dependency 2—$\omega_m > 0$, charge inhomogeneity decreases, $\dot{E}_m(0) = 0$. Distribution 3—$\omega_m < 0$, charge inhomogeneity increases, $\dot{E}_m(0) = 0$. For option 4—$\omega_m = 0$, charge inhomogeneity is constant, $\mathrm{Re}(\dot{E}_m(0)) \neq 0$. Curves 1, 2 and 3 correspond to the cathode contact with infinite conductivity $\sigma_c \to \infty$, following expression (13.5). Such a contact is called ohmic. In the virtual cathode region of vacuum devices, the same condition is fulfilled: $E_m(0) = 0$. Dependence 4 in Fig. 13.3 corresponds to the presence of an alternating field at the cathode $\mathrm{Re}(\dot{E}_m(0)) \neq 0$. Such a contact is non-ohmic, or barrier.

A qualitative interaction analysis of the field and charges in semiconductors makes it possible to draw a number of important conclusions:

- for samples with ohmic cathode contact $E_m(0) = 0$ and a positive or equal to zero value $\omega_m \geq 0$ it is impossible to obtain negative interaction power at any transit angle;
- for samples with ohmic cathode contact $E_m(0) = 0$ and negative relaxation frequency $\omega_m \leq 0$ it is possible to obtain power $P < 0$ with maximum (modulus) at transit angle $\Theta = \omega L / v_0 = 2\pi$ or at a multiple of this value;
- for samples with non-ohmic (barrier) cathode contact $\mathrm{Re}(\dot{E}_m(0)) \neq 0$ and differential frequency of relaxation equal to zero $\omega_m = 0$, it is possible to obtain $P < 0$. An optimum is observed for transit angles $\Theta = 2\pi - \Theta_i$. Value Θ_i determines the phase difference between the current and the field on the cathode.

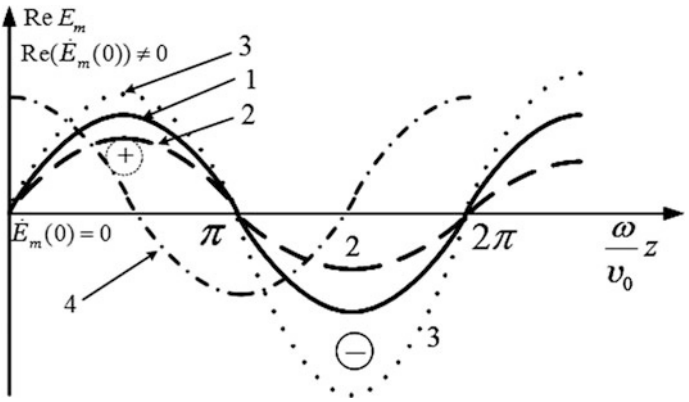


Fig. 13.3 Distribution of the real electric field component E from the transit angle for different values: 1—$\omega_m = 0$; 2—$\omega_m > 0$; 3—$\omega_m < 0$; 4—$\omega_m = 0$, $\mathrm{Re}(E_m(0) \neq 0)$

Analysis (13.8) allows us to draw some general conclusions about possible ways of obtaining negative dynamic resistance. The first term in formula (13.8) characterizes the effect of conditions on cathode contact $\sigma_c + i\omega\varepsilon$ combined with the effect of the transit angle and the properties of the semiconductor represented by factor $(1 - \exp(-i\Gamma L))$. For ohmic contact, this term is zero. The second term reflects the influence of semiconductor bulk properties: length L, dielectric constant ε and differential frequency of the Maxwellian relaxation ω_m. The third term reflects the effect of the transit angle and the properties of semiconductor bulk: ε, ω_m. Particularly, for $(\sigma_c + i\omega\varepsilon) = \varepsilon(\omega_m + i\omega)$, that is, in the case of matching (equal) contact conductivity and sample conductivity

$$R_s = \mathrm{Re}\left(\frac{L}{\varepsilon(\omega_m + i\omega)}\right) = \frac{L}{\varepsilon\left(\omega_m^2 + \omega^2\right)}\omega_m. \tag{13.11}$$

Formula (13.11) is valid for some region L allocated in the infinite sample. Condition $(\sigma_c + i\omega\varepsilon) = \varepsilon(\omega_m + i\omega)$ is naturally met there. Such an idealized sample with $\omega_m < 0$ ensures $R < 0$ for any transit angle, that is, at any frequency and for any sample length L.

If the conductivity is infinite: $\sigma_c \to \infty$ (ohmic contact), the value of R_s is expressed by the formula

$$R_s = \mathrm{Re}\left(\frac{L}{\varepsilon(\omega_m + i\omega)} + \frac{(-i\Gamma)^{-1}[1 - \exp(-i\Gamma L)]}{\varepsilon(\omega_m + i\omega)}\right). \tag{13.12}$$

For $\omega_m = 0$

$$R_s = \frac{L}{2\varepsilon v_0}\frac{\sin^2(\Theta/2)}{(\Theta/2)^2}, \tag{13.13}$$

where $\Theta = \omega L/v_0$ is the transit angle.

Formula (13.13) shows that for ohmic cathode contact in a sample with $\omega_m = 0$, the dynamic resistance for any transit angle is positive—$R_s \geq 0$.

Let us formulate this proposition in another way: amplification or generation of microwave oscillations in a sample with ohmic contact and $\omega_m = 0$ is impossible.

If $\omega_m < 0$, then to calculate the active part of the resistance it is necessary to use formula (13.12). An analysis of this formula shows that for some transit angles it is possible to obtain negative resistance.

If the contact is non-ohmic $\mathrm{Re}(\dot{E}_m(0)) \neq 0$, then the possibility of obtaining negative resistance expands. We rewrite (13.8), grouping the first and third terms. For $\omega_m = 0$ we obtain

$$Z = \left(\frac{L}{i\omega\varepsilon} + \frac{J_i}{J_m}(\omega C)^{-1}\frac{1 - \exp(-i\Theta)}{\Theta}\right)S^{-1},$$

where $C = \varepsilon/L$ is the sample capacity per unit of the cross-sectional area. The ratio

$$\frac{J_i}{J_m} = M_i \exp(-i\Theta_i), \tag{13.14}$$

where Θ_i is the phase shift of injected $J_i S$ and full $J_m S$; M_i is the ratio of their amplitudes (*the injection coefficient*). Parameter Θ_i is called the *injection delay angle*. Let us express the ratio of these currents through the corresponding conductivities:

$$\frac{J_i}{J_m} = \frac{\sigma_c}{\sigma_c + i\omega\varepsilon}. \tag{13.15}$$

It follows from (13.15) that for an ohmic contact with $\omega_c \to \infty$, injection coefficient $M_i = 1$, and injection delay angle $\Theta_i = 0$. For a barrier contact, values M_i and Θ_i are calculated according to (13.15). Taking into account (13.14), the expression for Z_s takes the following form

$$Z_s = \frac{1}{i\omega C} + M_i(\omega C)^{-1} \exp(-i\Theta_i) \frac{1 - \exp(-i\Theta)}{\Theta}. \tag{13.16}$$

So we have (13.17)

$$R_s = \frac{L^2}{2\varepsilon v_0} M_i \frac{\sin(\Theta_i + \Theta/2)\sin(\Theta/2)}{(\Theta/2)^2}. \tag{13.17}$$

Analyzing (13.17), we note that in the absence of injection delay ($\Theta_i = 0$), active resistance is *always* greater than or equal to zero for structures with ω_m = 0. For delay angle $\Theta_i \neq 0$ negative resistance $R_s < 0$ is observed in some range of the transit angles. Figure 13.4 shows dependencies $\mathrm{Re}(Z/R_{\mathrm{norm}})$ from the transit angle at four values of injection delay angle:

$$\Theta_i = 0, \quad \Theta_i = \frac{\pi}{2}, \quad \Theta_i = \frac{\pi}{4}, \quad \Theta_i = \pi.$$

$R_{\mathrm{norm}} = L^2/2\varepsilon v_0 S$ is chosen as normalizing resistance. Injection delay can occur due to various physical effects. The most well-known are the effects of avalanche multiplication, and the injection of charge carriers through the Schottky barrier.

The issues under consideration are important for understanding physical phenomena in semiconductor structures, and their differences and similarities in comparison with processes in vacuum devices.

Let us analyze the possibility of obtaining negative resistance by constructing a diagram of currents and voltages for a semiconductor sample. A similar diagram for a triode is presented in Sect. 5.2.

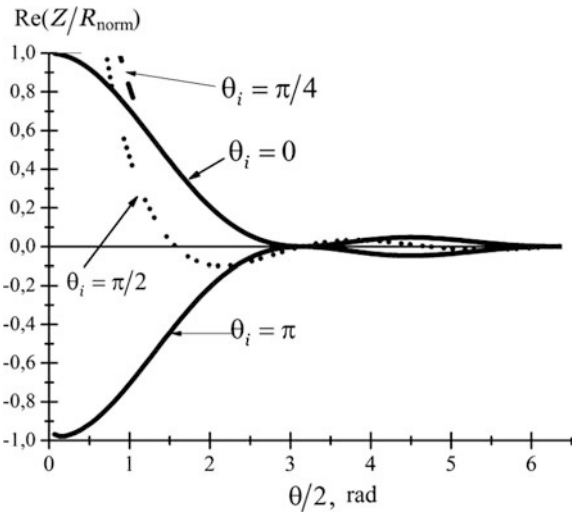


Fig. 13.4 Dependence of active resistance of the sample on transit angle for different injection delay angles

We write down the basic relations connecting the currents and the voltage on the sample, then we construct a diagram. The total current $\dot{I}_m$ has two components: induced current I_i and capacitive current $\dot{I}_c$, i.e.

$$\dot{I}_m = \dot{I}_i + \dot{I}_c. \tag{13.18}$$

Sample voltage $\dot{U}_m$ relates to capacitive current $\dot{I}_c$ as follows

$$I_c = C\frac{\mathrm{d}\dot{U}_m}{\mathrm{d}t} = i\omega C\dot{U}_m, \tag{13.19}$$

where C is the "cold" sample capacity, defined by the formula for a flat capacitor $C = \varepsilon S/L$.

Total current $\dot{I}_m$ in any sample cross-section is constant and consists of convection current $\dot{I}_k(z)$ and capacitive current $\dot{I}_{cq}(z)$, linked to charge transfer, i.e.

$$\dot{I}_m = \dot{I}_k(z) + \dot{I}_{cq}(z) = \dot{I}_{inj} + \dot{I}_{cq}(0), \tag{13.20}$$

where $\dot{I}_{inj}$ is the injection current, and $\dot{I}_{cq}(0) = i\omega\varepsilon\dot{E}_m(0)S$ is the capacitive current on the cathode. Induced current, according to the Shockley-Ramo theorem, is average in the gap convection current $\dot{I}_k(z)$:

$$\dot{I}_{ind} = L^{-1}\int_0^L \dot{I}_k(z)\mathrm{d}z. \tag{13.21}$$

If the velocity of charge carriers in the gap does not change, then in any section z current $I_k(z)$ lags behind injection current $\dot{I}_{inj}$ at angle $\omega z/v_0$, i.e. $\dot{I}_k = \dot{I}_{inj}\exp(-i\omega z/v_0)$. Then:

$$I_{ind} = L^{-1}\int_0^L \dot{I}_{inj}\exp(-i\omega z/v_0)\mathrm{d}z = \dot{I}_{inj}\frac{\sin(\Theta/2)}{\Theta/2}\exp(-i\Theta/2), \qquad (13.22)$$

i.e. the induced current lags behind the injected current by angle $\Theta/2$ and is $\sin(\Theta/2)/(\Theta/2)$ times less. This expression corresponds to the expression for the interaction coefficient M obtained for carrier motion in the vacuum gap.

We use the above expressions to plot the current and voltage diagrams of the sample. First we consider a sample with zero frequency of Maxwellian relaxation $\omega_m = 0$ and an ohmic cathode contact, when $\dot{E}_m(0) = 0$ and $\dot{I}_{inj} = \dot{I}_m$, according to (13.20). This corresponds to the absence of the injection delay in the cathode, i.e. $\Theta_i = 0$. For the sake of certainty, we set the transit angle to $\Theta = 3\pi/2$.

The drawing of diagram in polar coordinates (Fig. 13.5a) begins with the vector of total current $\dot{I}_m$. We assume that I_m has a zero initial phase. It represented by a radius vector along the real axis. According to (13.20), and the absence of injection delay, we draw vector $\dot{I}_i$ parallel to I_m (the dashed line). Convection current $\dot{I}_k$ within different sections of z is represented by the radius vector equal to $|\dot{I}_m|e^{-i\Theta z}$ of the transit angle Θ_z from 0 to $\Theta = 3\pi/2$. Then induced current $\dot{I}_{ind}$, according to (13.22), is represented by a radius vector having a phase of $\Theta = -3\pi/4$. Connecting the ends of vectors $\dot{I}_{ind}$ and $\dot{I}_m$, according to (13.18), we get current $\dot{I}_c$. Voltage $\dot{U}_m$, by expression (13.19), lags current $\dot{I}_c$ by $\pi/2$. As a result, we see that the shift between voltage $\dot{U}_m$ and current $\dot{I}_m$ is less than $\pi/2$, indicating positive real resistance component $\mathrm{Re}(Z) > 0$ and negative reactivity. For angles $\Theta > 2\pi$, the construction is repeated, but with a smaller magnitude of induced current, according to (13.22).

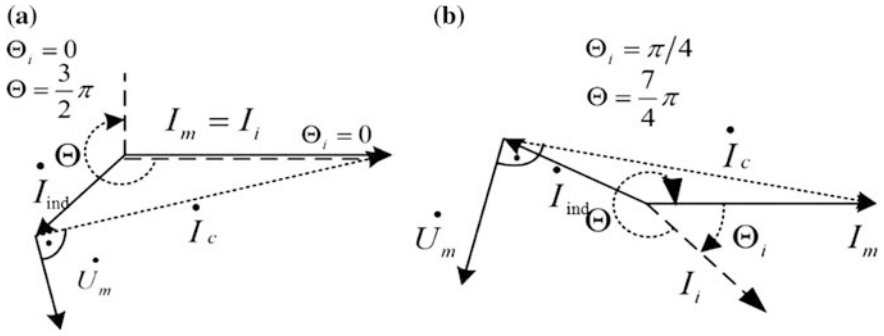


Fig. 13.5 Current and voltage diagram for a sample

Analysis of Fig. 13.5a shows that if the injection delay is provided, averaging the convection current produces an induced current in the second quadrant, resulting in a phase shift between the current $\dot{I}_m$ and voltage $\dot{U}_m$ greater than $\pi/2$. Then the actual part of the sample impedance is less than zero, that is, the sample can be used to amplify or generate oscillations. Figure 13.5b shows the diagram for a sample with non-ohmic contact $\Theta = 7\pi/4$ and transit angle $\Theta_i = 3\pi/4$. It is seen from the figure that voltage $\dot{U}_m$ is shifted relative to current $\dot{I}_m$ by an angle greater than $> \pi/2$.

This graphical analytic method is easily extended to structures with negative Maxwellian relaxation frequency $\omega_m < 0$. The increase in convection current $\dot{I}_k(z)$ contributes more to the resultant induced current from the regions closer to the anode. This leads to displacement of the induced current into the second quadrant and, accordingly, to a shift between current $\dot{I}_m$ and voltage $\dot{U}_m$ by an angle greater than $\pi/2$, even with an ohmic cathode contact.

13.3 Ways to Obtain an Alternating Convection Current in a Diode Structure

In the previous paragraph, the important roles of carrier injection delay, transit effects, and semiconductor properties in obtaining negative resistance was shown. In this paragraph we will analyze semiconductor structures in which these mechanisms can be realized. First of all, let us consider structures in which the charge inhomogeneities do not increase with drift, that is, $\omega_m = 0$. This mode of operation is realized at high static field strengths, i.e. $E_0 > E_{\text{th}}$.

One way to achieve injection delay is to use a barrier contact for which $E_m(0) \neq 0$. This property is possessed by the metal-semiconductor barrier (the Schottky barrier) biased in the reverse direction. Its equivalent circuit is a parallel connection of conductivity G_j and capacitance C_j (Fig. 13.6). Conductivity G_j determines the level of injection of charge carriers into the bulk of a semiconductor.

Figure 13.6 presents an equivalent circuit for the connection of such a barrier with a semiconductor bulk. The relation between the density of the injected and total currents is written as follows:

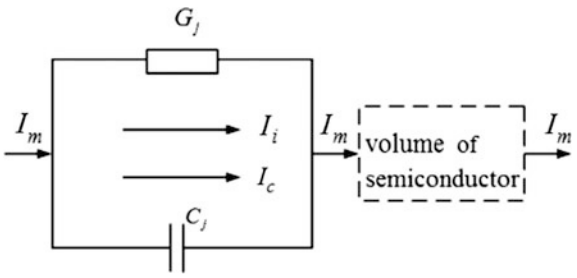


Fig. 13.6 Equivalent circuit for connecting a barrier contact and a semiconductor

$$\frac{J_i}{J_m} = \frac{G_j}{G_j + i\omega C_j} = \frac{G_j}{\sqrt{G_j^2 + \omega^2 C_j^2}} \exp\left(-i\text{arctg}\frac{\omega C_j}{G_j}\right), \tag{13.23}$$

where $G_i = \sigma_c$; and $C_j = \varepsilon/L$ are the specific barrier parameters (for S = 1).

Comparing (13.23) and (13.14), we write the expressions for the injection coefficient M_i and injection phase Θ_i:

$$M_i = G_j/\sqrt{G_j^2 + \omega^2 C_j^2}, \tag{13.24}$$

$$\Theta_i = \text{arctg}\left(\omega C_j/G_j\right). \tag{13.25}$$

It can be seen from (13.25) that the inverse Schottky barrier provides the maximum injection delay angle $\Theta_i = \pi/2$ under condition $\omega C_j \gg G_j$, but in this case $M_i \to 0$, i.e. the amplitude of the injected current is small. The increase in the injection ratio to the maximum is achieved under condition $G_j \gg \omega C_j$, then $M_i \to 1$ and $\Theta_i = 0$, corresponding to the ohmic contact. To obtain a barrier contact with a significant level of injection, we choose a compromise version: $\omega C_j = G_j$. Then $\Theta_i = \pi/4$ and M_i = 0.707.

The injection delay can also be achieved with ohmic contact. For this, the semiconductor sample is made from two dissimilar parts 1 and 2, as shown in Fig. 13.7. The first region on its right-hand boundary provides the necessary injection delay for region 2.

The properties of regions that ensure the injection delay and ultimately the negative real part of the impedance can be find using expression (13.16).

For the first region with ohmic contact, the expression takes the form

$$\text{Re } Z_1 = \frac{M_1}{\omega C_1} \frac{\sin^2(\Theta/2)}{(\Theta/2)^2}, \tag{13.26}$$

where $\Theta_1 = \omega L_1/v_{01}$ is the transit angle of carriers in the first region; M_1 is the injection coefficient; and $C_1 = \varepsilon_1 S_1/L_1$ is the capacity of the first region.

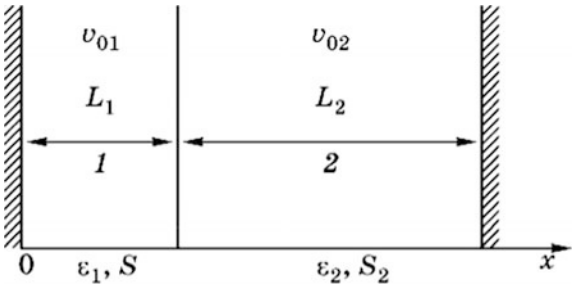


Fig. 13.7 Circuit for ensuring injection delay when contacting two dissimilar samples

For the second region

$$\mathrm{Re}\, Z_2 = \frac{M_2}{\omega C_2} \frac{\sin(\omega L_1/v_{01} + \Theta_2/2)\sin(\Theta_2/2)}{(\Theta_2/2)^2}, \tag{13.27}$$

where $\Theta_2 = \omega L_2/v_{02}$ is the transit angle of carriers in the second region; and $C_2 = \varepsilon_2 S_2/L_2$ is the capacity of the second region.

For the structure under consideration, on the left boundary of region 1 with ohmic contact $M_1 = 1$ and for region 2, under condition $\omega_{m2} = 0$, the injection coefficient is the same, i.e. $M_2 = 1$. For the transit angle of the first region $\Theta_1 = \pi/2$, the total resistance of the sample is written in the form

$$\mathrm{Re}\, Z = \mathrm{Re}(Z_1 + Z_2) = \frac{2}{\pi\omega C_1}\left(1 + \frac{\pi}{2}\frac{C_1}{C_2}\frac{\sin\Theta_2}{\Theta_2}\right). \tag{13.28}$$

For a transit angle in the second region $\pi < \Theta_2 < 2\pi$ and $C_1 \gg C_2$, the real part of the impedance takes negative values, that is, the structure represented in Fig. 13.7 has amplifying properties. Condition $C_1 \gg C_2$ can be met by selecting materials for the first and second regions. In particular, if the dielectric constant of the first region material ε_1 is significantly greater than the dielectric constant of the material of the second region $\varepsilon_1 \gg \varepsilon_2$, the required condition will be satisfied. Another, more hypothetical way, is to select materials with different drift velocities: $v_{01} \ll v_{02}$. In practice, it is difficult to provide such a condition, since the saturation velocities of materials differ by no more than 2.5 times. The third option, which is the most simple to implement, is to use regions of different cross sections, with $S_1 \gg S_2$. Strengthening properties of such heterostructures are manifested with a certain choice of material thickness and operating frequency. Such structures are called layered structures.

Figure 13.8 shows possible options for structures that provide negative dynamic resistance. The structure shown in Fig. 13.8a consists of layers with different values of ε. At the edges of the structure, there are ohmic contacts. The transit angle in each region is π radians. Distribution of the real field component $\mathrm{Re}\dot{E}_m(z)$ is presented below.

The structures shown in Fig. 13.8b, c have length-varying cross-section S. This variation can be step-wise (Fig. 13.8b) or smooth (Fig. 13.8c). Connecting the structures of Fig. 13.8c with their lateral surfaces in the limit, we obtain a cylindrical construction (Fig. 13.8d). The injecting surface (cathode) in this construction is the outer cylinder. Figure 13.8e shows a planar-type semiconductor structure, where the increase of capacitance in the injector region of device C_1 is achieved by introducing an additional metal electrode. This electrode is not supplied with an electrical potential, that is, it is a diode. If a control signal is applied on the additional electrode, the structure becomes a field-effect transistor.

The structure shown in Fig. 13.8a also refers to a layered structure. In practice, it is almost impossible to make such a structure. It is practically impossible to find

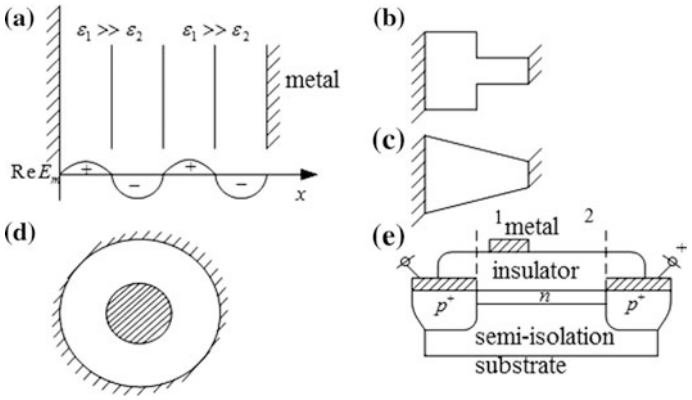


Fig. 13.8 Structures that realize negative dynamic resistance

materials consistent with a crystal lattice constant, but with different values of ε and v_0. However, layered structures can be realized in a *homogeneous* semiconductor with an *inhomogeneous* distribution of static field E_0. In this case, regions (layers) with various differential mobility values μ_d arise. The previously obtained formulas for calculating dynamic resistance are also suitable for use in this case. This approach is used in the determination of small-signal equivalent parameters of semiconductor diodes. Figure 13.9 shows, as an example, the distribution of field E_0 in a certain device. To find the layers with different values μ_d, an idealized field-velocity characteristic of the material is shown in the right-hand side of the figure. Setting lines for constant field strength E_{th} and E_s it is easy to differentiate layers 1, 2 and 3, for which differential mobility will differ: $\mu_{1d} > 0$, $\mu_{2d} < 0$ and $\mu_{3d} = 0$.

To calculate the equivalent resistance of such structures, one can use expression (13.8) for individual layers, "stitching" the layers according to coefficient M_i and injection delay angle Θ_i, as in a two-layered structure (see formula 13.28). Another example of this approach in calculating the equivalent resistance of a real device is presented in the next section.

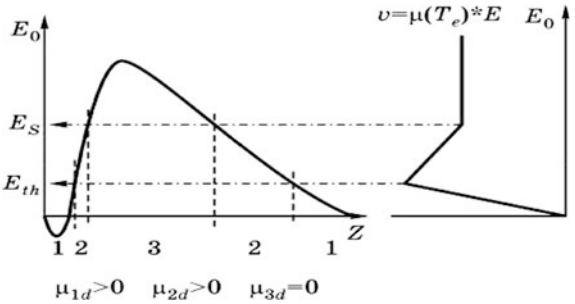


Fig. 13.9 Obtaining a layered structure of a homogeneous material with a varying field strength

13.4 IMPATT Diodes

13.4.1 Structure and Operation Principle of the IMPATT Diode

W.T. Read (Bell labs) propose a diode with a multi-layer structure as a microwave oscillator in 1958. This structure was called a Reed diode or IMPATT (Impact Avalanche and transit Time) diode. A.S. Tager (Istok JSC) observed microwave oscillations in diodes under avalanche breakdown conditions in 1959. Since then, Reed diodes have become one of the most widely used microwave generators.

Consider the principle of the IMPATT diode using the structure proposed by W. Read (Fig. 13.10a). It contains a p^+–n junction, a region of intrinsic conductivity (i-region), and high-doped contact regions. The distribution of the static field in such a structure with positive potential U_0 on region n^+ is shown in Fig. 13.10b. A characteristic feature of this distribution is the localization of a high field strength in a narrow region of the backward shifted p^+–n junction. Figure 13.10b shows an option where the external voltage provides a complete depletion of the n-region, and in the i-region the field is larger than the saturation field E_s. As voltage U_0 increases, the electric field strength in region l_a reaches the critical (breakdown) value E_{bd}, sufficient for the occurrence of avalanche breakdown. This area is called *the avalanche zone*. Electrons emerging from the avalanche zone drift under the action of an electric field in the n- and i-regions until an anode is reached. The length of the region, where the drift occurs at a constant velocity, is called *the drift space* and in Fig. 13.10b it is defined as l_d.

Let us consider the phenomena in the avalanche zone. Suppose that, along with bias voltage U_0, a harmonic alternating voltage acts on the diode, that is,

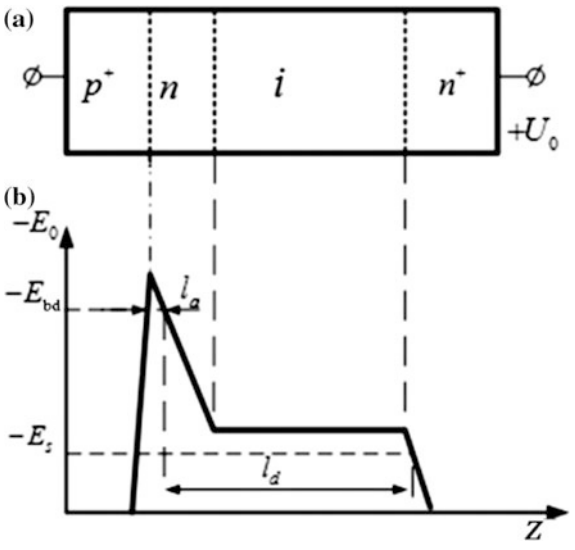


Fig. 13.10 Read diode. **a** Device structure; **b** distribution of electric field strength

$u = U_0 + u_m \sin \omega t$, where $u_m \ll U_0$. We also assume that the constant bias creates an electric field in the avalanche zone, close to breakdown: $E_0 \approx E_{bd}$. The dependence of the voltage on time is shown in Fig. 13.11. From the moment of time $t = t_1$ the field intensity in the avalanche zone becomes greater than the threshold value and the current through the transition begins to increase. The rate of current growth is determined by the difference $E - E_{bd}$ and is maximum at moment $t = t_2$. Then, the current growth rate decreases and becomes zero at time $t \approx t_3$ (Fig. 13.11). After this, the total voltage becomes less than the threshold value, and the convection current in the avalanche zone begins to decrease. At time $t = t_4$ the current becomes zero.

Thus, the maximum convection current for the ensemble of electrons emerging from the avalanche zone lags behind the maximum voltage by an angle of order $\pi/2$, that is, the avalanche zone behaves as an inductance. In other words, the avalanche zone provides a delay in the injection of the charge carriers, necessary to obtain a negative dynamic resistance. This fact corresponds to the waveform of induced current i_i i_i (Fig. 13.11).

In the avalanche zone, both electrons and holes are generated. Holes, moving to the left, recombine on the cathode contact and, in practice, do not induce current in the external circuit. The electrons, moving to the right, pass through the drift space, interacting with an alternating field in this region. The static field in this region is larger than E_s—saturation field, which corresponds to the zero Maxwellian

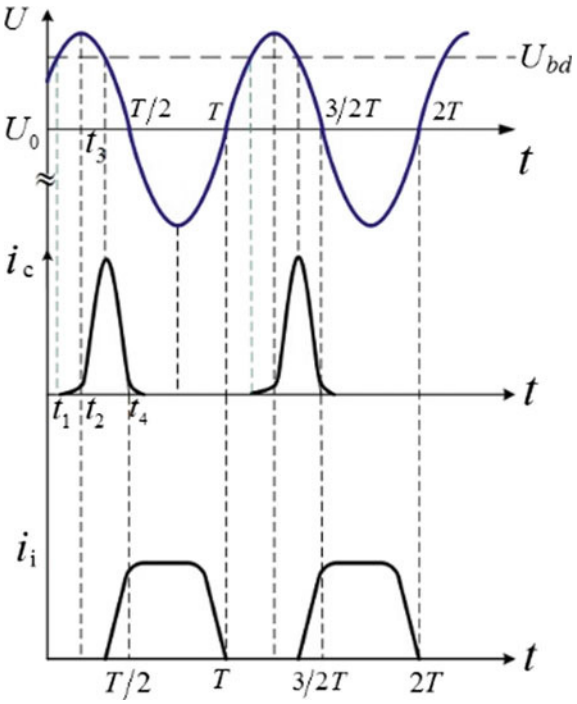


Fig. 13.11 Time dependencies of voltage and currents in IMPATT

relaxation frequency. Since the maximum current and field intensity in the avalanche zone are shifted by an angle $\pi/2$ (Fig. 13.11), the electron bunch center enters the drift space at the time when the high-frequency field changes from accelerating to decelerating. If the length of the drift space is chosen such that the transit angle of the electrons is equal to π, the bunch moves for a time $t \approx T/2$ in the deceleration phase of the alternating field, transferring to it a fraction of its kinetic energy. Since in the accelerating half-period, the number of electrons moving in the drift space is less than in the decelerating half-period, the energy balance is in favor of the alternating electromagnetic field, and therefore the structure has a negative dynamic resistance.

13.4.2 *Analysis of the Processes in the Avalanche Zone. Equivalent Resistance*

The general approach to the current transport processes in the IMPATT diode using the conservation equations (see Sect. 2.3) and the Poisson equation is too complex to obtain a general solution. We introduce a number of simplifying assumptions that will enable us to obtain an analytic solution for the avalanche zone. Consider the avalanche zone with length l_a (Fig. 13.12). We assume that $\partial/\partial x = \partial/\partial y = 0$ (one-dimensional model). Diffusion and thermal generation upon recombination are absent. Also $D_n = D_0 = 0$, $G_{\text{heat}} = 0$ (this assumption is justified because of the presence of a strong electric field in the zone and a high rate of impact ionization). The impact ionization coefficients for electrons and holes are: $\alpha_n = \alpha_p = \alpha$; the velocities of electrons and holes are equal to the saturation values, which, in turn, are equal to each other; $v_n = v_{ns}$, $v_p = v_{ps}$ and $v_{ns} = v_{ps} = v_0$.

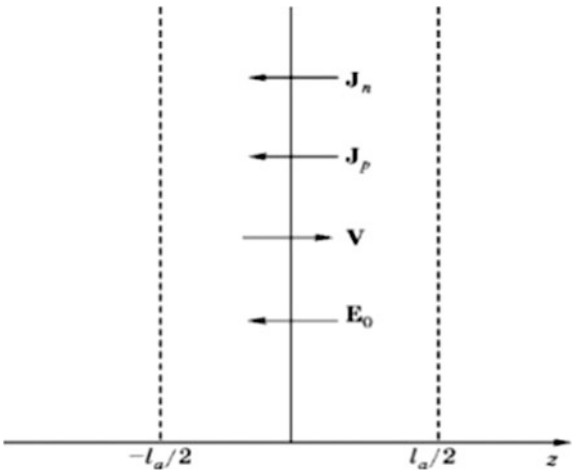


Fig. 13.12 Positive directions of currents, velocity and field strength in the avalanche zone

We choose positive directions of current densities, field strengths and velocities in such a way that $\mathbf{J}_n = -J_n\mathbf{e}_z$, $\mathbf{J}_p = -J_p\mathbf{e}_z$; $\mathbf{v} = v_0\mathbf{e}_z$; $\mathbf{E} = -E\mathbf{e}_z$, where $\mathbf{e}_z$ is the unit vector of axis z (Fig. 13.12).

Under these assumptions, the continuity equation (2.10) and the Poisson equation (2.8) are reduced to the following forms:

$$\frac{1}{v_0}\frac{\partial J_n}{\partial t} + \frac{\partial J_n}{\partial z} = \alpha\left(J_n + J_p\right), \tag{13.29}$$

$$\frac{1}{v_0}\frac{\partial J_p}{\partial t} + \frac{\partial J_p}{\partial z} = \alpha\left(J_n + J_p\right), \tag{13.30}$$

$$\varepsilon v_0\frac{\partial E}{\partial z} = J_n - J_p. \tag{13.31}$$

Define

$$J_n + J_p = J; \quad J_n - J_p = K \tag{13.32}$$

Adding and subtracting (13.29) and (13.30), we obtain:

$$1/v_0\frac{\partial J}{\partial t} + \frac{\partial K}{\partial z} = 2\alpha J, \tag{13.33}$$

$$1/v_0\frac{\partial K}{\partial t} + \frac{\partial J}{\partial z} = 0, \tag{13.34}$$

$$\varepsilon v_0\frac{\partial E}{\partial z} = K. \tag{13.35}$$

The current emerging from the left border of the avalanche zone consists purely of holes, and the current from the right boundary is purely electronic, so the boundary conditions for the resulting system of equations have the form

$$J = J_p; \quad K = -J; \quad x = -l_a/2; \quad J = J_n; \quad x = -l_a/2. \tag{13.36}$$

We introduce the small-signal approximation, i.e. assume that all the unknown quantities can be represented as a sum of a constant and a variable of the components:

$$J = J_0 + \widetilde{J}; \quad K = K_0 + \widetilde{K}; \quad E = E_0 + \widetilde{E}. \tag{13.37}$$

The variable components are much smaller in magnitude than the constants:

$$\widetilde{J} \ll J_0; \quad \widetilde{K} \ll K_0; \quad \widetilde{E} \ll E_0.$$

This approximation makes it possible to represent the ionization coefficient $\alpha(\boldsymbol{E})$ as an expansion in a Taylor series in the vicinity of $\boldsymbol{E}_0$, confining ourselves to the first two terms of the expansion:

$$\alpha(E) = \alpha(E_0) + (\mathrm{d}\alpha/\mathrm{d}E)\widetilde{E} + 0.5\left(\mathrm{d}^2\alpha/\mathrm{d}E^2\right)\widetilde{E}^2 + \cdots \approx \alpha_0 + \alpha_0'\widetilde{E}, \qquad (13.38)$$

where $\alpha_0 = \alpha(E_0)$; $\alpha_0' = (\mathrm{d}\alpha/\mathrm{d}E)_{E0}$.

Equations (13.33)–(13.35) must obviously be satisfied separately for constant and variable components. Substituting (13.38) in (13.33), we obtain two systems of equations: for constant components:

$$\partial K_0/\partial z = 2\alpha_0 J_0, \qquad (13.39)$$

$$\partial J_0/\partial z = 0, \qquad (13.40)$$

$$\varepsilon v_0 \frac{\partial E_0}{\partial z} = K_0$$

and for variable components:

$$1/v_0 \frac{\partial \widetilde{J}}{\partial t} + \frac{\partial \widetilde{K}}{\partial z} = 2\alpha_0\widetilde{J} + 2\alpha_0'\widetilde{E}J_0; \qquad (13.41)$$

$$1/v_0 \frac{\partial \widetilde{K}}{\partial t} + \frac{\partial \widetilde{J}}{\partial z} = 0; \qquad (13.42)$$

$$\varepsilon v_0 \frac{\partial \widetilde{E}}{\partial z} = \widetilde{K}. \qquad (13.43)$$

Integrating (13.39) in the range from $-l_a/2$ to $l_a/2$ considering (13.40), we obtain

$$2J_0 = 2J_0 \int_{-l_a/2}^{l_a/2} \alpha_0 \mathrm{d}z$$

or

$$\int_{-l_a/2}^{l_a/2} \alpha_0 \mathrm{d}z = 1. \qquad (13.44)$$

Thus, the steady-state mode is possible only if the avalanche breakdown condition is fully satisfied.

The density of the total current consists of convection current density J and displacement current density $J_c = -\varepsilon\partial\widetilde{E}/\partial t$. According to the law of total current

$$\frac{\partial \widetilde{J}_m}{\partial z} = \frac{\partial \widetilde{J}}{\partial z} - \varepsilon \frac{\partial^2 \widetilde{E}}{\partial z \partial t} = \frac{\partial \widetilde{J}}{\partial z} - \frac{\partial \widetilde{\rho}}{\partial z} = 0$$

If space charge $\widetilde{\rho}$ is not taken into account, it is easy to obtain $\partial \widetilde{J}/\partial z$. This approximation allows us to integrate (13.41) with respect to the coordinate:

$$\frac{l_a}{v_0}\frac{\partial \widetilde{J}}{\partial t} + 2\widetilde{J} = 2\widetilde{J} \int\limits_{-l_a/2}^{l_a/2} \alpha_0 \mathrm{d}z + 2J_0 \int\limits_{-l_a/2}^{l_a/2} \alpha'_{0e} \widetilde{E} \mathrm{d}z. \tag{13.45}$$

We define

$$\int\limits_{-l_a/2}^{l_a/2} \alpha'_0 \widetilde{E} \mathrm{d}z = \alpha'_{0e} \int\limits_{-l_a/2}^{l_a/2} \widetilde{E} \mathrm{d}z = \alpha'_{0e} \widetilde{u}_a, \tag{13.46}$$

where α'_{0e} is the affective value of the ionization coefficient derivative, obtained by averaging over the avalanche zone; and $\widetilde{u}_a$ is the alternating voltage of the avalanche zone. Substituting (13.44) and (13.46) in (13.45) and multiplying both sides of this equation by the cross-sectional area of the avalanche zone, we obtain

$$\frac{\partial \widetilde{I}}{\partial t} = \frac{2I_0 \alpha'_{0e}}{\tau_a} \widetilde{u}_a, \tag{13.47}$$

where $\tau_a = l_a/v_0$ is the carrier transit time through the avalanche zone.

It is well known that the relation between the current and the voltage on inductance L is determined by expression $u = L(\partial I/\partial t)$. Comparing this expression with (13.47), we see that equivalent inductance can be assigned to the avalanche zone:

$$L_a = \frac{\tau_a}{2I_0 \alpha'_{0e}}. \tag{13.48}$$

Let us consider the effect of space charge on processes in the avalanche zone. Suppose that at the initial moment of time, the distribution of concentrations of field and carriers in the zone is uniform. At some instant of time, the electric field in a thin layer inside the zone increases (for example, due to fluctuations). An increase in the field will cause an increase in the generation rate, creating an excess concentration of holes and electrons in the layer (Fig. 13.13a). Under the influence of the electric field, the holes move to the left and the electrons move to the right, and the concentration distribution of carriers takes the form shown in Fig. 13.13b. Between the excess charges there is an electric field directed toward the external field. That leads to the electric field weakening and a decrease in the generation rate (Fig. 13.13c).

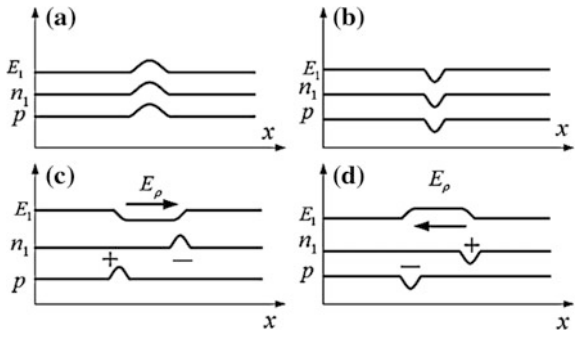


Fig. 13.13 Oscillations of charge carrier concentration and field strength in the avalanche zone

Due to the drift of carriers, regions with insufficient concentration of holes and electrons move in the avalanche zone (Fig. 13.13d). Between them there is an electric field directed in the same direction as the external field, leading to an increase in the field and the generation rate, etc. Thus, auto-oscillations arise in the avalanche zone.

A detail analysis taking into account changing of the convection current lengthwise of avalanche zone leads to the following expression for the total conductivity of the avalanche zone:

$$Y_a = i\omega C_a + (i\omega L_a - R_a)^{-1}, \tag{13.49}$$

where $C_a = \varepsilon S/l_a$ is the capacity of cross-section s; $L_a = \tau_a\left(3I_0\alpha'_{0e}\right)^{-1}$ is the avalanche zone inductivity; and $R_a = 5(I_0\alpha'_{0e})^{-1}$ is the active zone resistance.

Thus, the avalanche zone can be represented in the form of an oscillatory circuit, the inductance of which is one and a half times smaller than that calculated from the formula (13.48), obtained from the approximate theory. At some ratio of values, the sign of the active conductivity component can be negative. Thus, the avalanche zone can serve as a source of microwave energy. This is possible at frequencies close to avalanche frequency $\omega_a = (L_aC_a)^{-1/2} = (3J_0U_0\alpha'_{0e}/\varepsilon)^{1/2}$. This frequency does not depend on the width of the avalanche zone. It increases with an increase in current density J_0 and α'_{0e} in the sample and both of these quantities, in turn, depend on the bias voltage, making it possible to perform an electrical tuning of the avalanche frequency within a wide range.

However, to obtain the frequency ω_a lying in the microwave band, very large values of current density in the diode (about 10^3 A/cm^2) are required. Such conditions can be created, for example, in the *p–i–n* diode with sufficiently large reverse voltages. In the narrow avalanche zone, as, for example, in Read diodes, the oscillation properties of the avalanche are not practically manifested. Diodes that use oscillations in the avalanche zone to generate or amplify microwave oscillations are called the Misawa diodes, named after the Japanese scientist who first predicted the existence of such oscillations and found them experimentally.

13.4.3 Small-Signal Impedance of the IMPATT Diode

The impedance of the diode consists of three components: avalanche zone impedance Z_a, drift space impedance Z_d, and parasitic contact resistance R_s, i.e.

$$Z = Z_a + Z_d + R_s. \tag{13.50}$$

The equivalent scheme of the avalanche zone in the linear approximation consists of avalanche inductance L_a and avalanche zone capacity C_a. Hence, according to (13.49), for $R_a = 0$

$$Z_a = \frac{1}{i\omega C_a + (i\omega L_a)^{-1}} = \frac{i\omega L_a}{1 - (\omega/\omega_a)^2}$$

The avalanche zone is represented in the form of a parallel connection of capacity C_a and inductance L_a and it allows calculation of the ratio between injected current I_{inj} (current through L_a) and total current I_m (current through L_a and C_a):

$$\frac{I_{inj}}{I_m} = \frac{(i\omega L_a)^{-1}}{\left[i\omega C_a + (i\omega L_a)^{-1}\right]} = \frac{1}{1 - (\omega/\omega_a)^2}$$

Substituting this expression in (13.16) and using (13.50), we obtain

$$Z = R_s + \frac{l_d^2}{v_0 \varepsilon S\left(1 - \omega^2/\omega_a^2\right)} \frac{1 - \cos\Theta}{\Theta^2} + \frac{1}{i\omega C_d}\left[1 - \frac{\sin\Theta}{\Theta} + \frac{(\sin\Theta)/\Theta + l_a/l_d}{1 - \omega^2/\omega_a^2}\right]$$

where Θ is the transit angle in the drift space; and $C_d = \varepsilon S/l_d$ is its capacity. From the expression obtained, it can be seen that the active diode resistance can be negative when $\omega_a > \omega$ and with correct transit angle Θ. Figure 13.14 shows the dependence of the active and reactive components of diode resistance, normalized to resistance value R_0: $R_0 = (l_a/l_d)/\left(I_0 \alpha'_{0e}\right)$.

Maximum absolute negative resistance is observed at $\Theta = \pi$. It is proportional to the constant current component of the diode I_0 and the derivative of the ionization coefficient with respect to the electric field strength α'_{0e}.

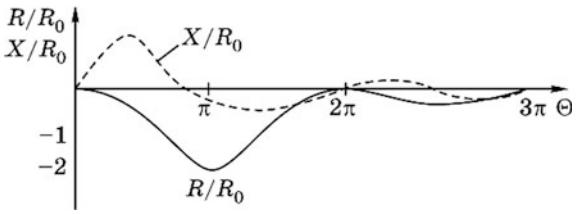


Fig. 13.14 Dependences of Read diode impedance on the transit angle in the drift space

13.4.4 Nonlinear Operating Mode of the IMPATT Diode

The aforementioned analysis in 13.11 did not take into account the influence of the space charge of bunches emerging from the avalanche zone. This influence is significant for large amplitudes of voltage and current in the diode. Consider changes in the electric field due to a bunch of electrons ρ emerging from the avalanche zone (Fig. 13.15). The field of the bunch is summed with the external field to the right and is subtracted with to the left, and as a result, field strength in the avalanche zone E_{left} decreases. This leads to premature termination of current growth (Fig. 13.15). The phase shift between the injection current and the total current at the entrance to the drift space becomes smaller than π by some angle φ. Hence, negative conductivity of the diode decreases. To the right of the bunch, space-charge field E_{right} grows as the bunch approaches the anode. At large current amplitudes, the field near anode can become more than the breakdown field and collision ionization will occur. The emerging holes move to the cathode in the accelerating HF field, taking energy from it, leading to a decrease in the negative conductivity of the diode. This undesirable effect makes it necessary to reduce the voltage and current amplitude on the diode, i.e. its output power and efficiency. The main difficulty in analyzing the operation of the IMPATT diode in the large signal mode is to solve nonlinear equations (13.33)–(13.35) describing processes in the avalanche zone. An exact solution of these equations is possible only by numerical methods. The result of the solution is the dependence of the convection current waveform upon escape from the avalanche zone, on the parameters of the zone and on the amplitude of the voltage on it.

Let us consider a number of approximations that allow us to obtain an analytic solution qualitatively consistent with the results of more accurate calculations. Integrating (13.33), under the assumption that current density J is independent of the coordinate and taking into account boundary conditions (13.36), we obtain

$$\frac{\tau_a}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\ln J) = \int_{-a/2}^{a/2} \alpha \mathrm{d}z - 1. \tag{13.51}$$

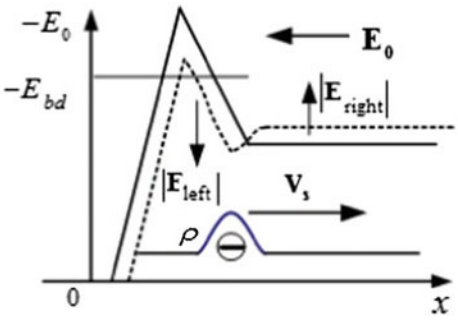


Fig. 13.15 Influence of space charge on the waveform of the field in the avalanche zone

The dependence of the ionization coefficient on the electric field strength is expressed by the formula

$$\alpha = A \exp[-(b/E_{\text{th}})]^m,$$

where A, b, m are parameters defined by the semiconductor material and the type of carriers. We approximate this dependence α on E near the point $\alpha_0 = \alpha(E_0)$, using the power function

$$\alpha = \alpha_0 (E/E_0)^n$$

The value of exponent n for silicon is six.

For an abrupt p^+–n^- junction, the field varies linearly from maximum value E_a for $z = -l_a/2$ to minimum E_d for $z = l_a/2$ (space charge is not taken into account). Obviously,

$$\partial E/\partial z = -(E_a - E_d)/l_a.$$

We use this expression and find an expression for the integral of $\int_{-l_a/2}^{l_a/2} \alpha \mathrm{d}z$ with the following condition $E_d \ll E_a$:

$$\int_{-l_a/2}^{l_a/2} \alpha \mathrm{d}z = \int_{E_a}^{E_d} \alpha \frac{\mathrm{d}z}{\mathrm{d}E} \mathrm{d}E = \frac{\alpha_0 l_a}{E_a - E_d} \left[\left(\frac{E_a}{E_0}\right)^{n+1} - \left(\frac{E_d}{E_0}\right)^{n+1} \right] \approx \left(\frac{E_a}{E_{a0}}\right)^{n+1}. \tag{13.52}$$

Value E_{a0} is related to value E_0, which is not known in advance. Taking into account, however, breakdown conditions (13.44), one can note that E_{a0} is the maximum field in the avalanche zone under steady-state breakdown conditions. Using (13.52), we rewrite (13.51):

$$\mathrm{d}(\ln J)/\mathrm{d}t = 2\left[(E_a/E_{a0})^{n+1} - 1\right]/\tau_a, \tag{13.53}$$

where $\tau_a = l_a/v_0$ is the transit time in the avalanche zone. The field strength in the avalanche zone can be represented as a sum of three components:

$$E_a = E_{a0} - E_c + E_m \sin \omega t. \tag{13.54}$$

The first component is the field establishing under the action of a constant voltage applied to the diode. The third component is the field establishing under the action of an alternating voltage across the diode. Field E_c is caused by the voltage establishing due to the flow of the rectified current through the diode.

Substituting (13.54) in (13.53), expanding the resulting expressions by the binomial formula and discarding terms of the third and higher orders of smallness, we obtain the small signal approximation ($E_c \ll E_a$)

$$\frac{\partial}{\partial t}(\ln J) = \frac{2(n+1)}{\tau_a}\left[-\frac{E_c}{E_{a0}} + \frac{E_m}{E_{a0}}\sin\omega t + \frac{\pi}{2}\left(\frac{E_m}{E_{a0}}\right)^2 \sin^2\omega t\right]$$

In order to prevent the current logarithm growing linearly with time, the first term and the constant component of the third term on the right-hand side of the equation must cancel out, that is:

$$\frac{E_c}{E_{a0}} = \frac{\pi}{4}\left(\frac{E_m}{E_{a0}}\right)^2$$

Taking this equality into account, and neglecting the terms containing the second and higher voltage harmonics, we obtain

$$\frac{\partial}{\partial t}(\ln J) = \frac{2(n+1)}{\tau_a}\frac{E_m}{E_{a0}}\sin\omega t$$

from which

$$J(t) = J_m \exp(-X\cos\omega t), \tag{13.55}$$

where $X = \frac{2(n+1)}{\Theta_a} E_m/E_{a0}$ is the dimensionless field intensity amplitude in the avalanche zone; and J_m is the integration constant, defined below. Using the well-known relation for modified Bessel functions, we rewrite expression (13.55) in the form

$$J(t) = J_m\left[I_0(X) + 2\sum_{k=1}^{\infty}(-1)^k I_k(X)\cos(k\omega t)\right]$$

From this expression, it follows that the constant component of the current density is $J_0 = J_m I_0(X)$. Thus,

$$J(t) = J_0\left[1 + \frac{2}{I_0(X)}\sum_{K=1}^{\infty}(-1)^K I_K(X)\cos(k\omega t)\right]. \tag{13.56}$$

Formulas (13.55) and (13.56) determine the form and spectral composition of the bunch emerging from the avalanche zone. The form of the bunch for different values of X is shown in Fig. 13.16.

It is obvious that as the electric field strength amplitude increases, the bunch contracts and its limit form tends to the δ-function. In accordance with (13.55), the first harmonic of the convection current density entering the drift space can be written in the form

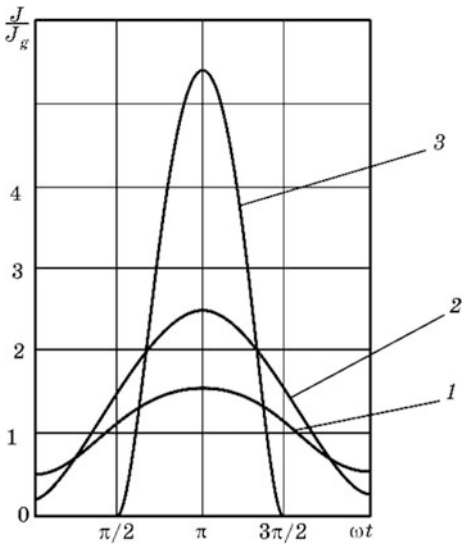


Fig. 13.16 Form of the bunch of electrons coming out of the avalanche zone: 1—$X = 0.5$; 2—$X = 1.5$; 1—$X = 5$

$$J_1 = 2J_0 \frac{I_1(X)}{I_0(X)} \cos \omega t.$$

In fact, the maximum value of the current is not achieved at time $\omega t = \pi$, as follows from the written expression, but at an earlier point in time. Indeed, it follows from (13.53) that the maximum current is reached when $E_a = E_{a0}$, that is, at the time when $E_m \sin \omega t_1 = E_c$. Such that $\omega t_1 = \pi - \delta$, where $\delta = \arcsin\left(\frac{\pi}{4}\frac{E_m}{E_{a0}}\right)$.

Consequently, the expression for the first harmonic amplitude of the avalanche current density must be written in the form

$$J_1 = -2J_0 \gamma(X) \exp(i\delta),$$

where $\gamma(X) = I_1(X)/I_0(X)$ is the ratio of the first kind of modified Bessel functions of the first and zero orders (Fig. 13.17). Multiplying J_1 by the area of the diode and setting the result in expression (13.22), we obtain the first harmonic amplitude of the induced current:

$$I_{1ind} = -2I_0 M_d \gamma(X) \exp(i\delta) \exp(-i\Theta/2),$$

where $M_d = \sin(\Theta/2)/(\Theta/2)$ is the interaction coefficient in the drift space; and I_0 is the constant component of the current through the diode. To determine power $P = 1/2\mathrm{Re}\left(U_m \overset{*}{I}_m\right)$ delivered by the diode, it is necessary to take into account phase shift $\pi/2$ between field and current. Thus,

$$P = -U_m I_0 M \gamma(X) \cos(-\delta + \Theta/2 - \pi/2).$$

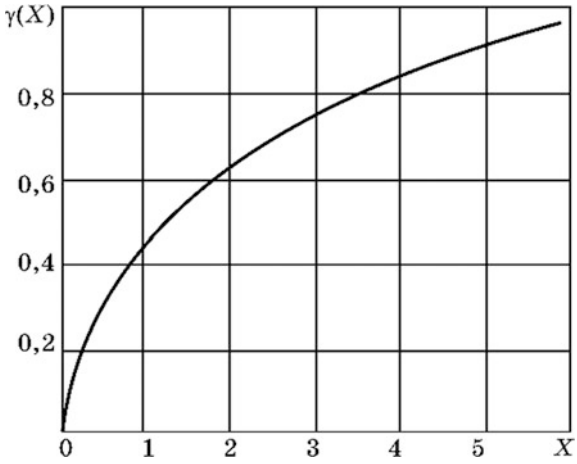


Fig. 13.17 Dependence graph $\gamma(X)$

We can see from the expression obtained that the optimal transit angle in the drift space $\Theta_{opt} = \pi + 2\delta$ depends on the amplitude of the voltage across the diode. As the amplitude increases, the optimal transit angle increases, leading to a decrease in the interaction coefficient.

Assuming the transit angle in the drift space to be equal to π, that is, optimal for small amplitudes of the voltage across the diode, we obtain

$$P = -0.64 U_m I_0 \gamma(X) \cos\delta.$$

The negative power sign indicates that the diode transmits high-frequency electromagnetic energy to the load. The efficiency of the device is given as:

$$\eta = \frac{|P|}{U_0 I_0} = 0.64 \frac{U_m}{U_0} \gamma(X) \cos\delta. \tag{13.57}$$

Since ratios $E_m/E_{a0} = U_m/U_0$, the device's efficiency is obviously determined by the ratio of the voltage amplitude across the diode to the constant bias voltage. Efficiency also increases with a decrease in the transit angle in the avalanche zone. The dependence of efficiency, calculated by formula (13.57), on ratio U_m/U_0 for different transit angles Θ_a is shown in Fig. 13.18. Obviously, the smaller Θ_a, the higher the maximum device efficiency and the lower the ratio of U_m/U_0, at which it is reached. The maximum diode efficiency can be estimated by assuming that in this case $X \gg 1$ and the form of the convection current bunch is close to the δ-function. Such bunches create an induced current in the form of a meander with an equal pulse and pause duration (at transit angle $\Theta = \pi$). For this current waveform, the ratio of the first harmonic amplitude to the constant component is $I_{m1}/I_0 = 4/\pi$. If we set $\delta = 0$, the shift of the induced current relative to the voltage is 180°. Then

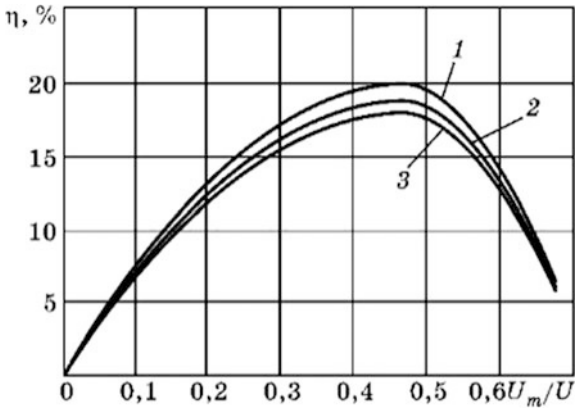


Fig. 13.18 Dependence of the IMPATT diode ECE on voltage amplitude when: 1—$\Theta_a = 0.1$; 2—$\Theta_a = 0.5$; 3—$\Theta_a = 0.9$

$$P = \frac{1}{2} U_m \overset{*}{I}_m = -\frac{2}{\pi} U_m I_0$$

from which

$$\eta = \frac{2}{\pi} \frac{U_m}{U_0}$$

The maximum value of ratio U_m/U_0 is limited on the one hand, by the possibility of impact ionization in the drift space, and on the other hand, by a decrease in the field, which causes a decrease in the drift velocity of the carriers below the saturation value. Thus, we have two conditions:

$$E_0 = E_{dm} \leq E_{bd}; \quad E_0 - E_{dm} \geq E_s$$

For a silicon diode $E_{\text{bd}} \approx 200$ kV/cm, $E_{\text{s}} \approx 10$ kV/cm, the maximum ratio E_{dm}/E_0 can be close to one, and then $\eta_{\max} \approx 2/\pi = 0.64$, considerably exceeding the results obtained with formula (13.57), since in this case the phase shift is not considered between the current and the voltage arising at large values of U_m/U_0. In addition, due to the space charge effect, ratio U_m/U_0 usually does not exceed 0.5–0.7 giving a $\eta_{\max}$ value of 30–35%. In practice, the maximum oscillator and amplifier efficiency in IMPATT diodes does not exceed 15–20%.

13.4.5 IMPATT Diodes Operating in Trapped Plasma Transit Mode (TRAPATT)

The operation mode of a Reed diode, considered in the previous section, supposed no impact ionization in the drift space. This mode was called *transit*, or IMPATT mode (according to the first letters of the English words IMPact Avalanche Transit

Time). In 1964, another mode of Reed diode operation was discovered, characterized by significantly higher efficiency, which was called the *plasma-trapping mode*, or TRAPATT (TRApped Plasma Avalanche Transit Time). For this operation mode, the non-sinusoidal form of the voltage across the diode in the form of a sequence of short pulses is typical, as shown in Fig. 13.19. Under the action of these pulses, dense bunches of electrons emerge from the avalanche zone. The field between the bunch and the cathode decreases, and in the region between the bunch and the anode, the field increases to values greater than the breakdown value. Impact ionization occurs in the maximum field region near the bunch that leads to a further increase in the field to the right of this region, etc. This leads to the appearance of an impact ionization wave that runs through the drift space at high speed leaving behind it an electron-hole plasma of high concentration. Such plasma has very high conductivity, causing a sharp decrease in the voltage across the diode after the wave passes. The voltage remains low while the current is large during the whole plasma resolution time, which is determined mainly by the external circuit. After plasma dissipation, current decreases sharply, diode resistance is restored, and the voltage in it sharply increases, leading to a new impact ionization wave. The idealized waveforms of voltage and current through the diode are shown in Fig. 13.19.

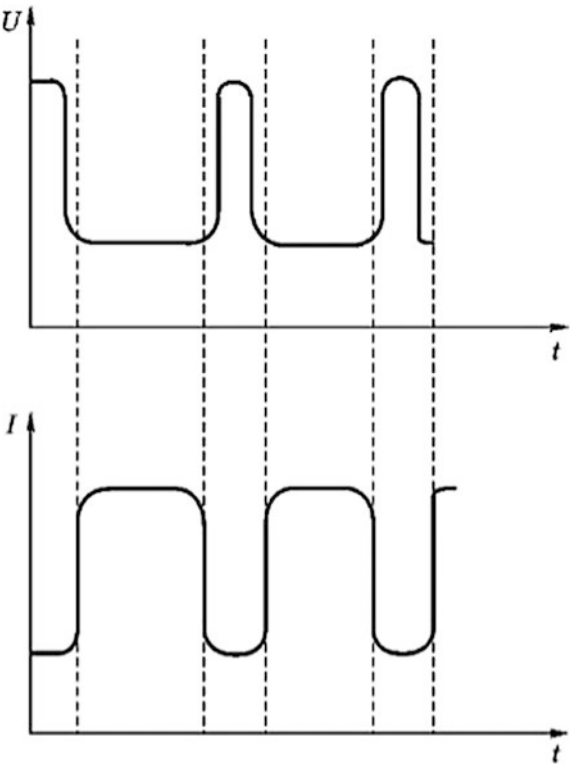


Fig. 13.19 Idealized voltage and current waveforms through the IMPATT diode in the TRAPATT mode

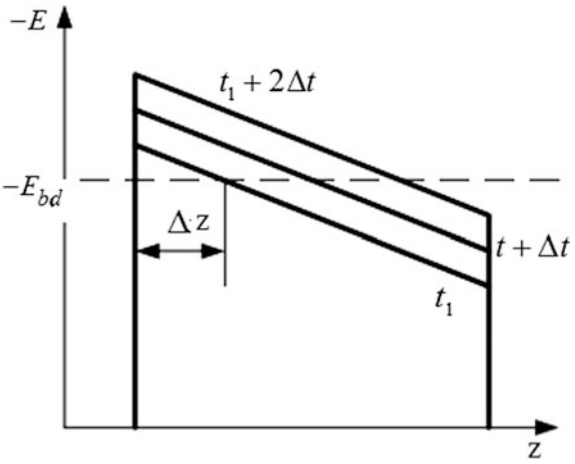


Fig. 13.20 Calculating the current density in IMPATT diodes in the TRAPATT mode

Since very weak current flows through a diode at high voltage, and large current flows at low voltage, device efficiency can be very large.

The process of impact ionization wave propagation can be analyzed qualitatively in the following example. Consider structure p^+–p^-–n–n^+, under reverse voltage. If the n-region is completely depleted of electrons, then the structure has a near-triangular electric field distribution (Fig. 13.20).

We apply to the structure current step I_0 ($I = 0$, $t < 0$; $I = I_0$, $t \geq 0$). Since in the p^--region there are no current carriers, displacement current $I_c = \varepsilon S \partial E/\partial t$ flows, i.e., the field strength increases with time. If at moment $t = 0$, field dependency is expressed as $E(z, 0) = E_m - qN_d z/\varepsilon$, then at an arbitrary moment of time t, this dependence takes the form

$$E(z,t) = E_m - qN_d z/\varepsilon + I_c t/(\varepsilon S)$$

From this expression, it follows that the point at which $E = E_m = E_{\text{bd}}$, shifts along the z axis at a velocity of $v_a = \mathrm{d}z/\mathrm{d}t = I_c N_d/(qS)$. At high current density, v_a is considerably (10 times or more) greater than the saturation velocity v_s. The diode field has to drop from values exceeding breakdown intensity E_{bd} to very small values after the impact ionization wave passes. To achieve this, it is apparently required that created charge $Q = \varepsilon E_{bd} S$ is absorbed during approximately half of the oscillation period T. Hence, $I_0 T/2 = Q$, where I_0 is the current in the external circuit. This determines the necessary current density through the diode: $J_0 \geq 2\varepsilon E_{\text{bd}} f$, where f is the oscillation frequency. Substituting the values of ε and E_{bd} for silicon in the expression ($\varepsilon = 11.7\varepsilon_0$, $E_{bd} \approx 2 \times 10^5$ V/cm), we obtain $J_0 \geq 10^{-6} f$, which for a frequency of 1 GHz gives a maximum current density of 1000 A/cm^2. The rapid growth of current density through the diode with increasing frequency limits the use of the TRAPATT mode at frequencies greater than 10–20 GHz.

The waveform of the voltage across the diode shown in Fig. 13.19 is idealized. In reality, a non-sinusoidal voltage waveform is obtained by creating a load for the diode

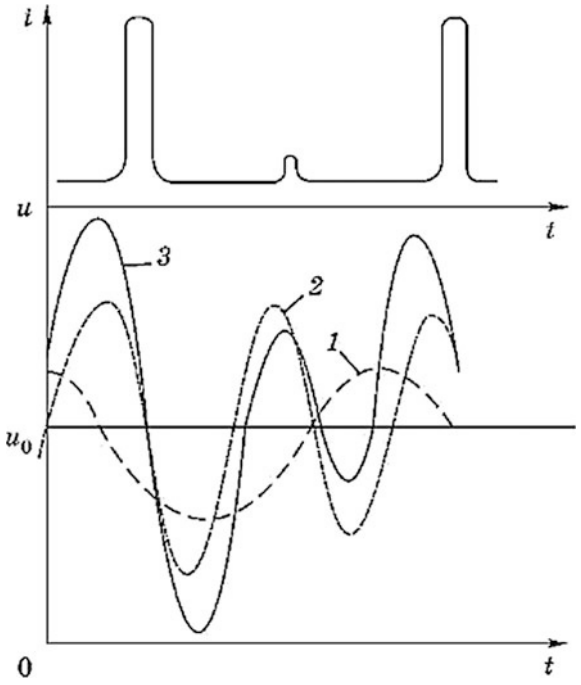


Fig. 13.21 Waveforms of the IMPATT diode voltage and current in the TRAPATT mode: 1—first voltage harmonic; 2—second harmonic; 3—resulting voltage

at several harmonics of transit frequency. So, if we take the first and second harmonics, as well as the constant bias, we can obtain the voltage waveform shown in Fig. 13.21.

Note that the amplitude of the second harmonic in this case exceeds the amplitude of the first. Hence, it is necessary to design a corresponding oscillation system.

An impact ionization wave is produced when a large voltage peak passes through the sample, causing a current pulse, the duration of which can be substantially longer than the transit time. A small voltage pulse with a rather high amplitude can provide diode operation in the usual transit-time mode. The duration of the current induced by this pulse is determined by the transit time. Usually the frequency of the second harmonic is chosen as equal to the transit-time frequency. The diode first starts to work in the transit-time mode at twice the frequency, but after the amplitude of the second harmonic reaches a certain value, it goes into the plasma-trapping mode. The quantitative analysis of the TRAPATT mode, in view of its significant nonlinearity, is possible only with the help of computers.

13.4.6 IMPATT Diode Structure and Design

The IMPATT diode structure, proposed by Read and considered in the previous sections, is not the only possible structure. Its advantage is the narrow avalanche zone, which increases the efficiency of the device. At the same time, the negative

resistance for this structure is observed only in a narrow region near transit-time frequency $f_{tt} \sim v_0/l_d$. Therefore, in addition to the Read structure, we also use structures like p^-–n–n^+, with a field distribution shown in Fig. 13.20. The avalanche zone in such a structure (when operating in the TRAPATT mode) is wider than that of the Read structure. This leads to a time spread for electrons to exit the avalanche zone.

Consequently, the maximum on the dependence curve for negative resistance on the transit angle is blurred, and the diode becomes capable of operating over a wide frequency range, albeit with lower efficiency. The dependence on frequency of active IMPATT diode conductivity with different structures is shown in Fig. 13.22.

Transit time and maximum current for these structures are the same. Recently, double-drift structures of type p^-–p–n–n^+ (Fig. 13.23a) have appeared, they use both electrons and holes. Such a structure can be represented as a serial connection of two diodes with a common avalanche zone, which allows double the output power due to increasing diode voltage. In addition, the efficiency increases, since one avalanche zone falls on two drift spaces. It is advisable to apply such diodes in the millimeter band as the drift space and diode voltage are very small.

When designing powerful devices, the problem of heat removal becomes especially important. The greatest heat dissipation in the IMPATT diode occurs in

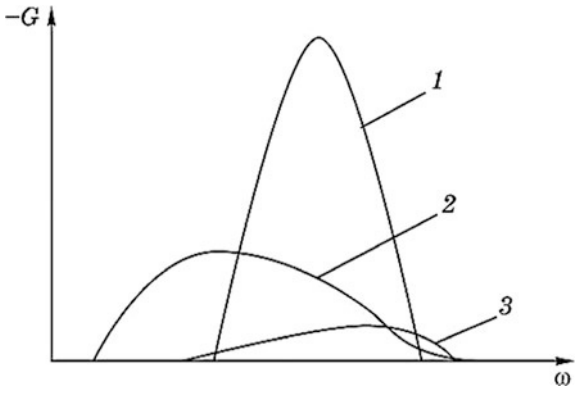


Fig. 13.22 Dependence of conductivity on frequency for different structures: 1—p^-–p–i–n^+; 2—p^-–n–n^+; 3—p^-–i–n^+

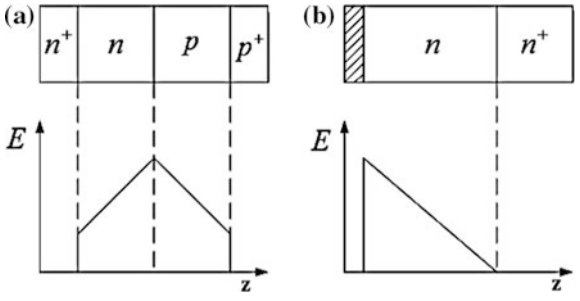


Fig. 13.23 IMPATT diode structures: **a** double-drift; **b** with a Schottky barrier

the avalanche zone, where the electric field strength is maximal. Therefore, it is important to reduce the heat path from this zone to the heat sink element. From this point of view, diodes with a Schottky barrier are promising. In the case of reverse mounting, the avalanche zone directly contacts the heat sink (Fig. 13.23b). Maximum continuous power is obtained with such diodes. We note, in conclusion, that all the above structures allow an inversion, that is, a replacement of *p*-regions by *n*-regions and vice versa.

Structurally, IMPATT diodes are, as a rule, in the form of a mesa-structure. The necessary doping profile is created through epitaxial growth, diffusion or ion implantation. Silicon is the material used for powerful IMPATT diodes in the centimeter band. Gallium arsenide is used in millimeter-wave devices. At the highest frequencies, single-drift silicon IMPATT diodes are used, since silicon allows the narrowest drift regions to be obtained. In order to protect it from the external environment, the semiconductor crystal is placed in a standardized case. Chip diodes are also used. The active region configuration of a typical IMPATT diode is shown in Fig. 13.24a. Epitaxial *n*-layer 3 is grown on the substrate 2. Then, p^+-layer 4 is formed by diffusion from the surface. After thermal contact 1 is applied, a film is etched to obtain a mesa-structure and is soldered to pin 5.

Figure 13.24b shows a design of a powerful GaAs IMPATT diode with a Schottky barrier. The diode contains four mesa-structures, mounted on a common copper gold-plated heat sink. The basis for the manufacture of the diode is a film with a doping profile, shown in Fig. 13.25a. Layers of Pt, Ti and Au, forming the Schottky barrier, are successively applied to the surface of the film through vacuum evaporation. To the gold layer of this junction, a gilded heat sink is welded through thermal compression. A metal anodic contact is applied to the other side of the film using the photolithography method, and the remaining part of the film is etched, leaving only the mesa-structures.

The selected doping profile ensures a rapid field decrease in the depletion layer of the Schottky barrier, reducing the length of the avalanche zone and the voltage drop across it. The distribution of the electric field in the diode is shown in Fig. 13.25b.

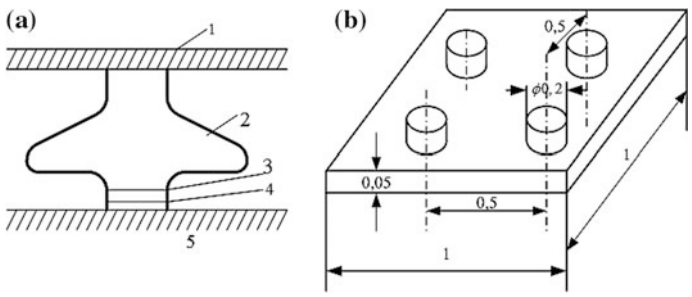


Fig. 13.24 Configuration of the active IMPATT diode region: **a** mesa-structure; **b** location of mesa-structures on the heat sink

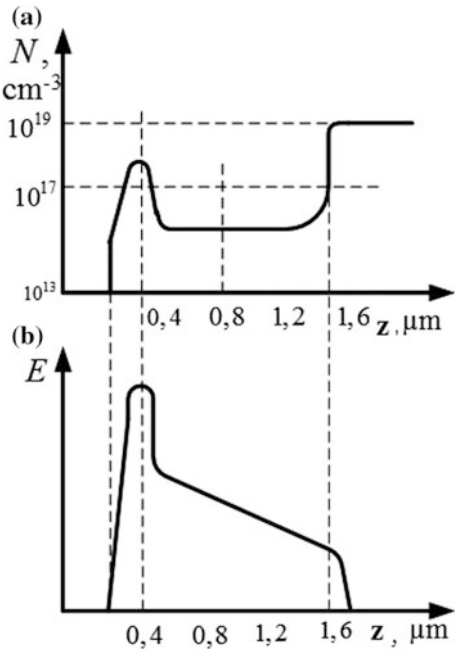


Fig. 13.25 Distribution of impurity concentration (**a**) and field strength (**b**) in the epitaxial layer of a diode with a Schottky barrier

Such a construction and doping profile dissipate heat much better than a single mesa-structure of equal area. Diodes of this design provide an output power of more than 10 W in a continuous mode with an efficiency of 22% at a frequency of 5 GHz. Diamond heat sinks are used nowadays in powerful IMPATT diodes.

At present, IMPATT diode oscillators are the most powerful and high-frequency solid-state generators operating in a continuous mode. The maximum oscillation frequency for industrially produced diodes is 300 GHz with an output power of 10 mW.

13.4.7 Structure and Parameters of IMPATT Diode Oscillators

Using IMPATT diodes as a one-port network with negative resistance, it is possible to build various microwave devices: generators, amplifiers, and converters, etc. The most widely used are IMPATT diode oscillators.

IMPATT diode amplifiers are used much less due to their comparatively large intrinsic noise caused by the stochastic nature of the avalanche process.

The construction of a typical IMPATT diode oscillator operating in centimeters band on a transit mode with electrical frequency tuning is shown in Fig. 13.26. The cylindrical cavity 1 and the IMPATT diode case 2 form a re-entrant resonator with

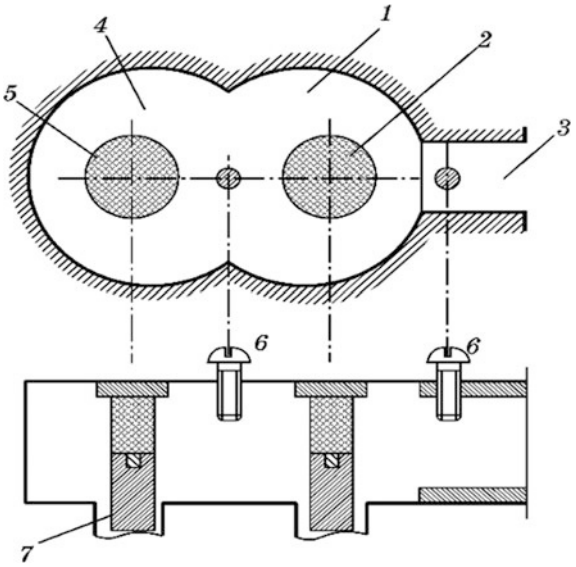


Fig. 13.26 Waveguide structure of an IMPATT diode oscillator operating in the transit mode

the semiconductor diode body in the gap. Slots couple the diode cavity with output waveguide 3 and cavity 4 with a varactor diode 5 adjusting the oscillation frequency. The coupling degree of the IMPATT diode cavity with the varactor and the load is regulated by means of screws 6. The diode DC supply circuits are made on coaxial transmission lines 7 with low impedance. This design limits the infiltration of microwave power through the terminals. The diode in the generator operates in a continuous mode, transmitting power in the order of 10 mW with 5–10% efficiency.

Figure 13.27 shows the microstrip structure of the Reed diode oscillator operating in TRAPATT mode. A silicon chip diode 1 with a diameter of 120 μm and a p^+–n–n^+ structure is fixed on a gilded copper base 2 to which an alumina ceramic substrate 3 of 0.5 mm thickness is attached. The upper terminal of the diode is connected to a segment of stripline 4 with a wave impedance of 30 Ω and a length

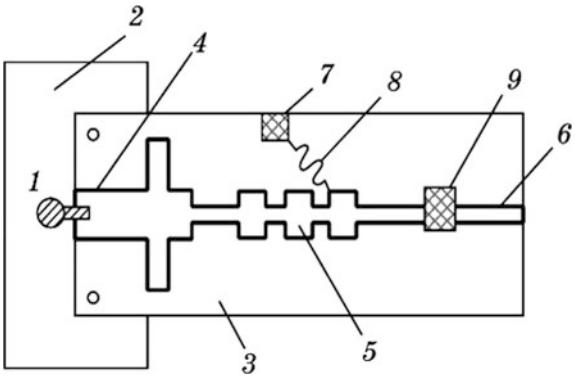


Fig. 13.27 Microstrip design of an IMPATT diode oscillator operating in the TRAPATT mode

of approximately $\lambda/2$, with a gold wire 25 μm in diameter, where λ is the wavelength of the generated oscillations. Thus, the segment is a resonator tuned to the first, second and higher harmonics of the operating frequency. A low-pass filter 5, with a cutoff frequency lying between the frequencies of the first and second harmonics, connects the resonator 4 to the output of the generator 6. Thus, the filter only allows fundamental frequency oscillations to the output, providing a resonator with high Q factor at harmonics and correspondingly increasing their amplitude. The input filter impedance is selected to provide phase relations between the amplitudes of the harmonic voltages necessary to form a voltage pulse on the diode.

Capacity 7, inductance 8 and a mounted separation capacitor 9 serve to supply voltage to the diode. Such an oscillator generates a power of 40 W with an efficiency of more than 30% at a frequency of 2–3 GHz when fed by pulses of 0.5 μs with a duty factor of 1000. The current density through the diode is 8000 A/cm^2.

Instead of a low-pass filter in the generator, a bandpass filter tuned to the second harmonic can be used. In this case, the oscillation frequency will be two times larger (4–6 GHz), since the filter allocates the second harmonic of the fundamental frequency, however, the power and efficiency of the oscillator decrease.

13.5 Injection-and-Transit-Time Diodes

Consider a semiconductor with a n^+–p–n^+ structure shown in Fig. 13.28. At a specified voltage polarity, the left (emitter) junction is shifted in the forward direction, and the right (collector) junction is in backward. There will be two

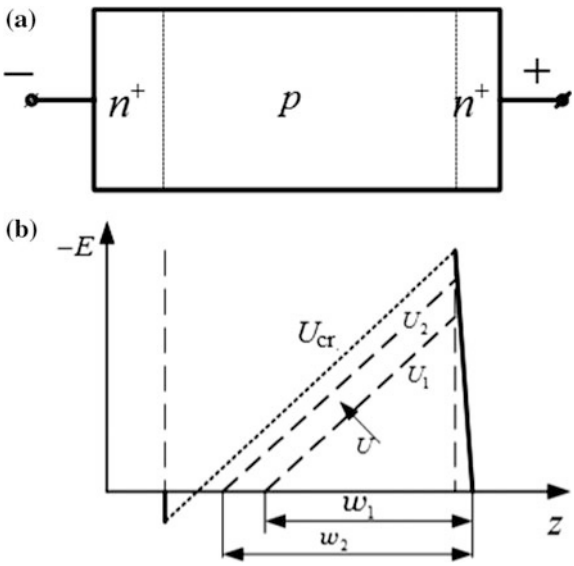


Fig. 13.28 The structure of the injection-and-transit-time diode (**a**) and the distribution of the electric field strength (**b**) within

depletion layers in the structure at low voltages U: near the left electrode due to contact potential difference φ_k, and at the right electrode due to φ_k and U. Increasing voltage U on the collector leads to an expansion of the depletion region in the p-layer at the right electrode. At voltage U_1, the width of the depletion layer is w_1, and for $U_2 > U_1$ the width of the layer increases, i.e. $w_2 > w_1$. Such a change in the width of the depletion layer is shown in Fig. 13.28. At some voltage U_{cr} a complete depletion of the p-layer will occur, and the depletion regions of the left and right electrodes unite. This situation is called the *p-layer puncture*. A further increase in the voltage will lead to a decrease in the barrier for the charges at the left electrode and to an injection of carriers from the emitter. If, with bias voltage U_{cr}, a small alternating voltage is applied, an alternating emitted current will emerge, which will lag behind the voltage, as shown in Sect. 13.3. Unlike IMPATT diodes, the injection delay angle will be smaller, but according to expression (13.17) at some transit angles, negative dynamic resistance can be obtained.

Such diodes have low output power, since they operate at small alternating voltage amplitudes. On the other hand, the absence of avalanche stochastic processes causes lower noise in comparison with IMPATT diodes. Such devices are called BARITT diodes (BARrier Injection Transit Time). The maximum negative conductivity of these diodes is observed at transit angles Θ close to $3\pi/2$. At the present time, these diodes are not in production (Sect. 13.6).

13.6 Transferred Electron Devices

13.6.1 The Gunn Effect. The Running High-Field Domain

History of transferred elecron devices began in 1961 when B.K. Ridley and T.B. Watkins pubished a paper where they predicted existance of negative differential resistance in some semiconductors. C. Hilsum showd in 1962 that GaAs and its alloys have neccessary zone structure for this effect. J.B. Gunn in 1963 observed current instabilities in GaAs and InP samples under high electric field. H. Kroemer explained these instabilities by effect Ridley and Wotkins theory.

The general principles of amplification and generation of oscillations in the semiconductor structures considered in Sect. 13.2, indicate the possibility of obtaining a negative dynamic resistance in semiconductors, with a region of negative differential mobility on the field-velocity characteristic $\mu_d < 0$. *A*IIIBV semiconductors such as GaAs, InP, GaN, etc. possess this property. The conduction band of these materials consists of several valleys, the electrons in which have different effective masses (and, correspondingly, different mobilities). This situation is shown schematically in Fig. 13.29. It shows the dependence of energy W on wave number k along the crystallographic axis [111] for GaAs. It can be considered as the dispersion characteristic of the de Broglie wave of an electron as it passes through a crystal lattice. The curvature of the level is proportional to the group

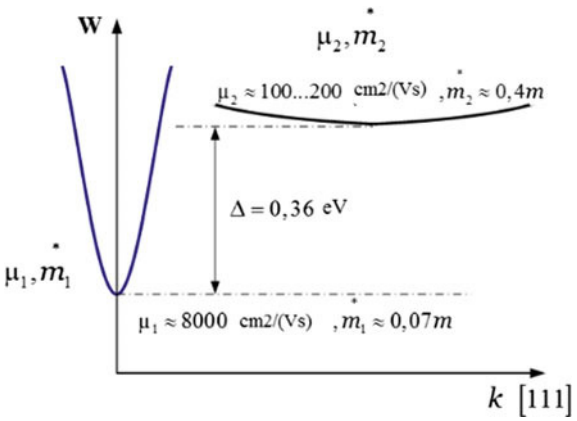


Fig. 13.29 Energy graph of the conduction band for GaAs

velocity of the wave, that is, the velocity of the electron. The energy gap Δ between the valleys is 0.36 eV. The difference in the population of the valleys n_2/n_1 is determined by the energy of the electrons kT_e and this gap:

$$n_2/n_1 = R\exp(-\Delta/kT_e), \tag{13.58}$$

where R is the ratio of the state density in the valleys (for GaAs it is 50). If an external "heating" field E_0 is applied, then the electron temperature rises, according to the equation of conservation of energy flux (2.4.8), which in the simplest case takes the form

$$qvE \approx \frac{3}{2}\frac{k(T_e - T_0)}{\tau_\varepsilon(T_e)}. \tag{13.59}$$

In expression (13.59), velocity v corresponds to the average carrier velocity in two valleys, that is,

$$v = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}E. \tag{13.60}$$

Using expressions (13.58)–(13.60), we can analyze and calculate the field velocity characteristic, given the value of the energy relaxation time $\tau_\varepsilon(T_e)$. An increase in the field strength causes an increase in the electron temperature T_e. This leads to an increase in the fraction of electrons passing into the upper valley and a decrease in the average speed of carriers. This decrease will be especially sharp in fields above a certain critical E_{th}. It follows from the calculations that the critical (threshold) field for GaAs is E_{th} = 3.4 kV/cm. With a further increase in the field, the velocity will tend to the value of $\mu_2 E$, since there will be many more electrons in the upper valley than in the lower one. Figure 13.30 shows the calculated curves for

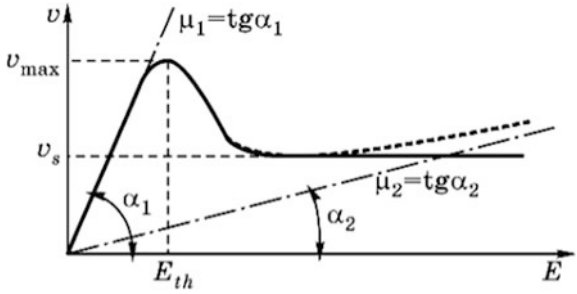


Fig. 13.30 Field velocity characteristic

the two variants: without taking into account the dependence of mobility on electron temperature $\mu_2(T_e)$ (dashed line) and taking it into account (solid line). In strong fields, the theory predicts a velocity increase (dotted curve), but in reality its saturation (solid curve) is observed. Negative differential mobility is observed in the range of 4–10 kV/cm. The ratio of maximum velocity $v_{\max}$ to saturation velocity v_s amounts to 1.5–2. The processes of inter-valley transfer of charge carriers not only significantly affects the shape of the field velocity characteristic, but also field dependence of another kinetic coefficient—diffusion coefficient D_e. Near the critical field, this dependence has a pronounced maximum, as shown in the experimental dependence in Fig. 13.31. Knowing function $D_e(T_e)$ is very important in determining the static distribution of the field and the operating mode of the device.

B.K. Ridley and T.B. Watkins predicted the described effect in 1961. J.B. Gunn (Bell laboratories) discovered the effect in 1962. Very soon, diodes based on this effect were realized. They were called *Gunn diodes* (GD) or *transferred electron diodes* (TED).

The considered mechanism for intervalley carrier transition makes it possible to explain the so-called Gunn effect. This effect consists in the appearance of periodic

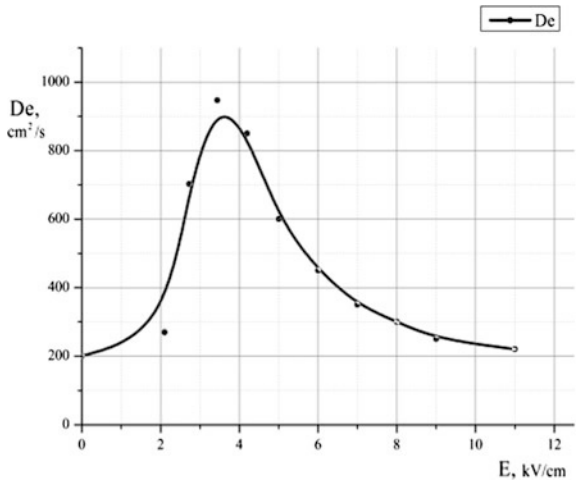


Fig. 13.31 The Experimental dependence of the diffusion coefficient on the field for GaAs

current pulses in the semiconductor structure of pulses with negative differential mobility. Conditions for the occurrence of pulses are as follows:

- the applied voltage must exceed a critical value $U_{th} \approx E_{th}L$;
- doping level N_d and the length of the active part L must exceed a certain critical value $N_dL > (N_dL)_{cr} = 2 \times 10^{11}$ cm^{-3}.

The structure of the diode is shown in Fig. 13.32, and time dependence of the diode current is shown in Fig. 13.33. The pulse repetition period is well approximated by the expression: $T = L/v_s$. Thus, by changing the length of the device L, it is easy to obtain the desired oscillation frequency. For example, a frequency of 10 GHz is generated with device length of 10 μm. Let us qualitatively analyze the described process, using the field-velocity characteristic (Fig. 13.30). Suppose that the doping level is homogeneous along the length of the sample. This is necessary in order to obtain a homogeneous static electric field in the range of applied voltages from 0 to U_{th}. At these voltage values, the current is proportional to the voltage according to expression

$$I = eN_dv(E)S_d = eN_dv(U_0/L)S_d. \tag{13.61}$$

It is obvious that the shape of the initial volt-ampere characteristic section repeats the shape of the corresponding section of the field-velocity characteristic.

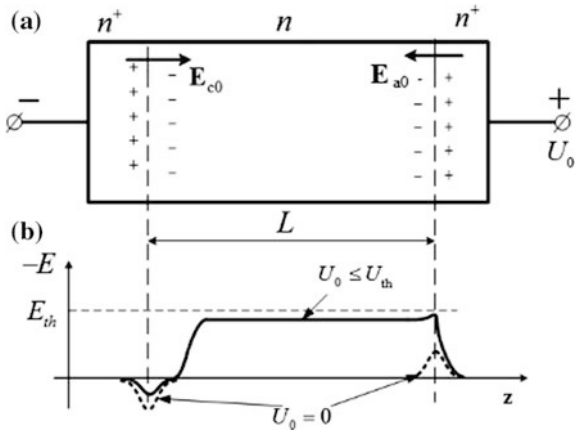


Fig. 13.32 Structure of a Gunn diode **a** and field distribution **b** at a voltage less than critical

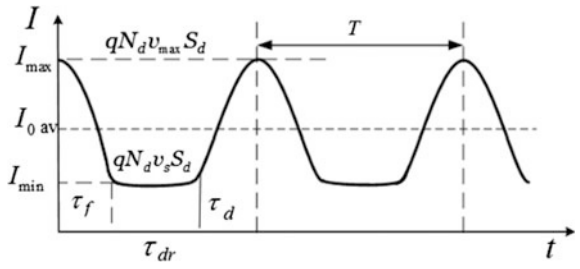


Fig. 13.33 Current waveform in a Gunn diode in the domain mode

The maximum value of the current in this mode will be equal to $I_{max} = eN_d v_{max} S_d$ (Fig. 13.33).

An increase in voltage above U_{th} leads to an increase in the field strength $E > E_{th}$. It can be assumed that the field will be at maximum value in a narrow region containing a small doping inhomogeneity. This assumption is based on the fact that in the cathode region there is an ohmic contact Me–n^+ to the n-region. In its vicinity, the level of doping varies greatly. For certainty, suppose that this inhomogeneity ("seed") is in the cathode part of the structure.

The field-velocity characteristic and field distribution at coordinate z are combined on one graph (Fig. 13.34). This allows us to determine the limits of changing of physical parameters. As soon as the field in some area exceeds the critical value, the velocity of the carriers in this region becomes less than the velocity of carriers to the right and left of it, leading to an increase in electron concentration on the left and electrons shifting to the right relative to the coordinate of the maximum field. As a result, a double charge layer arises in the inhomogeneity region, called the *high-field domain* (Fig. 13.35). In the center of this domain, the intensity of the electric field increases. We note an important circumstance: the escape of electrons from the domain front violates the quasi-electroneutrality condition: the positive charge of the donors is not compensated by the negative electron charge. The density of the positive charge in the front of the domain will be qN_d. Any field increase in the domain region leads to a further drop in electron velocity in the central region of the domain, and hence to an increase in the accumulation of charges on the left and the departure of electrons from the left.

As a result, the domain field increases. The field outside the domain will drop, since the common integral of the field respective to the coordinate remains constant and equal to the applied voltage U_0. This will lead to a decrease in the velocity of electrons to the right and left of the domain, and consequently to a decrease in the current through the diode. The domain grows until the electron velocities in all parts of the domain are equal. Furthermore, the form of the domain does not change.

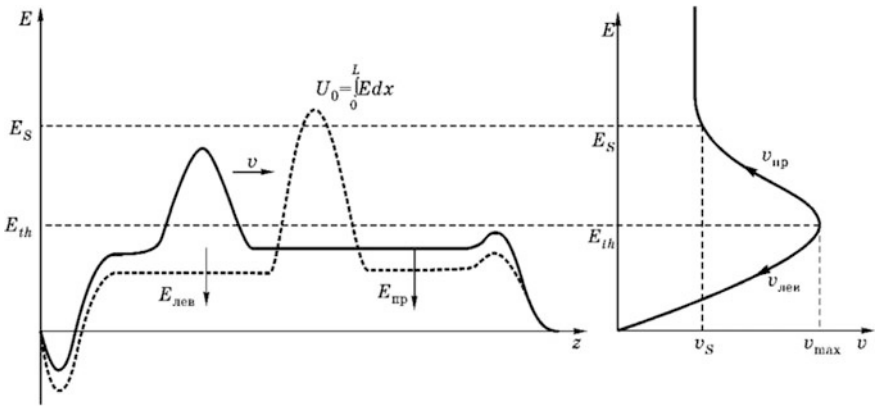


Fig. 13.34 The process of forming a running domain

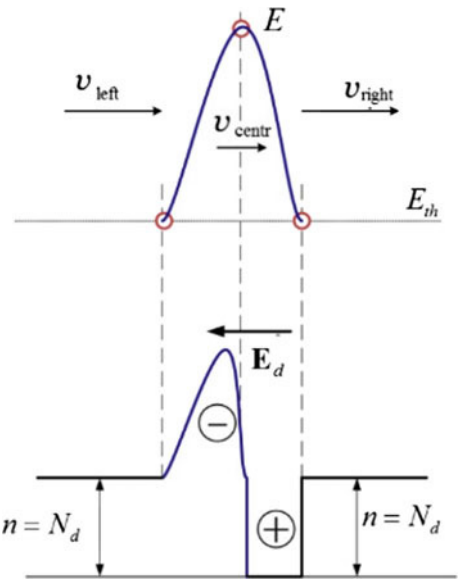


Fig. 13.35 Distribution of field, velocity and charge in domain region

After reaching the anode, the domain disappears, the field in the sample increases and the whole process and situation repeats. This causes periodic oscillations of the current (Fig. 13.33). It should be noted that diffusion significantly affects the process of charge entry into the anode. Significant gradients in the distribution of charges in the domain increase the role of diffusion fluxes, which for certain parameters of the diode structure N_d, L lead to the emergence of dynamic equilibrium, when the growth of the domain ceases and it "sticks" to the anode. There are no periodic pulsations of current in this case. This mode is known as the anode static domain mode.

13.6.2 Distribution of Static Field in the Gunn Diode

In order to thoroughly consider the described processes, it is necessary to solve the system of particles, momentum and energy flux conservation equations for given values of momentum and energy relaxation time constants, together with the Poisson equation (see Chap. 2). Let us analyze the general results of such a simulation, taking into account the nonlinear dependencies of the mobility $\mu(E)$ and the diffusion coefficient $D(E)$ on the field. First of all, we note that the joint solution of the Poisson equation and the equation for the current density yields three typical solutions for the field distribution in the structure. Two of them are stable, and the third is unstable. First—static distribution with the maximum field at the anode (anodic static domain), second—near the cathode (cathodic static domain) and third—the running domain. These distributions are shown in Fig. 13.36. The specific form of field distribution

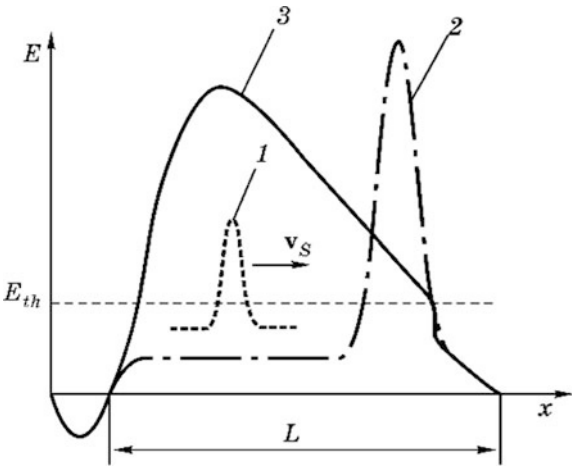


Fig. 13.36 Field distribution for a running domain mode (1), anodic static domain mode (2) and cathodic static domain mode (3)

depends on the boundary conditions at the cathode and anode, the length of the sample, the doping profile and the applied voltage. It is also necessary to take into account changes in the sample temperature due to the flow of current through it. These parameters affect not only the process of domain formation, but also external characteristics, in particular, the volt-ampere characteristics of the diode.

Figure 13.37 shows the possible static volt-ampere characteristics of a Gunn diode operating in different modes. The difference in dependencies 3 and 4 is determined by the nature of static field distribution in the device: 3 corresponds to the anode static domain, and 4 corresponds to the cathode static domain. The mode of a running domain (or domains) is most likely near the threshold voltage. This is manifested in stochastic oscillations of the current. On the graph, this mode is marked with number 5. The initial section of all characteristics is the same and its form is determined by the initial branch of the field velocity characteristics, as follows from expression (13.61). Mode 1 with a running domain is manifested in a characteristic current step at a voltage greater than the critical value $U \geq U_{\text{th}}$.

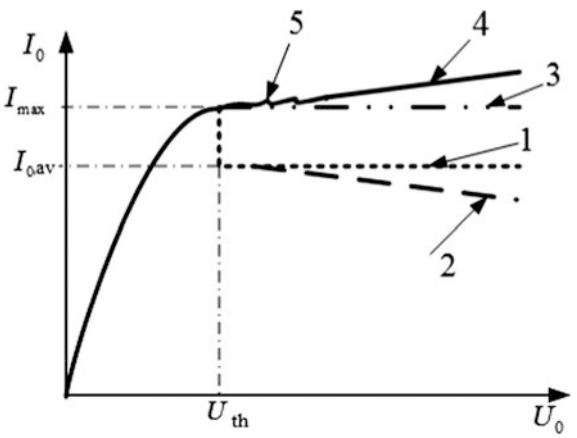


Fig. 13.37 Volt-ampere characteristics of the Gunn diode in different modes: 1—with a running domain; 2—considering structure heating; 3, 4—static domain in the pulse mode; 5—unstable region

Domain formation leads to a decrease in the field outside the domain, and hence to a drop in the current. With a further increase in voltage, the current remains constant, since the field outside the domain remains at the level of the saturation field E_s (Fig. 13.34). Such a situation is observed in the pulsed operating mode, when the heating of the diode can be ignored. In the continuous mode, structure heating leads to a decrease in current due to a decrease in the saturation rate v_s. This leads to a falling section on the volt-ampere characteristic (line 2). However, this section with negative differential resistance cannot be used in the microwave band, since it is caused by slow heating processes in the structure.

When the theory and practice of devices with intervalley transfer was first developed, attention was mainly placed on the domain mode. However, obtaining stable operating modes for such devices caused great technological difficulties. In particular, high process sensitivity to the parameters of the "seed" and its location in the cathode region was revealed. In addition, the rate of domain growth is significantly affected by the external circuit resistance, which limits the current. These effects significantly complicated the domain mode application. As a result, structures with a stable static field distribution, i.e., with cathodic and anodic static domains, have been used in microwave generators and amplifiers. On an alternating signal, such structures exhibit negative dynamic resistance in a certain frequency range. To calculate the equivalent diode resistance, depending on the material properties and the mode parameters, the general expressions obtained in Sect. 13.2, and particularly expression (13.8), are used. The diode in this calculation has a layered structure, similar to that shown in Fig. 13.9.

Gunn diodes and IMPATT diodes are used as amplifiers and generators. An important feature of Gunn diodes in comparison with IMPATT diodes is a low level of phase noise. This is due to the fact that the diode is made of a homogeneous material. Its principle of operation is based on processes occurring within, rather than on the surface or any boundary. There are no stochastic processes of avalanche formation, and there is no shot noise, since there are no barriers, etc. In general, this leads to a low level of flicker noise. Using such devices in the generator circuits provides low phase noise. However, the output power level of Gunn diode generators is lower than that of the IMPATT diode generator. The modern level of design makes it possible to obtain an output power of about 1 mW at a frequency of 100 GHz. The use of gallium nitride as a material provides a significant increase in output power (by an order of magnitude) and maximum working frequency.

At present, GDs are mostly used in heterodynes and low-power tunable oscillators.

13.7 Tunnel Diode

13.7.1 Structure and Operating Principle

Historically, the first semiconductor device with negative dynamic resistance to be invented was the tunnel diode (TD). In 1958, Japanese researcher L. Esaki discovered on a static volt-ampere characteristic of a germanium diode, a region in

which the current decreases with increasing applied voltage. For the development of the theory of this phenomenon and its experimental detection, Esaki was awarded the Nobel Prize in Physics in 1973.

Figure 13.38 shows the voltage-current characteristic for a diode structure p^{++}–n^{++} with doping levels of the donor and acceptor regions of 10^{19}–10^{20} cm^{-3}. The negative slope of the curve is explained by the quantum tunneling of electrons through a narrow potential barrier. At the mentioned doping levels, the thickness of the depletion layer (see formula 11.3.3) is 10–50 nm. At such distances, the probability of tunneling through the barrier increases. Figure 13.39 shows the band diagram of the contact in question, p^{++}–n^{++}. The dashed line shows the Fermi level F. Usually the degeneracy level (the "depth" of the Fermi level in the corresponding band) qU_n and qU_p is in the order of several kT units. The band diagram (Fig. 13.39) corresponds to the zero voltage on the structure and to the absence of current in the diode. When the forward bias is applied to the p^{++}-region, there is a shift in the levels at which electrons from the conduction band of the n-semiconductor are at the same energy level with empty states in the p-semiconductor. In this case, a tunnel current arises that increases with increasing overlapping of the levels. The dependence of this current on the voltage is shown by the small dotted line in Fig. 13.38.

The current reaches a maximum value I_{max} when the bottom of the conductivity band of the n-region and the Fermi level of the p-region coincide. A further increase in the voltage leads to a decrease in the band overlap and a decrease in the current to the value I_{min}. Hence, negative differential resistance is observed. If the voltage is increased further, then the "ordinary" diffusion over barrier current, marked on the volt-ampere characteristic by a large dotted line, will pass through the structure. These two trends in the current change determine the resulting voltage-current characteristic, shown in Fig. 13.38 by a solid line. With a decrease in the doping level, the "hump" with forward bias becomes invisible, and the current increase

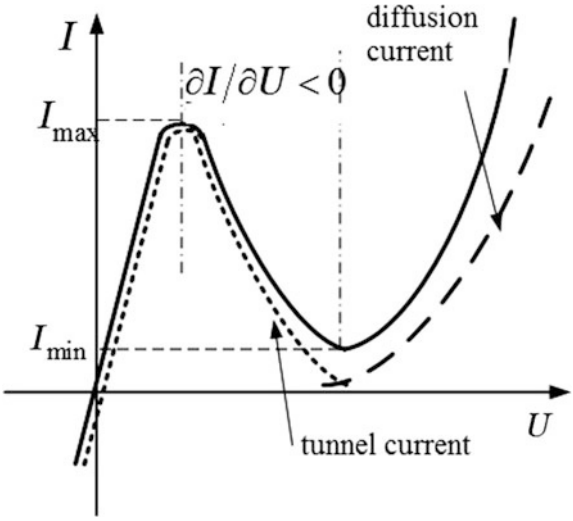


Fig. 13.38 Volt-ampere characteristic of a tunnel diode

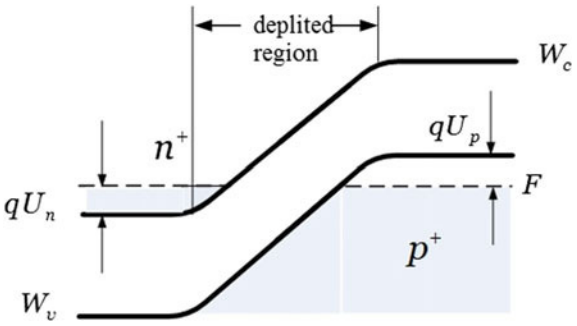


Fig. 13.39 Band structure of a tunnel diode

with reverse voltage is preserved. A device with this characteristic is called a *reversed tunnel diode*. It does not have an NDR region. It is used in detector and mixing devices.

The type of volt-ampere characteristic will not change at high frequencies, since the tunneling process is very quick (10^{-13} s). Such a property allows implementing devices in the microwave band, including the terahertz frequency range. The presence of negative differential resistance (NDR) allows the use of tunnel diodes in generators and reflective amplifiers.

With a lack of real competition in the 1960s from other devices, the TD was the main active component in the construction of micro miniature receivers of microwave radio-electronic systems. The optimism of the developers regarding the application of TDs was not only associated with the simplicity of the device design, but also with the establishment of the manufacturing technology, which ensured the high reliability of the device and its low cost. However, the difficulty of matching the diode to the output line, the low level of output power and competition from IMPATT diodes, Gunn diodes, and subsequently transistors, limited the scope of the TD. Currently, such diodes are used to create low-power generators in the terahertz frequency range.

13.7.2 Equivalent Circuit. Features of Use in the Microwave Band

It is convenient to analyze the high-frequency characteristics of TD using its equivalent circuit. It is similar to that of a detector diode (Fig. 12.4) and is shown in Fig. 13.40.

The constant bias voltage provides a mode with NDR: $R = (\partial I/\partial U)^{-1} < 0$. The structure itself is represented by a parallel connection of the negative resistance R and the barrier capacitance C_j. Parasitic parameters of the structure are presented by series contacts resistor R_s and inductance L_s. The velocity of carrier tunneling across the barrier, through charge (discharge) of reactive elements C_j, L_s determines

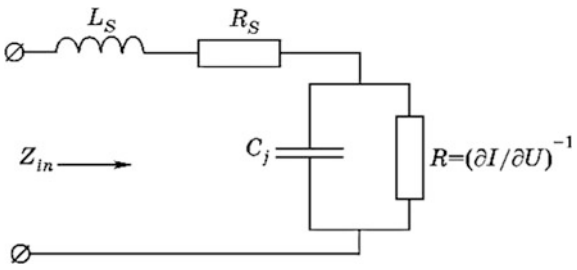


Fig. 13.40 Equivalent TD circuit

the device's speed of operation. For real structures based on GaAs, the maximum oscillation frequency in such devices reaches 200–300 GHz.

Let us analyze the operation of TDs in microwave schemes, using a voltage-current characteristic and an equivalent circuit. The main parameter of the diode is the ratio I_{max}/I_{min} and the voltage range in which NDR is observed. In existing diodes, the current ratio can be 3–4, and the NDR voltage range can be 0.2–0.3 V. The value I_{max} affects the rate of capacitance C_j charging and the saturation power. In real structures, it is possible to obtain a current density of 10^4 A/cm^2, which provides 30 μW saturation power in the centimeter waveband. Low saturation power and a narrow range of input voltages dictate a small dynamic range of amplifiers built on the TD.

As an example, Fig. 13.41 shows a reflective amplifier circuit on a centimeter band TD. In the circuit, circulator C decouple input and output of the amplifier. The matching transformer MTr provides the resonance conditions at the operating frequency and allows the changing of the HF voltage on the tunnel diode. The bias circuit with a filter system based on lumped (L_{1f}, L_{2f}, C_{2f}, R_1, R_2) and distributed

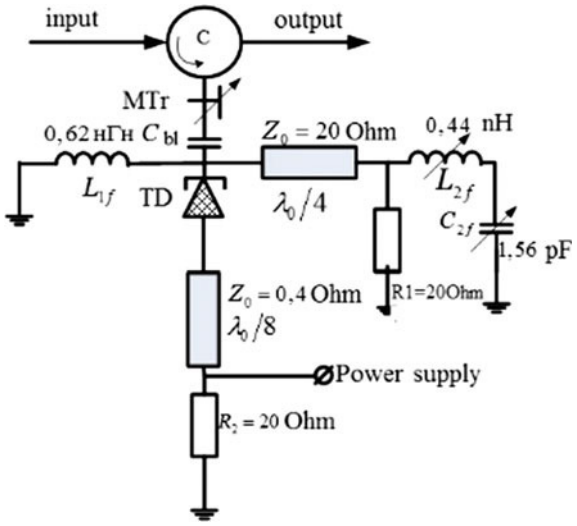


Fig. 13.41 Reflective amplifier circuit on a TD

elements ($\lambda_0/8$ and $\lambda_0/4$) provides the necessary DC bias voltage and limits the loss of HF signal.

Despite the fact that a lot of attention was given by the TD developers, they did not find proper recognition. Theoretically, they promised to become low-noise amplifiers, simple and cheap generators. However, due to the low saturation power, those optimistic forecasts were not realized. This disadvantage can be tolerated in many applications, however the TD has had many competitors, such as parametric amplifiers and multipliers on varactor diodes, transistor amplifiers and generators.

The area where practical interest remains in the TD, is the creation of terahertz-band low-power generators.

13.7.3 Resonance Tunnel Diode (RTD)

With the development of molecular beam epitaxy technology, it was possible to fabricate nanoscale heterostructures with a regulated thickness and height of energy barriers. If we fabricate a heterostructure with two thin 2–10 nm barriers separated by a narrow region of width 5–10 nm, then NDR is observed in it at certain bias voltages, as in a tunnel diode. The mentioned semiconductor structure was called a *resonance tunnel diode* (RTD).

Consider the mechanism of NDR formation in such a structure. Figure 13.42 schematically shows the conduction band bottom profile of the structure with two barriers separated by a region of length L. These barriers form a quantum "well" between them. Tunneling of electrons in such a structure can be regarded as the passage of the electron wave function Ψ through the resonating element in the form of two barriers. In the quantum "well" the allowed energy states of the electrons correspond to the resonance of the wave function, that is,

$$2L = (n - \varphi_b/2\pi)\lambda, \tag{13.62}$$

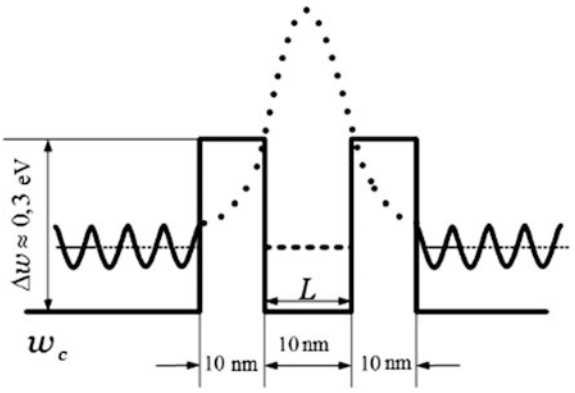


Fig. 13.42 Double barrier and quantum well of a resonance tunnel diode

where n is a whole number, 1, 2, 3 …, L is the length of the well; φ_b is the effective reflection phase of the wave function from the barrier; and λ is the wavelength of the electron wave function. Tunneling of electrons through such a structure is likely if their energy outside the barriers corresponds to the energy of the allowed levels between the barriers.

In such a structure, the voltage-current characteristic has a falling section, like the current-voltage characteristic of a tunnel diode. We can explain this phenomenon by considering the representation of different diode bias modes and corresponding shifts in the conduction band, as shown in Fig. 13.43. With zero bias on the diode ($U_b = 0$) there is no current, which corresponds to the initial point of the volt-ampere characteristic. As bias voltage increases, the energy of the electrons increases, but it is insufficient to overcome the barrier. However, it partially corresponds to the allowed levels in the potential well. If these levels coincide $U_{\text{bias}} = U_{I\text{max}}$, tunnel current $I_{\max}$ passes through the structure. Further bias increase till the $U_{\text{bias}} = U_{I\text{min}}$ point leads to a fall in tunnel current to $I_{\min}$ due to electron energy "discrepancies" and allowed levels in the "well". In the ideal case, the current must fall to zero. However, the presence of the above-barrier current component and the non-ideality of the structure lead to finite values of $I_{\min}$. An increase in the bias voltage above the point $U_{\text{bias}} = U_{I\text{min}}$ leads to an increase in the role of the above-barrier emission and current increases with rising voltage.

An important advantage of such a structure over an ordinary tunnel diode is the ability to purposefully change the structure's parameters to obtain the required diode properties. In particular, by varying the height of the barriers and the thickness of the quantum "well", it is possible to increase the ratio $I_{\max}/I_{\min}$ and lower resistance.

As an example, Fig. 13.44a shows the energy profile of a heterostructure with NDR. The contact regions within are made of indium arsenide and the inner regions are made of gallium and aluminum antimonide. Similar structures are also made with solid solutions of AlGaAs.

Having completed ohmic contacts in this structure, we obtain a resonance tunnel diode shown in Fig. 13.44b. It should be noted that the left (emitter) part has a significantly higher level of doping than the right one—the collector. Figure 13.44c shows the change in energy level of the bottom of the conduction band in the NDR region.

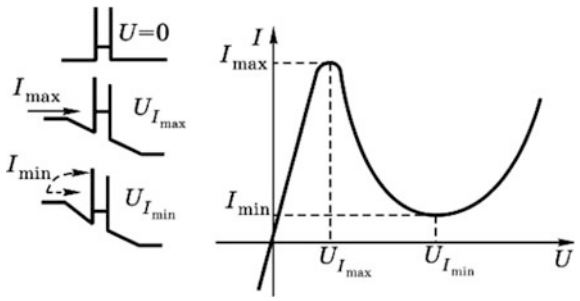


Fig. 13.43 Obtaining NDR in a structure with two barriers

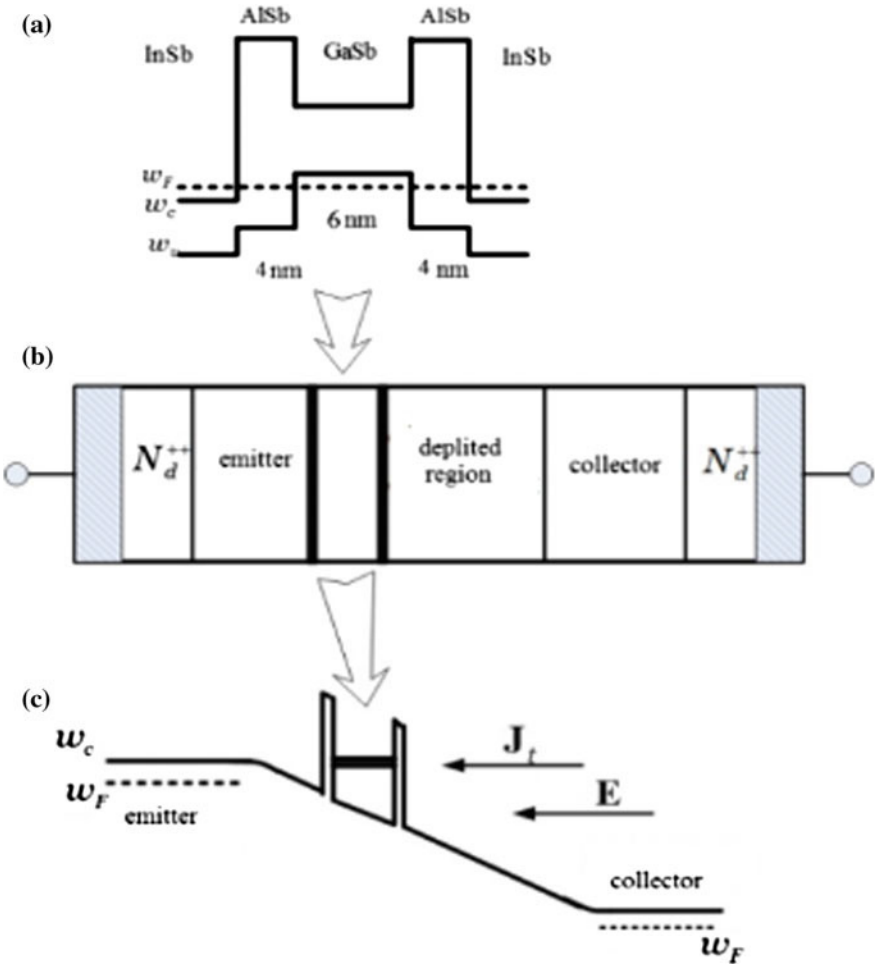


Fig. 13.44 Energy graph (**a**), diode structure (**b**), and operation circuit (**c**)

The experimental static volt-ampere characteristic of such a structure with an active-region diameter of 50 μm and a current density 10^3 A/cm^2 is shown in Fig. 13.45. The achieved ratio of maximum current to minimum is 3.8, and the NDR region exists in the bias voltage range of 0.35–0.55 V. Such a diode makes it possible to obtain a power of 20 μW at a frequency of 443 GHz. The construction of a microwave generator with an output power of 10 μW at a frequency of 1.3 THz is shown in Fig. 13.46 (*a*—circuit, *b*—appearance).

The main advantage of such devices is the simplicity of their production and the low cost of radio-electronic systems created on their basis, in particular, high-speed wireless networks for general use.

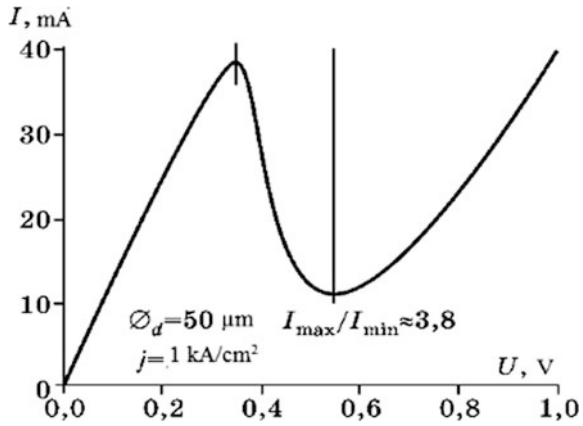


Fig. 13.45 Real volt-ampere characteristic of a resonance tunnel diode

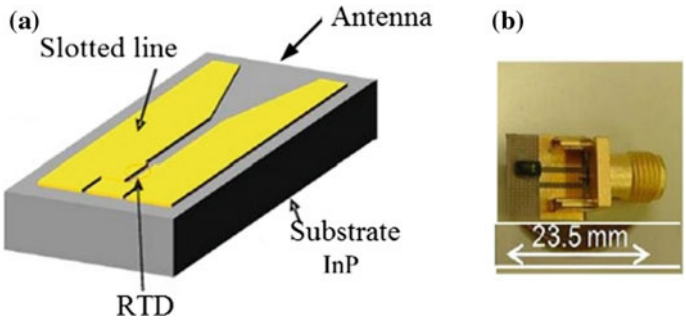


Fig. 13.46 Generator circuit on the slotted line (**a**), generator structure (**b**)

Advancement Questions

1. Which devices belong to the class of devices with negative dynamic resistance? What is the difference between differential and dynamic resistance?
2. Analyze the methods of obtaining a negative dynamic resistance for structures with different values of differential Maxwellian relaxation frequency, at different transit angles and with different boundary conditions.
3. How do charge carriers transmit energy to the microwave field if their drift velocity is constant?
4. Draw constructive options to obtain a negative dynamic resistance in semiconductor structures.
5. How does the IMPATT diode structure provide a location for the avalanche zone?
6. Is it possible to "force" not only electrons, but also holes to work in IMPATT diodes?

7. How do you calculate low-signal resistance in IMPATT diodes?
8. How does the space charge of a bunch affect the distribution of a static field in a nonlinear mode of operation?
9. What is the TRAPATT mode in the IMPATT diode? What are the requirements for load?
10. What is the difference in structure, principle of operation and parameters between the IMPATT diode and the BARITT diode?
11. Give an example of building a tunable oscillator on an IMPATT diode.
12. What is the Gunn effect?
13. What modes of field distribution exist in Gunn diodes? Describe the running domain mode.
14. Draw and explain the static volt-ampere characteristic of a Gunn diode.
15. How to calculate the low-signal impedance of a Gunn diode?
16. Draw an amplifier circuit on a GD.
17. What is the difference in the structure and operating principle of a conventional and a resonance TD?
18. Is there a difference between the differential and dynamic resistances of a TD?
19. What are the frequency range limits of a RTD?
20. Compare the limit parameters of devices with negative dynamic resistance in terms of frequency-power parameters.

Chapter 14
Microwave Transistors

14.1 Field Effect Transistors

14.1.1 Structure of the Schottky Field Effect Transistor

A Schottky field effect transistor (MESFET) is a unique device, the idea of which was expressed in 1920, i.e., much earlier than the bipolar device, and it was manufactured for the first time much later than the bipolar device. The technology of that time did not allow the creation of a contact with a low density of surface states. These states shielded the gate from the channel, greatly reducing the controllability of the transistor. The first field-effect transistor only appeared in 1953. GaAs was used as the main material. The high carrier, low-field mobility of this material made it possible to obtain small parasitic resistances. Decreasing the length of the gate significantly reduced the drift time in the device. This led to a significant increase in speed in comparison with silicon bipolar devices on Si.

A MESFET contains a semiconductor body, an emitter of charged particles (*source*) and a collector (*drain*) (see Fig. 14.1). Metal contacts placed on the semiconductor surface form a Schottky barriers with the body. The height h of the depletion regions under the gate depends on the voltage gate-source. Carriers can move from source to drain in the region between depletion regions (*channel*). The width of this region $2a$ and consequently its resistance depends on h. By changing gate voltage, we can change channel resistance and drain current.

MESFET can be considered as a conventional vacuum triode, working in a semiconductor medium. The *source* in such a device plays the role of a cathode, the *gate* plays the role of a grid and the *drain* plays the role of an anode. Figure 14.1 shows a longitudinal section of a symmetrical field-effect transistor implemented on an n-type semiconductor.

Comparing the grid in the triode and the gate in the field effect transistor, it should be noted that unlike the triode, motion under the gate takes a significant part of the total time of carrier motion in the active region of the transistor.

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6_14

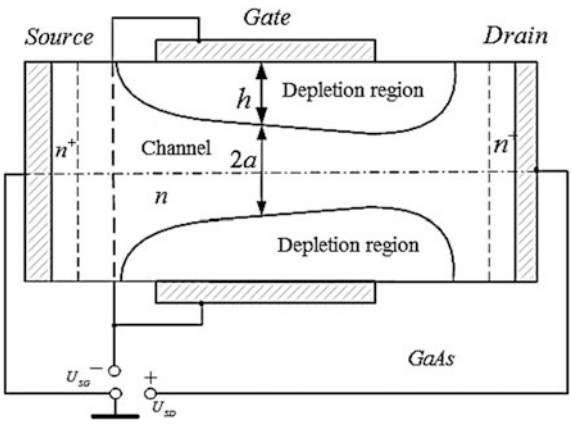


Fig. 14.1 Symmetrical construction of a MESFET

In the vacuum triode, transit time via the grid itself is negligibly small compared to the transit times in gaps: cathode-grid and grid-anode. At present, semiconductor devices with a gate length of <0.1 μm are already being created. The transport of carriers in such devices becomes practically ballistic (collisionless), which emphasizes the generality of the physical phenomena in the MESFET and triode, and not just their analogy.

On the other hand, MESFET can be compared to a resistor in which resistance is controlled by an electric field with a direction normal to the direction of current flow. For transistors with a gate length of more than 0.5–1 μm, such an analogy is closer in nature than a comparison with a vacuum triode.

The noticeable length of the gate leads to shielding of the source from the drain potential, which causes significant differences in characteristics compared with a vacuum triode. Let's consider the simplest example of these differences in a symmetric MESFET in Fig. 14.1. Let us make some important remarks on the design of the device and the features of current transfer:

- Source and drain have ohmic contacts. Hence, they are indistinguishable from the point of view of obtaining the electron flux. In such a design, unlike in a triode, when the polarity of the applied voltages is reversed, the device will retain the properties inherent in MESFET;
- At the metal-semiconductor boundary, a depletion layer (DL) is formed. The width of this layer is determined by the potential difference between the current channel (CC) and the gate, according to the expression (11.7). The flow of current causes a change in the potential along the length of the channel, which leads to a change in the width of the depletion region.

The transistor shown in Fig. 14.1 has not been produced due to technological difficulties in its manufacture and integration into the case. Figure 14.2 shows the design of a planar MESFET, widely used in practice. It introduces two additional layers: a semi-insulating substrate and a buffer layer. In the substrate, current practically does not flow due to its high resistance. It limits the current channel from

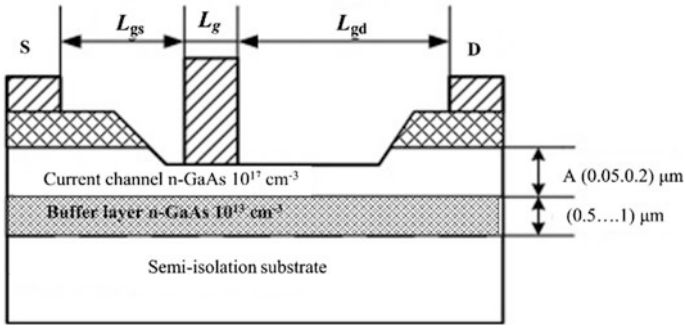


Fig. 14.2 Cross section of a typical microwave transistor

below. Letter *A* denotes the height of the current channel, and letter *Z* denotes gate width. To exclude the migration of impure atoms from the substrate to the current channel, a thin buffer layer of low-doped GaAs is introduced into the structure. Typical doping levels and dimensions are shown in the figure. Figure 14.3a shows the topology of the transistor, and Fig. 14.3b shows a powerful transistor that consists of several low-power transistors connected in parallel. On the top of the photo there are four drain pads. Wide light platforms at the bottom are source contacts, and small elliptical ones are gate contacts. The periphery of the photo shows dimensions in μm. Note that even when a large number of transistors are connected (Fig. 14.3b), the total width of the gates is only 1.13 mm. Thin gold conductors connect contact areas of the device to the external pins of a special case. It allows easy plugging of the transistor into a microstrip line. Such a design increases the parasitic reactive parameters of the device as a whole and narrows its frequency range of operation. A limited set of similar case structures is used as shown in Fig. 14.4, with a millimeter scale in the corner of the image.

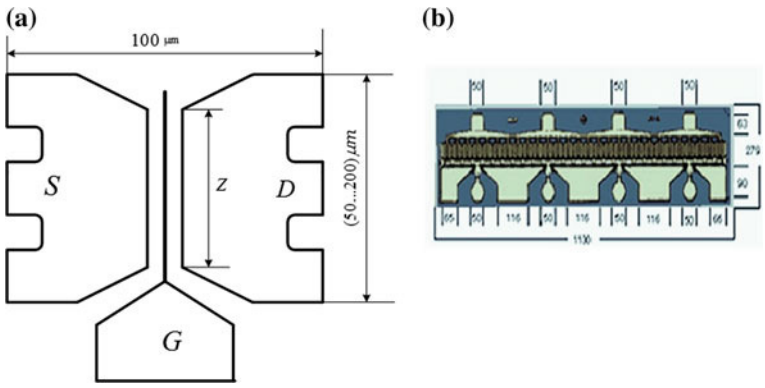


Fig. 14.3 MESFET topology: **a** low-power transistor; **b** powerful transistor

Fig. 14.4 Centimeter band transistor cases: the unit of the attached rulers is 1 mm (manufacturer —«Mitsubishi»)

In microwave integrated circuits, the problem of plugging a transistor in a circuit is simplified; the parasitic parameters of contacts are reduced. This extends the frequency range of circuit operation.

14.1.2 Static Characteristics of Schottky Field Effect Transistors

The dependence of the drain current I_d on source-gate voltage U_{gs} at constant drain voltage U_{ds}, i.e. $I_d = f(U_{gs})$, is called the *input characteristic*. The dependence of the drain current I_d on drain-source voltage U_{ds} at constant gate voltage U_{gs}, i.e. $I_d = f(U_{ds})$, is called the *output characteristic*. When constructing these dependencies, we use the similarity of this device to the resistor in which resistance (cross section) is changed by an electric field applied to a gate. To calculate the dependence, we use formula (11.7) to calculate the thickness of the Schottky barrier depletion layer h:

$$h(x) = \sqrt{\frac{2\varepsilon\Delta U}{eN_d}} = \sqrt{\frac{2\varepsilon(U_{ch}(x) + \varphi_k - U_g)}{eN_d}}, \tag{14.1}$$

where $U_{ch}(x)$ is the potential at point x of the current channel; and φ_k is the contact potential difference. In formula (14.1), there is no term associated with the thermal expansion of the channel. This means that the process of heating up the carriers due to the field is omitted from this consideration. In this approximation, the boundary between the depletion layer and the current channel is sharp.

The potential difference $\Delta U = U_{ch}(x) + \varphi_k - U_g$, which ensures the condition for a complete channel overlap, is called the *full overlap voltage* U_p (also named "pinch-of" voltage). From (14.1) for $h(x) = A$ it follows that

$$U_p = \frac{1}{2\varepsilon} eN_d A^2. \tag{14.2}$$

With closed channels, the current is zero, meaning that $U_{ch}(x) = 0$. Then the gate voltage, at which the channel will be blocked, is determined from the expression

$$U_{ip} = -U_{gp} + \varphi_k \tag{14.3}$$

where U_{gp} is the external overlap voltage at the gate.

It should be noted that for a small thickness of channel A, the voltage U_{gp}, calculated by formula (14.2), can be less than φ_k . Then the external voltage of the overlap $U_{gp} \geq 0$. This mode is used in field-effect transistors operating in digital electronic circuits.

Potential distribution in the channel $U_{ch}(x)$ depends on the flowing current I_d, hence in expression (14.1) this voltage should be changed to

$$U_{ch}(x) = U_{ds} - I_d R_{xd}$$

where R_{xd} is the channel resistance from point x to the drain.

We write down the drain current, using the expression for current $I_d = \rho v S \approx eN_d v(A - h)Z$, where $S = (A - h)Z$ is the cross-sectional area of the channel and Z is the width of the gate. Substituting the relations obtained above, we arrive at the following:

$$I_d = eN_d v(E_{ch}(x)) \left(A - \sqrt{\frac{2\varepsilon(U_{ds} - I_d R_{xd} + \varphi_k - U_g)}{eN_d}} \right) Z \tag{14.4}$$

The solution of this transcendental equation for specific approximations is given in the next chapters. We shall analyze the qualitatively possible form of the solution, using a number of assumptions. First, we find a typical shape of output characteristic for a structure without a gate. In this case, we obtain a two-electrode arsenide-gallium structure with ohmic contacts, that is, we have a Gunn diode. For small voltages, the volt-ampere characteristic of such a device is described by linear dependency, specific to resistors, and when the critical field is reached, current saturation occurs (Fig. 13.6.9). Such a characteristic is shown in Fig. 14.5 and defined as number 1. If a metal electrode is deposited on the side surface of such a double-electrode structure, the cross-section of the current channel will be reduced due to the formation of a depletion layer. This depleted layer (DL) is formed by the

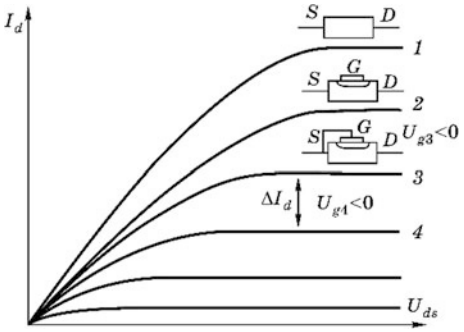


Fig. 14.5 Set of SFET output characteristics and DL form at current saturation

contact potential φ_k of the semiconductor and the metal. A reduction of cross section causes a decrease in the saturation current, as shown in the curve 2.

Increasing the gate-source voltage increases the width of the depletion layer h, and hence the height of the current channel $a = A - h$ (curves 3 and 4 in Fig. 14.5). The dependencies shown in Fig. 14.5 are typical for devices with a gate length of more than 1–2 μm.

Current saturation in the transistor is caused not only by the Gunn effect, but also by the internal negative current feedback. An increase in the drain current raises the voltage drop across the channel, which causes the current channel to narrow [see formula (14.4)]. This effect causes a decrease in the current increment. As a result, the output characteristics of the transistor have a small slope in the saturation region.

In this sense, a MESFET differs from a triode in having a significant slope of anode characteristic. At the same time, the anode characteristics of the pentode are very similar in shape to the output characteristics of the field-effect transistor. In the pentode, the introduction of additional grids reduces the influence of the anode voltage on the field at the cathode, and hence on the anode current. The transistor has one gate, but its length does not allow the drain field to strongly *deflect* through a narrow current channel. Reducing the length of the gate will cause an increase in the slope of the characteristics in the saturation region.

Figure 14.6 shows an experimentally derived MESFET characteristic with a gate length of 0.25 μm. The shape of experimental dependencies differs from the dependencies shown in Fig. 14.5. First of all, this concerns the slope of the output characteristics in the current saturation region.

The input characteristic of the transistor is shown in Fig. 14.7. When external voltage U_{gp} is equal to the full overlap voltage, drain current is equal to zero: $I_d = 0$. As the voltage across the gate increases, the source current and the drain current increase respectively, reaching a maximum at a positive voltage in the order of φ_k. A further increase in voltage leads to a decrease in the drain current. This is caused by the branching of a part of the source current to the gate. This corresponds

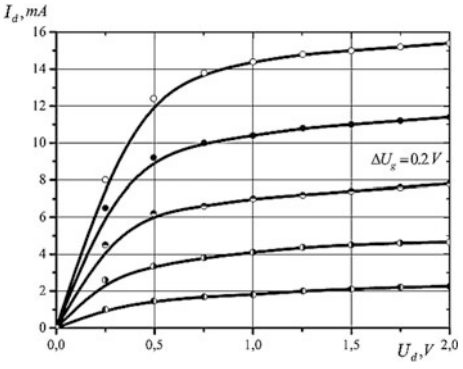


Fig. 14.6 Experimental SFET volt-ampere characteristics with a gate length of 0.25 μm

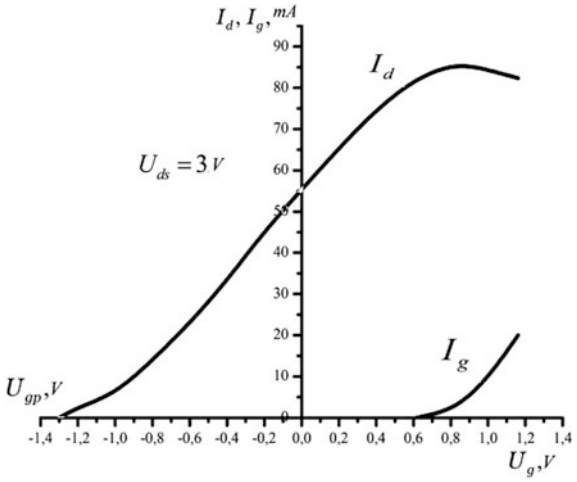


Fig. 14.7 Input characteristic of the transistor

to the current distribution in the triode, when a positive potential is applied to the grid. In fact, the gate current does not appear at voltage $U_g = \varphi_k$, but at

$$U_g = \varphi_k + I_s R_s,$$

where R_s is the "parasite" resistance of the resistive source-gate region. Under this condition, the depletion layer disappears at the gate point closest to the source. As the voltage on the gate increases, the gate area to which part of the source current is moving increases and the depletion region shifts toward the drain.

14.1.3 Small-Signal Parameters and Equivalent MESFET Circuit

The static characteristics considered above do not give a complete idea of transistor behavior with an alternating signal, especially in cases when the voltage period is commensurable with carrier transit time under the gate. Additional difficulties in describing the processes arise if the operating mode of the device is nonlinear.

However, with the help of static characteristics, it is possible to accurately describe the behavior of the transistor in low frequency band. In this case, lumped equivalent circuits can be used, the parameters of which are determined from the static characteristics. Small device dimensions, in comparison with the signal wavelength, small transit angle explain the legitimacy of using lumped elements to describe the behavior of the transistor in the microwave range.

Let's determine the response of the transistor to the effect of small deviations of supply voltages. Initially static voltages U_{g0}, U_{sd0} are applied to the electrodes, causing drain current I_{d0}. Voltage increments at the electrodes ΔU_{gs}, ΔU_{sd} can be

expressed through the differential parameters of the transistor. For this we use the small-signal (linear) approximation, that is, we set $\Delta U_{gs} \ll \varphi_k, \Delta U_{sd} \ll U_{sd0}$. Then current increment

$$\Delta I_d = \frac{\partial I_d}{\partial U_{gs}} \Delta U_{gs} + \frac{\partial I_d}{\partial U_{sd}} \Delta U_{sd}, \tag{14.5}$$

where $\partial I_d / \partial U_{gs} = G$ is parameter characterizing the slope of the input characteristic. It is called *transconductance*; $\partial I_d / \partial U_{sd} = G_{ds}$ is the parameter characterizing the slope of the output characteristic for given static values U_{g0}, U_{ds0}. It is called *output admittance*.

Suppose that alternating signals $U_{gm} \sin \omega t$ and $U_{sd} \sin \omega t$ are added to the static voltages U_{g0}, U_{sd0}. We assume that $U_{gm} \ll \varphi_k$ and $U_{sdm} \ll U_{d0}$. Then, to find the precise response of the device, it is necessary to take into account the relationship between the voltage period $T = 2\pi/\omega$ and transit time. Using a complex representation of quantities, we rewrite (14.5):

$$\dot{I}_{dm} = \dot{G}\dot{U}_{gsm} + \dot{G}_d \dot{U}_{sdm}, \tag{14.6}$$

where $\dot{G}$, $\dot{G}_d$ are the complex parameters of the device, which at low frequencies are equal to the previously introduced differential parameters.

Figure 14.8a shows the electrical circuit satisfying (14.6). It consists of an ideal current generator $\dot{G}\dot{U}_{gsm}$, controlled by the input voltage U_{gsm}, and the output conductivity $\dot{G}_d$. This is the simplest equivalent circuit of the transistor. Such a circuit does not reflect the charge exchange of the depletion layer capacitance when the polarity of the alternating voltage is changed.

We introduce additional parameters into the equivalent circuit that take these processes into account. To do this, we write down the change in the depletion layer charge through the voltage increment at the electrodes:

$$\Delta Q = \partial Q / \partial U_{gs} \Delta U_{gs} + \partial Q / \partial U_{gd} \Delta U_{gd}$$

or

$$\Delta Q = C_{gs} \Delta U_{gs} + C_{gd} \Delta U_{gd} \tag{14.7}$$

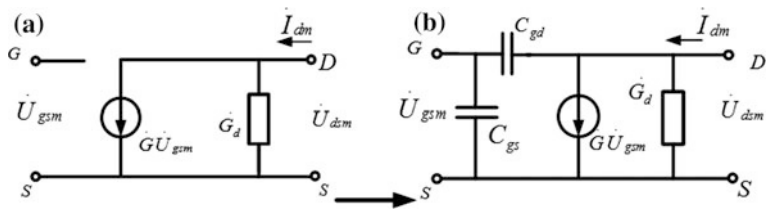


Fig. 14.8 Small-signal equivalent circuit of a field effect transistor

where $C_{gs} = \partial Q/\partial U_{gs}$ is the differential source-drain capacitance; and $C_{gd} = \partial Q/\partial U_{gd}$ is the differential gate-drain capacitance.

When these capacities are introduced into an equivalent circuit (Fig. 14.8a), the latter is converted to the form shown in Fig. 14.8b. Capacity C_{gs} is often called the input capacitance and C_{gd} is the feedback capacitance. Note an important detail. When writing (14.7), a differential gate-drain capacitance is used, rather than the source-drain capacitance expected by analogy with (14.5). This is due to the fact that when the voltage on the drain changes, the boundary of the depletion layer changes predominantly near the gate end, facing the drain contact. This is due to the weak deflection of the drain potential in the narrow subgate region of the transistor. The change in the depletion layer charge in the region between the gate and the drain is mainly determined by the gate-drain voltage difference.

The equivalent circuit (Fig. 14.8) characterizes the properties of the transistor itself. The practice of using such a circuit has shown that to describe the HF properties of the device it is necessary to introduce in the circuit an additional resistance channel R_i, which is inserted in series with capacitance C_{gs}. This capacitance is charged through it. We use voltage on the capacitance C_{gs} as voltage on the current source, not full voltage U_{gs}.

Using this circuit, we determine at what frequency the current gain will be equal to one. This parameter is called the boundary frequency f_b. Suppose there is no feedback, $C_{gd} = 0$ and the output conductivity is zero Input current $\dot{I}_{\text{in}\,m} = i\omega C_{gs}\dot{U}_{gsm}$. Assuming that $\left|\dot{I}_{\text{in}\,m}\right| = \left|\dot{I}_{\text{out}\,m}\right|$ we obtain

$$f_b = \frac{G}{2\pi \cdot C_{gs}} \tag{14.8}$$

It is necessary to add to this circuit external "parasitic" elements that determine the influence of passive areas of the structure and its connections to the case. Figure 14.9 shows these elements and their relations to the topology of the transistor.

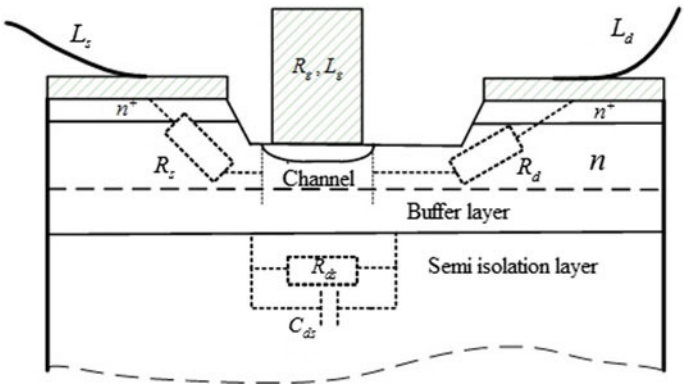


Fig. 14.9 Structure of a transistor with "parasite" elements

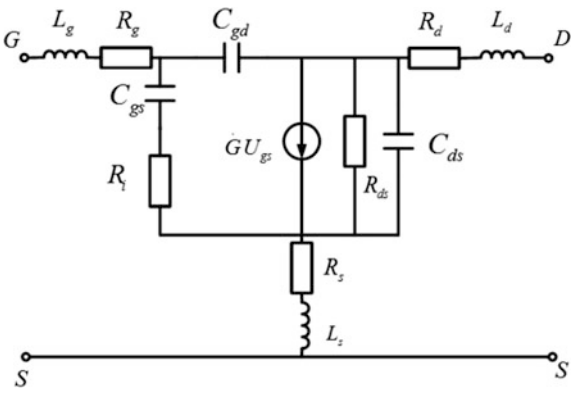


Fig. 14.10 Equivalent circuit of a field effect transistor

Resistances of ohmics contact and parts of the neutral semiconductor from the contact to the depletion layer are denoted by R_s and R_d. Resistance R_g is the ohmic resistance of the gate metallization. At high frequencies, when calculating R_g, it is necessary to take into account the skin effect. Inductances L_g, L_s, L_d are defined by the dimensions and shape of the contacting conductors.

Finally, the equivalent circuit takes the form shown in Fig. 14.10. Elements of the circuit reflect the properties of the physical processes of current transfer and, accordingly, can be calculated using physic-topologic models. On the other hand, the parameters of this circuit can be measured experimentally. Part of the parameters is measured on direct current. The greatest difficulty is defining channel resistance R_i and output resistance $R_{ds} = 1/G_{ds}$. When using the circuit in the upper part of the operating frequency range, the numerical values of the circuit elements are corrected by parametric optimization methods, using measured scattering parameters (see Fig. 14.10) as the goal function. Similar equivalent circuits are successfully used for nonlinear analysis. The dependence of the parameters on the applied voltage can be expressed analytically or approximated using the measurement results.

In the high-frequency region, it is important to take into account the time delay between the voltage and the response. To do this, we introduce a complex transconductivity $\dot{G}$ which can be written in the form $\dot{G} = g_m e^{-i\omega\tau}$ where τ is the characteristic delay (transit) time. An approximate estimate of τ is given by the formula

$$\tau \sim L_g / v_s,$$

where v_s is the saturation rate.

The circuit can be supplemented with elements related to unaccounted physical aspects of current transfer. Figure 14.11 shows the equivalent circuit of the TriQuint TGF2023-2-01 transistor with R_{dg} and R_{gs} elements, which improve the accuracy of the description of the device characteristics. This transistor is made of GaN on a SiC substrate, has a total gate width 1.2 mm, a working frequency range

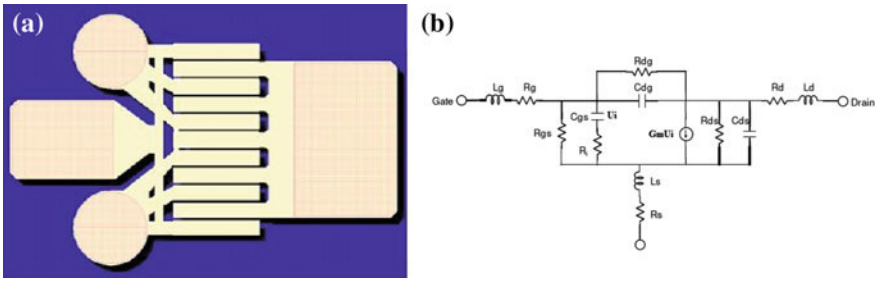


Fig. 14.11 Transistor TGF2023-2-01: **a** topology; **b** equivalent circuit

Table 14.1 Equivalent circuit parameters of transistor TGF2023-2-01

Voltage on drain U = 28 V				
Parameter	Value			Unit
I_{ds}	25	62.5	125	mA
R_g	0.57	0.58	0.64	Ω
R_s	0.03	0.03	0.04	Ω
R_d	1.46	1.51	1.06	Ω
g_m	0.152	0.249	0.289	Sym
C_{gs}	1.619	1.831	1.903	pF
R_i	0.07	0.07	0.12	Ω
C_{ds}	0.282	0.273	0.303	pF
R_{ds}	242	166	134	Ω
C_{gd}	0.056	0.052	0.052	pF
τ	3.38	3.320	3.010	Ps
L_s	0.014	0.014	0.009	nH
L_g	0.034	0.034	0.060	nH
L_d	0.004	0.008	0.0004	nH
R_{gs}	336	307	154	kΩ
R_{gd}	1380	1360	421	kΩ

of 0–18 GHz, an output power of 6 W at a frequency of 3 GHz, and a chip size of 0.82 × 0.66 × 0.1 mm. Table 14.1 contains the parameters of the equivalent circuit and their dependence on the drain current I_{ds}. Some parameters, as expected, depend little on drain current, but capacitance C_{gs}, transition conductivity G and output resistance R_{ds} vary significantly.

The small-signal equivalent circuit considered describes the parameters in a wide frequency range, but does not reflect the presence of negative current feedback. In addition, the formal introduction of two capacitances, C_{gs} and C_{gd}, does not give a clear boundary of the depletion layer regions corresponding to these capacitances. Introducing resistance R_i to the circuit is well interpreted from a physical point of view, but there are difficulties with its calculation. Use of physical topological models for the direct calculation of quantities $\dot{G}$, $\dot{G}_{ds}$ and C_{gs}, C_{gd} according to

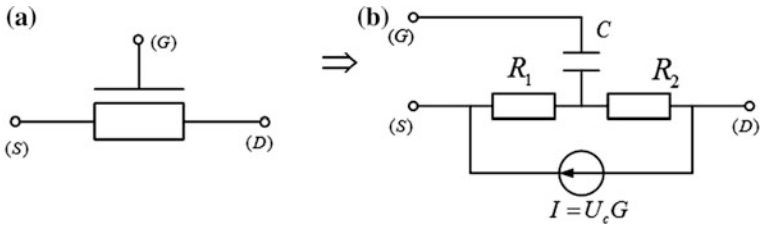


Fig. 14.12 New equivalent circuit of FET

(14.5) and (14.7) for given voltage increments, does not give the correct value of R_i. The value of this parameter should be found with the help of additional numerical experiments.

The circuit shown in Fig. 14.12b is partially free from these disadvantages. The topology of the circuit resembles the structure of the transistor in Fig. 14.12a. Resistance R_1 corresponds to the resistance of the input part of the current channel. It preserves internal negative current feedback even at constant current. Resistance R_2 corresponds to the resistance of the output part of the current channel. Usually, $R_2 \gg R_1$. The influence (feedback) of the drain voltage on the control voltage U_c occurs through resistance R_2. Capacity C reflects the properties of the charge of the entire depletion layer. We call this circuit a T-circuit.

Equations like (14.5)–(14.7) for the T-circuit are written as follows:

$$I_d = GU_c + \frac{U_{sd}}{R_1 + R_2} = G(U_{gs} - U_{ch}) + \frac{U_{sd}}{R_1 + R_2}$$

Here, the value of the "acting" voltage in the channel U_{ch} is entered and is determined by the expression

$$U_{ch} = \frac{U_{sd}}{R_1 + R_2} R_1$$

Then

$$I_d = GU_{gs} + \frac{U_{sd}}{R_1 + R_2}(1 - GR_1)$$

The current increment is written as follows:

$$\Delta I_d = G\Delta U_{gs} + \frac{(1 - GR_1)}{R_1 + R_2}\Delta U_{sd}$$

Capacity C is determined by the ratio of the increment of the entire DL charge ΔQ_{OC} to the voltage increment ΔU_c, i.e.,

$$C = \Delta Q_{oc}/\Delta U_c$$

Calculations and investigations have shown that at low frequencies including zero frequency, a more accurate description is given by a T-circuit. It is especially convenient to use this circuit when calculating transistor noises, which will be explained in Sect. 14.7.

A universal way of describing the linear operating mode of a transistor, as an element of the microwave circuit, is to use scattering parameters—*s*-parameters. With this description, the transistor is represented by a two-port network with an unknown internal structure (black box). Such a representation is called a formal model.

The description of transistors with *s*-parameters became possible with the development of measurement techniques and with the help of automated vector network analyzers (VNA). The high accuracy and rapidity of measurements make it possible to use *s*-parameters as a universal tool in the development of microwave devices.

s-parameters determine the relations between the normalized incident wave a, and reflected wave b, at the input and output ports of the transistor:

$$\left\| \begin{matrix} \dot{b}_1 \\ \dot{b}_2 \end{matrix} \right\| = \left| \begin{matrix} \dot{s}_{11} & \dot{s}_{12} \\ \dot{s}_{21} & \dot{s}_{22} \end{matrix} \right| \left\| \begin{matrix} \dot{a}_1 \\ \dot{a}_2 \end{matrix} \right\|$$

Scattering parameters have a simple physical sense. Parameter $\dot{s}_{11} = \dot{b}_1/\dot{a}_1$ is the reflection coefficient from the input for matched output $(\dot{a}_2 = 0)$. Parameter $\dot{s}_{22} = \dot{b}_2/\dot{a}_2$ is the reflection coefficient from the output for matched input $(\dot{a}_1 = 0)$. Parameter $\dot{s}_{21} = \dot{b}_2/\dot{a}_1$ is the transmission coefficient from input to output for matched output $(a_2 = 0)$. Parameter $\dot{s}_{12} = \dot{b}_1/\dot{a}_2$ is the transmission coefficient from output to input (feedback) with matched input $(\dot{a}_1 = 0)$.

The scattering parameters are measured in a regular transmission line with wave impedance Z_0 (usually 50 Ω) relative to the selected reference planes, as shown in Fig. 14.13.

In modern practice, transistor manufacturers, together with the suppliers of devices, provide developers with the dependence of *s*-parameters on frequency. These characteristics are usually represented in the form of hodographs on a Smith

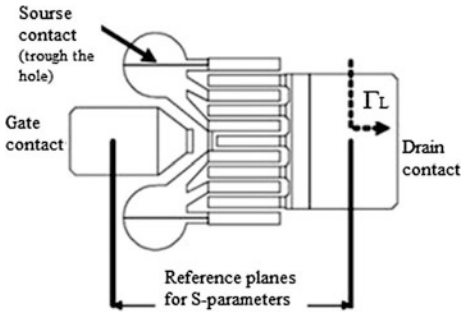


Fig. 14.13 Reference planes for measuring scattering parameters

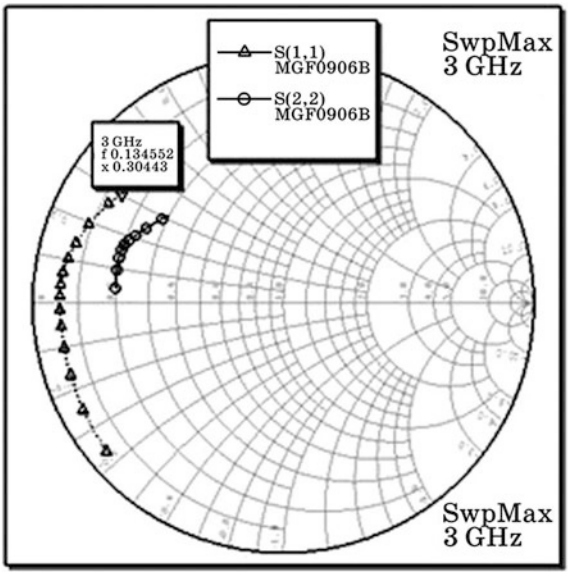


Fig. 14.14 Representation of scattering parameters on a Smith chart

chart (Fig. 14.14). It should be noted that only the reflection coefficients $\dot{s}_{11}$ and $\dot{s}_{22}$ can be displayed directly on the chart. However, due to capacitive input and output impedances of the transistor, often the hodographs of these quantities are only in one-half of the diagram. Hence, the second part of the diagram can be used to represent the coefficients $\dot{s}_{21}$ and $\dot{s}_{12}$ in a polar coordinate system with a convenient scale to represent the moduli of these essentially different quantities.

14.1.4 Modelling of Field Effect Transistors

To describe the properties transistors, we use a number of physico-topological models. These models make it possible to calculate parameters of equivalent circuits using the device topology and material properties. In addition, they allow analysis of the features of current transfer in the structure. The information obtained in this case is important for further improvement of devices and the circuits based on them. This is especially true in the high frequency band, where the physical picture of the processes in the device becomes more complicated.

The Shockley model was widely used for a long time, from the creation of the field effect transistor, up to the creation of a transistor with a gate length of 1–2 μm. The model gave a good level of accuracy for the description of volt-ampere characteristics and satisfactory accuracy when calculating the parameters of a small-signal equivalent circuit despite a number of obvious rough approximations. Considering the wide distribution of this model, we describe its main provisions, and estimate the limits of its application.

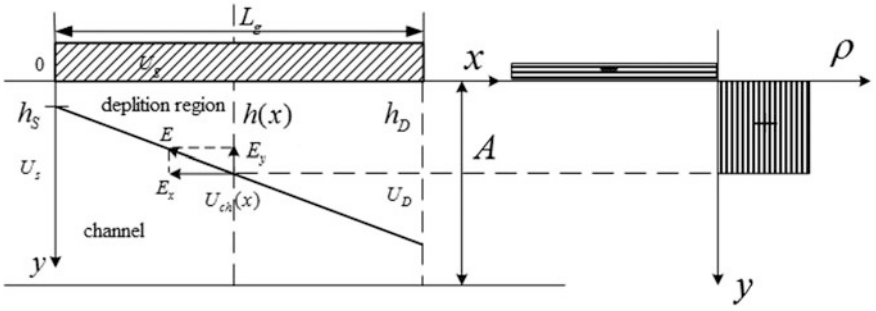


Fig. 14.15 Shockley model

The Shockley Model

This model describes the behavior of the transistor subgate region. The topology of the region is shown in Fig. 14.15. The substrate of the transistor is supposed to be insulating.

The basic approximations are as follows.

Assumption 1 The model uses the local field approximation (LFA), when the drift velocity at the given point is determined by the electric field strength at the same point: $v(\mathbf{r}, t) = \mu_0 E(\mathbf{r}, t)$. Moreover, it is assumed that mobility μ_0 does not depend on the field intensity E. This approximation is not quite correct, but its use allowed the obtaining of analytical solutions describing properties of transistors with acceptable accuracy.

Assumption 2 The diffusion coefficient is zero: $D(E) = 0$. This means that there is no Debye "blurring" of the boundary between the current channel and the depletion layer.

Assumption 3 Approximation of the "gradient" channel, when the length of the gate is much greater than the thickness of the current channel, i.e. $L_g/A \gg 1$. This model is often called the smooth channel model.

We find the expression for the drain current I_d as a function of voltages on the electrodes $I_d = f(U_g, U_d, L_g, Z)$ for a uniformly doped current channel, that is, $N_d = \text{const}$.

To find the drain current, it is necessary to know the velocity of carriers in the current channel and the width of this channel. To find these quantities, we first of all determine the distribution of the electric field intensity using the Poisson equation:

$$\text{div}\mathbf{E} = \rho/\varepsilon. \tag{14.9}$$

If the boundary between the current channel (CC) and the depletion layer (DL) is sharp (Assumption 2), then the carriers move tangentially to this boundary. In this case, the force acting on the charge is $\mathbf{F} = eE_x\mathbf{e}_x + eE_y\mathbf{e}_y$. For the smooth channel $(L_g/A \gg 1)$ we can put $E_y \approx 0$. The static charges of the DL and the opposite

charges on the gate create a field normal to the metallic surface of the gate, i.e., $E_y\mathbf{e}_y$ (Fig. 14.15). Under these conditions, we write (14.9) separately for DL

$$\frac{\mathrm{d}E_y}{\mathrm{d}y} = \frac{eN_d}{\varepsilon}, \tag{14.10}$$

and for CC

$$\frac{\mathrm{d}E_x}{\mathrm{d}x} = \frac{e(N_d - n)}{\varepsilon}. \tag{14.11}$$

First we solve (14.10) for the following boundary conditions: $E_y|_{y=h} = 0, U_{y=0} = U_g, U_{g=h} = U_{ch}(x)$. Let's suppose that the field E_y does not penetrate into the gate material, since the gate is made of metal with high conductivity. Figure 14.16 shows the distribution of charge in the gate material and in the DL. For metal, the thickness of the layer of negative charges δ tends to zero.

The total charge in the metal is equal in absolute value to the total charge of positive donors in the depletion layer.

Solving expression (14.10) for the given boundary conditions gives

$$U_{ch}(x) + \varphi_k - U_g = \frac{eN_d}{2\varepsilon}(y - h(x))^2, \tag{14.12}$$

where $h(x)$ is the height of the DL. In this expression, the left part $U_{ch}(x) + \varphi_k - U_g$ is the voltage acting between the gate and the considered point of the channel.

The full overlap voltage U_p, of the channel with height $h = A$, according to (14.12), is defined by the following equation

$$U_P = \frac{1}{2}\frac{eN_d}{\varepsilon}A^2 \tag{14.13}$$

Value $U_{gp} = U_p - \varphi_k$ is called the *external overlap voltage*.

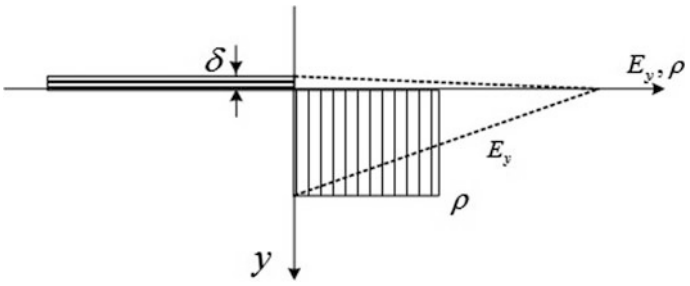


Fig. 14.16 Distribution of charges near the gate

Using the results obtained, we find the expression for the drain current:

$$\begin{aligned} I_d &= envZ(A-h)Z = eN_d\mu_0 E_x(A-h)Z \\ &= eN_d\mu_0(-\partial U_{ch}/\partial x)(A-h)Z \\ &= eN_d\mu_0(-\partial U_{ch}/\partial h)\frac{\partial h}{\partial x}(A-h)Z. \end{aligned}$$

In this expression, derivative $\partial U_{ch}/\partial x$ is expressed through derivative $\partial U_{ch}/\partial h$ and $\partial h/\partial x$.

Using (14.12), we get

$$\partial U_{ch}/\partial h = h\frac{eN_d}{\varepsilon}.$$

Finally, the expression for the drain current takes the form

$$I_d \mathrm{d}x = e^2 N_d^2 \mu Z(A-h)h\mathrm{d}h.$$

We integrate this expression along the length of the device. Coordinate x varies from 0 to L_g, whereby the depletion layer changes from a certain value h_s on the source side to a value h_d on the drain side:

$$I_d \int_0^{L_g} \mathrm{d}x = -\frac{e^2 N_d^2}{\varepsilon}\mu_0 Z \int_{h_s}^{h_d} h(A-h)\mathrm{d}h.$$

We rewrite the resulting expression in terms of dimensionless quantities. Normalization of linear dimensions h_s and h_d is logically realized to the height of channel A, voltage—to overlap voltage U_p, current I_d—to value $I_P = eN_d(\mu_0 U_p/L_g)AZ$.

Defining

$$s = h_s/A = \sqrt{(U_s - U_g + \varphi_k)/U_p}, \quad p = h_d/A = \sqrt{(U_d - U_g + \varphi_k)/U_p},$$

we obtain

$$I_d = I_p\left((p^2 - s^2) - \frac{2}{3}(p^3 - s^3)\right) \tag{14.14}$$

Parameter s determines the degree of overlapping of the channel at source. On the other hand, this is the normalized potential of the source. p is the degree of overlap on the drain and the normalized potential of the drain.

The voltage drop across the transistor in accordance with the accepted notations is expressed by the expression

$$U_{ds} = U_p\left(p^2 - s^2\right). \tag{14.15}$$

Using (14.14), we show graphically the dependence of the normalized current on the normalized voltages s and p, i.e., $I_d = f(s,p)$. Figure 14.17 shows a set of such characteristics, where s is a parameter. It should be noted that physically interpreted results correspond to inequalities $0 \le s \le 1$ and $0 \le p \le 1$. From the analysis of the obtained expression, it follows that if we obtain the dependence $I_d = f(0,p)$, then dependence $I_d = f(s_1,p)$ for other values of s is obtained by a simple displacement of the original $I_d = f(0,p)$ along the abscissa by the value s_1.

The obtained set represents the initial branch of voltage-current characteristic in the normalized form not considering current saturation, i.e., $p \le 1$. The experiment shows that when the voltage on the drain increases, the current remains practically unchanged even for $p > 1$. In order to know the volt-ampere characteristic for any voltage, another assumption is introduced into the Shockley model. If the normalized potential $p > 1$, then the current value is assumed to be equal to the current for $p = 1$. This approximation corresponds to the dashed lines in Fig. 14.17.

Current calculations for the smooth channel model are in good agreement with the results of the voltage-current characteristic measurements for gate lengths $L_g > 2$ μm. The existing differences are manifested in the slope of the characteristics in the saturation region: the experimental dependences have a small slope, calculated dependences don't.

In spite of the satisfactorily matched, measured, and calculated characteristics, the physical picture of the current transfer process cannot be considered acceptable. In the Shockley model, maximum current is observed when the current channel is completely overlapped at the drain end. If the channel is blocked in at least one section, current cannot flow. A formal explanation of the possibility for current flow in this case is the possibility of an infinite increase in velocity when the channel is

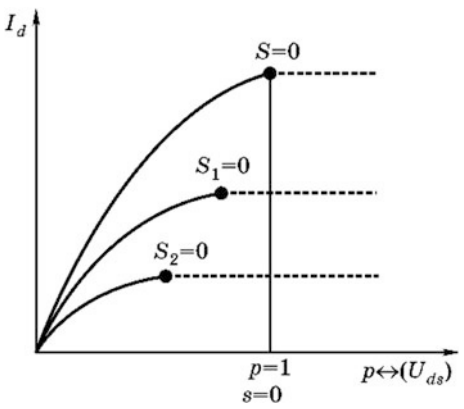


Fig. 14.17 Calculated volt-ampere characteristic on the Shockley model

narrowed due to an increase in the "pulling" field and the direct proportionality of the velocity and the field $v = \mu_0 E$. Then the product of an infinitely small thickness of the current channel with an infinitely high velocity, will give a finite value of the current.

For gates with a length of 1–0.5 μm or less, the discrepancy of the measured and calculated values has become unacceptably large. To obtain an acceptable correspondence between the experiment and the theory, a simple and intuitive model was developed, called the model of two regions.

The Model of Two Regions

The main difference of this model from the smooth channel model is the piecewise linear approximation of the field-velocity characteristic, as shown in Fig. 14.18. In the first section, the velocity and the field are proportional: $v = \mu E$, and in the second, the velocity is constant and equal to the saturation value: $v = v_s$. This approximation reflects the typical field-velocity characteristic (see Chaps. 2 and 4).

As in the Shockley model, the two regions model assumes a local coupling between the velocity and the field, i.e., the local field approximation is applied.

Consider only the intrinsic transistor, without parasite elements. Figure 14.19 shows the topology of the region under consideration, the field-velocity characteristic (FVC), and the supposed distribution of electric field $E(x)$. The coordinate x_{th}, where the field $E(x_{th}) = E_{th}$, divides the subgate region of the transistor into two regions: the input region 1 with length L_{g1}, where the drift velocity of carriers and the field are connected linearly $v = \mu_0 E$, as in the Shockley model, and the output area 2 with length L_{g2}, where the velocity of carriers is constant and equal to the saturation value $v = v_s$. Such an approach involves the analysis of current transfer in the first region using the Shockley model with a gate length of the transistor L_{g1}.

In the second region L_{g2} it is assumed that the height of the current channel is constant in order to maintain a constant current. The model parameters are

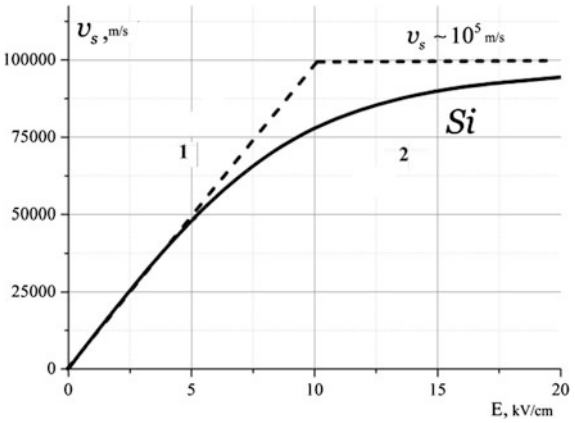


Fig. 14.18 Piecewise linear approximation of the FVC

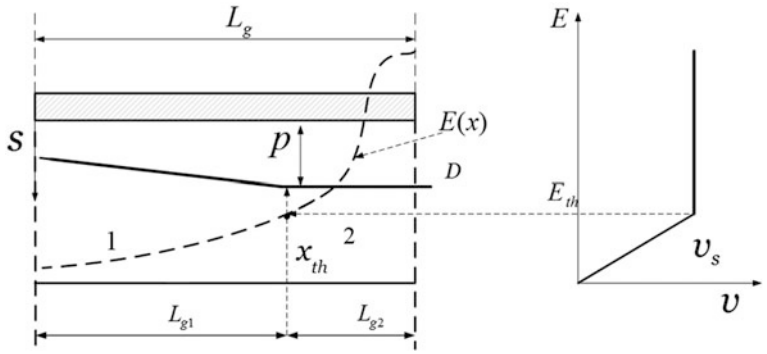


Fig. 14.19 The two regions model

determined by "stitching" the solutions for both regions. The current at the interface must be continuous, and the voltage drop at the first region U_{Lg1} when added to the voltage of the second region U_{Lg2} must be equal to the applied voltage between the source and the drain, i.e.

$$U_{L_{g1}} + U_{L_{g2}} = U_{ds}. \tag{14.16}$$

The voltage drop across region 1, according to the smooth channel model (14.15), will be written as

$$U_{L_{g1}} = U_p\left(p^2 - s^2\right),$$

where p is an unknown quantity. The voltage drop in region 2 is determined from the analytical solution of the Poisson equation for the region:

$$U_{L_{g2}} \cong U_p \cos\frac{\pi y}{2A} sh\frac{L_{g2}}{2A}. \tag{14.17}$$

Total voltage across the transistor

$$U_{ds} \cong U_p\left(p^2 - s^2\right) + U_p \cos\frac{\pi y}{2A} sh\frac{L_{g2}}{2A}. \tag{14.18}$$

Currents in both parts of the transistor must be equal. Then, using (14.14) and writing current in region 2 as

$$I_d = eN_d v_3 AZ(1-p) = I_{\max}(1-p).$$

we obtain an equation connecting the unknown quantities:

$$I_{\max}(1-p) = I_p\left(L_{g1}\right) \cdot \left(p^2 - s^2 - 2/3\left(p^3 - s^3\right)\right), \tag{14.19}$$

where $I_{\max} = eN_d v_s AZ$ is the maximum possible current through the structure.

Solving the system of (14.18) and (14.19) with respect to unknown quantities p and L_{g1}, we obtain the dependence of the drain current I_d on voltages U_{ds} and U_{gs} i.e. the input and output characteristics of the transistor. The model puts no restrictions on the values of the applied voltages. The current calculation algorithm makes it easy to take into account the influence of the source and drain parasite resistances, adding to the voltage drop on the internal regions U_{ds} the voltage drops on R_s and R_d:

$$U_{ds\Sigma} = I_d R_s + U_{ds} + I_d R_d,$$

where $U_{ds\Sigma}$ is the external voltage on the transistor.

This equation is included in the (14.18) and (14.19) system and they are jointly solved using the method of successive approximations.

From the dependences obtained, it is possible to determine the differential parameters of the transistor, in particular for the transconductivity:

$$G = I_p / U_p \left(p^2 - s^2\right)$$

The correspondence between the calculated and measured values of current and transconductivity for transistors with gate lengths up to 1 μm is much better than in the Shockley model. The greatest discrepancy between the experimental and measured values is observed for the output conductivity G_d. The measured values are 3–5 times smaller than the calculated ones.

14.1.5 Peculiarities of Mathematical Modeling of Field Effect Transistors

The expansion of the transistor operation frequency range is stipulated, first of all, by decreasing gate length. At present, minimum gate length is 0.25–0.025 μm, which is comparable with the energy relaxation length of the carrier l_ε. For these devices, the local field approximation (LFA) is not correct. In this connection, one needs to solve the conservation equations system (2.39)–(2.41) and the Poisson equation (2.8) for current transfer analysis and calculation of its external characteristics. The complexity of solving these equations forces the developers to use a number of approximations that allow the preservation of important physical aspects of the process and the obtaining of analytical expressions for the integral parameters of the device.

First, consider conservation of momentum and energy equations, allowing us to find the velocity of carriers v and the electronic temperature T_e under conditions of rapidly varying field intensity. It is shown in Sect. 2.3.3, that with a rapid change of

the field E in time (or a sharp change in space), the phenomenon of overshooting arises (Fig. 2.8). We rewrite (2.40) and (2.41) in the approximation of the slow change of energy due to the microwave field and weak spatial divergence of the heat flux.

The equation of flux momentum conservation is transformed to the form

$$v = \mu(T_e)E - D(T_e)\frac{\nabla n}{n}. \tag{14.20}$$

Electron temperature T_e is found from the energy conservation equation

$$\mathbf{jE} \approx \frac{3}{2}\frac{nk(T_e - T_0)}{\tau_\varepsilon(T_e)}. \tag{14.21}$$

To solve this system, it is necessary to know how the relaxation time $\tau_\varepsilon(T_e)$ and the diffusion coefficient $D(T_e)$ vary with temperature T_e. These values are determined using higher-level models, in particular, the particle-in-cell (PIC) model, or are measured experimentally. However, the complex nonlinear character of the dependences $\tau_\varepsilon(T_e)$, $D(T_e)$ and the need for an iterative process to find T_e make it impossible to obtain solutions explicitly, even with the simplifications introduced.

We will use computer modeling results to describe a physical picture of current transfer. Figure 14.20 shows typical field dependencies in channel $E(x)$, electron temperature $T_e(x)$, and width of the depletion layer $h(x)$ for a transistor with gate length of 0.5 μm. As can be seen, the maxima $E(x)$ and $T_e(x)$ are shifted to distance l_ε. The localization of the maximum field E at the gate drain edge is explained by the fact that in this section the current channel is the narrowest. Hence, the specific resistance and voltage drop in the considered region are maximized. The electron temperature continues to grow but at a lower speed, since the "heating" field E decreases. As a result, the maximum of $T_e(x)$ is shifted to the right by a distance l_ε.

Using function $T_e(x)$, it is possible to calculate the blurring of the depletion layer boundary. For this, we take into account the thermal potential $U_T = kT_e/e$ in the formula for calculating the depletion layer (14.12):

$$h(x) = \sqrt{\frac{2\varepsilon(U_{ch}(x) - U_g + \varphi_k - \frac{kT_e(x)}{e})}{eN_d}}. \tag{14.22}$$

Due to the increase in electron temperature from the beginning of the transistor to its end, the degree of blurring, determined by the Debye length, $L_D = \sqrt{kT_e(x)/(\varepsilon eN_d)}$ increases. It should be noted that without taking this factor into account, it is impossible to explain the principle of the action of the field effect transistor for high voltages on the drain.

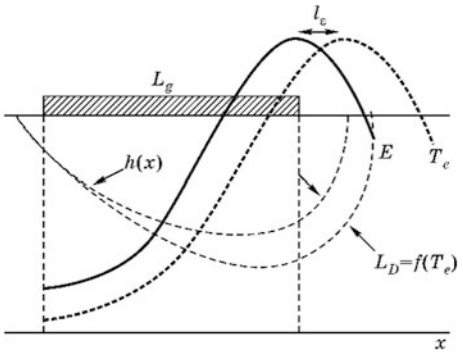


Fig. 14.20 Typical distributions of field, electron temperature, and depleted layer width in a channel

Analysis of the distribution of carrier velocity in the channel is more complicated than the field and the electron temperature distributions. We consider a model of the transistor that allows us to solve these problems, below.

14.1.6 Quasi-Two-Dimensional Temperature Model of MESFET

In the described versions of the current transfer analysis, the two-dimensional nature of the field distribution under the gate of the transistor was not taken into account. The Poisson equation for real structures has to be solved in a two-dimensional approximation. However, in modeling practice, a simplified quasi-two-dimensional approach is used that allows an accurate reducing of the problem to a one-dimensional nature.

The essence of this model lies in the fact that the two-dimensional Poisson equation is written with respect to the averaged values of the electric field strength and the charge density, separately for the channel and the depletion layer. In the current channel, the determining quantity is the x-component of field **E**, and in the depleted channel it is the y-component of field **E**. To obtain the integral characteristics of the device as a whole, it is not always necessary to have detailed information on the distribution of the field and current along the channel cross-section. It is sufficient to know the mean values of the unknown quantities.

For the current channel. The average $\overline{E}_x$ value for the current channel cross section

$$\overline{E}_x = \frac{1}{A-h} \cdot \int_h^{A-h} E_x(y)\mathrm{d}y.$$

Average value of doping level $\overline{N}_d$:

$$\overline{N}_d = \frac{1}{A-h} \cdot \int_h^{A-h} N_d(y)\mathrm{d}y.$$

With a known channel height $a(x) = A - h(x)$, the Poisson equation with respect to the mean values is reduced to the form

$$\frac{\mathrm{d}(\overline{E}_x a)}{\mathrm{d}x} = \frac{e(\overline{N}_d - \bar{n})a}{\varepsilon}. \tag{14.23}$$

For the depletion layer. We calculate $h(x)$, solving the Poisson equation for the y-component of the electric field:

$$\frac{\mathrm{d}E_y}{\mathrm{d}y} = \frac{eN_d(y)}{\varepsilon}. \tag{14.24}$$

The stitching of solutions is carried out at the boundary of the current channel to the depletion layer by the ratio of the fields:

$$E_{ysp} \approx \overline{E}_x \cdot a'(x), \tag{14.25}$$

where $a'(x) = \partial a/\partial x$. It is equal to the slope of the tangent to the boundary of the current channel (Fig. 14.21). Condition (14.25) means that electrons move along the boundary and do not enter the depletion layer. We rewrite the equation for the width of the depletion layer (14.22), taking into account the y-component of the field at the boundary.

Let us note the fact that the term "effective" should be added to the depletion thickness $h(x)$, calculated using (14.26), since it takes into account the blurring of the border due to T_e. The obtained expressions can be used not only in the subgate region, as noted at the beginning, but also in regions outside the gate. To do this, we consider the length of the line connecting the point on the boundary with the nearest

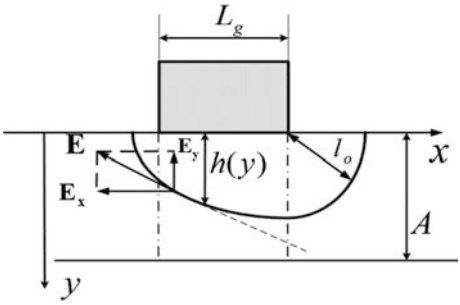


Fig. 14.21 Taking into account the transverse component of the field at the boundary of the current channel

point on the gate $l_0(x)$ as the depletion region size. For regions outside the gate, such points are the source and drain gate edges (Fig. 14.21). The Poisson equation is written along this direction: $\mathrm{d}E_l/\mathrm{d}l = eN_d(l)/\varepsilon$. The depletion region length, similar to (14.26), is written as follows:

$$l_o(x) = -\frac{E_{lb}\cdot\varepsilon}{eN_d} + \sqrt{\left(\frac{E_{lb}\cdot\varepsilon}{eN_d}\right)^2 + \frac{2\varepsilon\left(U_{ch}(x) - U_g + \varphi_k\right)}{eN_d}}, \tag{14.26}$$

where E_{lb} is the boundary field in the line direction l.

It is necessary to include the continuity equation in this system of equations. It allows determination of the average drift velocity and the average carrier concentration $\bar{n}$.

Let us compare simulation results obtained by a local field model (model of two regions) and a temperature model of devices with different gate lengths. We take required dependencies $\tau_\varepsilon(T_e), D(T_e), v(E_0)$ from the PIC model. We compare both the voltage-current characteristics and the distribution of physical quantities inside the devices. For the analysis, transistors on GaAs were chosen.

Figure 14.22 shows a comparison of the output characteristics obtained experimentally with respect to the local field (LFA) model and the temperature model. The length of the transistor gate is 0.5 μm. The presented dependences clearly demonstrate the good agreement of the experimental dependences and those calculated from the temperature model over the entire range of variation of the applied voltages. At the same time, it can be seen that the local field approximation (LFA) gives a satisfactory correspondence with experimental data only in the voltage range of the drain-source 0–0.5 V.

$$h(x) = -\frac{E_{ysp}\varepsilon}{e\overline{N}_d} + \sqrt{\left(\frac{E_{ysp}\varepsilon}{e\overline{N}_d}\right)^2 + \frac{2\varepsilon\left(U_{ch}(x) - U_g + \varphi_k - kT_e(x)/e\right)}{e\overline{N}_d}} \tag{14.27}$$

Figure 14.23 shows the distribution of the average drift velocity v/v_s, longitudinal relative electric field intensity E_x/E_{th} and electron temperature T_e/T_0 for the same device. We note, first, the shift of the field and temperature maxima ("cooling" length $l_\varepsilon \approx \tau_\varepsilon v_s$), as well as a gradual increase in the drift velocity up to the maximum of the field, that is specific for the temperature model (TM). The local field approximation (LFA) is characterized by a decrease in the velocity in the region of the maximum field, in accordance with the field-velocity characteristic (FVC).

It should be noted that the distribution of the field E is affected not only by the narrowing of the current channel, but also by the appearance of a high-field domain, as in homogeneous Gunn diodes (see 13.6.1).

It is not possible to distinguish these components, but it is necessary to take both factors into account.

The comparison of transistor calculation results with a gate length of 1 and 0.25 μm makes it possible to understand the difference in current transfer processes in these devices. For comparison, GaAs devices with an experimentally measured

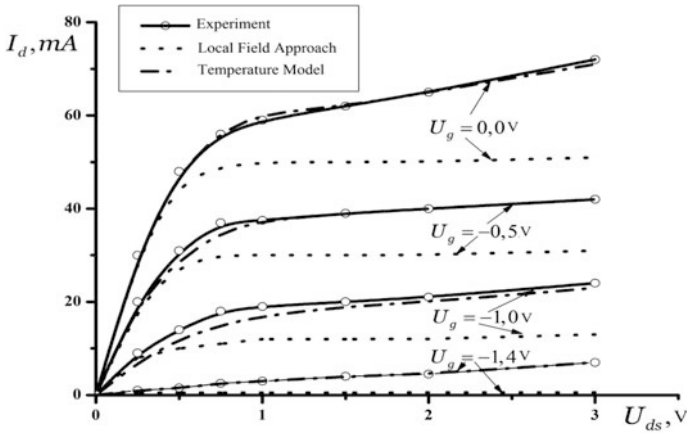


Fig. 14.22 Output characteristics of a transistor, obtained experimentally, according to the local field (LFA) and temperature models

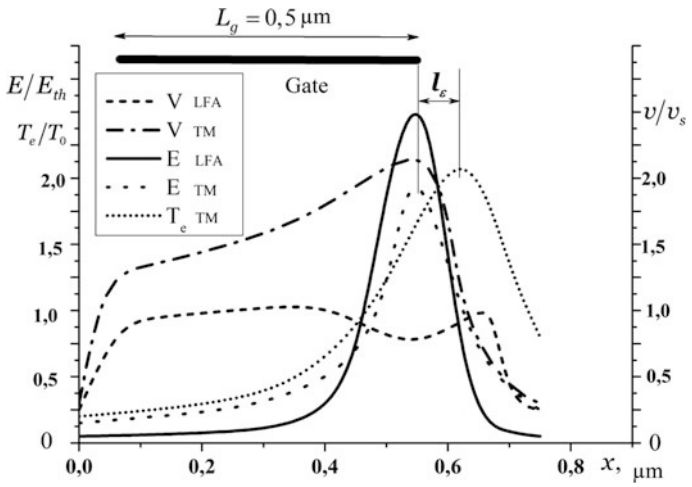


Fig. 14.23 Comparison of the temperature and local field models for the length of the transistor 0.5 μm

doping profile were chosen for the comparison. The values of the main parameters being: $\overline{N}_d = 1.78 \times 10^{17}$ cm^{-3}, $A = 0.142$ μm, and saturation velocity $v_s = 1.35 \times 10^5$ m/s. Taking into account the inhomogeneity of the doping profile it is important to consider the phenomena of current transfer in the buffer layer. In Fig. 14.24a, b the distributions of the average values of the velocity, field strength, and electron temperature along the transistor channel obtained by the temperature model are shown. With a gate length $L_g = 1$ μm velocity increases monotonously along the channel, reaching a value of 1.5 v_s at the edge of the drain gate. For

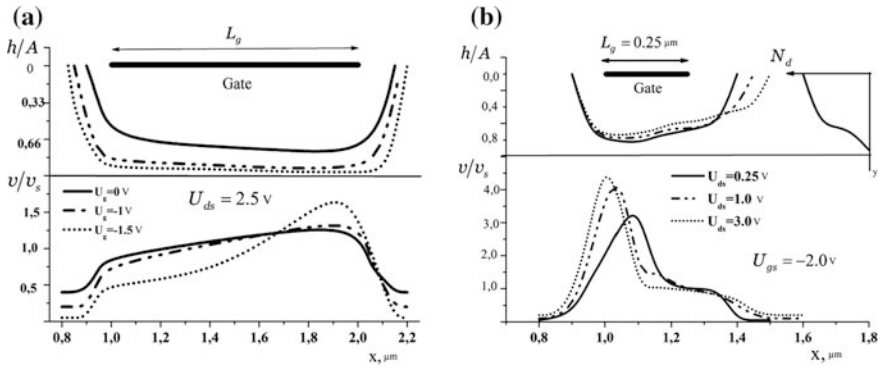


Fig. 14.24 Current channel shape and velocity distribution for: **a** L_g = 1 μm; **b** L_g = 0.25 μm

devices with $L_g = 0.25$ μm the overshooting effect is noticeable. The maximum velocity is 4.5 v_s, which is 2.5 times higher than the maximum velocity for static FVC. The localization of this maximum at the beginning of the sub-gate region is apparently caused by a sharp change in the field in this region (a lower level of doping, a narrow channel). In accordance with the velocity, the effective thickness of the current channel also changes.

The results obtained from the simulation demonstrate the possibility of calculating both the integral parameters of microwave transistors and analyzing the transport processes within the structure with a high degree of accuracy. Currently, the temperature model is used to calculate the parameters of the equivalent circuit and the noise parameters of the transistors.

14.1.7 Noise Characteristics of Field Effect Transistors

During its history, the field transistor has gone from the noisiest device to the least noisy. In the first FETs, the low-frequency noise of the external boundaries dominated: noises of the Schottky barrier and of the boundaries from the current channel to the semi-insulating substrate. Technological imperfections in the body of the current channel and parasite contact resistance created a high level of thermal noise.

If we take into account the fact that the first devices had a small gain, it becomes clear why MESFET was not used in low-noise amplifiers. Later, noise reduction was provided by reducing the gate length, improving the technology of obtaining a metal-semiconductor barrier and improving the contact of the current channel to the semi-insulating substrate. The improvement of the Schottky barrier manufacturing technology led to a decrease in the density of the surface states, which reduced the flicker noise level. The introduction of a buffer layer between the semi-insulating substrate and the channel restricted the migration of trap centers from the semi-insulating substrate to the channel. In connection with this, the temporal

stability of the amplifying properties and noise characteristics during operation was improved.

In modern transistors, the noise level is largely determined by the thermal noise of the channel and the thermal noise of the parasite gate and source resistances. The contribution to the noise factor of these resistances is up to 40% of the total noise coefficient.

Let us analyze the noise model of a MESFET using the basic equations obtained in Sect. 2.4 and the temperature model that allows us to find the electron temperature of carriers. The purpose of the analysis should be to determine the dependences of the device noise factor on the device's physico-topological parameters, at the applied DC voltages.

First, consider the noise of the transistor itself without taking into account the contribution of the parasite source and gate resistances. The analysis is made feasible by replacing the transistor with a connection of individual partial transistors, as shown in Fig. 14.25. Such a circuit can be considered as a cascade connection of amplifiers or as a distributed amplifier. The contribution to the total noise figure from each of the partial transistors will be different. This is due not only to the difference in the electronic temperature of the carriers in each partial transistor. A more significant factor is the difference in the total gain for each of the considered layers (partial transistors). Input layers, therefore, make the maximum contribution to the noise factor.

To calculate the noise sources in each partial transistor, we use the van der Ziel formula (2.48) for the selected section Δx at electron temperature $T_e(x)$, diffusion coefficient $D(T_e)$ and current channel thickness $a(x)$:

$$\left\langle i^2_{n\Delta x} \right\rangle = 4e^2 D(T_e) naZ \frac{1}{\Delta x} \cdot \Delta f, \tag{14.28}$$

where Z is the width of the channel.

This noise fluctuation of the current will cause a change in the noise voltage, i.e.

$$\left\langle e^2_{n\Delta x} \right\rangle = \left\langle i^2_{n\Delta x} \right\rangle \cdot R^2_{\Delta x}.$$

Differential resistance $R_{\Delta x}$ can be calculated by dividing voltage increment ΔU_d in the region by drain current change ΔI_d: $R_{\Delta x} = \Delta U_{\Delta x}/\Delta I_d$. Then, the noise voltage on element Δx:

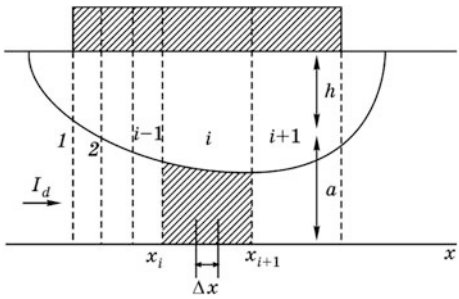


Fig. 14.25 Method of noise presentation in a current channel

$$\left\langle e_{n\Delta x}^2 \right\rangle = 4e^2 D(T_e) naZ \cdot \frac{1}{\Delta x} \Delta f \left(\frac{\Delta U_{\Delta x}}{\Delta I_d} \right)^2 = 4e^2 D(T_e) naZ \frac{1}{(\Delta I_d)^2} \left(\frac{\Delta U_{\Delta x}}{\Delta x} \right)^2 \Delta x \Delta f.$$

In this expression, the factor $(\Delta U_{\Delta x}/\Delta x)$ determines a certain field, called *the impedance field*. To determine the total voltage across the entire length of the partial transistor, it is necessary to integrate the right part from x_i to x_{i+1}:

$$\left\langle e_n^2 \right\rangle_i = \int_{x_i}^{x_{i+1}} 4e^2 D(T_e) naZ \frac{1}{(\Delta I_d)^2} \left(\frac{\Delta U_{\Delta x}}{\Delta x} \right)^2 \Delta f \mathrm{d}x. \tag{14.29}$$

This method of calculating noise voltage is called the *impedance field method*.

To calculate the noise at the transistor output, we represent each partial transistor by its equivalent circuit, as shown in Fig. 14.26. The source of the noise voltage for each of the transistors is connected in series with the differential channel resistance R_i.

The recalculation of this circuit using the equivalent source method makes it possible to obtain the circuit shown in Fig. 14.27. In this circuit, equivalent noise sources are represented as two current noise sources at the input $\left\langle i_{ng}^2 \right\rangle$ and output $\left\langle i_{nd}^2 \right\rangle$. In the general case, these sources are correlated, i.e., the correlation coefficient

$$C_c = \frac{\left\langle i_{ng} i_{nd} \right\rangle}{\sqrt{\left\langle i_{ng}^2 \right\rangle \left\langle i_{nd}^2 \right\rangle}} \neq 0.$$

Numerical analysis of noise parameters and their comparison with experimental data have shown that in most cases one noise source at the input is enough for practical accuracy.

In general, the RMS value of noise current can be written in the form

$$\left\langle i_{ng}^2 \right\rangle = 4kT_o G_{ng} \Delta f, \left\langle i_{nd}^2 \right\rangle = 4kT_o G_{nd} \Delta f,$$

where G_{ng} and G_{nd} are the equivalent noise conductivities of the transistor field. Calculating the minimum noise figure for the actual transistor gives the following relationship:

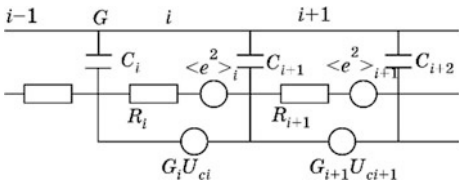


Fig. 14.26 Noise model of neighboring partial transistors

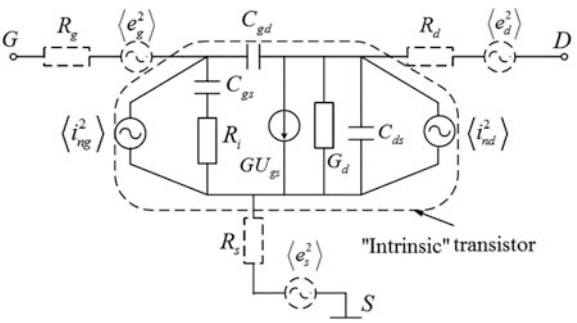


Fig. 14.27 Noise equivalent circuit of the transistor

$$N_{Fmin} = 1 + \frac{\sqrt{G_{ng} G_{nd}}}{G}, \tag{14.30}$$

where G is the transconductivity. Formula (14.30) determines the noise for some optimal parameters of the input circuit matching:

$$G_{\text{opt}} = G \cdot \sqrt{\frac{G_{nd}}{G_{ng}}},$$

$$B_{\text{opt}} = -(\omega C_{gs} + \omega C_{gd}).$$

To find the noise figure for the entire transistor, noise sources of parasite resistances are added to the circuit. To determine them, we can use the Nyquist formula:

$$\left\langle e_{nR_s}^2 \right\rangle = 4kT_0 R_s \Delta f,$$
$$\left\langle e_{nR_g}^2 \right\rangle = 4kT_0 R_g \Delta f,$$

where R_s is the parasite resistance of the source region (Figs. 14.9 and 14.10); and R_g is the gate metallization resistance.

The noise source for drain resistance R_d can be ignored, because it is at the output, where the signal is already amplified and additional noise from R_d is negligible.

14.1.8 *Noise Parameters of the Transistor as a Function of the Working Regime*

Let us consider the dependence of the noise coefficient upon the device's mode of operation. To do this, we shall use the van der Ziel formula. We write it for an arbitrary part of transistor Δx and carry out the transformations:

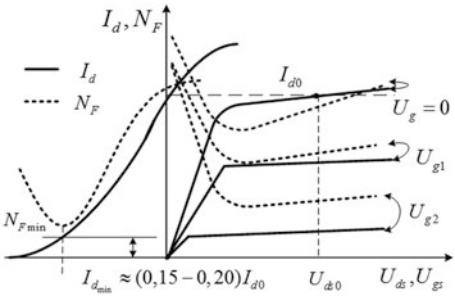


Fig. 14.28 Dependence of the noise coefficient on the operating mode of the transistor

$$\left\langle i_n^2 \right\rangle = 4e^2 D(T_e) naZ \frac{1}{\Delta x} \Delta f = 4envaZ \frac{D(T_e)}{v} \frac{e}{\Delta x} \Delta f = 4e \frac{D(T_e)}{v} I_d \frac{1}{\Delta x} \cdot \Delta f. \tag{14.31}$$

In formula (14.31), the parameters that depend on the mode are drain current I_d and the ratio of diffusion coefficient to carrier velocity $D(T_e)/v$.

First of all, we estimate the change in these quantities for the input part of the transistor. In the operating mode, when the drain current reaches saturation, the velocity in the gate part of the transistor depends little on the gate voltage and drain voltage. It tends to saturation velocity (Fig. 14.24). The diffusion coefficient varies, depending on the electron temperature, but not significantly. Then ratio $D(T_e)/v$ will change insignificantly. Turning to formula (14.31), we see that the smaller the current I_d, the less the noise current. This thesis received experimental confirmation. In Fig. 14.28, against the background of the input characteristic and the family of output, the dotted line shows the experimental dependences of the noise coefficient N_F. A specific feature of the experimental dependence $N_F = f(I_d)$ is the presence of a minimum noise coefficient at current $I_{dN_{F\min}} = (0.15 - 0.2) I_{d0}$, where I_{d0} is the current at zero voltage on the gate. This fact contradicts the stated provision on noise reduction with decreasing current, according to (14.31). The contradiction is eliminated by taking into account the influence of the buffer layer. At low currents (narrowed channel), the electrons are already warmed up in the initial part of the transistor and get the opportunity to overcome the barrier and to drift in the buffer layer. This leads to a decrease of transconductance, a decrease in gain, and to an increase in the noise coefficient accordingly. From this point of view, the minimal noise should have a symmetric transistor (Fig. 14.1), in which there is no current flow into the buffer layer.

Today we have the noise factor N_F = 1.5–2 dB in the 10–12 GHz band in classical MESFETs.

14.1.9 High Electron Mobility Field Effect Transistor

One of the ways to increase the maximum frequency of the transistor, according to (14.8), is to increase the transconductivity G. For a given gate length, this is made possible by reducing the thickness of channel A, while increasing the doping

level N_d. However, this reduces the low-field mobility, according to the empirical formula

$$\mu_0 = \frac{0.8}{\sqrt{1 + N_d[\mathrm{cm}^{-3}] \times 10^{-17}}} \left[\mathrm{m}^2/(\mathrm{V} \times \mathrm{s})\right].$$

At the same time, the saturation velocity also decreases. As a result, the desired increase in transconductance does not occur.

This mobility behavior around a temperature of 300 K is due to the predominant contribution of Coulomb scattering to dopant-impurity ions over other scattering mechanisms. To preserve the mobility of charge carriers as the doping level increases, it is necessary to conduct a spatial separation of the free carrier charge and the corresponding donor impurities.

Figure 14.29 shows the band diagram for the contact of two semiconductors with different forbidden gap widths. The broadband material is doped more strongly than the narrow band one. Due to the diffusion, part of the free carriers will pass from the region of strong doping to the region of weak doping. With such a junction, a potential barrier is formed on the boundary, which prevents further diffusion. In a low-doped region, the transferred electrons form a potential well, which limits their motion in the direction normal to the junction plane. This ensemble of charge carriers is called a 2D gas. The width of the potential well that is formed can be comparable with the de Broglie electron wavelength, which causes the quantization of energy levels in the well. The population density for different energy levels will be different. If an external voltage is applied to the structure under consideration, then the position of the energy level on the boundary and the number of charges in the 2D gas will change. The mobility of carriers in the low-doped region can substantially exceed mobility in the heavily doped region. Thus, in gallium arsenide with the doping levels indicated in Fig. 14.29, mobilities of $\mu \sim 6\ \mathrm{m}^2/(\mathrm{V\ s})$ for $T = 300$ K and $\mu \sim 100\ \mathrm{m}^2/(\mathrm{V\ s})$ for $T = 77$ K can be achieved.

Transistors using this effect were almost simultaneously invented in different countries and as such received different names: HEMT (High Electron Mobility Transistors), MoD FET (Modulated Doped Field Effect Transistors), 2DEG FET (Two Dimension Electron Gas Transistors).

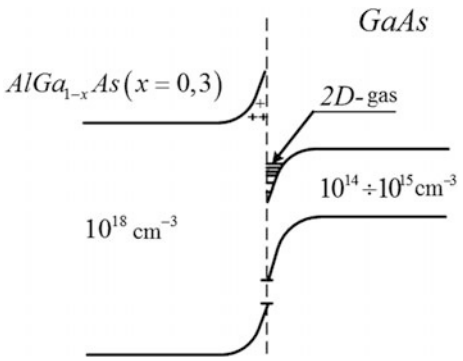


Fig. 14.29 Band diagram of a heterojunction with different doping levels

Figure 14.30 shows a possible structure with physico-topological parameters of the layers $A_k \sim 6$ nm, $N_d \sim 10^{18}$ cm^{-3}, $A_{sp} \sim$ (1–3) nm, $A_{GaAs} \sim 1$ μm. A_{sp} is the thickness of the wide-gap low-doped "spacer" layer, which reduces Coulomb scattering of the 2D gas. These layers are usually formed by molecular beam epitaxy.

Figure 14.31 shows the band diagram of the transistor structure under consideration. The energy profiles for two gate voltages are drawn. The first (dashed line) corresponds to the voltage at the gate, which does not ensure complete depletion of the channel in the high-doped region. The second (solid line) corresponds to the voltage, which completely depletes the current channel Ch_1. In this case possible act on the potential well with 2D gas with help of gate potential. The current in the resulting transistor can flow through two parallel channels: a conventional classical transistor channel Ch_1 with a high doping level—current I_{d1}, and channel Ch_2 with 2D gas—current I_{d2}. By changing the voltage on the gate, it is possible to change the fraction of the current flowing in the channels mentioned. Figure 14.32 shows a typical set of output voltage-current characteristics at different voltages at the gate. The step for changing the voltage is the same. To exclude the flow of current in part of the device corresponding to a conventional transistor, a negative voltage is applied to the gate, sufficient to completely deplete the current channel of this transistor. The figure shows that the transconductivity of the HEMP device is higher than when transistor operates in the usual MESFET mode.

In practice, thickness A_k and its doping level N_{dA_k} are selected from the condition

$$\varphi_k = \frac{1}{2} \cdot \frac{eN_{dA_k}}{\varepsilon} \cdot A_k^2. \tag{14.32}$$

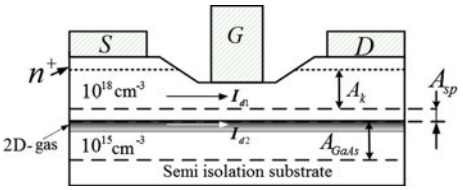


Fig. 14.30 Structure of a HEMT

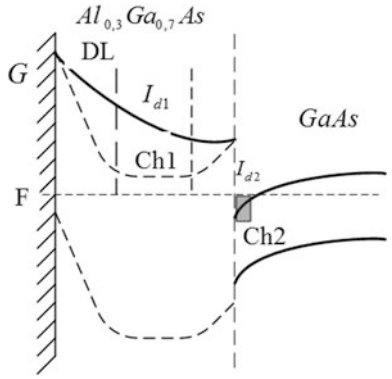


Fig. 14.31 Band structure of a HEMT

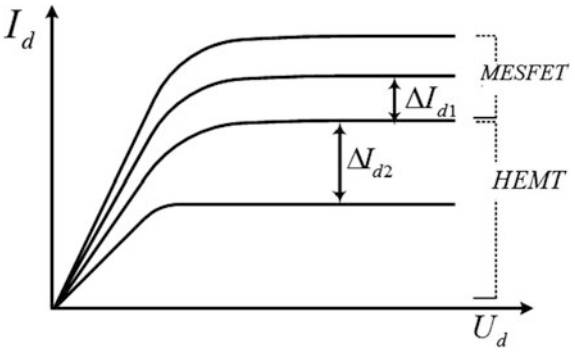


Fig. 14.32 Set of output volt-ampere characteristics for a hetero-structure FET

Formula (14.32) shows that current channel depletion of the usual transistor can be achieved even at zero voltage on the gate due to the potential difference of the contact.

One of the main technological difficulties in the manufacture of HEMTs is the provision of small contact resistances in the region of source R_s and drain R_d. Usually this is done by ion implantation. The use of such transistors makes it possible to obtain $N_{F\,\min} = 0.5-0.7$ dB at a frequency of 11 GHz.

14.1.10 *Developmental Prospects of Microwave Field Effect Transistors*

In recent years, the technology of manufacturing microwave transistors has made significant progress due to the use of new semiconductor materials. These include wide-band semiconductors and heterostructures that are based on them. Let's compare the basic physical parameters of these materials, using the data from Table 11.1 and FVC in Fig. 11.3. For clarity, Fig. 14.33 shows the band gap,

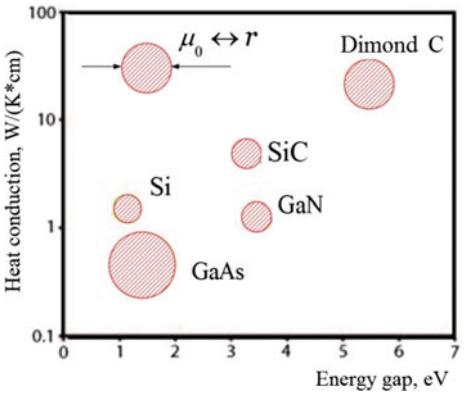


Fig. 14.33 Property comparison of semiconductors

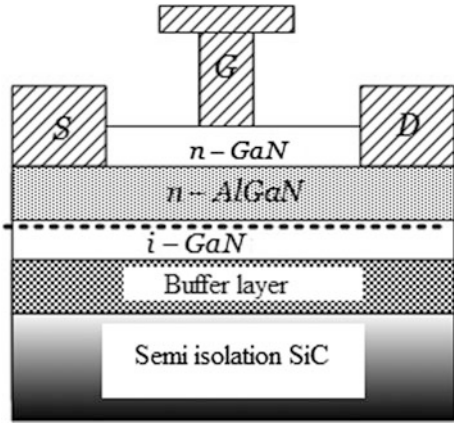


Fig. 14.34 Structure of a modern GaN-based transistor

thermal conductivity and low field mobility of some semiconductors, presented on a single graph. The mobility value is proportional to the radius of the circle, the center of which is located at the intersection of coordinates corresponding to the thermal conductivity and the width of the forbidden gap.

It follows from the graph that the preferred material for use in creating high power high frequency transistors is gallium nitride. It has a high saturation velocity $v_s \approx 2.7 \times 10^5$ m/s, and a wide energy gap. It also has a low-field mobility that is lower than that of gallium arsenide, but is substantially higher than that of silicon carbide.

In practice, heterostructures are used in a combination of the observed materials. In particular, the most promising is the structure of GaN, AlGaN/GaN on the SIC substrate, shown in Fig. 14.34. The use of a silicon carbide substrate improves heat dissipation in powerful devices. When priority is given to the price of the device, a structure on a silicon substrate is used.

Figure 14.35 shows a comparison of the working characteristics of transistors based on silicon, gallium arsenide and gallium nitride. They clearly show the fact that the admissible amplitudes of current and voltage in GaN-based devices are much higher than those for silicon and gallium arsenide transistors.

The same figure shows the hyperbola of the critical mode corresponding to the limiting permissible scattering power $P_{\max} = I_0 U_0$. This curve limits the range of permissible voltages and currents.

Success in the cultivation of artificial diamonds made it possible to use this material in microelectronics. Usually, we refer to diamond as an insulator. The use of this material was limited to the production of substrates with a unique thermal conductivity of 22 W/(K cm) (see Table 11.1). However, in recent years this material has been used to create MESFETs with maximum frequencies of up to 150 GHz. Semiconducting properties are arised in diamond not only by adding doping impurities to the volume of the material, which is technologically very difficult, but by modifying surface properties. One of the methods of modification is the hydrogenation of the surface using microwave hydrogen plasma. At the same time,

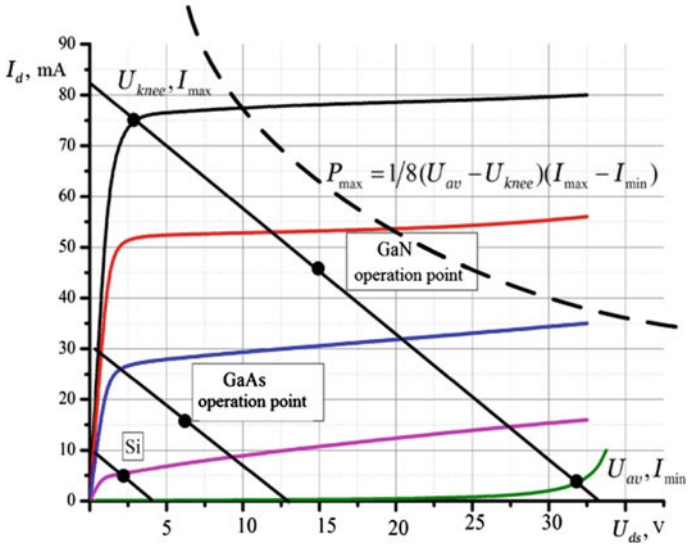


Fig. 14.35 Comparison of typical operating modes of Si-, GaAs- and GaN-based transistors

the C–H bonds formed on the surface significantly decrease electron affinity. This leads to the escape of electrons from the volume to the surface region. The resulting bending of the valence band is sufficient to form hole conduction. Indeed, experiments have shown the presence of a layer of *p*-type conductivity. Surface conductivity increases by an order of four. The thickness of this *p*-layer is 1–2 nm and the depth of "deposition" is 10 nm. Figure 14.36 shows the structure (a) and a

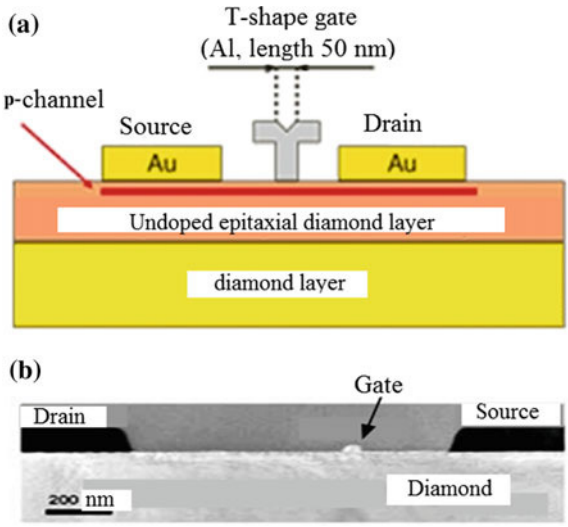


Fig. 14.36 Transistor based on diamond

photograph of such a transistor obtained using a scanning electronic microscope (b). The gate length is 50 nm. Recently, technologies for alloying diamonds with boron (an acceptor impurity) and vanadium (a donor impurity) have been developed.

14.2 Microwave Bipolar Transistors

14.2.1 Structure and Operating Principle

In the lower part of the microwave band (less than 15 GHz) classical bipolar transistors (BT) are used as power amplifiers and generators with low frequency noises. The main advantages of such devices are their low cost, low flicker noise and high reliability, determined by the debugging of silicon technology.

Let us consider the features of a bipolar device operating in the microwave band. As already mentioned in Sect. 14.1.1, the improvement of the inertial characteristics of such devices is achieved by reducing the size, especially the length of the base ω_b. However, this increases the probability of base puncture, i.e., the closing of the depletion layers of the emitter-base and the base-collector. To avoid puncture, it is necessary to increase the level of doping in the base. As a consequence, the emitter efficiency factor falls. Possible ways to resolve these contradictions are discussed in Sect. 14.2.3.

Let us analyze what levels of doping layers, sizes and materials allow an increase to the boundary frequency of bipolar transistors. To do this, consider the typical structure of the device, the cross-section of which is shown schematically in Fig. 14.37. In the figure, resistors and capacitances characterizing the distributed (in space) nature of the current flow between the emitter and collector base are shown in the base region. Different parts of the base have different potentials relative to the emitter. The end base regions work more efficiently than the middle ones, because of the voltage drop on the resistive elements of the base. That is why the current lines in Fig. 14.37 have a large density along the edges of the base. The effect of

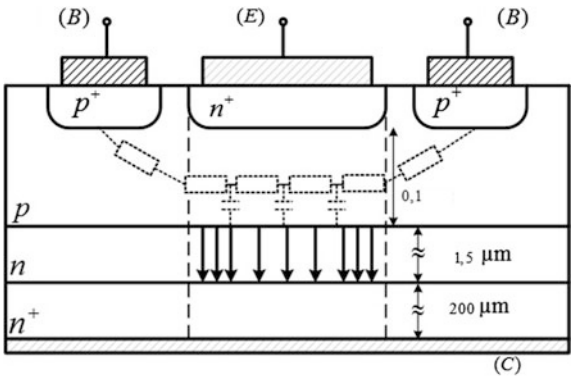


Fig. 14.37 Transverse section of a bipolar device

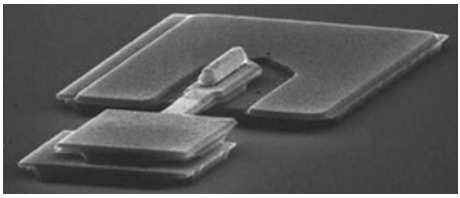


Fig. 14.38 Cell of a comb-structure transistor

current concentration along the edges of the emitter electrode is most noticeable in microwave devices, when the thickness of the base should be reduced, which causes an increase in its resistance. To reduce this effect, the ratio of the area of the emitter electrode to its perimeter is minimized. In this regard, powerful microwave bipolar transistors are built in the form of cellular or comb structures. The width of the emitter electrode in such devices is 1–5 μm. The topology of such a device is illustrated in Fig. 14.38. It shows a photograph of one cell of the transistor. The central electrode (emitter) is surrounded along almost its entire perimeter by the base electrode. A structure consisting of several cells connected in parallel is called the comb structure.

Typical concentration levels of mobile carriers for a microwave *n-p-n*-transistor are shown in Fig. 14.39. Thin dashed lines show the boundaries of the depletion zones at the emitter-base and collector-base boundaries. The figure also shows the external circuit, which provides height change of the potential barriers formed by the fields in the depletion layers.

In general, a transistor can be represented by a connection of two counter-connected diodes. The main feature of such a connection is a thin neutral base through which the mobile carriers (in this case electrons) can *diffuse* from the emitter-base region into the collector-base depletion layer. An accelerating electric field exists in this layer. As a result, the bulk of the charge carriers from the emitter

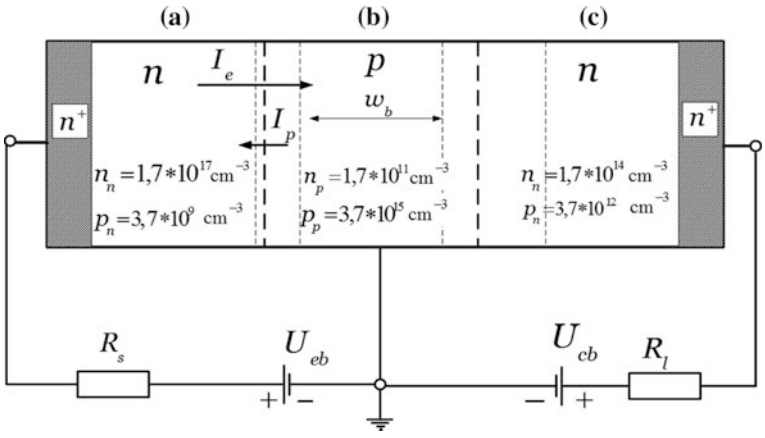


Fig. 14.39 Typical doping levels of a microwave BJT

reaches the collector in time τ_{ec}. Thus, for the efficient operation of the bipolar transistor, it is necessary that the diffusion length of the carriers be greater than the length of the base, i.e. $l_d > \omega_b$.

Bipolar transistors have three electrodes and can be included in three different kinds of circuit, as well as vacuum triodes (see Sect. 6.4). The possible switching circuits are: with a common emitter (analog with a common cathode), with a common base (analog with a common grid), and with a common collector (analog with a common anode). In the microwave band, a common emitter circuit is generally used, providing maximum gain.

14.2.2 Equivalent Circuits and HF Parameters of BT

Analysis of the high-frequency properties of a bipolar transistor, as in the case of field transistors, is carried out using an equivalent circuit. Figure 14.40 shows the equivalent circuit of the transistor, based on its topology (Fig. 14.37). The physical interpretation of the individual elements of the circuit with this approach is fairly obvious. Nonlinear resistance R_{eb} and capacitance C_{eb} characterize the properties of the emitter-based barrier. Elements r_{b_i} and C_{dr} correspond to the previously mentioned distributed character of current transfer in the base.

These elements allow us to take into account the change in the control potential along the base layer. The current source αI_e denotes the injection into the depletion layer of the collector-base emitter current part I_e. The parameter α is called the current transfer coefficient in the common-emitter circuit. Together with this value, we use the value $\beta = \alpha/(1-\alpha)$, which is called the current amplification factor.

Using electrical engineering methods, a distributed circuit can be brought to the simplified circuit shown in Fig. 14.40b.

Let us analyze what parameters of the structure in question affect the limiting and maximum frequency of the transistor. The total transit time consists of base transit time τ_b, transit time of the emitter-base junction τ_{eb}, transit time of

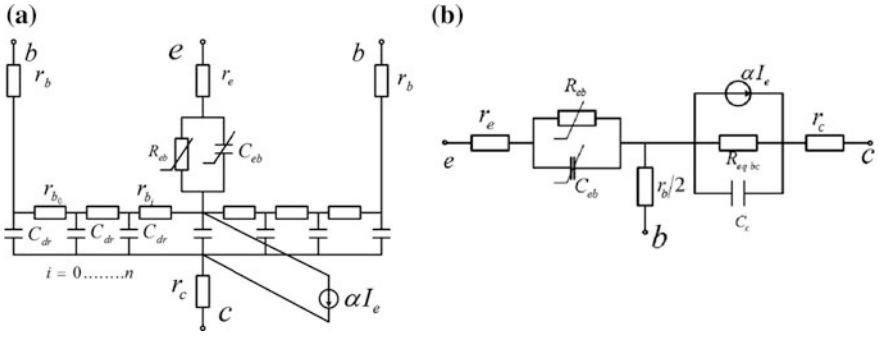


Fig. 14.40 Equivalent bipolar transistor circuits: **a** in-depth; **b** simplified

the base-collector junction—τ_{bc} and the transit time of the collector depletion layer τ_c, i.e.

$$\tau_\Sigma = \tau_{eb} + \tau_b + \tau_{bc} + \tau_c.$$

In a properly constructed transistor, τ_b plays the main role in delay. However, τ_{bc} also plays an important role.

High-frequency transistor parameters are usually characterized by the cut-off frequency (the critical frequency f_t or $\omega_t = 2\pi f_t$). It is defined as the frequency at which the current gain in the short circuit mode of the common emitter circuit is equal to one. The critical frequency is related to the delay time τ_Σ:

$$f_t = 1/2\pi\tau_\Sigma.$$

Another important parameter is the maximum frequency $f_{\max}$, on which the power gain in the circuit with a common emitter is equal to one. To find $f_{\max}$, we simplify the circuit (Fig. 14.40b), leaving the most important elements: the base resistance r_b at the input, and collector capacitance C_c.

The maximum power in the active load G_L is observed when the currents through capacitance C_c and through the load are equal. This condition is fulfilled only at one frequency. Suppose that it is satisfied at the critical frequency, that is $G_L = \omega_t C_c$. If the input is matched, the power gain factor will be written in the form

$$\frac{P_{\text{out}}}{P_{\text{in}}} = \frac{\left(\frac{\beta I_{\text{in}}}{2}\right)^2 / G_L}{I_{\text{in}}^2 r_b} = \frac{\beta^2}{4 r_b G_L} = \frac{(\omega_t/\omega)^2}{4 r_b C_c \omega_t} = \frac{\omega_t}{4 r_b C_c \omega^2}. \tag{14.33}$$

For $\omega = \omega_{\max}$, $P_{\text{out}}/P_{\text{in}} = 1$ then

$$\frac{\omega_t}{4 r_b C_c \omega_{\max}^2} = 1$$

and

$$\omega_{\max} = \sqrt{\frac{\omega_t}{4 r_b C_c}}. \tag{14.34}$$

If we represent the dependence of the power gain coefficient on frequency on a logarithmic scale, we obtain

$$10 \lg \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \lg \frac{1}{4 r_b C_c \omega_t} + 10 \lg \left(\frac{\omega_t}{\omega}\right)^2 = G_0 + 20 \lg \frac{\omega_t}{\omega}.$$

This dependence is shown in Fig. 14.41. It is characterized by a drop in gain of 6 dB for a frequency octave in the frequency region above f_t.

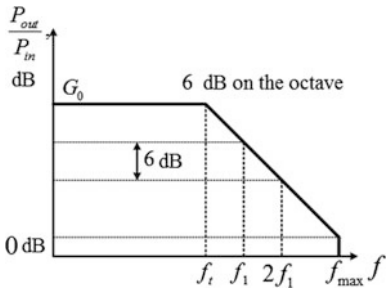


Fig. 14.41 Typical frequency response of a bipolar device

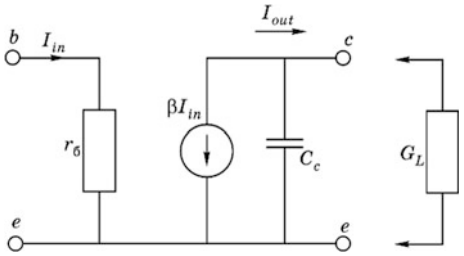


Fig. 14.42 Hybrid equivalent π-circuit of a BT

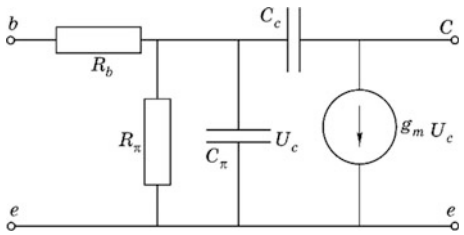


Fig. 14.43 Calculating the maximum frequency of a transistor

When describing the high-frequency characteristics of bipolar transistors, a so-called hybrid π-circuit is often used in practice (Fig. 14.42). Unlike the circuit in Fig. 14.43, the output current source of this circuit is controlled by the emitter-base voltage, and not by the base current. In this variant, as in the field transistor, one can apply the concept of transconductivity g_m, defining it by the following expression:

$$g_m = \partial I_c / \partial U_{eb}. \tag{14.35}$$

The value of g_m is easy to estimate using expression (11.15) for current I_e through the emitter-base junction:

$$I_e = I_s[1 - \exp(eU_{eb}/kT)].$$

Taking into account the fact that the collector current is approximately equal to the emitter current, the expression for steepness (14.35) leads to the form

$$g_m \approx \frac{e}{kT}\exp(eU_{eb}/kT) \approx \frac{e}{kT}I_e = \frac{I_e[\text{mA}]}{26}. \tag{14.36}$$

Formula (14.36) is valid for temperature $T = 300$ K.

The remaining parameters of the π-circuit can be found from the recalculation of the distributed circuit (Fig. 14.40a) or by using the measured s-parameters. As an example, the characteristic values for a transistor in the centimeter wave band are presented below:

$$R_b = 7\ \Omega, \quad R_\pi = 110\ \Omega, \quad C_\pi = 18\ \text{pF}, \quad C_c = 18\ \text{pF}, \quad g_m = 900\ \text{ms}$$

Just as for a field transistor, the critical frequency is given by the expression

$$f_t = g_m/2\pi C_\pi$$

For the presented values of the π-circuit parameters

$$f_t = 900 \times 10^{-3}/(2\pi \times 18 \times 10^{-12}) \approx 16\ \text{GHz}$$

The equivalent circuits of bipolar transistors, just like the circuits of field transistors, are particularly convenient for analyzing nonlinear operation modes. In this regard, many firms that manufacture such devices provide, along with s-parameters, linear and nonlinear transistor models.

14.2.3 Heterojunction Bipolar Transistors

The main way to increase the critical frequency of bipolar transistors is to reduce the width of the base while increasing the doping level. The limiting factor in this approach is the decreasing of emitter efficiency: $\gamma_e = I_n/(I_n + I_p)$. The value γ_e is determined by the fraction of electrons going to the collector, in the total emitter current. When the efficiency of the emitter falls, the values of α and β decrease. As a result, no expected improvement is observed in the HF parameters of the device. In order to increase the doping level in the base without reducing the emitter efficiency, a heterostructure emitter-base junction is used. A bipolar transistor with such layers is called a heterojunction BT (HBT).

The parameters of the layers to be contacted are selected in such a way as to provide a different height to the potential barrier for electrons and holes. Figure 14.44 shows one of the options for implementing such a structure. The figure shows the band diagram of a n-p-n-transistor with a wide-band emitter and narrow-band base and collector: $W_{ge} > W_{gb}, W_{gc}$. The ratio of the hole current I_p, to

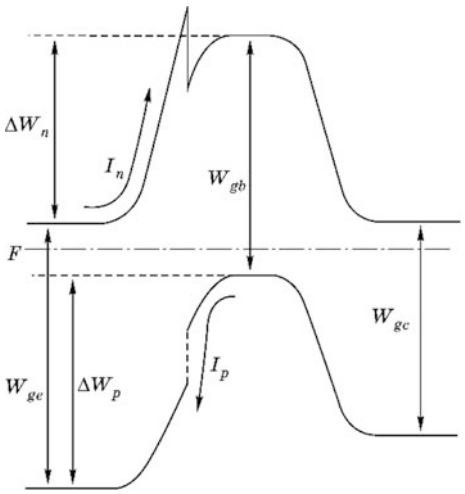


Fig. 14.44 Band diagram of a transistor with a wide-band emitter

the electron current I_n, is determined by the difference in the energy barriers and is written in the form

$$I_p/I_n = \mathrm{e}^{-\frac{\Delta W_p - \Delta W_n}{kT}}, \tag{14.37}$$

where W_n and W_p are the heights of the potential barriers for electrons and holes respectively. Using this approach, it was possible to raise the level of doping in the base by 2–3 orders. Such an increase in the level of doping allows not only the reducing of the thickness of the base, but also significantly decrease its parasite resistance. In general, this leads to an increase in the critical frequency.

Similar structures are made of materials such as AlGaAs and InP. The most widespread devices in recent years are those using indium phosphide. This material is characterized by a big low-field mobility and a higher saturation velocity than that of GaAs (see Table 11.1). The use of InP in these devices makes it possible to obtain a large difference in ΔW_n and ΔW_p.

The cross section of the heterojunction transistor is shown in Fig. 14.45. For a base doping level of 6×10^{19} cm^{-3}, an emitter doping level of 4×10^{19} cm^{-3}, a base thickness of 25 nm, and an emitter size of 0.35×12 μm, a critical frequency of about 500 GHz is obtained. At a current of 1 mA, the breakdown voltage was 2.7 V.

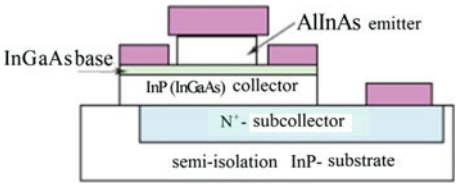


Fig. 14.45 Structure of a transistor with a wide-band emitter

14.3 Microwave Transistor Specifics

14.3.1 Physical and Technological Limitations of Creating Microwave Transistors

Transistors, just like vacuum triodes (see Chap. 6), are devices with quasi-static control. For their operation in the microwave band, it is necessary to reduce the transit angle of carriers in the space of interaction of charge carriers with the field and to reduce the parasite parameters of the structure as a whole. Technological methods for fulfilling these conditions in bipolar and field effect transistors vary considerably.

Let's analyze the physical and technological limitations when creating microwave transistors. Figure 14.46 presents simplified schemes of bipolar (a) and field effect (b) transistors without parasitic elements. Note that the current in the bipolar device runs across the layer, and in the field device it runs along the semiconductor layer. The key time parameters that determine the high-frequency properties of the devices are the carrier diffusion time through the base of the bipolar transistor $\tau_b \approx \omega_b/(D\nabla n/n)$ and the drift time under the gate $\tau_g \approx L_g/v_s$ respectively. In the given expression for the transfer time τ_b, value $D\nabla n/n$ determines the rate of carrier diffusion through the base (see formula (2.42)). Here, the influence of the drift component is not taken into account, which is manifested only in the special transistor with a variable band gap in matereal of base (*varyzone transistors*). In the first approximation, we assume the value $\nabla n \approx n/\omega_b$. Then the expression for the transfer time in the base is written as follows:

$$\tau_b \approx \omega_b^2/D.$$

Technologically, it is easier to make a thin base than to make a short gate using lithography. Therefore, the bipolar device has certain application potential in the microwave region of frequencies. The limitations to its application are associated with large parasite resistances, due to the low mobility of electrons in silicon. The use of a material with high carrier mobility, in particular gallium arsenide, in such a structure is difficult, due to the lack of stable compounds providing a *p*-type semiconductor. The field transistor, being unipolar, can be fabricated using only *n*-type GaAs and a controlling Schottky barrier.

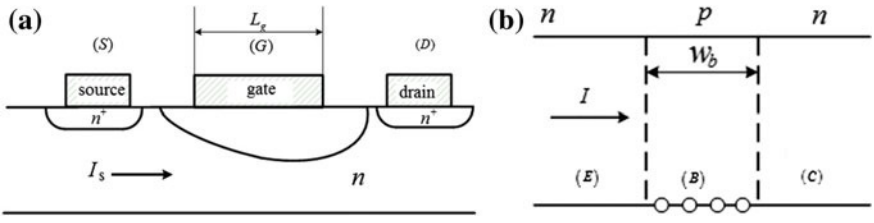


Fig. 14.46 Structure of field and bipolar transistors

Let us estimate the order of the quantities characterizing the base thickness and the gate length when the device is operated at a frequency of 100 GHz. Set the transit angle $\theta = \pi/2$ radians and the diffusion coefficient $D = 0.01$ m^2/s. Then

$$\omega_b \approx \sqrt{(\theta/\omega)D} = \sqrt{0.25TD} = \sqrt{0.25 \times 10^{-11} \times 0.01} \approx 0.16 \text{ μm};$$

$$L_g \approx v_s \tau_g = 10^5 \times 0.25T = 10^5 \times 0.25 \times 10^{-11} = 0.25 \text{ μm}.$$

Modern technological processes make it possible to fabricate BTs with base thicknesses of 0.01 μm and a MESFET gate length of 0.1–1 μm, which allow bipolar transistors and field transistors to operate at frequencies of around 250 GHz.

With the indicated dimensions and a voltage drop of 5 V in the depletion zone, the field strength can be 320 kV/cm, exceeding the breakdown field for silicon.

The submicron dimensions of the devices not only determine the technological difficulties in the manufacture of such structures, but they also impose higher requirements on the analytic methods for physical processes and design. In particular, it is necessary to take into account the non-local nature of velocity and field strength coupling (overshooting), and the quantum approach to describing the transport of carriers in narrow potential barriers that arise in heterostructures (see Chaps. 2 and 4). For example, with a base length of 0.01 μm and a doping level in it of 10^{16} cm^{-3}, it contains only one donor layer. In this situation, the application of the hydrodynamic approximation and corresponding equations will only give approximate results for the device models.

14.3.2 Transistor "Family Tree"

We systematize the variety of existing transistors, and the changing topology of the simplest device structure. This establishes the genealogical relationships of various devices and helps to understand the direction of transistor development. We begin the analysis with a *planar* field-effect transistor with a Schottky barrier (Fig. 14.47). A MESFET belongs to the class of unipolar devices, working on only one type of charge carrier principle: electrons or holes.

Consider the *n*-type channel. The ohmic contacts on the source and drain are made in the form of a degenerate semiconductor layer n^+ and with a metal (see Sect. 11.5). Let's make two device sandwiches, connecting them with metal electrodes to each other. Connecting in parallel several sandwich structures we get the

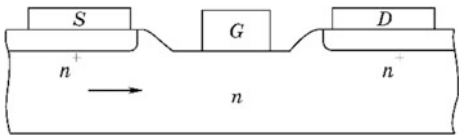


Fig. 14.47 Horizontal MESFET structure

device shown in Fig. 14.48. It is a volumetric structure of a field transistor with multiple gates.

Turning the resulting structure through 90°, we get the device shown in Fig. 14.49. If we now make the gate in the form of a thin metal grid, we get the device shown in Fig. 14.50. This device is called a permeable base transistor. In this case, the gate represents a certain barrier and the current in the structure flows perpendicular to the layers of the semiconductor.

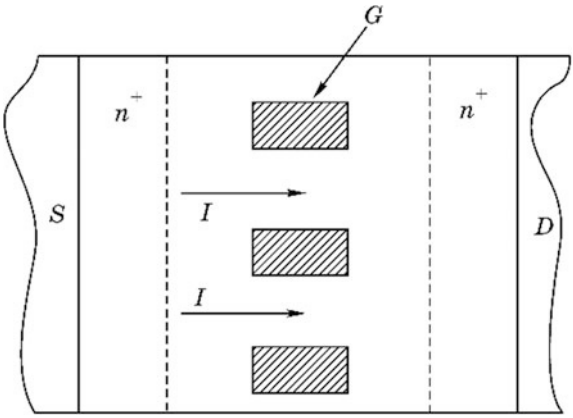


Fig. 14.48 Structure with several gates

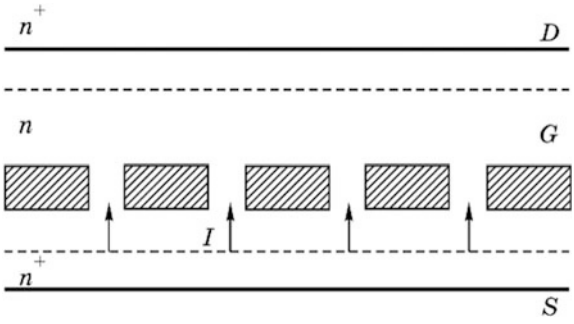


Fig. 14.49 Structure with vertical (cross-sectional) current direction relative to active layers

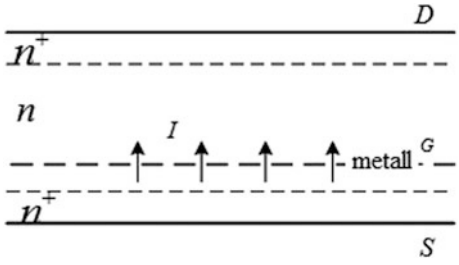


Fig. 14.50 Permeable base transistor

In all considered structures, metal is used for shielding the region of control from the region where there is an extraction of energy from the flux of the carriers. In view of the high concentration of free electrons in metal, such a screening can be realized by a very thin layer of this metal. This screening function can also be provided by a material of a different type of conductivity, that is, of *p*-type. As a real semiconductor has a smaller concentration of holes than the concentration of electrons in the metal, the thickness of the shielding region should be increased. Its minimum thickness corresponds to the Debye length (see Appendix A).

Figure 14.51 shows a structure that is called a bipolar transistor. The electrode name is changed to that accepted in technical literature: a drain here is called a collector, a source is called an emitter, and a gate is called a base.

In contrast to the field transistor, in this device, both holes and electrons are used simultaneously. Hence, the name "bipolar".

The considered topological prototypes can be with the help of various technological methods. Devices sometimes receive their own characteristic names. For example, Fig. 14.52 shows the structure of the device, which the developers called the SIT—Static Inductor Transistor. This device is similar to the parallel connection of vacuum triodes, in which quasi-static control is used. By minimizing the size of such a structure, it is possible to achieve conditions of collisionless carrier transport under the gate. Such a current transfer is called ballistic transport, and the devices are known as BTT—Ballistic Transport Transistors.

Figure 14.53 shows the classification of modern transistors that are widely used in the microwave band. The main types of transistors in this scheme correspond to the constructed "family tree". Let us single out the field effect transistor with a Schottky barrier. The abbreviation of such a device is MESFET (MEtal-Semiconductor Field Effect Transistor). This type of transistor is most common in microwave electronics.

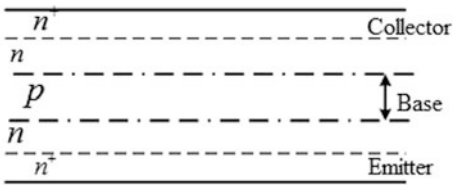


Fig. 14.51 Bipolar junction transistor

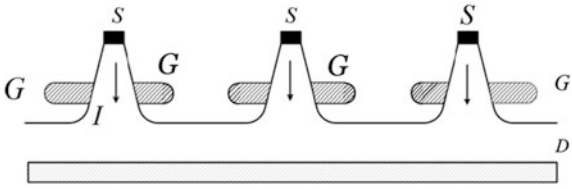


Fig. 14.52 Structure of a transistor with static induction

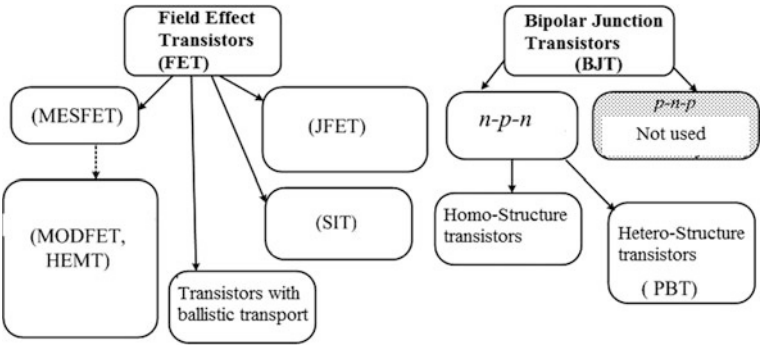


Fig. 14.53 Classification of microwave transistors

14.3.3 Comparison of Transistor Speeds

Let's compare the limiting parameters of the main types of transistors. Conditions for the comparison are as follows: the devices must be made of the same material and have the same active cross-sectional area (for current flow) S.

Table 14.2 shows the following data: total delay time of the signal, critical frequency f_t, and maximum operating frequency $f_{\max}$ for gallium arsenide with $S = 100\ \mu m^2$. Remember that the critical frequency is the frequency at which the current gain is equal to one and the maximum frequency corresponds to the power gain equal to one in the common emitter (or common source) scheme. The table shows both calculated and experimental data for four types of transistor: MESFET, BT, permeable base transistor and ballistic transport transistor.

Note that comparison takes place for an equal area of current flow, and not for a given gate length or base thickness. For a field transistor this value is $S = AZ$, where A is the thickness of the channel, and Z is the width of the gate. For a bipolar transistor: emitter length d_e, area $S = d_eZ$. The comparison shows the potential of using transistors with a permeable base and devices with ballistic transport.

Table 14.2 Comparison of fast-switching transistors

Transistors	τ (ps)	f_c (GHz)		$f_{\max}$ (GHz)	
		Theory	Experiment	Theory	Experiment
Bipolar transistor (BT)	26	5	5	6.7	6.7
Field transistor (FT)	7.1	10	10	13	13
Permeable base transistor	1.9	36	36	58	58
Ballistic transport transistor	0.2	400	?	?	?

14.3.4 New Type of Transistors: Graphene FET

Progress in studying the properties of graphene made it possible to use this material for designing field effect transistors. High electron mobility and saturation velocity provide theoretically very high working frequencies lying in the terahertz band. The common scheme of field-effect transistors (FET) cannot be applied to a graphene transistor since graphene sheets have no band-gap. Hence, a transistor has no cutoff at negative source-drain voltages. This feature does not negate the building of analog RF transistors, but without cutoff, they cannot be effective.

As was written in Sect. 11.2, a finite band-gap forms in graphene nanoribbons with an armchair edges structure. However, numerous lattice defects near the edges prevent the creation of transistors on GNR.

L. Ponomarenko and colleagues at the University of Manchester proposed another approach for designing a transistor with a high on/off ratio. They used two graphene layers, separated by a thin (several atoms thick) layer of boron nitride (BN) or molybdenum disulphide (MoS_2). The structure of this transistor with vertical transport is shown on Fig. 14.54. The source contact lays on the lower graphene sheet, and the drain contact is connected to the upper graphene sheet. Applying a voltage between source and drain, we stimulate electrons to tunnel through the insulating layer. The probability of tunneling depends on the gate potential. The gate contact is isolated from graphene by a thin silicon oxide layer. Because of the very fast tunneling process, this transistor can have a very high cut-off frequency.

The first graphene transistors had low cut-off frequencies, not exceeding 10 GHz, because of low electron mobility caused by scattering from lattice defects, impurities and substrate influence. However, technological progress and development of new types of transistor structure provide fast enhancement of transistor parameters. Figure 14.55 shows the structure of a MESFET graphene transistor on a thin diamond-like sheet, placed on a standard silicon substrate. Diamond is a non-polar dielectric, so it does not trap and scatter charge carriers as silicon does. With a gate length of about 40 μm, these transistor have a cut-off frequency of

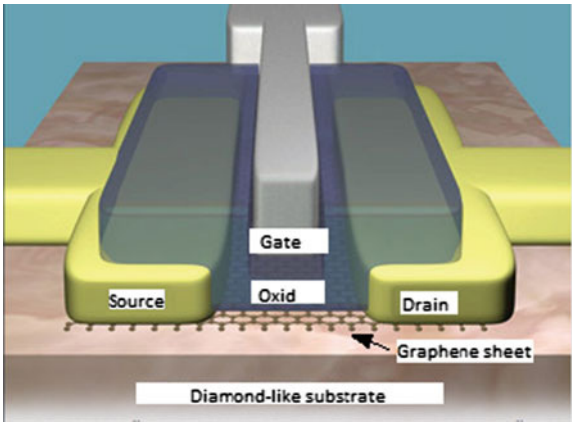


Fig. 14.54 Graphene FET with 40 μm gate length

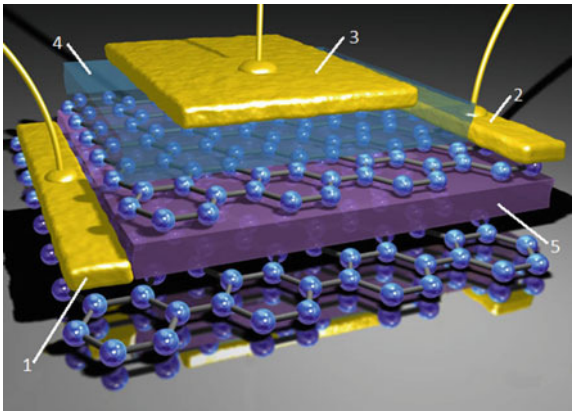


Fig. 14.55 Graphene FET with vertical transport

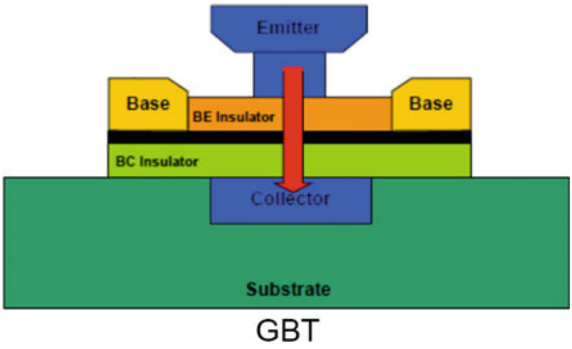


Fig. 14.56 Graphene bipolar transistor. A graphene sheet is shown in black

155 GHz. Further cut-off frequency increase can be obtain by using graphene with higher mobility and shortened gate width.

Bipolar graphene transistors are also candidates for terahertz application. One of possible structures of these transistors is shown in Fig. 14.56. Charge carriers in these transistors move perpendicular to the graphene surface through an insulating boron nitride layer thanks to the quantum tunneling.

Both research on, and the technology of, graphene transistors are developing fast and soon terahertz graphene devices will be available on the market.

14.4 Using Transistors in Hybrid and Monolithic IC in the Microwave Band

The creation of hybrid and monolithic integrated circuits is the natural development of microwave devices and systems from the point of view of micro-miniaturization of radio electronic devices. Hybrid circuits are made on substrates of various

Fig. 14.57 Hybrid microwave band circuit

materials: duroid, Teflon, and polycor (ceramic Al_2O_3). The required functional structure is created on these substrates by mounting the active and passive elements. Figure 14.57 shows a similar hybrid circuit with coaxial-to-microstrip adapters.

A characteristic feature of such circuits is the possibility of using tuned elements that improve the characteristics of the circuit as a whole. Small dielectric losses in the substrates make it possible to obtain good noise parameters. However, the technology of assembly and installation of such circuits is complex for mass production.

As an alternative to such IC, there are circuits with conventional integrated circuit technology, when both active and passive elements are produced in a single technological process. Such circuits are called *monolithic microwave integrated circuits* (MMIC). As a substrate material, silicon or gallium arsenide is commonly used. Figure 14.58 shows an integrated circuit with GaAs. It shows an amplifier with matching circuits created on the lumped elements L and C. Transceiver modules are manufactured using this technology. Under conditions of mass production, the price of the module becomes acceptable to the market, despite the high cost of materials and technology. An important advantage of such circuits is their miniature size, the feature that is very important in mobile communication systems.

Achievements in the creation of MMIC can be illustrated by an example: a transceiver module in the 30 GHz band has dimensions of $4 \times 8 \times 3$ mm.

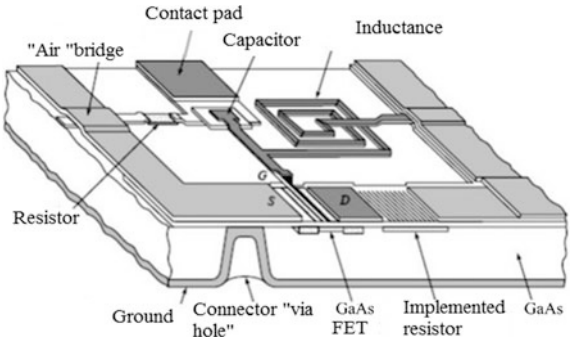


Fig. 14.58 MMIC fragment of microwave band for GaAs

Advancement Questions

1. What physical factors limit the speed of field and bipolar transistors?
2. Why is gallium arsenide preferred when choosing materials?
3. Illustrate the "genealogical" connection between field and bipolar transistors. Include a vacuum triode in the structures under consideration.
4. Compare the performance of different types of transistors.
5. Draw a cross section of the planar FT. Why does the FT need a buffer layer?
6. How can the power of the FT be increased?
7. Draw the input and output characteristics of the FT. How do you define differential parameters? What is the difference between the differential parameters and the small-signal parameters of the equivalent circuit?
8. Draw equivalent circuit of the FT and find dependes of its parameters from the physico-topological parameters of the structure. What kind of circuits do you know? What is the cutoff frequency? What is the maximum?
9. What is the physical meaning of scattering parameters? How are scattering parameters related to the parameters of the equivalent circuit?
10. What are the main assumptions in the Shockley model? Why do rough approximations give acceptable results? What is the local-field approximation?
11. What is the difference between the Shockley model and the model of two regions?
12. What are the main points of the transistor temperature model? Explain the effect of super-velocity.
13. Explain the FT operating principle from the viewpoint of the temperature model. Draw the electric field distribution and the electron temperature distribution. What is the shape of the depletion layer and the current channel?
14. What is the nature (sources) of noise in a MESFET? Why do the transistors have small noises, while the charge carriers in the current channel have a high electron temperature?
15. Explain the impedance field method when calculating the noise of a FT.
16. How do device noise parameters depend on the power mode?
17. What is the structure of a HEMT? Why does it have a high steepness?
18. What is the advantage of using modern materials: GaN, SiC, and C?
19. Compare the load characteristics of transistors made of different semiconductors.
20. Draw structure of the classical microwave bipolar transistor. What are the doping levels of the main regions? What are the ways to increase the working frequency of BT?
21. Show the relationship of the equivalent circuit elements with the structure of the device. How are the cutoff and maximum frequency of BT determined?
22. What is the structure of a transistor with a wide-band emitter? What is the efficiency of the emitter?
23. Explain the frequency response of BT.
24. What is the difference between hybrid and monolithic microwave IC?

Appendix A
Time and Space Intervals Defining the Behavior of Charged Particles

The motion of charged carriers in semiconductors occurs in a medium filled with charged ions and neutral atoms. Under equilibrium conditions, any macroscopic volume is quasi-neutral. Since charges experience thermalfluctuations, the neutrality condition is satisfied on average for a certain time interval and in a certain volume. Plasma gas has similar properties. By analogy with it in semiconductors, a system of charged particles and neutral atoms is called *solid-stateplasma*. The properties of the system can be determined by perturbing it externally and observing the process of return to a stationary state (relaxation). Such an effect may be obtained with inhomogeneous heating, injection of charge carriers, and the effect of electromagnetic field, etc. The process of returning to the initial state occurs due to internal processes inherent to the system under consideration (recombination, dissipation of energy in scattering processes, etc.). Specific relaxation *time* scales quite fully characterize plasma. A perturbation in time has a spatial localization scale for the perturbation associated with it. The nature of the system reaction to the external electromagnetic action will depend on the ratio of time (oscillation period T) and space (wavelength) parameters of this action in the investigated solid-state plasma.

A.1 Relaxation Time of Momentum and Energy

Relaxation timeτ_p of momentum $\mathbf{p} = m\mathbf{v}$ characterizes the rate of change (decrease) in the momentum of the particle when perturbation is removed. According to the approximation of the relaxation time, we write

$$\partial \mathbf{p}/\partial t = -\mathbf{p}/\tau_p.$$

The solution of this equation has the form:

$$\mathbf{p}(t) = \mathbf{p}(0)\exp(-t/\tau_p).$$

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6

The value of τ_p is the mean free path time. If this value is multiplied by the root-mean-squareof thermal velocity v_T, then we obtain the mean free path of the particle l_p:

$$l_p \approx v_T \tau_p. \tag{A.1.1}$$

Another relaxation constant characterizes energy relaxation. When scattered on ions and neutral atoms, the particle (electron, hole) loses a small amount of energy because of the large difference in the masses of the colliding particles. Therefore, for a noticeable energy loss of the particle, it is necessary to make a large number of collisions and the energy relaxation time is much longer than the pulse relaxation time. We rewrite the equation of conservation of energy (2.4.9) when the external influence is removed and without taking into account the spatial inhomogeneity of the energy flux. In this case

$$\frac{\partial(T_e - T_0)}{\partial t} = -\frac{(T_e - T_0)}{\tau_\varepsilon(T_e)}, \tag{A.1.2}$$

where T_e is the actual electron temperature; T_0 is the equilibrium temperature, equal to the temperature of crystal lattice; and $\tau_\varepsilon(T_e)$ is the energy relaxation time. The solution of this equation is $\Delta T(t) = T_e(t) - T_0 = \Delta T(0)\exp(-t/\tau_\varepsilon)$

It should be noted that the use of the *electron temperature* concept is valid only when the energy equilibrium inside the electron gas is reached quicker than the energy equilibrium in the lattice. This can only be the case when the frequency of electron-electron collisions exceeds electron-lattice collision frequency or $\tau_{ee} \ll \tau_\varepsilon$. Here τ_{ee} is the characteristic duration of electron-electron collisions.

By analogy with the mean free path, we assume a certain distance at which the energy is relaxed. Let's call this distance "the cooling length"l_ε. When applying an external electric field, it is more logical to call this value "the heating length". We estimate this length as follows:

$$l_\varepsilon \approx v_T \tau_\varepsilon. \tag{A.1.3}$$

The use of this concept is important for the analysis of current transfer features in structures with submicron dimensions.

Consider the characteristic values τ_p and τ_ε, as well as l_p and l_ε. It should be noted that the introduction of a single relaxation time for a variety of relaxation mechanisms, such as phonon scattering, ionized atoms, neutral particles, etc., is not rigorous. In addition, the particle energy also has a significant effect on the value of the parameters considered. Therefore, we can talk about some average (effective) values and ranges of their change. In semiconductor structures, the momentum relaxation time $\tau_p \sim 10^{-13}$–10^{-14} s, and energy $\tau_\varepsilon \sim 10^{-11}$–$10^{-12}$ s, i.e., τ_ε. Sometimes it is said that the "electron memory" for momentum is much shorter than the "electron memory" for energy. In other words: the relaxation process of the heated electron gas in plasma occurs in such a way that the directed carrier velocity is first lost, and then the lattice temperature and carriers are equalized.

Thus, under the conditions of elastic scattering $l_\varepsilon \gg l_p$. For T = 300 K RMS velocity $v_T \sim 10^5$m/s. Then $l_p \sim 10^{-8}$–10^{-9} m, and $l_\varepsilon \sim 10^{-6}$–$10^{-7}$ m. Note that the average length of heating (cooling) can be of the order of 0.1–1 μm, which is comparable with the size of the active regions of modern microwave devices.

A.2 Time and Length of Charge Relaxation

The work of any electronic device is associated with the creation of excess charges in a certain space. Naturally, due to electrical interaction forces, such a charge inhomogeneity will change in time. The characteristic charge relaxation parameters can be found using the Maxwell equations. This effect explains the use of the terms *Maxwellian chargerelaxation time* and *relaxation length*.

Consider a medium with conductivity σ and dielectric constant ε, into which a charge with volume density ρ is placed. Calculating the divergence of the right and left parts of the first Maxwell equation (2.2.2) and the Poisson equation (2.2.4), we obtain the equation for ρ:

$$\frac{\partial \rho}{\partial t} = -\frac{\rho\sigma}{\varepsilon}, \tag{A.2.1}$$

from which

$$\rho = \rho_0 \mathrm{e}^{-t/\tau_m}, \tag{A.2.2}$$

where

$$\tau_m = \varepsilon/\sigma = \varepsilon/(en\mu). \tag{A.2.3}$$

is the Maxwellian relaxation time; ρ_0 is the charge at time $t = 0$; n is the charged particle concentration; and μ is the charged particle mobility.

It follows from (A.2.2) that charge inhomogeneity will decrease (dissipate) under ordinary conditions if $\mu > 0$. Such an expected and natural reaction of the medium to an excess charge in some region is substantially complicated if not only is a small variable field, but also a sufficiently large static (heating) electric field are applied to the sample. In this case, the kinetic coefficients of current transfer (mobility, diffusion) change, and accordingly, the nature of the excess charge relaxation also changes. Figure A.1 shows a typical A_3B_5 semiconductor field velocity characteristic with three regions of differential carrier mobility$\mu_d = \mathrm{d}v/\mathrm{d}E$.

The first region is that of weak fields $E_0 < E_{\text{th}}$, where $\mu_d > 0$, the second region $E_{\text{th}} < E_0 < E_s$, where $\mu_d < 0$ and the third region $E_0 > E_s$, where $\mu_d \approx 0$.

From (A.2.1), we obtain

$$\frac{\partial \rho_m}{\partial t} = \frac{\rho_m}{\varepsilon/\sigma_d}, \tag{A.2.4}$$

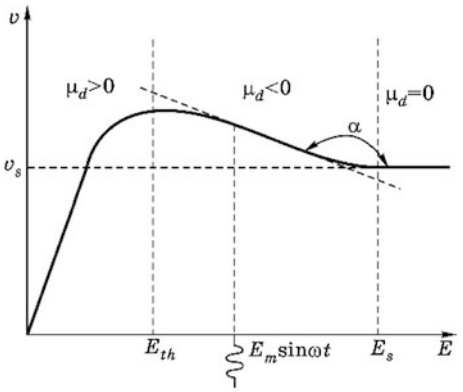


Fig. A.1 Explaining the differential mobility of charge carriers

where ρ_m is the charge inhomogeneity amplitude; and $\sigma_d = qn\mu_d$ is the differential conductivity. Then, by analogy with (A.2.3) we introduce the concept of differential Maxwellian relaxation time:

$$\tau_{md} = \frac{\varepsilon}{en\mu_d} \tag{A.2.5}$$

Then (A.2.2) is rewritten in the form $\rho_m = \rho_{mo}e^{-t/\tau_d}$, where ρ_{mo} is the amplitude of charge inhomogeneity at the initial moment. Hence, it follows that the amplitude of the charge inhomogeneity can decrease ($\mu_d > 0$), remain constant ($\mu_d = 0$) or even increase ($\mu_d < 0$). The correctness of this conclusion is confirmed by the existence of such an amazing physical effect as the Gunn effect. Figure A.2 shows possible variations in the amplitude of the charge inhomogeneity.

Along with the relaxation time, we introduce the concept of differential Maxwellian relaxation frequency:

$$\omega_{md} = \frac{en\mu_d}{\varepsilon} \tag{A.2.6}$$

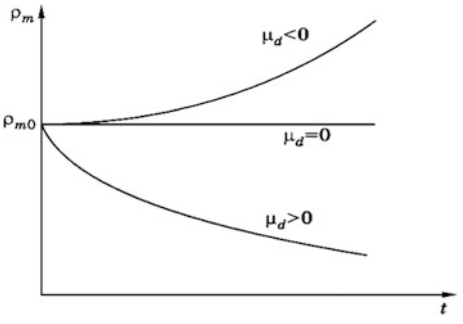


Fig. A.2 The change in charge inhomogeneity amplitude for various medium parameters

If the charge inhomogeneity moves in space, then along with the Maxwellian relaxation time, the length at which this relaxation occurs is introduced. Knowledge of this parameter is especially important for the analysis of processes when carriers drift through the region of a device with small fields. An example of such a region is the quasi-neutral region of the bipolar transistor base. Carriers are transported by diffusion force with no pulling external field. We define the Maxwellian relaxation length as

$$l_m = \sqrt{D\tau_m}. \tag{A.2.7}$$

If the width of the neutral part of the base is greater than this parameter, that is, $w_b > l_m$, then the efficiency of energy extraction from such a charge will be significantly reduced.

A.3 Period of Plasma Oscillation, Debye Length

The complexity of analyzing the interactions of particles in plasma implies the use of an integrated energy approach to these processes. Let us demonstrate the fruitfulness of this approach using the example of plasma oscillations. In semiconductors in a state of thermodynamic equilibrium with temperature T_0, there are free mobile charges (electrons or holes), as well as stationary charges (ionized donors and acceptors). A condition of electroneutrality is kept: $e(N_d + p - n - N_a) = 0$. It is clear that this condition is violated if the volume is small. In addition, the electroneutrality may be violated when mobile carriers change their location randomly due to their thermal velocities. When the particles are displaced relative to the equilibrium state, Coulomb forces arise, forming a force that causes a shift to the equilibrium position, as shown in Fig. A.3. Mobile carriers pass equilibrium positions by inertia on their way back, causing the appearance of a restoring force. The process acquires an oscillatory character and is called *plasma oscillations*.

Let us determine the period and spatial amplitude (range) of such an oscillatory process. We use Newton's and Poisson equations for this purpose. However, within the framework of this discussion, we find these values from a comparison of particle energy expressed in terms of classical mechanics, thermodynamics and electrodynamics. It is natural to assume that the average energy of a particle, determined from thermodynamic relationships, must be equal to the kinetic energy of the particle and the energy determined by the laws of electrodynamics, that is, $W_{\text{thern}} = W_{\text{kinem}} = W_{\text{eldyn}}$, or

$$\frac{3}{2}kT_0 = \frac{mv_T^2}{2} = eU_T \tag{A.3.1}$$

where v_T is the RMS thermal velocity; and U_T is the characteristic thermal potential. Let us determine the spatial distribution of this thermal potential to

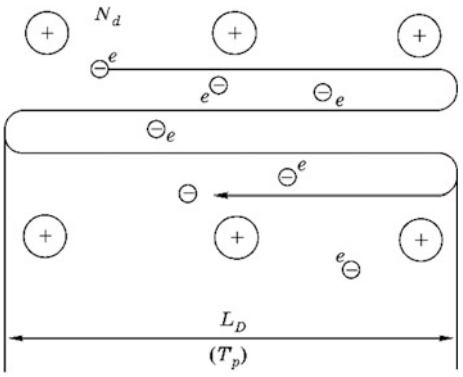


Fig. A.3 Explaining the nature of plasma oscillations

separate donors and electrons by a certain distance x, using the Poisson equation for a one-dimensional case:

$$-\frac{\mathrm{d}^2\mathrm{U}}{\mathrm{dx}^2}=\frac{\mathrm{d}E}{\mathrm{d}x}=\frac{eN_d}{\varepsilon}.$$

Assuming that the level of doping does not vary with respect to the coordinate, i.e., $N_d(x) = \text{const}$, we obtain the potential difference ΔU between points at a distance x: $\Delta U = (eN_d/2\varepsilon)x^2$, or $x = \sqrt{2\varepsilon\Delta U/(eN_d)}$. Equating the potential difference to the value of the thermal potential $\Delta U = U_T$, we obtain the maximum distance to which the electron shifts:

$$L_D = \sqrt{\frac{2\varepsilon U_T}{eN_d}}. \tag{A.3.2}$$

Replacing in this expression the value of U_T in terms of energy, according to (A.3.1), we obtain:

$$L_D = \sqrt{\frac{3\varepsilon kT_0}{e^2N_d}} \tag{A.3.3}$$

This quantity is called the Debye length. We note that when electrons are heated by field their energy is increased and instead of the temperature T_0 it is necessary to use T_e. Figure A.4 shows the dependence of the Debye length on donor concentration for different values of T_e.

Let us determine the period of plasma oscillations using relation $T_p \approx L_D/v_T$ and expression (A.3.3). We obtain:

$$T_p \approx \sqrt{\frac{\varepsilon m}{e^2N_d}}. \tag{A.3.4}$$

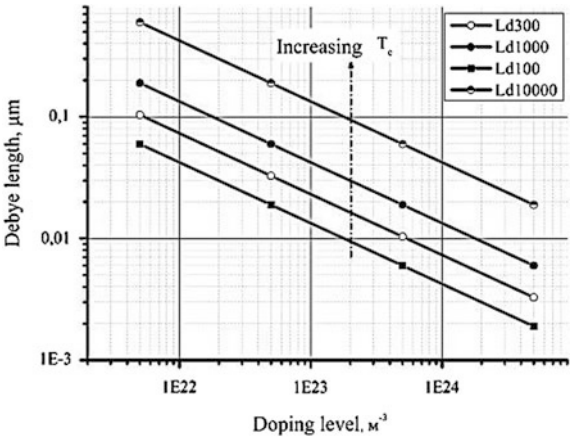


Fig. A.4 Dependence of Debye length on the doping level at electron temperatures of 100, 300, 1000, and 10,000 K in gallium arsenide

In practice, the plasma angular frequency is often used:

$$\omega_p = \frac{2\pi}{T_p} = 2\pi\sqrt{\frac{e^2 N_d}{\varepsilon m}}. \tag{A.3.5}$$

This analysis allows us to conclude that charges are periodically shifted relative to the equilibrium position. Such a shift leads to a local violation of electroneutrality. If we choose a volume larger than L_D^3 as the object, then the electroneutrality will overall be fulfilled. Given this circumstance, we cannot speak of sharp boundaries in the distribution of charges in semiconductors. The boundary can be determined with an accuracy in the order of L_D. It is often called the spatial scale of charge separation.

It is important to note the dependence of this parameter on the temperature of carriers and doping concentration. The effect of the external electric field leads to an increase in the energy of mobile carriers and, accordingly, to an increase in L_D. This fact must be taken into account when considering the physics of current transfer in a field effect transistor, especially in modes close to the cut-off regime of the device. The obtained L_D values are in the order of 0.01–0.1 μm, that is comparable with the size of the current channel in field effect transistors (0.1–0.2 μm), even with weak heating of carriers.

With increasing heating fields, it is impossible to neglect the blurring of the channel boundaries, even from the point of view of explaining the physical principles of the device, in particular, the process of overlapping the channel in the drain region. Figure A.5 shows the typical shape of the current transistor channel and the nature of its change, taking into account the heating up of carriers. It is important to note the increase in L_D to the drain end of the gate and its comparability with the thickness of the active region A. The plasmafrequency in the transistor channel is in the order of 10^{12}–10^{13} Hz. This means that plasma oscillates with a greater frequency than a typical microwaves signal (frequency 10^9–10^{11} Hz).

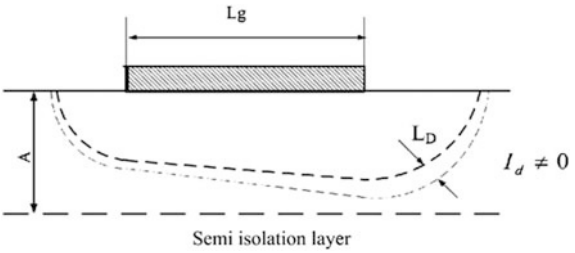


Fig. A.5 Current channel shape in FET taking into account Debye length

Vacuum devices also use the concept of plasmafrequency in the analysis of space charge forces. The occurrence of longitudinal oscillations with a certain characteristic frequency is possible in electron beam. The process leading to the displacement of carriers relative to a certain equilibrium position can be velocity modulation due to external influences. In Sect. 3.1, expression (3.1.3) was defined as follows: $\omega_p = \sqrt{e\rho/m\varepsilon_0}$. This expression differs from (A.3.5) by a factor of 2π. However, it should be noted that in the vacuum case we have longitudinal oscillations due to the Coulomb forces of interaction between similar charges. Therefore, it is not entirely correct to talk about plasma. In semiconductors, oscillations arise due to the forces between different charges, and the initial displacement of charges arises from the presence of thermal velocities. In this case, the concept of solid-state plasma is applied.

What is the practical need of introducing this parameter in the vacuum case? To answer this question, we estimate the value of the plasma frequency and the corresponding oscillation amplitude for the typical values of device current I_0, accelerating voltage U_0 and diameter of the beam D. Figure A.6 shows the calculated concentration of electrons in the beam with parameters $I_0 = 0.1$ A, $U_0 = 3000$ V for three values of beam diameter $D = 1$; 1.5; 2 mm.

Comparing the concentrations of charge carriers in a vacuum device $\sim 5 \times 10^{16}\,\mathrm{m}^{-3}$ with a characteristic level of doping in a semiconductor device $\sim 1 \times 10^{23}\,\mathrm{m}^{-3}$, we see a difference of almost six orders.

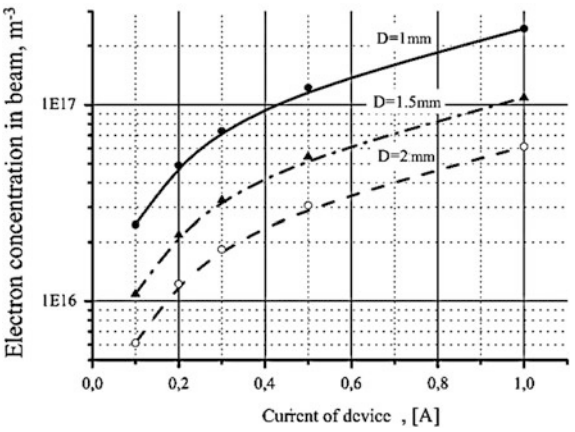


Fig. A.6 Concentration of electrons in the flow of a typical vacuum device for three flow diameter values

The possibility of obtaining a high electron density in a semiconductor plasma is associated with the compensating action of a positive donor charge: the semiconductor is quasi-neutral. Note that the calculation of the plasma frequency in a vacuum gives a value in the order of $\sim 10^9$ Hz, which corresponds to the lower border of the microwave band. In this case, the characteristic distance between the peaks of the longitudinal oscillations (*plasma wavelength*) λ_{pl} will be

$$\lambda_{pl} \approx v_0 T_p = \sqrt{2\frac{e}{m}U_0} \cdot T_p,$$

where v_0 is the drift velocity. So, we obtain $\lambda_{pl} \sim 5-20$ mm. This is comparable to the characteristic physical dimensions of the vacuum device. The analysis shows that it is important to take into account plasma oscillations in these devices, both in the time domain and in the space domain.

During the motion of charges in semiconductors, relaxation processes occur, and they are associated with the generation and recombination of charges. In order for the devices to operate in the microwave band, it is necessary to preserve the particle for at least one oscillation period $\tau_n \geq T$, or better $\tau_n \gg T$. In the space domain, the diffusion lengthl_d should be much greater than the characteristic size of the device $l_d \gg l_{\text{device}}$.

Practical measurements of the lifetime of charge carriers in semiconductors have shown the value of $\tau_n \sim 10^{-6}$–10^{-8} s. The spatial interval associated with this quantity, called the diffusion lengthl_d is calculated using the relation

$$l_d = \sqrt{D\tau_n}.$$

where D is the diffusion coefficient.

Appendix B
Electron-Optical Systems of Microwave Devices

B.1 General Properties and Parameters of Electron Beams Used in Microwave Devices

Electron beams in electronic microwave devices are used to convert the energy of external power sources to the energy of electromagnetic waves in the microwave range. Devices must provide the greatest energy conversion efficiency, high microwave power, efficiency and gain. Hence, microwave devices use electron beams with a high current density, at which the intrinsic space charge of the beam significantly affects the nature of electron motion. Electronic beams, in the analysis of which the forces of Coulomb repulsion cannot be ignored, are conventionally called *intense*.

The electron beam current limited by the space charge at the cathode, is determined by the formula

$$I = pU_0^{3/2},$$

where p is the parameter characterizing the degree of space charge influence on the beam motion, called the *perveance*. It is known from experience that the influence of the space charge becomes noticeable at perveance values $p > 0.1\,\mu\mathrm{A}/\mathrm{V}^{3/2}$. Microwave electronic devices use beams with perveances of up to 20 $\mu\mathrm{A}/\mathrm{V}^{3/2}$ and even higher, which makes it necessary to take into account and compensate the forces of Coulomb repulsion.

In addition to these parameters, the "quality" of beam includes such parameters as beam phase response, the phase ellipse and beam emittance. They describe the "quality" of the beam, the degree of its structural ordering and the distribution of transverse velocities. The phase response of the beam is the set of points in the transverse phase space r and r', where r and $r' = \mathrm{d}r/\mathrm{d}z$ are the radial coordinate and slope of each of the electron trajectories forming the electron beam.

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6

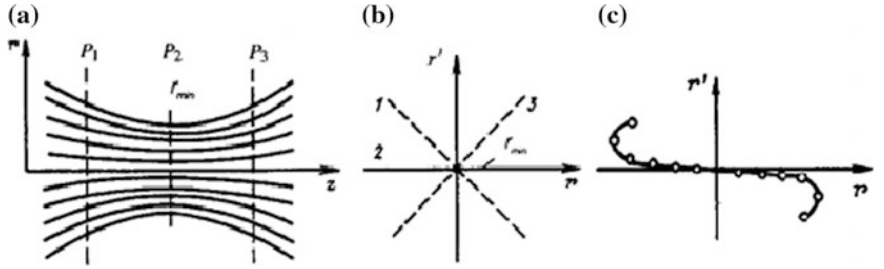


Fig. B.1 Electron trajectories (**a**), linear phase characteristics (**b**), nonlinear phase characteristic (**c**)

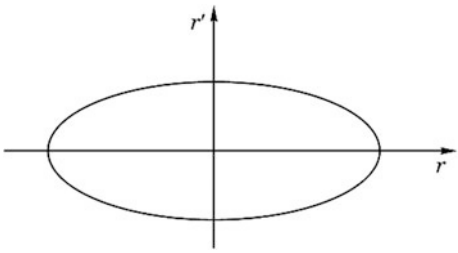
Figure B.1a shows the electron trajectories in an ideally formed electron beam: the beam is uniformly compressed, reaches a minimum cross section in plane P_2 and expands under further action of Coulomb forces. Figure B.1b shows the phase characteristics of this beam for planes P_1, P_2 and P_3. Characteristic 1 corresponds to a uniformly converging beam, and the slope of trajectory r' is proportional to their radial coordinate. In the minimal section plane P_2 the beam trajectories are parallel to axis z, and consequently, $r' = 0$ and phase response 2 is located on axis r of the phase space. Phase response 3 corresponds to a uniformly divergent beam. The linearity of these characteristics corresponds to an ideally formed electron beam.

The aberrations of the electron gun lead to nonlinearity in the phase characteristics. An example of a nonlinear phase characteristic in the plane of the minimum beam cross section is shown in Fig. B.1b. This kind of characteristic reflects the effect of intersection of electron trajectories, since the same radial coordinate values correspond to different inclination angles of the trajectories.

The thermal spread of electron velocities leads to the fact that on the phase plane the electron beam is not represented by a line, but by some figure with a finite area. In the plane of the minimum cross-section of the beam (crossover), this figure has the form of a straight ellipse (Fig. B.2). Each radial coordinate of the beam r corresponds to a set of r' values. The area of the phase ellipse A, divided by π, is known as the beam *emittance*:

$$\varepsilon = A/\pi$$

Fig. B.2 A phase ellipse



For a number of applications, brightness is an important characteristic of the charged particle beam. The brightness of a continuous axially symmetric beam is given by

$$B = I/(\pi b^2 \Omega),$$

where I is beam current with cross section radius a, and Ω is a solid angle, which determines convergence (or divergence) of the electron beam. High requirements for the brightness of an intense electron beam are, particularly, in free-electron lasers, where it reaches values of the order of $10^6 \mathrm{A/cm}^2$ sr.

Electron optical systems (EOS) of microwave devices generally include three parts. One of them initially forms the electron beam of the specified electric and geometric parameters and is called a beam formation system or, more often, an electron gun. The second is intended for the transfer of an electron beam through a channel, the extent of which considerably exceeds its transverse dimensions. This system is usually called a focusing system, although this name does not accurately reflect the functions it performs. The third part is a collector system that must ensure the reception of the spent electron flow, dissipation of the exuded heat, and in some cases the recovery of the electron beam's residual energy.

B.2 Systems for Electron Beam Formation

In modern microwave electronic devices, various electron beams of different spatial configuration are used.

Devices with O-type dynamic control (klystrons, traveling-wave tubes, etc.) based on long electrons existence in the electron-field interaction space use extended electron beams with a sharply delineated boundary. In devices with quasi-static control, the use of electronic fluxes in the form of sharply outlined beams is not mandatory, but it has opened new possibilities in the development of superpower triodes and tetrodes. Thus, the formation and focusing of intense electron beams is one of the main problems solved in the development of modern microwave electronic devices.

Methods for the formation and focusing of electron beams are usually associated with the principle of controlling them. This is of major importance in those devices where the elements of the electron optical devices directly enter the design of the resonator or slow wave systems. Nevertheless, there are a number of general requirements to these systems. For a clear understanding of these requirements, we briefly consider the main types of electron optical systems used in microwave electronic devices. Let us begin this consideration with electron beam initial formation systems, or electron guns.

The main goal of the electron gun, as noted above, is to form an intense electron beam of a certain configuration with a given electron current and velocity and, if possible, with laminar electron motion.

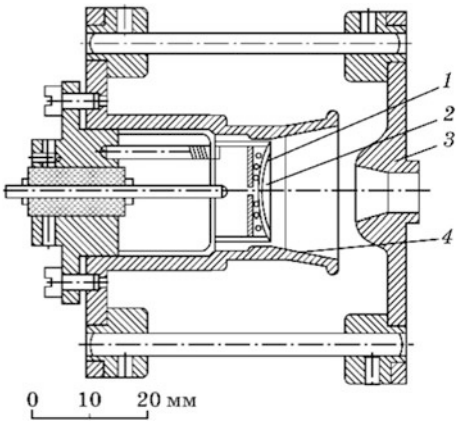


Fig. B.3 Pierce electron gun with micro perveance 0.5 $\mu A/V^{3/2}$

J.R. Pierce made an important contribution to the solution of this problem when he proposed a method for the formation of rectilinear laminar electron beams with simple configurations: sheet, cylindrical and conical. Based on this method, high-efficiency electron guns were developed, and are widely known under the name of Pierce guns.

The Pierce type of electron gun (Fig. B.3) consists of a concave spherical equipotential cathode 2 with a heater 1, a focusing electrode 4 and an anode 3 with a central hole. Usually, the cathode electrode has a potential of the cathode, and is located such that its surface is kind of an extension of the cathode surface. This explains the term diode for this type of gun.

By appropriately calculating the shape of the electrodes produced by an analytical method or by mathematical modeling, an electric field configuration is created in the gun in which electrons from the entire cathode surface converge uniformly into a narrow electron beam passing through the anode hole. In particular, to obtain an axially symmetric flow with trajectories parallel to the axis of the gun, the focusing electrode slope angle should be equal to 67.5°.

The degree of convergence of electrons is characterized by the so-calledcoefficient of convergence (compression). There is a coefficient of convergence in current density (or cross-sectional area of the beam) termed C_j, equal to the ratio of the maximum current density in the electron beam to the cathode current density. The convergence coefficient C_r, along the radius, determined by the ratio of the cathode radius to the radius of the minimum cross section (crossover) of the beam is also used. Obviously, with a small cathode curvature $C_r \simeq \sqrt{C_j}$.

As the coefficient of convergence in the beam increases, the Coulomb forces increase, preventing the beam from contracting. Consequently, the coefficient of convergence depends on the space charge in the beam being formed, which is determined by its perveance.

It should be emphasized that in Pierce diode guns, which usually operate in space charge mode, the value of perveance p does not depend on the anode voltage

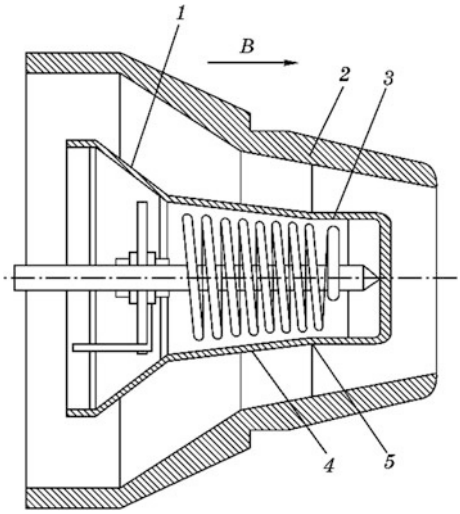


Fig. B.4 Magnetron gun

and as follows from the law "power 3/2" it is determined only by the geometric dimensions of the diode gun. Therefore perveance, being a parameter of the electron beam, and a measure of its intensity, is simultaneously a parameter of the gun itself, that is, its design.

Using the Pearce gun, we can form converged electron beams with perveance $p \leq 1\mu A/V$. For small values of perveance, the coefficient of convergence in current density can be 100 or more. For large values of p, the coefficient of convergence does not exceed several units. To obtain converging beams with a higher perveance, various modifications of Pierce guns are used.

If the initial axially symmetric Pearce gun is rotated with respect to an axis parallel to its symmetry axis, it is possible to obtain a toroidal gun capable of forming a hollow cylindrical or hollow conical beam.

Electron guns of the magnetron type also form axially symmetric hollow electron beams (Fig. B.4). In such a gun, beam formation takes place in crossed electric and magnetic fields. The gun consists of a cold cathode electrode 1 with the heater 3 and anode 2. The cathode ring between 4 and 5 cross-sections is coated with the emitting substance. The electric field is created due to the potential difference between the cathode 1 and the anode 2 electrodes. It has both radial and longitudinal components. External solenoids create the magnetic field. The magnetic induction lines of this field are directed approximately parallel to the surface of the cathode. The value of the magnetic induction exceeds the critical value (see Chap. 8). Therefore, the electrons emitted by the cathode ring do not reach the anode and create an electron cloud in the near-cathode region. Under the action of the longitudinal component of the electric field, they move in the longitudinal direction and leave the cathode-anode space, forming a hollow electron stream.

Crossed fields are also used to form sheet electron beams in M-type devices. Figure B.5 shows the design of an electron gun forming a sheet electron beam. The

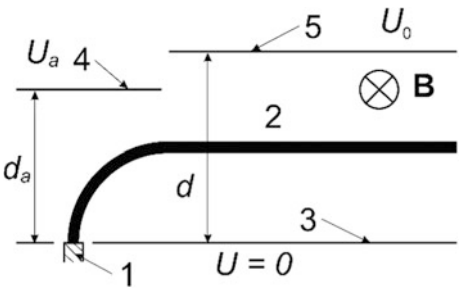


Fig. B.5 A magnetron gun with short optics

gun consists of a thermal cathode 1 that generates an electron beam 2, a cold cathode 3, a control electrode 4 and an anode electrode 5, the role of which is usually played a slow wave system.

The whole system is in a homogeneous magnetic field with induction B. The thermo cathode and the cold cathode are under zero potential, and the anode electrode is under positive potential U_0.

The electrons leave the cathode and are accelerated by the electric field of the accelerating electrode $E_a = U_a/d_a$, simultaneously deviating under the action of a magnetic field. When the electron is at the top of its cycloid trajectory, its velocity $v_y = 2E_a/B$, and it enters the interaction space. If its velocity is equal to velocity $v_e = E_0/B$, then the electron will move along a rectilinear trajectory, since the initial conditions for rectilinear motion in the crossed fields will be satisfied. To do this, the electric field created by the control electrode must be two times less than in the interaction space, since at the top of the cycloid the electron has a velocity twice as high as the velocity of the center. The disadvantage of such a gun, also known as short optics, is the impossibility of creating rectilinear flows from long cathodes, since the conditions of rectilinear motion are satisfied only for one trajectory.

In klystrons and TWT, guns with a control electrode are used for low-voltage electron beam current modulation.

The simplest of these is the conventional Pierce electron gun, in which the focusing electrode is isolated from the cathode and used as a control electrode. However, as shown by special studies, beam control with the help of a focusing electrode voltage is ineffective, especially in guns with large perveance. So, termination of cathode current in a gun with perveance 1 $\mu A/V^{3/2}$ require the application of a negative voltage to the focusing electrode $U_{cth} \simeq 0.5U_0$, and in the electron gun with perveance 3.6 $\mu A/V^{3/2}$—voltage $U_{cth} \simeq U_0$. The decrease in control efficiency with increasing perveance is due to the increasing influence of the anode potential on the field at the cathode.

Significantly better control parameters are achieved in guns with a control grid. By choosing a grid with a small transparency, the ratio U_{cth}/U_0 can be reduced. However, the fraction of the beam current intercepted by the positively charged grid will also increase. This disadvantage can be eliminated if two identical grids are used instead of one and the grid closest to the cathode is connected to it. Then the interception of the beam current by a second control grid with the same positive

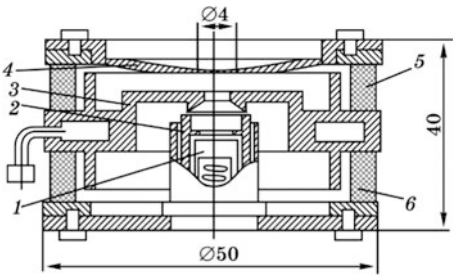


Fig. B.6 Electron gun with longitudinal beam compression *1* cathode; *2* pre-cathode electrode; *3* first anode; *4* second anode; *5*, *6* isolators

potentials on it, as in the previous case, can be reduced by a factor of two orders (up to 0.1%). Such guns are known as electron guns with a shadow grid.

Considering guns with a control electrode, i.e. triode guns, it should be emphasized that they allow the formation of electron beams with a higher output perveance, than the original diode guns. This is because two electrodes with adjustable potentials U_c and U_0 not only make it possible to vary the current I, but also the energy of the electron beam at the electron gun output, determined by the value U_0. Changes in I and U_0 can be almost independent of each other, implying two ways to adjust the perveance of the triode gun.

Triode guns which slow down the beam to increase its perveance in the interval between the first and second anodes are known as *guns with longitudinalbeam compression*.

Figure B.6 shows one of the experimental designs of such guns, designed to form axially symmetric electron beams with a diameter of less than 4 mm. At a potential of about 7 kV in the first anode, the cathode current is approximately 1 A, which corresponds to the beam perveance of 1.5 $\mu A/V^{3/2}$. With a potential at the second anode equal to 1.5 kV, the beam current does not practically change, since the current transfer coefficient is 99%, but its perveance increases in magnitude, i.e. up to 15 $\mu A/V^{3/2}$. There are modes of operation that allow the obtaining of an exit perveance beam of up to 50 $\mu A/V^{3/2}$.

B.3 Focusing (Transporting) Systems of Microwave Devices

The forces of the Coulomb interaction that arise in the electron beam during its formation cause the beam to expand, change its configuration, which ultimately lead to current losses on the electrodes surrounding the beam. Therefore, it is possible to pass an electron beam with minimal losses through the interelectrode space of a device due to the focusing properties of the electron gun only in those cases when the length of the interelectrode space is relatively small and the electron beam has a relatively small perveance. Hence, in most cases it is necessary to use additional magnetic or electric focusing systems that ensure the conservation of transverse flow within the specified limits over its entire length.

Magnetic fields created by solenoids or permanent magnets are used to focus (transport) electron beams in electrodynamic system channels. Systems with permanent magnets are preferable, since they do not require an additional power source.

According to the nature of the axial distribution of magnetic fields, magnetic systems are divided into three main types: systems with a homogeneous field, systems with a reversible field and magnetic periodic systems. The use of reversible and periodic fields provides a significant gain in mass and dimensions of the magnetic systems.

The simplest way to limit electron beam transverse dimensions is by means of systems that create a uniform longitudinal magnetic field.

Such systems include:

- screened solenoids, which are mounted directly on the electronic device;
- systems based on permanent magnets with pole pieces made of magnetically soft material.

The use of screens and pole pieces allows an increase in the uniformity of the field in the working region and provide the required degree of screening to the gun and collector region.

An electron beam formed by an electron gun, before it enters a homogeneous longitudinal magnetic field, passes throw a transition region, where along with the longitudinal component of the magnetic field there is a transverse (radial) component. Because of interaction with the field in the transition region, the beam electrons undergo rotational motion about the axis of symmetry of the magnetic system. Electron rotation in the homogeneous field region leads to the appearance of a magnetic force directed to the system axis.

The transverse motion of electrons in the homogeneous magnetic field zone is determined by the action of two forces: Coulomb interaction force F_ρ and magnetic focusing force F_μ. The first one is directed from the system's axis of symmetry and defocuses the beam, and the second, directed to the system axis, must balance the action of the first. This is achieved by selecting the value of the longitudinal magnetic field. As a result, the beam will perform an equilibrium motion, retaining the original diameter over its entire length. Such an ideal beam is called a Brillouin beam.

Theoretical analysis gives the following expressions for the indicated forces acting on the boundary electrons of the beam:

$$F_\rho = \frac{eI}{2\pi\varepsilon_0 r\sqrt{2eU_a/m}}, \tag{B.3.1}$$

$$F_m = -\frac{1}{4}\frac{e^2}{m}B_z^2 r, \tag{B.3.2}$$

where I is the electron beam current; U_a is the accelerating voltage; r is the beam radius; and B_z is the longitudinal component of the magnetic induction.

The equation of motion for the boundary electrons of the beam in this case has the form

$$\dot{z}^2 \frac{\mathrm{d}^2 r}{\mathrm{d}z^2} - \frac{eI}{2\pi\varepsilon_0 m r \dot{z}} + \frac{1}{4}\left(\frac{e}{m}\right)^2 B_z^2 r = 0, \tag{B.3.3}$$

where the longitudinal component of electron velocity $\dot{z} = \sqrt{1(e/m)U_a}$.

Assuming $\mathrm{d}^2 r/\mathrm{d}z^2 = 0$, we find the balance condition of defocusing and focusing forces:

$$B = \frac{1}{r}\sqrt{\frac{2I}{\pi(e/m)\varepsilon_0\sqrt{2(e/m)U_a}}}. \tag{B.3.4}$$

Substituting the numerical values of the constants in this formula gives

$$B = \frac{0.083}{r}\sqrt{\frac{I}{\sqrt{U_a}}}, \tag{B.3.5}$$

where B is Brillouin magnetic field induction, T; r is the beam radius, cm; I is the electron beam current, A; and U_a is the accelerating voltage, V.

To realize a beam of particles under strict equilibrium, it is necessary to fulfill a number of conditions that are practically difficult to fulfill. For example, it is difficult to ensure the introduction of an electron beam into the region of homogeneous field with zero radial velocities and a specified radius. As a result, real beams have a pulsating boundary.

An approximate solution of (B.3.3) gives the following expression for the boundary trajectory of the electron beam:

$$r = r_b + (r_0 - r_b)\cos\left(\sqrt{2}\frac{\omega_L}{\dot{z}}z\right) + \frac{r_0'\dot{z}}{\sqrt{2}\omega_L}\sin\left(\sqrt{2}\frac{\omega_L}{\dot{z}}z\right),$$

where r_b is the equilibrium radius, the value of which is determined from formula (B.3.4); r_0 is the initial radius of flow; $\omega_L = e/(2m)B_z$ is the parameter proportional to the induction of the magnetic field B_z, known as the Larmor frequency; and r_0' is the initial boundary trajectory slope.

It follows from the solution obtained that the boundary particle, in its motion along the z axis, performs periodic oscillations with respect to the equilibrium radius r_b. The amplitude Δr_m and the wavelength of the oscillations λ are determined by the expressions

$$\Delta r_m = \left[(r_0 - r_b)^2 + \frac{(r_0'\dot{z})^2}{2\omega_L^2}\right]^{1/2},$$

$$\lambda = \frac{2\pi\dot{z}}{\sqrt{2}\omega_L}.$$

With an ideal beam input into a homogeneous field $(r_0 = r_b, r_0' = 0)$, there are no beam boundary oscillations.

The main disadvantage of focusing (transporting) systems with a homogeneous magnetic field is the low efficiency of using a magnetic field. Indeed, in order to ensure a sufficient degree of homogeneity of the field over a section of length l, the transverse dimension (diameter) of pole pieces should be of the same order. Consequently, the total volume V occupied by the field of such a system will have a value of the order of $V \sim l^3$, while the useful working volume has the following value $V_u = \pi r_a^2 l$, where r_a is the radius of the channel in which the electron beam propagates.

An increase in the concentration of the magnetic field in the working volume can be achieved by applying reversible and periodic focusing systems, which allow a sharp reduction in the mass and dimensions of the focusing systems.

Reversemagnetic focusing. From (B.3.2) it follows that the radial magnetic force is determined by the square of the magnetic induction and, consequently, does not depend on the direction of the magnetic field. This allows us to apply focusing systems with magnetic field reversal, which are characterized by the fact that during system focusing, the magnetic field reverses direction once or several times.

Figure B.7 shows magnetic induction distributions (B-curves) for ideal and real focusing systems with a single reverse. B-curves in Fig. B.7a correspond to the ideal reversed field, when the field instantly changes polarity and the length of the reverse is zero. If such a field could be realized, the radial motion of electrons in this field would occur in the same way as in a homogeneous magnetic field of the same intensity.

Real reverse systems have two regions: a region of homogeneous field with a length L_0 and a reverse region of length L_p (Fig. B.7b). Since the magnetic induction in the region of the reverse is less than in the region of the homogeneous field $B_{z0} < B$, then, crossing this region, the electron beam experiences perturbation. In particular, the equilibrium beam, having passed the reversal zone, begins pulsate.

This effect can be significantly reduced in a reversible system with compensating peaks, the effect of which is explained as follows. Passing the emission zone, the electrons receive some excess radial momentum directed toward the axis of

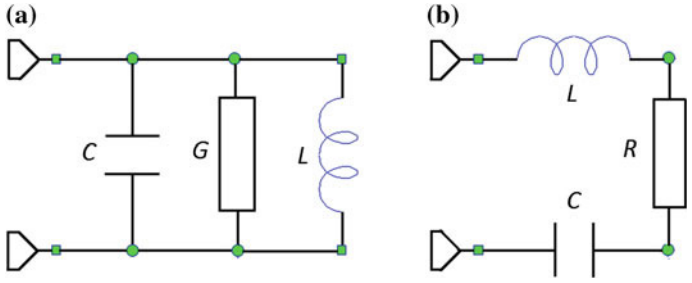


Fig. B.7 Axial distribution of the magnetic field in ideal (**a**) and real (**b**) reverse magnetic systems

symmetry, which compensates for the decrease of the magnetic focusing force in the region of the reverse. In the first approximation, the compensating peaks are selected in such a way that the average square value of the magnetic induction in the region of the reverse is equal to the induction of a homogeneous field B_r:

$$\bar{B}_r^2 = \frac{1}{L_r}\int_{L_r} B_{z0}^2 \mathrm{d}z = B_z^2.$$

Reverses allow a significant increase in the coefficient of magnetic field use. For a single reverse, the region with a homogeneous field has a length $L_0 \approx L/2$, where L is the total length of the focusing system. In this case, the field utilization factor increases approximately four-fold as compared to a system without a reverse having the same total length,

$$K_{b1} \approx r_a^2/L_0^2 = 4r_a^2/l^2.$$

An even greater gain is obtained with a multiple field reverse. For a system with N reverses, we get $K_{BN} = r_a^2/L_0^2 = (N+1)^2 r_a^2/l^2$. More efficient use of the magnetic field in reversible systems allows a substantial decrease in the mass and dimension of the focusing system, approximately $1/(N+1)^2$ times.

With a large number of reversals, the size of a single section gradually decreases, and the length of the region with the homogeneous field is reduced. The reversible system degenerates into a periodic focusing system with an axial distribution of the magnetic field close to the cosine wave.

Periodic system with a cosine-shaped magnetic field distribution. Among periodic focusing systems, the system with permanent magnets has the greatest practical application (Fig. B.8). The source of the magnetic field is an annular magnet 1 magnetized in the longitudinal direction. The required distribution of the field in the channel, where the electron beam passes, is formed with the help of pole pieces 2.

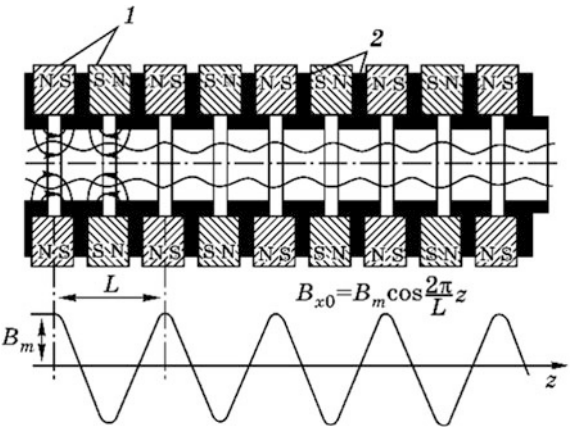


Fig. B.8 Periodic magnetic system of a TWT with permanent magnets: *1* magnetic rings; *2* pole pieces

The axial distribution of the magnetic field in most of these systems is described with sufficient accuracy by the cosine law:

$$B_{z0} = B_m \cos \frac{2\pi z}{L},$$

where L is the system period.

The use of such a system makes it possible to sharply reduce the mass and dimensions of the focusing system as compared to a system with a homogeneous field.

Since B_z is a function of axis z, the magnetic periodic system cannot provide accurate balance to the Coulomb force F_ρ and the magnetic force F_μ throughout the entire focusing system. However, under certain conditions, it is possible to balance these forces on average over the period of the focusing system. The condition of such a balance can be written in the form

$$\frac{1}{L}\int_0^L F_\rho \mathrm{d}z = \frac{1}{L}\int_0^L F_m \mathrm{d}z.$$

Assuming approximately that on the segment of axis z, equal to the field length L, the radius of the electronic flow varies little ($r \approx \text{const}$), taking into account the formulas (B.3.1) and (B.3.2), we get

$$\frac{2I}{\varepsilon_0 (e/m)\sqrt{2(e/m)U_a r^2}} = \frac{1}{L}\int_0^L B_{z0}^2 \mathrm{d}z. \tag{B.3.6}$$

This ratio establishes a relationship between beam parameters I, U_a, r and magnetic induction B_{z0}, which balances the Coulomb and magnetic forces on average over the period of the system.

Assuming that B_z varies according to a harmonic law, from (B.3.6) we deduce

$$B_m = \frac{\sqrt{2}}{r}\sqrt{\frac{2I}{\pi (e/m)\varepsilon_0 \sqrt{2(e/m)U_a}}},$$

or, after substituting the numerical values of the constants,

$$B_m = \frac{1174}{r}\sqrt{\frac{I}{\sqrt{U_a}}}. \tag{B.3.7}$$

The quantities included in this formula have the same dimensions as the values in formula (B.3.5).

Qualitatively, the nature of electron beam motion when the Coulomb and focusing force balance is ensured on average over the period can be described as follows. In regions of small magnetic induction values, the beam expands under the action of Coulomb interaction forces. However, falling into the region of large magnetic induction values, it experiences the predominant effect of the magnetic force and begins to contract. At the end of the half-period, the beam radius approaches its initial value.

In the next half-periods the nature of beam motion is repeated. Periodic changes in beam radius with a period λ equal to half the period of the magnetic field, $\lambda = L/2$, are a characteristic feature of beam motion in a magnetic periodic focusing system. These periodic changes in beam radius are called beam waviness.

If the conditions for the balance of forces are not fulfilled on average over the period, then the beam motion is more complex. Beam boundary pulsations are added to the wave ripple, depending on the ratio of the beam parameters and the magnitude of the magnetic induction.

Calculation of the beam motion generally requires solution of the differential equation for the boundary beam trajectory B.3.3.

$$\dot{z}^2 \frac{\mathrm{d}^2 r}{\mathrm{d}z^2} - \frac{eI}{2\pi\varepsilon_0 mr\dot{z}} + \frac{1}{4}\left(\frac{e}{m}\right)^2 B_{z0}^2 r = 0.$$

In this equation, the second term takes into account the defocusing effect of the electron beam space charge and the third term—the focusing effect of the magnetic field.

We reduce this equation to the form in which it is usually used in the periodic focusing theory. To do this, we introduce the normalized variables $R = r/r_0$, $Z = 2\pi z/L$, $b(Z) = B_{z0}/B_m = \cos Z$, where normalizing quantities include: r_0 is the initial beam radius at the entrance to the regular part of the focusing system, L is the field period, and B_m is the field amplitude.

The equation of boundary electron motion, written in the normalized variables, takes the form

$$\frac{\mathrm{d}^2 R}{\mathrm{d}Z^2} + 2\alpha b^2(Z) - \frac{\beta}{R} = 0,$$

or

$$\frac{\mathrm{d}^2 R}{\mathrm{d}Z^2} + \alpha[1 + \cos(2Z)]R - \frac{\beta}{R} = 0,$$

where

$$\alpha = \frac{1}{32\pi}\left(\frac{e}{m}\right)^2 \frac{B_m^2 L^2}{\dot{z}^2}$$

is the magnetic field parameter.

$$\beta = \frac{1}{8\pi^3 \varepsilon_0} \frac{e}{m} \frac{IL^2}{r_0^2 \dot{z}^3}$$

is the space-charge parameter.

This equation is a nonlinear differential equation with a periodic coefficient of the Mathieu type. In the general case, its solution can be obtained only by numerical methods. The analysis of numerical solutions yields the following main results:

- solution characters depend on the value of the magnetic field parameter: they can be either limited or increasing;
- for values $\alpha \leq 0.4$ (the first region of stable solutions) solutions are bounded and periodic, which corresponds to the boundary trajectory of the electron beam with periodically changing radial coordinate $R(Z)$; With condition $\alpha = \beta$ the amplitude of the ripple of the boundary trajectory ΔR_m is minimum, the pulsation form is determined by a simple harmonic law:

$$\Delta R = \Delta R_m \cos 2Z = \Delta R_m \cos(4\pi z/L)$$

When the condition $\alpha = \beta$ is satisfied, the force caused by the magnetic field balances the space-charge force on the average over the period of the field L. From the equality $\alpha = \beta$, a formula can be obtained for determining the optimum amplitude of the magnetic induction, at which quasi-equilibrium motion of the electron beam is ensured. It coincides with formula (B.3.7).

Comparing formulas (B.3.6) and (B.3.7), we find $B_m = \sqrt{2}B$. In fact, this means that the RMS value of the magnetic induction during the field period L, is numerically equal to the equilibrium value of the magnetic induction of a homogeneous field $B_{ms} = B_z$.

$$B_{ms}^2 = \frac{1}{L} \int_0^L B_m^2 \cos^2\left(\frac{2\pi}{L} z\right) \mathrm{d}z = \frac{B_m^2}{2}$$

The magnetic periodic system presented in Fig. B.8, is located outside the vacuum envelope of the device. In this version, it is widely used in traveling wave tubes with a spiral slow wave system.

In a TWT with a coupled cavity slow wave system, the focusing system is integrated into the slow wave system (Fig. B.9). The ring magnets 1 are located outside the vacuum envelope, and the pole pieces 2 enter the vacuum envelope, while simultaneously being the walls of the cavities. In order to ensure high conductivity of the walls, the surfaces of the pole pieces are covered with a thin layer of copper.

In magnetic periodic systems, barium ferrites, neodymium-iron-boron, and samarium-cobalt are used as magnetically hard materials (permanent magnets).

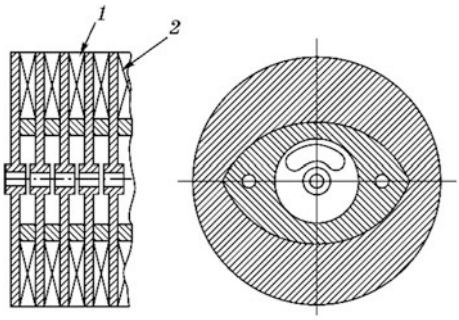


Fig. B.9 Integrated periodic magnetic system: *1* ring magnets, *2* pole pieces

The pole pieces are made of soft magnetic material, for example, of technically pure iron (Armco iron).

B.4 Electron Beam Energy Recovery Systems

In microwave vacuum devices, the electron beam is an instrument (intermediary) used to convert the energy of a power source into microwave energy. The efficiency of energy conversion is characterized by an electronic efficiency. This parameter is defined as the ratio of the power transmitted by the electron beam to the electrodynamic system field P_e, to the total power of the electron beam $P_0 = U_a I_0$:

$$\eta_e = P_e / P_0.$$

The power that is stored in the spent electron beam is called residual power P_{rez}

$$P_{\text{res}} = P_0 - P_e = P_0(1 - \eta_e).$$

It is possible to return a part of this power to the device power source and thereby increase the efficiency of energy conversion. The process of returning the power (energy) of the spent electron beam to a power source is called the *recuperationof the electron beam energy*. This process can be realized by decelerating the spent electron flow in the region of the electron collector. If all the electrons in the beam have the same velocities, they can be braked to a stop. As a result, the electronic efficiency of the device will be equal to 100%.

In real beams there is a scatter of electrons in terms of energy. It can be caused by a number of factors: thermal energy spread, the result of collisions of fast electrons (the Boers effect) and the result of the interaction of electrons with the electrodynamic system field (EDS). The last of these is of primary importance for microwave devices. As a result of interaction with alternating electric fields of the EDS, the electrons entering the collector can have a wide energy spectrum, less or more than eU_a.

Recuperation of the electron beam energy in a single-stage collector. A single-stage collector is shown schematically in Fig. B.10. The first electrode of the collector 1 has a potential equal to the potential of the electrodynamic system (EDS), which is assumed to be zero, $U_1 = U_{es} = 0$. A negative potential is applied to the second electrode 2 $U_2 = -\alpha U_a$, where $\alpha < 1$ is the numerical coefficient, and U_a is the anode voltage of the electron gun. It is assumed that the anode potential U_a is equal to potential EDS: $U_A = U_{es} = 0$, and the cathode potential U_κ is negative and equal to $U_c = -U_a$. If the value of the coefficient is one, then the potential of the second collector electrode is equal to the cathode potential: $U_2 = U_c = -U_a$. Difference in potentials of the first and second electrodes in an electric field creates a decelerating field.

The electrons entering into the collector have a spectrum of velocities. As the electrons move toward the second electrode, the electron velocity decreases. Slow electrons can be completely retarded and their motion direction can be inversed. Fast ones can overcome the braking field and land on the second electrode of the collector.

Efficiency of recuperation is defined as the ratio of the power returned to the power source P_{rec}, to the residual beam power P_{resT}:

$$\eta_r = P_{rec}/P_{res}.$$

Recuperation efficiency of a single-stage collector is usually small.

Better results are obtained from two-stage collectors, the construction of which is shown in Fig. B.11. The first electrode of the collector 1 is simultaneously a manifold body. The second 2 and the third 3 electrodes are fixed inside this case with ring isolators 4.

The first electrode has a zero potential, and negative potentials are applied to the second and third electrodes $U_2 = -\alpha_2 U_a, U_3 = -\alpha_3 U_a$, where α_2 and α_3 are numerical coefficients, $\alpha_2 < \alpha_3 < 1$. By the appropriate choice of the coefficients α_2

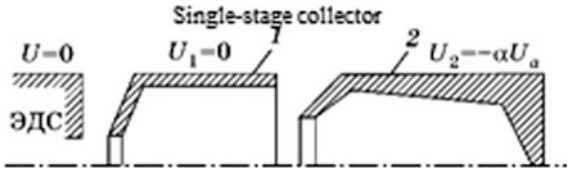


Fig. B.10 Single-stage collector: *1* the first collector electrode; *2* the second collector electrode

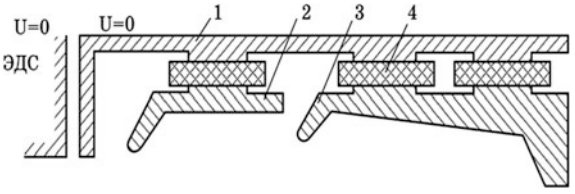


Fig. B.11 Two-stage collector

and α_3, the process of energy recuperation can be optimized and the maximum value of recuperation efficiency η_c can be obtained. It is obvious that in the two-stage collector it is possible to obtain a recuperation efficiency greater than in the single-stage collector.

In the practical implementation of collectors with energy recovery, it is necessary to take into account secondary electrons that arise as a result of bombardment of the collector electrode surface. Under the action of the electric fields existing in the collector, these electrons are accelerated and move along complex trajectories in the direction of increasing potential, i.e. towards the inlet to the collector.

They can be deposited on collector electrodes having a higher potential than the electrode potential from which they are emitted, or penetrate into the channel of the electrodynamic system. The acceleration of these electrons consumes the energy of the power source and thereby reduces recuperation efficiency. In addition, as a result of the bombardment of the collector by secondary electrons, its thermal load increases. The penetration of secondary electrons into the channel of the collector can promote self-excitation of the device.

The effect of secondary electrons can be reduced by applying anti-emission coatings to the receiving surface of electrodes and by selecting the geometry of the electrodes.

Multi-stage collectors with soft electrons landing. The problem of secondary electrons was most radically solved in multi-stage collectors, which have been called collectors with a soft electron deposit.

One of the variants of such collectors is the Kosmal collector, shown in Fig. B.12. The collector contains four stages, which are formed by five electrodes.

The first electrode has zero potential $U_1 = 0$, negative potentials are given to the others, respectively $U_2 = -0.55U_a$, $U_3 = -0.75U_a$, $U_4 = -0.95U_a$, $U_5 = -U_a$. The fifth electrode with a potential close to the cathode potential has an axial projection, which is called the reflector needle.

The configuration of the electrodes and the distribution of their potentials are chosen in such a way that the electrons deviate from the axis of the collector and, as a result of the braking process, are deposited on the rear surfaces of the electrodes.

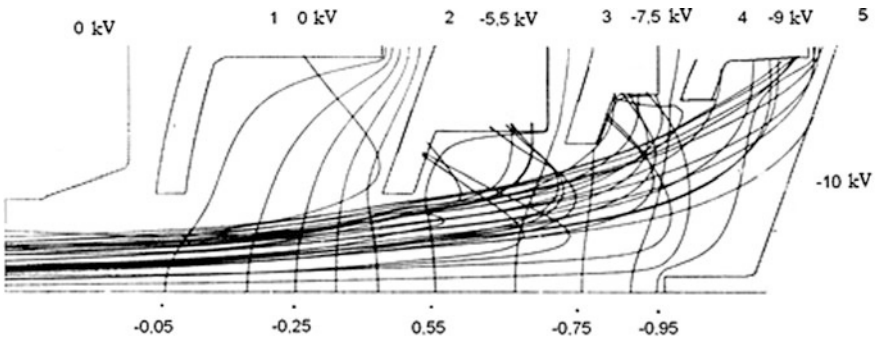


Fig. B.12 Multi-stage collector design. *1–5* Electrodes with indicated potentials

In this case, the secondary electrons knocked out by them fall into the retarding electric field and the main mass returns to the electrode.

For collectors containing 4–5 steps, the efficiency of the collector can reach 60–80% depending on the type of device (klystron, TWT, etc.) and its operating mode.

Appendix C
Electrodynamic Systems of Microwave Electron Devices

C.1 Cavity Resonators

C.1.1 Elements of the General Theory of Cavity Resonators

In the absence of external currents, the electromagnetic field in the cavity resonator (CR) is described by the Helmholtz equation:

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = 0, \tag{C.1.1}$$

where $\mathbf{E}$ is electric field intensity, $k = \omega/c$—wave number in free space, ω—angular frequency of field oscillations. This equation has an infinite countable set of solutions, each of them having its own eigen wave number k_j (eigen frequency $\omega_j = c/k_j$) and eigen function $\mathbf{E}_j$. The k_j and $\mathbf{E}_j$ together determine the cavity *eigen mode*. At each mode, the cavity is characterized by the following basic parameters:

1. The complex eigen angular frequency of the j-th oscillation mode $\omega_{0j} = \omega'_{0j} + j\omega''_{0j}$, which is determined by the formula

$$\omega_{0j}^2 = c^2 \frac{\int_V \varepsilon_r^{-1}\left|\nabla \times \mathbf{H}_j\right|^2 \mathrm{d}V + \mathrm{j}\omega_{oj} \oint_S \left(\mathbf{E}_j \times \mathbf{H}_j^*\right)\mathrm{d}\mathbf{S}}{\int_V \mu_r \left|\mathbf{H}_j\right|^2 \mathrm{d}V}.$$

The second term in the numerator of this expression expresses the power dissipated in the cavity and radiated from it. If this power is zero and there is no energy loss in the medium filling the resonator ($\varepsilon_r'' = \mu_r'' = 0$), the eigen frequency is real, otherwise it is complex. The real part of the angular frequency determines the

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6

frequency of the oscillations mode, and the imaginary part determines the rate of their damping:

$$\mathbf{E}_j(t) = \mathbf{E}_j(0)\mathrm{e}^{-\omega''_{0j}t}\mathrm{e}^{j\omega'_{0j}t}, \quad \mathbf{H}_j(t) = \mathbf{H}_j(0)\mathrm{e}^{-\omega''_{0j}t}\mathrm{e}^{j\omega'_{0j}t}.$$

2. Internal Q-factor $Q_{0j} = \omega'_{0j}/(2\omega''_{0j})$.
3. Wave (characteristic) impedance

$$\rho_j = \frac{\left|U_{ej}\right|^2}{2\omega'_{0j}W_j}, \tag{C.1.2}$$

where $U_{ej} = -\int_1^2 \mathbf{E}_j \mathrm{d}\mathbf{l}$ is the equivalent voltage, defined as a linear integral on the given line l between points 1 and 2, located on the surface of the resonator, W_j is the energy stored in the resonator at a given mode of oscillation.

In electronics, the integration path is usually chosen as the charged particle trajectory, so that equivalent voltage determines the intensity of the cavity electric field effect on the charged particle.

4. Equivalent impedance and equivalent admittance

$$R_{ej} = \rho_j Q_{0j}, \quad G_{ej} = 1/R_{ej}.$$

An important parameter of the resonator is frequency separation, that is, the difference between the frequencies of the working and neighboring oscillation modes. Frequency separation is characterized by the parameter

$$\delta f = \frac{|f_0 - f_1|}{f_0}$$

where f_0 is the eigen frequency of the working mode; and f_1 is the eigen frequency of the closest mode to the working one.

All these parameters can be measured or calculated with the help of special computer programs (Ansis HFSS, CST Microwave Studio, RFS, etc.).

C.1.2 Types and Structures of Cavity Resonators

The large frequency band in which microwave systems operate gives rise to a wide variety of types and structures of cavity resonators that are used in these devices. In microwave electronics, the cavities are mainly used for providing interaction of their electromagnetic field with beams of charged particles.

Cavity resonators (CRs) are divided into closed and open types. Closed cavity examples have the closed envelope impenetrable to the electromagnetic field, while

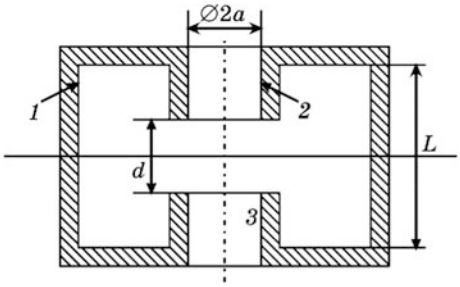


Fig. C.1 Scheme of a re-entrant resonator

open cavity examples have no such envelope. To increase the efficiency of interaction, the electric field of the closed cavities should be concentrated in the zone where a charged particles beam passes through the cavity. Consequently, the magnetic field is concentrated in other parts of the resonator. These parts can be attributed to capacity and inductance, that is, such resonators are well described by equivalent circuits. They are called resonators with quasi-lumped parameters.

One of the most common quasi-lumped cavities is the *re-entrant* (*toroidal*) *cavity*, the scheme of which is shown in Fig. C.1. It includes a cylinder 1 closed by covers with bushes 2. In the bushes, there are axial holes, forming a transit canal 3. The space between the bushes forms a high-frequency gap in which the electric field of the working mode of oscillations is concentrated. Holes can be tightened by metal grids (*grid gap*) or be without grids (*gridless gap*).

Consider an electric field in the gap. For the longitudinal component of the electric field of an azimuthally homogeneous mode in the cylindrical coordinate system, (C.1.1) assumes the form

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_z}{\partial r}\right)+\frac{\partial^2 E_z}{\partial z^2}+k^2E_z=0. \tag{C.1.3}$$

In the gridded gap, the electric field can be assumed to be homogeneous with a sufficient degree of accuracy, that is $\partial^2 E_z/\partial z^2 = 0$. Equation (C.1.2) in this case turns into the Bessel equation, the solution of which has the form

$$E_z = AJ_0(kr), \tag{C.1.4}$$

where $J_0(x)$ is the Bessel function of the first kind of zero order. The graphs of the E_z distribution along the radius for various ka values is shown in Fig. C.2. We can see that for $ka = 1$, field distribution is rather homogeneous.

Electric field distribution in symmetry planes of the re-entrant cavity, obtained with the help of a simulation program, are shown in Fig. C.3. As can be seen, the maximum value of the longitudinal component of the electric field placed near the bushes ends.

Figure C.1 shows the *H*-shaped cavity. However, one of its symmetrical halves (the *U*-shaped cavity) can also be successfully used (under other equal conditions, the *H*-form cavity has the advantage of a higher wave resistance).

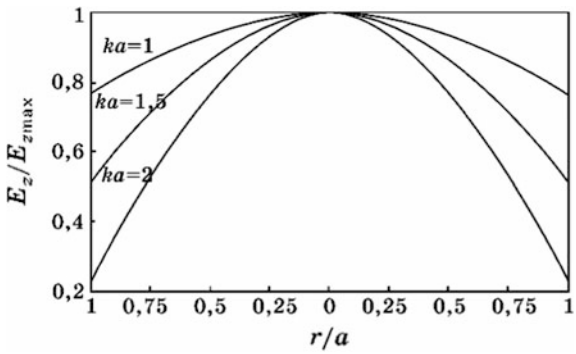


Fig. C.2 Distribution of the electric field along the radius of a resonator with a grid gap

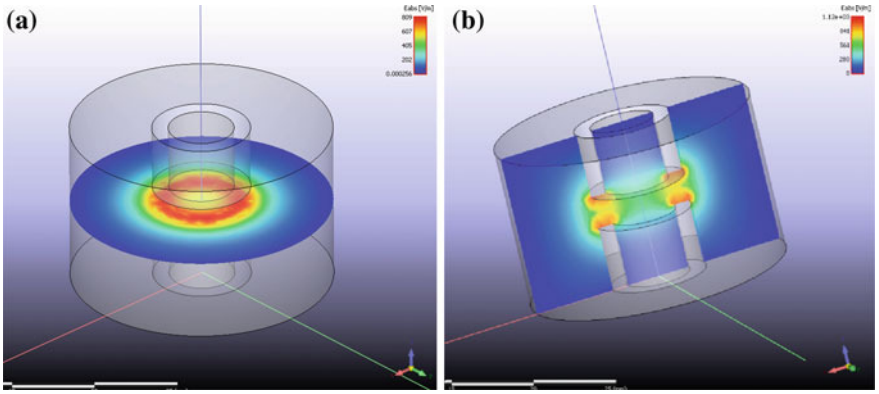


Fig. C.3 Distribution of the electric field in the transit channel of a toroidal resonator

Cavities with gridless gaps are used mainly in high-power devices, in which grids do not withstand the thermal load that arises from the electron bombardment. A magnetic field is concentrated in the peripheral part of the resonator. An electric field in a gridless gap depends on coordinate z. Hence, all components should be conserved in (C.1.2). To solve this equation, we represent the field strength as the product of two functions:

$$E_z(r,z) = \psi(r)\zeta(z).$$

Substituting in (C.1.2), after simple transformations we obtain

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\psi}{\mathrm{d}r} + k_c^2\psi = 0, \tag{C.1.5}$$

$$\zeta'' + \beta_e^2\zeta = 0, \tag{C.1.6}$$

where with dashes denoting the differentiation with respect to coordinate z and $\beta_e^2 + k_c^2 = k^2$.

The solution of the equation C.1.6 is

$$\zeta(z) = A\cos\beta_e z + B\sin\beta_e z, \quad \beta_e = \omega/v_e.$$

Admissible electron transition angle in gap $\beta_e d = 2\pi/3$ (coupling factor 0.825). From this $\beta_e = \omega/v_e = 2\pi/(3d)$, where v_e is the electron velocity in the resonator gap, and d is the length of the gap. Since electron velocity is always less than the speed of light in a vacuum, $\beta_e > k$ and $k_c^2 < 0$. Consequently, k is an imaginary number: $k_c = j\gamma$, where γ is the real number. Considering that, (C.1.5) can be rewritten:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\psi}{\mathrm{d}r} - \gamma^2\psi = 0. \tag{C.1.7}$$

The solution of this equation with a finite value on the cavity axis, is written in the form of a modified Bessel function of the first kind of zero order:

$$\psi(r) = AI_0(\gamma r).$$

Thus, the distribution of the electric field in the transverse cross-section of the transit canal is described by the formula

$$E_z(r) = E_{z0}\frac{I_0(\gamma r)}{I_0(\gamma a)}, \tag{C.1.8}$$

where E_{z0} is the longitudinal component value of the electric field intensity at the boundary of the transit canal. The graph of the function $E_z(z)$ is shown in Fig. C.4 for different values of γa. As can be seen, on the axis of the transit canal the field intensity is less than on its boundary. The non-uniformity of the field leads to a non-uniform interaction of electrons with the field of the resonator.

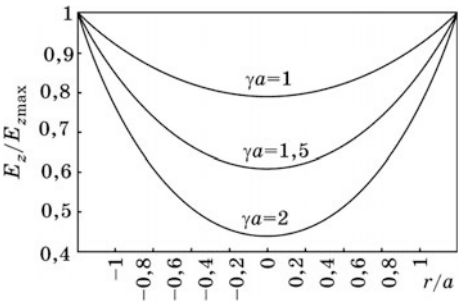


Fig. C.4 Distribution of the electric field along the radius in a resonator with a gridless gap

In order to reduce the negative effects of this effect, the transit-channel radius is usually chosen from the condition $\gamma a \approx 0.9$, which corresponds to the ratio $E_z(0)/E_z(a) \approx 0.8$.

An analogue of the re-entrant cavity is the brick cavity with bushes, designed to interact with a sheet electron beam (Fig. C.5). The resonator consists of a parallelepiped volume with lamellae 3. The transit canal 1 and gap 2 have a rectangular shape. The distance between lamellae and side walls is approximately $\lambda_c/4$, where λ_c is the cutoff wavelength for the rectangular waveguide. Hence, magnetic wall boundary conditions exist at the ends of the lamellae. This provides homogeneous field distribution along lamellae.

Consider the field distribution in the transit canal of the cavity. Assuming that the field is homogeneous along the y-axis, we write the Helmholtz equation:

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial z^2} + k^2 E_z = 0.$$

Performing transformations analogous to those used to analyze the field in the re-entrant cavity, we obtain an expression for the function describing the distribution of the field in the cross section of the transit channel:

$$\psi(r) = A\cosh(\gamma r).$$

The graphs of this function for different values of γa are shown in Fig. C.6. Usually the ratio $E_z(0)/E_z(a)$ should not be greater than 1.25. Hence, the condition $\gamma a \leq 0.7$ should be satisfied. As can be seen, this is a more stringent condition than for a cylindrical channel.

Figure C.7 shows the distribution of the electric field in the H-shaped resonator along the x, y, z axes calculated using the simulation program. As can be seen, the distribution of the field along the x axis is close to uniform.

To interact with hollow electron fluxes, cavities with an annular gap are used. The scheme of such a cavity and the distribution of the field in it are shown in Fig. C.8. Ring cavities (Fig. C.9) are used for the same purposes. These resonators can also be used in multi-beam devices.

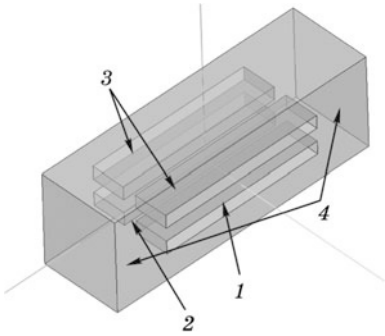


Fig. C.5 H-shaped resonator for interaction with a sheet beam

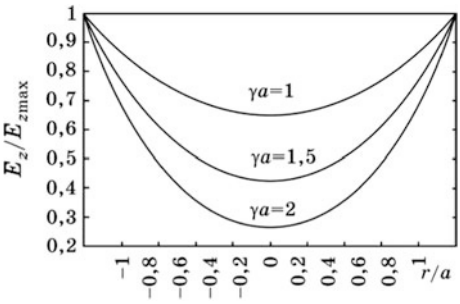


Fig. C.6 Field distribution in the gap of a gridless cavity

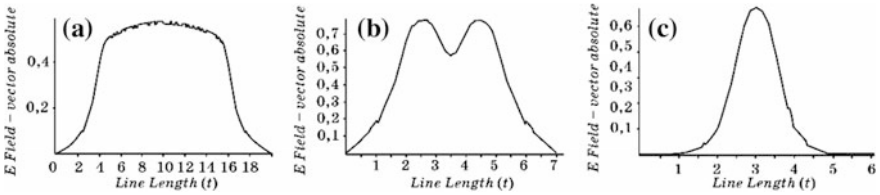


Fig. C.7 The calculated field distribution in the gap of the flat H-shaped resonator: **a** along the *x-axis*; **b** along the *z-axis*; **c** along the *y-axis*

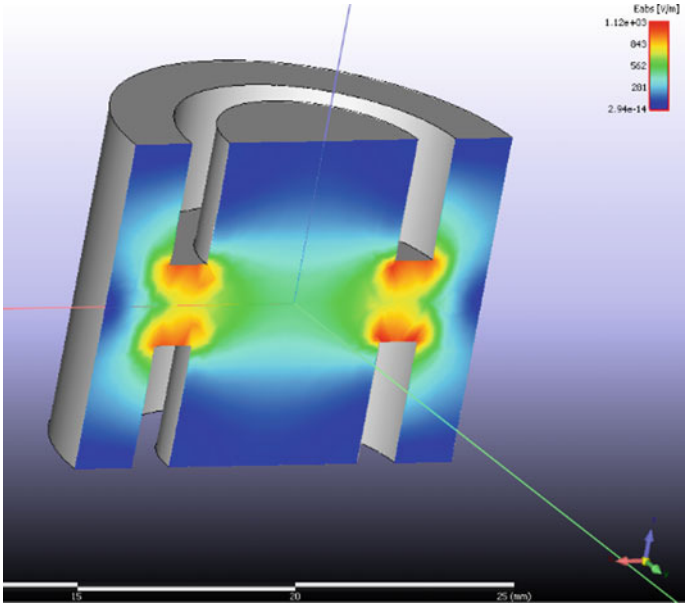


Fig. C.8 A re-entrant cavity with an annular gap and the distribution of the electric field in it

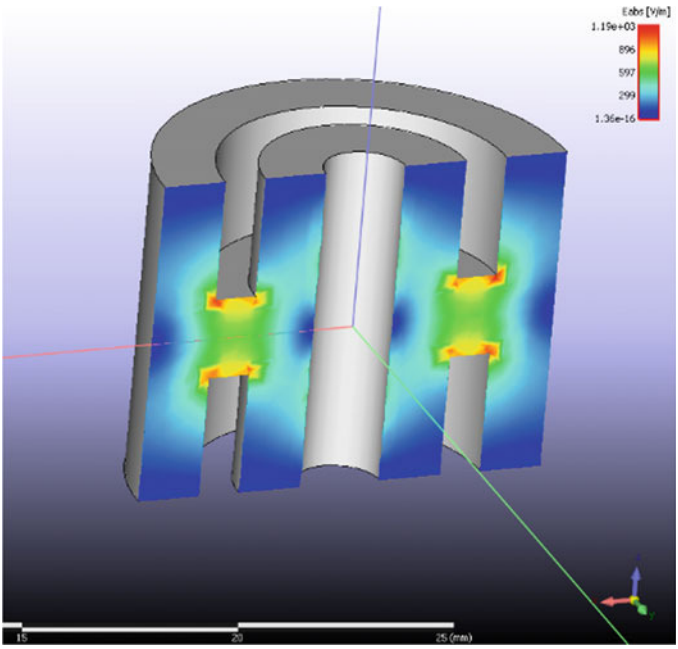


Fig. C.9 A ring cavity and the distribution of the electric field in it

The desire to increase the wave impedance of the resonator has led to the appearance of *multi-gap resonators. Let us imagine that, using the coupling elements, N single-gap resonators are* combined into one so that the electron beam sequentially penetrates all the gaps. The distance between the gaps is chosen such that the electron beam passes each of them in the same phase of the field. It is obvious that in this case the energy stored in the resonator increases by a factor of N, and the total voltage acting on the electron beam also increases by a factor of N. But, since in the formula for the wave resistances the voltage is squared, the wave resistance increases by a factor of N compared with its value for a single-gap resonator. In fact, the gain is less, since part of the energy is stored additionally in the communication elements, but it is still significant.

The most commonly used is a two-gap resonator based on a coaxial or strip line. The line is shorted at one end and contains a hole for the electron beam to pass on the other end. The same holes are also made in the bushes on the outer case of the resonator (Fig. C.10). The length of the line is close to a quarter of the wavelength. The electric fields in the gaps are directed in opposite directions, so the transition angle between the centers of the gaps should be 180°. The wave impedance of such resonators can reach 200 Ω.

Resonators with a large number of gaps are formed from the short-circuited section of the slow wave system (SWS). Usually, a SWS of coupled cavity type, a comb-like SWS or an interdigital SWS are used. Section C.3 a contains a more detailed description of the properties of these cavities.

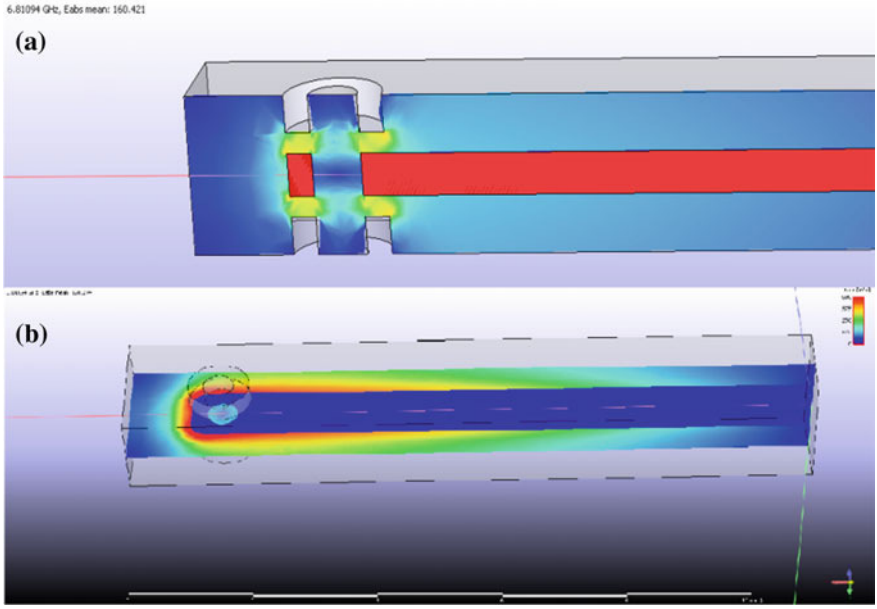


Fig. C.10 Two gaps strip-line cavity

In multi-barrel devices, special resonators are used, the designs of which can be quite complex. One such design, developed at the Saratov Technical University and intended for a four-barrelmulti-beam klystron is shown in Fig. C.11. The two-gap resonator consists of a cylindrical body 1 with a rod 2, on which radial rods with blocks of passing tubes 3 are located. The inductive loop 4 is intended to couple the resonator to the external transmission line. Unfortunately, the complexity of the design does not allow the use of similar resonators in the short-wave part of the microwave band.

Open waveguide resonators (OWR) are used in gyro-resonance devices. Such a resonator consists of a circular inhomogeneous waveguide segment 1 (Fig. C.12a), one end of which is connected to a canal 2 serving to inject an electron beam, and the other end is connected via a diaphragm 3 to a waveguide 4 through which energy is output. This waveguide serves as a collector for waste electrons. The shape of the resonator, i.e., the dependence of the radius of waveguide 1 on coordinate z, is chosen such as to obtain the optimum field distribution along the length of the resonator, ensuring the most effective interaction with the electron beam (Fig C.12b). Diaphragm 3 provides the necessary degree of resonator coupling with the waveguide and determines its loaded Q-factor. In powerful devices, it may be absent.

Since in gyro-resonance devices, electrons transfer their rotational motion energy to the field, OWR uses oscillations of the form TE_{mnp}, without longitudinal

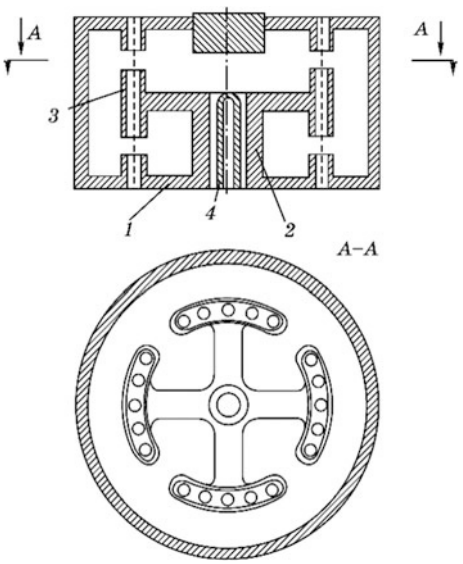


Fig. C.11 Two-gap resonator for a multi-barrel klystron

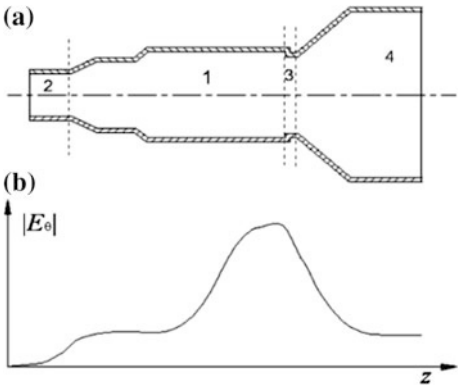


Fig. C.12 Waveguide open cavity. **a** shape, **b** field distribution along axis

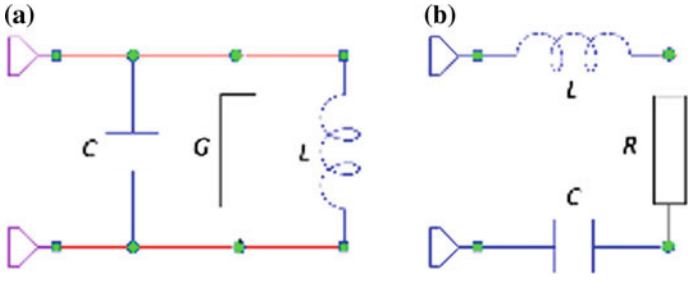


Fig. C.13 Equivalent circuit of a resonator near the eigen frequency: **a** parallel, **b** series

electric field components. The longitudinal index p is usually chosen as equal to one, and because cavity length $L \gg \lambda$, the eigen frequency of the resonator is approximately equal to the cutoff frequency of the waveguide averaged over its length. The second index determining the distribution of the field along the radius is chosen such as to provide maximum value of the azimuthal component of the electric field in the region of the electron beam. The first index determines the number of variations of the field along the azimuth. Often, it is chosen to be zero, since oscillations of the form H_{0np} have the greatest quality factor.

C.1.3 Excitation of Cavity Resonators

Near the eigen frequency, the resonator can be represented by an equivalent circuit in the form of a parallel (Fig. C.14a) or sequential (Fig. C.14b) LC oscillatory circuit, the parameters of which are determined by the electrodynamics parameters of the resonator for a given mode of oscillation:

$$L_e = \rho/\omega_0', \quad C_e = 1/(\omega'\rho).$$

For a parallel circuit

$$G_e = 1/(Q_0\rho),$$

for a series circuit

$$R_e = \rho/Q_0$$

The input admittance of a parallel equivalent circuit near the eigen frequency:

$$Y_{\text{in}} = G_{\text{in}} + jB_{\text{in}} = G_e + j[\omega C_e - 1/(\omega L_e)] = G_e(1 + j\xi),$$

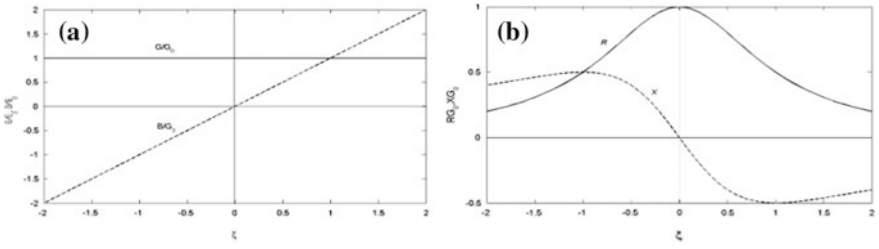


Fig. C.14 Dependence of admittance (**a**) and impedance (**b**) on frequency of the parallel equivalent resonator circuit

where

$$\xi = Q_0\left(\frac{\omega}{\omega_0'} - \frac{\omega_0'}{\omega}\right) \approx 2Q_0(\omega - \omega_0')$$

is the generalized detuning. The overall impedance of this circuit: $Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}} = 1/Y_{\text{in}}$. The graphs of these dependencies are shown in Fig. C.14. As we see, the active conductivity near the resonance remains constant, and the reactive conductivity varies linearly. The greatest value of active impedance $R_{\text{in}} = 1/G_e$ is observed at resonance (ξ = 0). The active impedance of the resonator decreases by a factor of two at $\xi = \pm 1$, which determines the bandwidth of the resonator.

Analogous, the input resistance of a serial circuit $Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}} = R_e + j[\omega L_e - 1/(\omega C_e)] = R_e(1 + j\xi)$.

The graphs of dependencies R_{in} and X_{in} are similar to the graphs of Fig. C.15 substituting Y with Z.

If the resonator is connected to an external circuit, part of the oscillation energy goes into the load of this circuit (or is radiated into the free space). In this case, an external and loaded Q-factor of the resonator are introduced:

$$Q_{\text{out}} = \frac{W}{\omega_0' P_{\text{out}}}, \quad Q_l = \frac{W}{\omega_0'(P_{\text{out}} + P_0)},$$

where P_0 and P_{out} are power dissipated in the resonator and its external circuit. Obviously the ratio is valid

$$\frac{1}{Q_l} = \frac{1}{Q_0} + \frac{1}{Q_{\text{out}}}$$

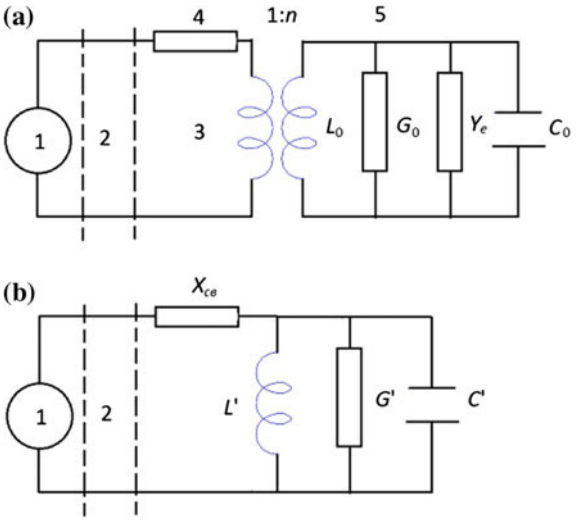


Fig. C.15 Cavity loaded with external transmission line through the coupling element (**a**). Equivalent circuit recalculated in the primary circuit (**b**)

Various circuits are used to connect resonators to external transmission lines—inductive loops, pins and holes in the walls of the resonator. Sometimes the tip of the pin is connected to some element of the resonator structure. Such a connection is called conductive. The equivalent circuit of any coupling element can be represented as an ideal transformer, in series or in parallel with the primary winding of which (connected to the resonator) reactivity of the coupling element is connected. Loops and pins are used, as a rule, to couple with the coaxial transmission line and holes (diaphragm)—with waveguides. In open resonators so-called diffraction coupling is also used, where part of the resonator energy excites the wave in the free space through the hole in the mirror.

Consider the excitation of the resonator from an external source, using an equivalent circuit of the resonator in the form of a parallel LC circuit. Most interesting is the dependence of the voltage on the capacitor (on the resonator gap) on the excitation frequency. The entire equivalent circuit is shown in Fig. C.15a. It consists of a generator 1, a transmission line 2 with wave impedance Z_0, a coupling element consisting of an ideal transformer 3 with a transformer coefficient n, of the coupling element intrinsic reactivity 4 and the resonator 5 with parameters L_0, C_0, G_0. The gap of the resonator corresponds to the capacitor C_0. In parallel, electronic conductivity is connected $Y_e = G_e + B_e$ (see Sect. 4.2.4). With small changes in frequency, the active part of the electron conductivity can be considered independent of frequency, and the reactive part, proportional to it. In this case, it is possible to simplify the equivalent resonator circuit by entering parameters $G = G_0 + G_e$, $C = C_0 + B_e/\omega$. We note that if amplitude and phase conditions are satisfied at a certain frequency $G_0 + G_e \leq 0$, $B_0 + B_e = 0$, the resonator is self-excited at this frequency (if there is no coupling with the load).

After recalculating the circuit parameters in the primary circuit of the transformer (Fig. C.15b), the input resistance of the coupling element is determined by the formula

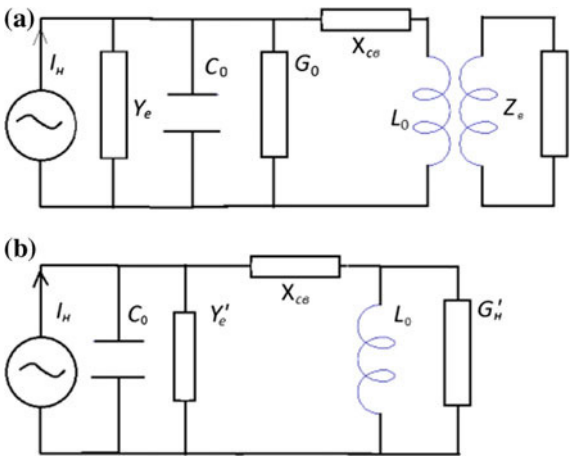


Fig. C.16 Equivalent circuit of a resonator excited by electron beam (**a**). Circuit with the load recalculated in the primary circuit (**b**)

$$Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}} = X + \frac{1}{Y'} = X + \frac{R'_e}{1+j\xi} = X + \frac{R'_e(1-j\xi)}{1+\xi^2},$$

where $R'_e = \rho' Q = n^2 R_e$. To fully transfer power from the generator to the load, it is necessary that

$$R_{\text{in}} = \frac{n^2 R_e}{1+\xi^2} = Z_0,$$

$$X_{\text{in}} = X - \xi \frac{n^2 R_e}{1+\xi^2} = 0.$$

From these equations, we obtain a generalized detuning necessary for matching the resonator with the generator:

$$\xi_{opt} = X/Z_0.$$

The voltage on the resonator gap:

$$U = I_2 Z = I_2 R_e \frac{1-j\xi}{1+\xi^2} = I_2 \frac{R_e}{1+\xi^2} e^{j\varphi},$$

where $\varphi = \arctan(1/\xi)$. Since $I_2 = I_1/n$ and $I_1 = \sqrt{2P/Z_0}$ for the gap voltage modulus we obtain the expression

$$|U| = \sqrt{\frac{R_e}{Z_0}} \frac{1}{1+\xi^2} = \sqrt{2PR_e} \frac{R_e Z_0}{Z_0^2 + X^2},$$

where P is the excitation power. At the same power supply, the voltage on the gap increases as the equivalent resonator impedance increases and as the reactive resistance of the coupling element decreases.

A bunched electron beam can excite the oscillations in the resonator. In this case, the equivalent circuit of the device has the form shown in Fig. C.16a. The electron beam creates an induced current in the resonator circuit $I_{\text{н}}$. The effect of the electron beam is also taken into account by electronic conductivity Y_e, parallel to the gap capacity C_0. A transmission line with a wave resistance is connected with the resonator $Z_{\text{в}}$. The same scheme with TL parameters recalculated in the resonator circuit is shown in Fig. C.16b.

The load power is given by

$$P_l = \frac{1}{2} |U|^2 G'_g$$

where $G'_g = n^2/Z_g$. Ignoring the reactive resistance of the coupler, we find the voltage on the load:

$$U = \frac{I}{G_0 + G_e + G'_g + j[\omega C_e - 1/(\omega L_e) + B_e]}.$$

Obviously, the maximum voltage corresponds to the frequency at which the imaginary part of the denominator of the last formula is zero:

$$P_{l\max} = \frac{|I_H|^2}{2G'_g}\left(1 + \frac{G_0 + G_e}{G'_g}\right)^2.$$

The output power increases when the coupling of the resonator with the load decreases (transformation coefficient n). However, power growth stops when the voltage reaches value U_0. Further reduction of the load leads to the appearance of the rejection effect, as a result of which the induced current begins to decrease.

Filter systems. To expand the device bandwidth coupled resonators are often used in the output circuit of the klystron. The active cavity (excited by an electronic current) is coupled with a passive resonator located in the output transmission line. As a rule, a passive resonator has the form of a waveguide segment, bounded by two inductive apertures or rods.

Figure C.17 shows one of the variants of coupled cavities. The toroidal resonator 1 through diaphragm 2 is connected to a rectangular waveguide 3. Usually a non-standard waveguide with a reduced size of a narrow wall is used. A waveguide segment located between inductive irises 4 forms the passive resonator. The waveguide 5 is connected to a vacuum window (not shown in the figure). The distance from the active resonator to the passive resonator must be an integer number of half-waves, such that the transformation ratio of this segment is equal to one.

Figure C.18 shows frequency response characteristics of a single output resonator (curve 1) and an output resonator with a passive resonator (curve 2). As can be seen, the use of a passive resonator allows a significant increase in the bandwidth of the output system due to a certain decrease in the amplitude of the output voltage.

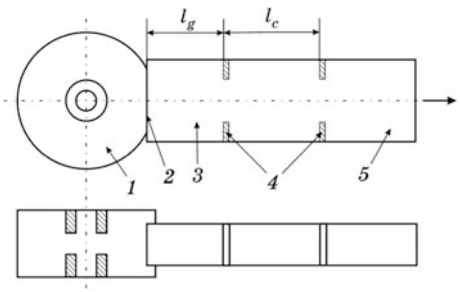


Fig. C.17 Output circuit with passive cavity

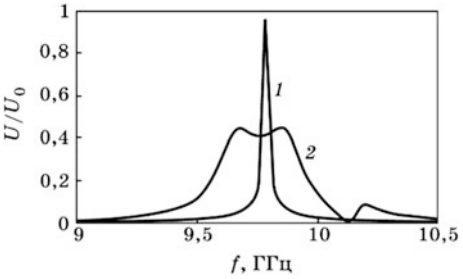


Fig. C.18 FRC of a single output resonator (*1*) and an output resonator with a passive resonator (*2*)

C.2 Slow Wave Systems

C.2.1 Characteristics and Parameters of SWS

Slow wave systems (SWS) are used in O-type TWT, twistrons, M-type devices, orotrons and some other microwave devices. Short-circuited or closed in ring SWS segments form multi-gap cavities, used in klystrons with distributed interaction and magnetrons.

The purpose of SWSs is to slow down the phase velocity of the wave propagating in it n_p times compared to the speed of light in vacuum. This allows a long-term interaction of the SWS field with the electron beam, because the wave in SWS and the electrons in the beam move with almost identical velocities. SWS is a transmission line (TL), which coincides with itself during displacement (translation) by a distance that is a multiple of a certain minimum distance, called the SWS period$D$ (periodic TL). There exist one-dimensional SWS in which alignment is possible for translation in only one direction, and two-dimensional periodic SWS, in which alignment occurs upon translation in two independent directions. The most common are one-dimensional SWS, which are considered in this appendix.

In SWS, as in any transmission lines, the existence of different wave modes is possible. Based on Floquet's theorem, the electromagnetic field of each mode can be represented as a superposition of spatial harmonics:

$$\mathbf{E}(z) = \sum_{p=-\infty}^{\infty} \mathbf{E}_p \mathrm{e}^{-\alpha z} \mathrm{e}^{-j\beta_e z}, \quad \mathbf{H}(z) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p \mathrm{e}^{-\alpha z} \mathrm{e}^{-j\beta_e z},$$

where p is the harmonic number; $\mathbf{E}_p, \mathbf{H}_p$ are complex amplitudes of its electric and magnetic fields intensities; $\beta_p = (\varphi + 2\pi p)/D$ is the harmonic phase constant; $-\pi \le \varphi \le \pi$ is the phase shift on the SWS period; and α is the attenuation constant, common to all harmonics. The phase velocity of the harmonic is given by

$$v_p = \frac{\omega}{\beta_p} = \frac{\omega D}{\varphi + 2\pi p}. \tag{C.2.1}$$

Group velocity

$$v_{gp} = \left(\frac{\mathrm{d}\beta_p}{\mathrm{d}\omega}\right)^{-1} = D\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\omega}\right)^{-1},$$

obviously is common for all harmonics.

Amplitudes of spatial harmonics are defined as the coefficients of the Fourier series:

$$\mathbf{E}_p = \frac{1}{D}\int_{-D/2}^{D/2} \mathbf{E}(z)\mathrm{e}^{j\beta_p z}\mathrm{d}z, \quad \mathbf{H}_p = \frac{1}{D}\int_{-D/2}^{D/2} \mathbf{H}(z)\mathrm{e}^{j\beta_p z}\mathrm{d}z.$$

Usually, harmonics with small absolute numbers $|p|$ have the largest amplitude. The spatial harmonic having the greatest phase velocity is called the *fundamental harmonic*.

The dispersion characteristic of a SWS is constructed as the dependence of the phase velocity deceleration $n_p = c/v_p$ on wavelength in free space. Another form of the dispersion characteristic is the dependence of the wave number in free space on the phase constant. The dispersion characteristic allows the determination of the type of spatial harmonic dispersion: (normal positive, anomalous positive or anomalous negative), phase and group velocities.

In addition to dispersion, an important characteristic of SWS is the *coupling impedance* of the spatial harmonic, defined by formula

$$R_{cp} = \frac{|E_{zp}|^2}{2\beta_p^2 P}, \tag{C.2.2}$$

where E_{zp} is the amplitude of the electric field intensity of the p-th spatial harmonic; and P is the power transmitted by this mode. Coupling impedance depends on the spatial harmonic number and on the frequency.

In the millimeter band, the attenuation constant of SWS α, is important and can significantly affect the parameters of the device.

C.2.2 Main Types of SWS

Helical SWS. A helical SWS was used in the first TWT created by Kompfner, and is still widely used in these devices. The simplest helical SWS consists of copper or molybdenum wire, coiled into a spiral and placed in a metal shell (screen) (Fig. C.19a). The spiral conductor can have a rectangular cross-section (ribbon spiral) 1 (Fig. C.19b). The spiral inside the screen 2 is supported by ceramic support rods 3. In SWS with a circular conductor, the amplitudes of higher spatial

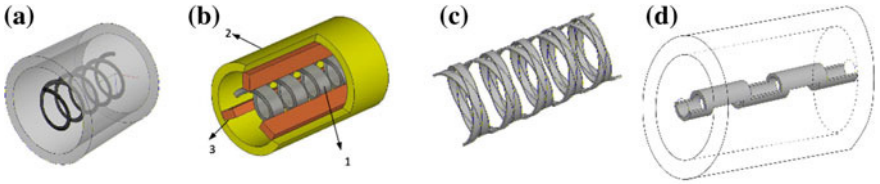


Fig. C.19 Spiral SWS. **a** With round spiral wire; **b** with a ribbon spiral; **c** with double helix counter winding; **d** ring-rod type

harmonics ($|p| > 0$) are insignificant compared to the fundamental ($p = 0$). In SWS with a ribbon spiral, the amplitudes of the higher spatial harmonics with $p = \pm 1$ are comparable with the fundamental harmonic, which makes it possible to use such SWS in the BWO.

Figure C.19c shows a SWS with two spirals with counter-coiling and Fig. C.19d shows a ring-rod SWS type, which can be considered as a kind of helical SWS with counter-coiling. This type of SWS is characterized by a rigid construction and is used in high-power TWTs. A SWS with a double helix is distinguished by strong dispersion and a narrower bandwidth in comparison with a conventional helical SWS.

In the roughest approximation, we can assume that the wave in a helical SWS propagates along the wire at the speed of light. Then the deceleration of the phase velocity (slowing factor) of the basic spatial harmonic can be found as the ratio of the turn length to the period:

$$n_{p0} = \frac{\sqrt{(2\pi a)^2 + D^2}}{D} = \sqrt{1 + \left(\frac{2\pi a}{D}\right)^2} = \frac{1}{\sin \xi},$$

where $\xi = \operatorname{arctg}[D/(2\pi a)]$ is the winding angle of the spiral. As can be seen, this quantity is independent of frequency, that is, there is no dispersion in the helical SWS. In fact, the dispersion of the fundamental spatial harmonic in a helix SWS is observed, but in a very wide frequency band it is sufficiently small (Fig. C.20). To control the shape of the dispersion characteristic, radial metal fins are connected to the screen (Fig. C.21). By adjusting the height and number of fins, it is possible to control the shape of the dispersion characteristic.

The coupling impedance of the fundamental harmonic in a helical SWS has a value in the order of 100 Ω and depends, comparatively weakly, on the frequency. Supporting rods and radial fins (if any) strongly influence the value of coupling impedance. To increase R_c, it is necessary to choose a rod material with a low dielectric constant.

A helical SWS has a low heat dissipating capacity. Therefore, it cannot be used in powerful devices. To improve the thermal characteristics, the rods are sometimes made of ceramics based on beryllium oxide, which has good thermal conductivity. Sometimes metal rods with diamond endings are used to make thermal contact with

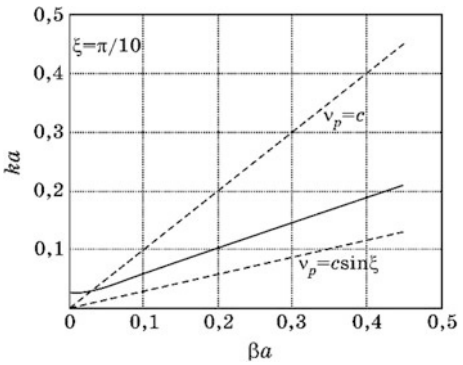


Fig. C.20 Dispersion characteristic of the helix SWS

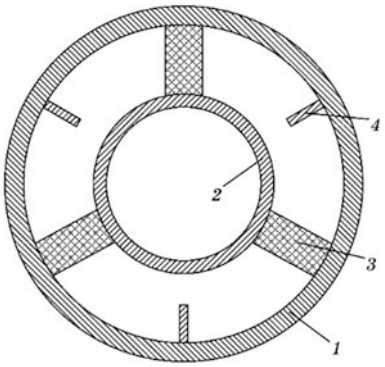


Fig. C.21 Helix SWS with rods. *1* metallic shell; *2* spiral; *3* dielectric rods, *4* metal fins

the spiral. Such holders simultaneously correct the dispersion characteristic of the SWS.

Coupled cavity chain typeof SWS(CCC-SWS). They have a rigid structure providing good heat dissipation. Therefore, they are used in high-power O-type TWT. SWS design is shown in Fig. C.22. The circular waveguide 1 is divided by diaphragms 2 into separate resonators. Slots 4 serve to couple the resonators with each other. The number of slots in the diaphragm, their dimensions, and the angle of rotation relative to adjacent slots can vary. Usually, one slot is used, rotated at an angle of 180° with respect to neighboring slots, as shown in Fig. C.22.

Dispersion of the fundamental spatial harmonic CCC-SWS ($p = 0$) is negative, therefore in TWT, the negative first harmonic is used as the working one ($p = -1$) which has positive dispersion. Figure C.23 shows dispersion characteristics and coupling impedance of the minus one spatial harmonic of a typical CCC-SWS in the two lowest-frequency pass bands, resonator and slot. These pass bands are so named because the boundary frequency of one of them practically coincides with the eigen frequency of the resonator without the coupling slots, and the boundary frequency of another pass band (in this case, a higher one) coincides with the eigen frequency of the slot.

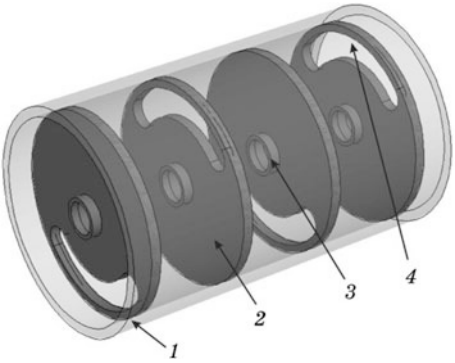


Fig. C.22 Coupled cavity SWS. *1* round waveguide; *2* diaphragm; *3* bushes with drift holes, *4* slots

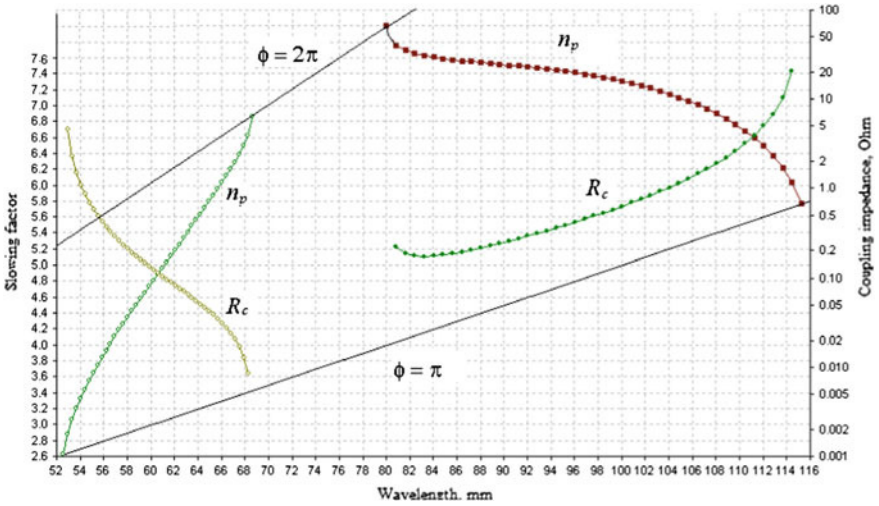


Fig. C.23 Dispersion characteristic of a CCC-type SWS in two bandwidths

The relative width of the resonant bandwidth is 36%, the coupling impedance in this band varies from 0.2 to 10 Ω. By choosing the geometric parameters of the SWS, it is possible to connect the resonator and slotted bandwidths, creating one wide band.

COMB SWS are used in TWTO of the millimeter band due to their simple and rigid design. The use of them is particularly effective in devices with a sheet electronic beam. Designs of some varieties of comb SWS are shown in Fig. C.24. Figure C.25a shows a simple open comb with grooves of rectangular shape. The electron flow is passed between the ends of the ridges and the screen. Dispersion

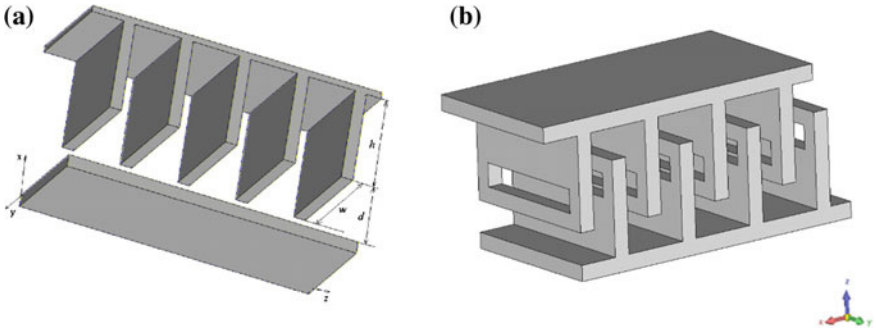


Fig. C.24 Crested SWS: **a** comb with a screen; **b** interdigital SWS

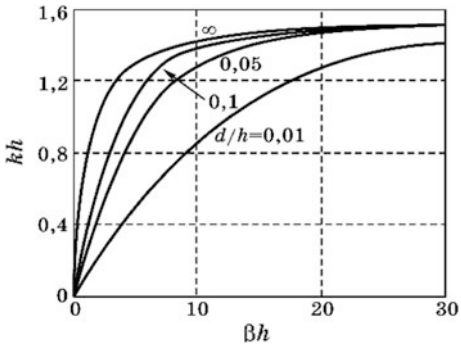


Fig. C.25 Dispersion characteristics of a comb SWS

characteristics of the comb for different ratios of the transit canal height d to the groove depth h are shown in Fig. C.25.

Dispersion increases with increasing ratio d/h, however, this also increases the coupling impedance. In electronic devices, metal walls close the comb on its sides. The walls are nearly a quarter of the wavelength from the ridges. Such a screen has little effect on the dispersion characteristic of the comb. Comb grooves can have different shapes and sizes. More about this is written in the section on multi-cavity magnetrons.

Figure C.25b shows an interdigital comb SWS. In such a system, the field of each wave mode can be represented as the sum of two sets of spatial harmonics, called symmetric and anti-symmetric components. If for one of the components the phase shift for the SWS period is equal to φ, then for the other it is equal to $\pi - \varphi$. The electric field of the symmetric component is symmetric with respect to the sliding plane symmetry SWS, and the electric field of the antisymmetric component is antisymmetric and equals zero in the plane of symmetry. Therefore, the electron beam passing through the transit canal in the symmetry plane interacts only with the symmetric component of the total field.

The dispersion characteristic of the zero spatial harmonic of the interdigital SWS is shown in Fig. C.26. The symmetric component (curve 1) has a negative dispersion, which makes it possible to use this SWS in wide-band BWO.

Stub SWS are used in type-O TWT and BWO, as well as in M-type devices (mitrons and magnetrons). Figure C.27 shows the most common types of rod SWS: ladder SWS with a ledge (Fig. C.27a) and groove (Fig. C.27b), interdigital stubs

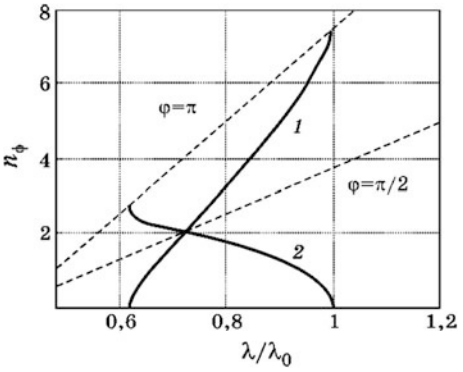


Fig. C.26 Dispersion characteristic of an interdigital comb SWS

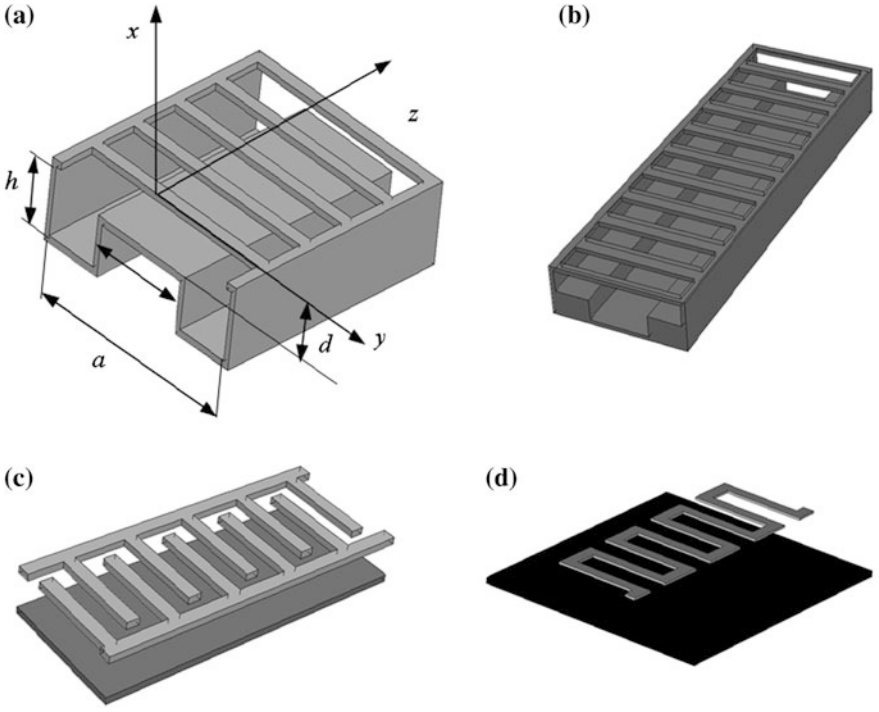


Fig. C.27 Stub SWS. **a** Ladder-type with a ledge; **b** ladder-type with a groove; **c** interdigital type; **d** meander type

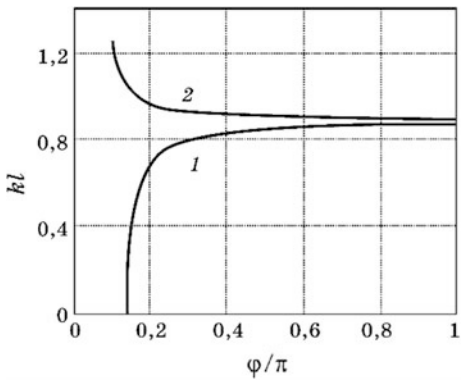


Fig. C.28 Dispersion characteristics of a ladder-type SWS: *1* with ledges; *2* with grooves

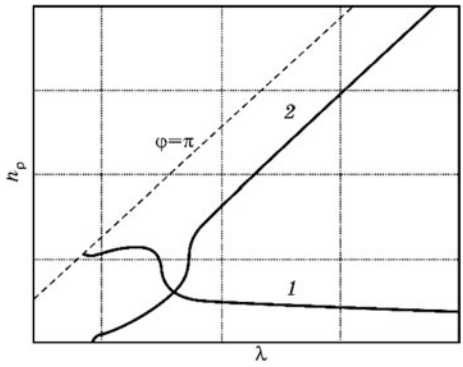


Fig. C.29 Dispersion characteristics of a meander-type SWS: *1* symmetric component; *2*anti-symmetric

SWS (Fig. C.27c) and meander (Fig. C.27d). The dispersion characteristics of ladder SWS are shown in Fig. C.28. As can be seen, they have a strong dispersion, but a simple design allows them to be used in TWT of the millimeter band.

The interdigital SWS have a dispersion characteristic similar to the SWS counter-type combs (Fig. C.27). They are used in the BWO. The SWS dispersion characteristic of the meander type is shown in Fig. C.29. The symmetrical component of this SWS has a small dispersion at $\varphi \leq 0.3\pi$, which allows these systems to be used in broadband TWTs. Coupling impedance of the symmetric component of the main spatial harmonic in the rod SWS amounts to tens of Ohms.

Note that the pin structure of this type of SWS can be located on a dielectric substrate (also made in the form of a printed circuit board). Such systems provide a greater deceleration with decreasing coupling impedance.

C.3 SWS-Based Resonators

The SWS segments can be used as resonators. There are two types of resonators - linear, made of the rectilinear SWS segment, short-circuited at both ends, and annular ones made of a SWS segment, closed in a ring.

The resonance condition in the linear resonator is the addition of a wave reflected twice from the ends, with the original wave. This condition is written as follows:

$$2\varphi N = 2\pi m + 2\pi, \tag{C.3.1}$$

where N is the number of periods of the SWS; m = 0, 1, … is the integer number, the last term on the right-hand side takes into account that when reflected from an ideally conducting wall, the phase of the wave changes by 180°. From (A.3.34) it follows that the phase shift on the SWS period can take fixed values

$$\varphi_m = \frac{\pi m}{N}.$$

Each value of m corresponds to a particular resonator mode of oscillation, which has its own eigen frequency ω_m. Since the phase shifts lay in the range $0 \leq \varphi \leq \pi$, in this type of resonator there can exist $N + 1$modes corresponding to $m = 0, 1, \ldots, N$. The first mode $(m = 0)$ of oscillation is usually called the zero mode, and the last is called the π-mode. These types are, as a rule, operable.

It should be noted that, since for π-type oscillations the longitudinal components of the electric field in adjacent periods have opposite directions, it is necessary to use boundary conditions of a magnetic wall type on the ends of the resonator. Since this is impossible, the π-type oscillations in the resonators are not excited, and the last type is with $m = N - 1$. It is not difficult to see that a field of this type also has an opposite direction in neighboring periods, However, the amplitude of the field varies from gap to gap, taking the maximum value in the middle of the resonator (Fig. C.30). The dashed curve in this figure shows the envelope of the field amplitude in the gaps, and the solid curve shows the amplitude of the longitudinal component of the electric field in the gaps. To correct the distribution of the field, SWS periods are made unequal.

Calculating the wave impedance of a multi-gap resonator according to the formula (C.1.1) can produce a result of zero, as, for example, for a two-gap resonator with identical gaps operating on the π-mode.

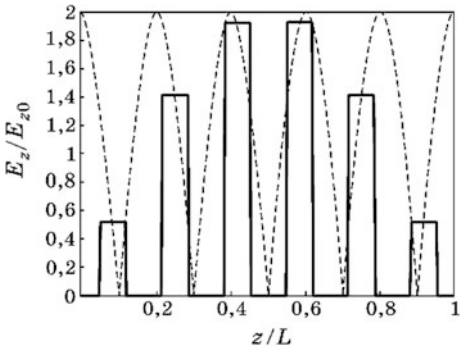


Fig. C.30 The distribution of the longitudinal component of the electric field in the gaps of the multi-gap resonator on π-mode

It should be taken in mind that in microwave electronics, wave impedance characterizes the degree of the resonator field effect on a charged particle. In order for this effect to be maximum, the velocity of the particle must be chosen such that each gap it transits is in the same field phase. Then the action of all the gaps on the particle is summed. Therefore, in formula (C.1.1) the equivalent voltage must be calculated from formula

$$U_e = \int_0^L |E_z(z)|\mathrm{d}z,$$

taking the absolute value of the field strength in each gap.

In ring resonators based on SWS, resonance is observed if the wave, passing through the resonator, coincides in phase with the original wave, i.e.,

$$\varphi N = 2\pi m, \quad m = 0, 1, \ldots, N/2.$$

Hence, the values of the phase shift angle corresponding to different modes of oscillations in the ring resonator:

$$\varphi_m = 2\pi m/N.$$

Each mode has its own eigen frequency, which can be determined from the dispersion characteristic of the SWS. As a working type, the π-type ($m = N/2$) is often used.

C.4 Vacuum Windows

Energy input and output from the vacuum device must be carried out through vacuum-tight transitions, which are called *vacuum windows*. The design of vacuum windows depends on the type of transmission lines into which they are integrated and the level of transmitted power.

Low-power coaxial vacuum windows consist of a section of a coaxial transmission line into which a ceramic insulator is soldered (Fig. C.31). The diameter of the inner conductor at the location of the insulator is reduced such that the wave impedance of the line remains constant.

Waveguide windows consist of a segment of a rectangular waveguide 1 (Fig. C.32), into which the metal diaphragm 2 is soldered. The diaphragm is made of metal, the thermal expansion coefficient (TEC) of which is close to the TEC of ceramic.

The ceramic insulator 3 is soldered into the diaphragm 2. The dimensions of the aperture in the diaphragm are selected such that the window has almost zero reflection coefficient at the operating frequency of the device. The formula for

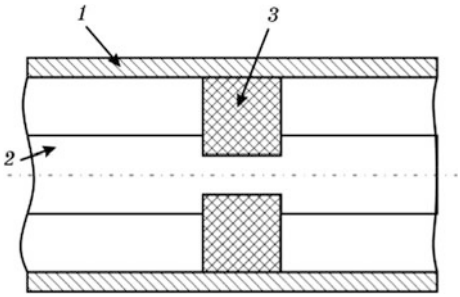


Fig. C.31 Coaxial vacuum window: *1* external conductor; *2* internal conductor; *3* ceramic isolator

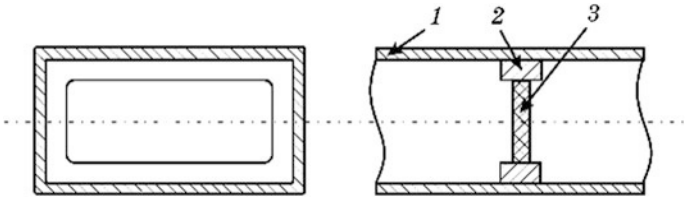


Fig. C.32 Waveguide vacuum window: *1* waveguide; *2* diaphragm; *3* ceramic insulator

determining ceramic window dimensions a', b' is obtained using equality of wave impedances of the waveguide section filled by air and ceramic:

$$\frac{b'}{b} = \frac{a'}{a}\sqrt{\frac{\varepsilon_r}{\mu_r}}\sqrt{1-\left(\frac{\lambda}{2a'}\right)^2\frac{1}{\varepsilon_r\mu_r}}\left[\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}\right]^{-1},$$

where a, b are the empty waveguide cross-section dimensions, ε_r, μ_r—relative permittivity and permeability of the ceramic.

To output energy from high power devices, can windows are used (Fig. C.33). They consist of a circular waveguide segment 2 built into a rectangular waveguide 1. A wave of H_{10} mode in a rectangular waveguide excites a H_{11} mode in a circular waveguide. At the other end of the circular waveguide, the H_{11} mode is converted back to an H_{10} wave type in a rectangular waveguide. The length of the circular waveguide is chosen as equal to half the wavelength. If the resonator formed by a circular waveguide segment is weakly coupled to rectangular waveguides, an almost standing wave is established in it, and in the middle plane, where the vacuum window 4 is located, the electric field intensity decreases almost to zero, which sharply reduces losses in ceramics and heat release in it . Conditions for the occurrence of electrical breakdown on the surface of ceramics are also hindered. However, the bandwidth of such a window is very small, so the coupling with the

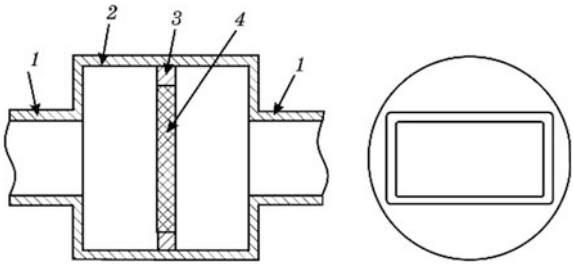


Fig. C.33 A can waveguide window: *1* rectangular waveguide; *2* circular waveguide; *3* diaphragm; *4* ceramic window

waveguides is done sufficiently strong and the field in the ceramic region is attenuated 1.5–2 times compared to the traveling wave field in a circular waveguide of the same cross section. Nevertheless, such windows can transmit large power, up to several megawatts. If necessary, the external surface of the circular waveguide is forced-cooled with air or liquid. Ceramic based on Al_2O_3, BeO, quartz and synthetic diamond is used as a material.

It should be noted that the dielectric sheet itself is a resonator, with resonant frequencies that can lie in the operating frequency band of the device. Therefore, the parameters of the sheet (thickness, radius, dielectric permittivity) should be chosen so that there are no parasite resonances in the operating frequency band of the device.

The windows with conical insulators (Fig. C.34) can have even greater impulse power, since they have a longer path for surface breakdown. However, the heat dissipation path from ceramics in such windows is longer than in planar ones, so it is inappropriate to use them to pass large average power. In modern high-power devices, conical windows are rarely used.

Modern window structures can transmit power of up to 10 MW in pulsed mode, and hundreds of kW in continuous mode in the three-centimeter wavelength band. The operating bandwidth is usually 10–15%, reaching up to 50% in special designs. Nevertheless, in high-power microwave devices, the power transmitted by the window is sometimes insufficient, and several power outputs must be made so that the output power is shared between them.

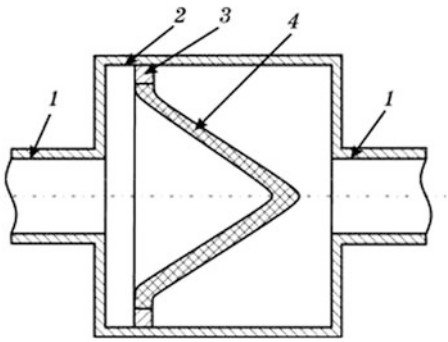


Fig. C.34 A can waveguide window with conical insulator: *1* rectangular waveguide; *2* circular waveguide; *3* diaphragm; *4* ceramic window

Bibliography List

Textbooks on Microwave Electronics

T.G. Roer, Microwave Electronic Devices. Microwave Technology Series. Springer US, 1994, 340 pp. doi: 10.1007/978-1-4615-2500-4

D.M. Pozar, *Microwave Engineering*, 4th edn. (Wiley, 2012), 736 pp

N.C. Luhmann, *Modern Microwave and Millimeter-Wave Power Electronics*, in ed. by N.C. Luhmann, J.H. Booske, R.J. Barker. (Wiley, *2005*), 872 pp

M.L. Sisodia, *Microwave Active Devices*. *Vacuum and Solid State*. (New Delhi: New Age Int. Publishers, 2004), 524 pp

D.I. Trubetskov, *Lectures on Microwave Electronics for Physicists*, vols. 1 and 2. M.: Physmatlit, 2003 (in Russian)

Microwave Vacuum Electronics Devices

J.C. Slater, *Microwave Electronics*, 3rd edn. (D. Van Norstand, 1959), 406 pp

J.E. Rowe, *Nonlinear Electron-Wave Interaction Phenomena*. (Academic Press, NY, 1965), 591 pp

J.R. Pierse, *Travelling Wave Tubes*, in D. Prinston, (Van Nostrand, NJ, 1950), 228 pp*

W.I. Kleen, *Electronics of Microwave Tubes* (Academic Press, NY, 1958)

W. Kleen, Poschl K. Einfuhrung in die Mikrowellen-Electronik. Teil II. Stuttgart: S. Hisel Verlag, 1958, 271 pp.

S.E. Tsimring, *Electron Beams and Microwave Vacuum Electronics* (Wiley, Hoboken NJ, 2007), 576 pp.

V.N. Shevchik, Fundamentals of microwave electronics, in *International Series of Monographs on Electronics and Instrumentation,* ed. by W.A. Gambling (Elsevier, 2014), 288 pp

A.S. Gilmour, *Klystrons, Traveling Wave Tubes, Magnetrons, Crossed-Field Amplifiers, and Gyrotrons* (Artech House, 2011), 859 pp

Electron Optical Systems of Electron Devices

J.R. Pierce, *Theory and Design of Electron Beams*, 2nd edn (D. Van Nostrand, NY, 1954), 222 pp

P.T. Kirstein, *Space-Charge Flow*, in P.T. Kirstein, G.S. Kino, W.E. Waters (McGraw-Hill, NY, 1967), 600 pp

S.I. Molokovsky, A.D. Sushkov, *Intense Electron and Ions Beams* (Springer, Berlin, 2005), p. 281

P.W. Hawkes, E. Kasper, *Principles of Electron Optics*, vol. 3 (Academic Press, NY, 1996), 3710 pp

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6

Electromagnetic Systems of Electron Devices

R.M. Bevensee, *Electromagnetic Slow Wave Systems* (Wiley, NY, 1964), 464 pp
R.E. Collin, *Foundations for Microwave Engineering*, 2nd edn. (IEEE Press, New York, NY, 2001), p. 944
S. Staras, *Wide-BandSlow-Wave Systems*. S. Staras, R. Martavicius, J. Skudutis et al. (CRC Press, 2012), 438 pp
A.D. Grigoriev, *Microwave Resonators and Resonator Slow Wave Systems*, A.D. Grigoriev, V.B. M. Yankevich, (Radio i svyaz, 1984), 247 pp. (in Russian)
A.D. Grigoriev, *Electrodynamics and Microwave Technique* (SPb.: Lan, 2007), 703 pp (in Russian)

Physical Phenomena in Semiconductors

S.M. Sze, K.K. Ng, *Physics of Semiconductor Devices*, Vols. 1, 2, 3rd edn. (Wiley, 2006), 832 pp

Microwave Semiconductor Devices

Microwave Semiconductor Devices and their Circuit Applications. ed. by H.A. Watson. (McGraw-Hill, NY, 1969), 617 pp*
J.E. Carrol, *Hot Electron Microwave Generators* (Edward Arnold Ltd., London, 1970)
S. Yngvesson, *Microwave Semiconductor Devices* (Springer, 1991), 471 pp. doi: 10.1007/978-1-4615-3970-4
F. Schwierz, J.J. Liou, *Modern Microwave Transistors: Theory, Design and Performance* (Wiley, 2003), 486 pp
M. Shur, *GaAs Devices and Circuits* (Springer, 1987), 670 pp. doi:10.1007/978-4899-1989-2

Circuits of Microwave Semiconductor Devices

M.J. Howes, D.V. Morgan (Eds.), *Microwave Devices. Device Circuit Interaction* (Wiley-Interscience, New York, 1976), 414 pp
S. Iezekiel (Eds.), *Microwave Photonics. Devices and Applications* (Wiley, 2009), 360 pp
M.J. Howes, D.V. Morgan (Eds.), *Microwave Devices. Device Circuit Interaction* (Wiley-Interscience, New York, 1976), 414 pp

Noise in Microwave Electron Devices

A. Van der Ziel, *Noise; sources, characterization, measurement* (Prentice Hall, 1970), 184 pp
M.J. Buckingham, *Noise in Electronic Devices and Systems* (Wiley, 1985), 372 pp*

Index

A. D. Grigoriev et al., *Microwave Electronics*, Springer Series in Advanced Microelectronics 61, https://doi.org/10.1007/978-3-319-68891-6