

THE SERVICES
TEXTBOOK OF RADIO

Volume 3
ELECTRONICS

ALL OVER THE WORLD



ERNEST TURNER

instruments are specified and relied upon for precision

ELECTRICAL MEASUREMENT in high quality ELECTRONIC
and COMMUNICATIONS EQUIPMENT in the fields of:—

RADIO • TELEVISION • PUBLIC UTILITY SERVICES
INDUSTRY • RESEARCH

Contractors to:

H. M. Admiralty, Ministry of Supply, National Coal Board, Post Office, etc.

Manufacturers of:

Electrical Measuring Instruments and Equipment for Specific Requirements

ERNEST TURNER

ELECTRICAL
INSTRUMENTS
LIMITED

CHILTERN WORKS - HIGH WYCOMBE - BUCKS

Telephone:
HIGH WYCOMBE 1301-2-3

Telegrams:
"GORGEOUS HIGH WYCOMBE"

Rwd' Ing

2

20,000 ohms per volt

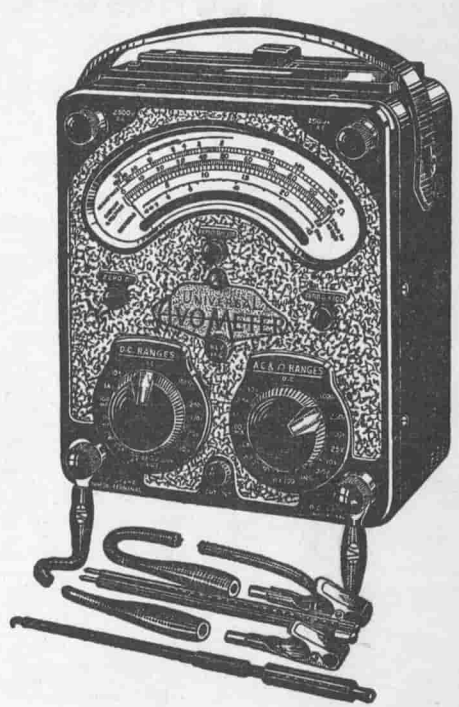
**MODEL 8 UNIVERSAL
Avo Meter**

Produced in response to a demand for a high sensitivity version of the world-famous Universal AvoMeter, this model incorporates the traditional design features of its predecessors, so highly valued for simplicity of operation and compact portability.

It has a sensitivity of 20,000 ohms per volt on all D.C. voltage ranges and 1,000 ohms per volt on A.C. ranges from 100 V. upwards. A decibel scale is provided for audio frequency tests. In addition, a press button has been incorporated which reverses the direction of current through the moving coil, and thus obviates the inconvenience of changing over test leads when the current direction reverses. It also simplifies the testing of potentials both positive and negative about a common reference point.

A wide range of resistance measurements can be made using internal batteries, separate zero adjustment being provided for each range.

It is of importance to note that this model incorporates the "AVO" automatic cut-out for protection against inadvertent overloads.



Size $8\frac{1}{8}'' \times 7\frac{1}{4}'' \times 4\frac{1}{2}''$
Weight $6\frac{1}{2}$ lbs. (including leads)

The following accessories are available to widen still further the ranges of the instrument: A Resistance Range Extension Unit to extend the limits of measurements from 0.025 ohm to 200MΩ, a 10kV. D.C. multiplier and a number of A.C. current transformers.

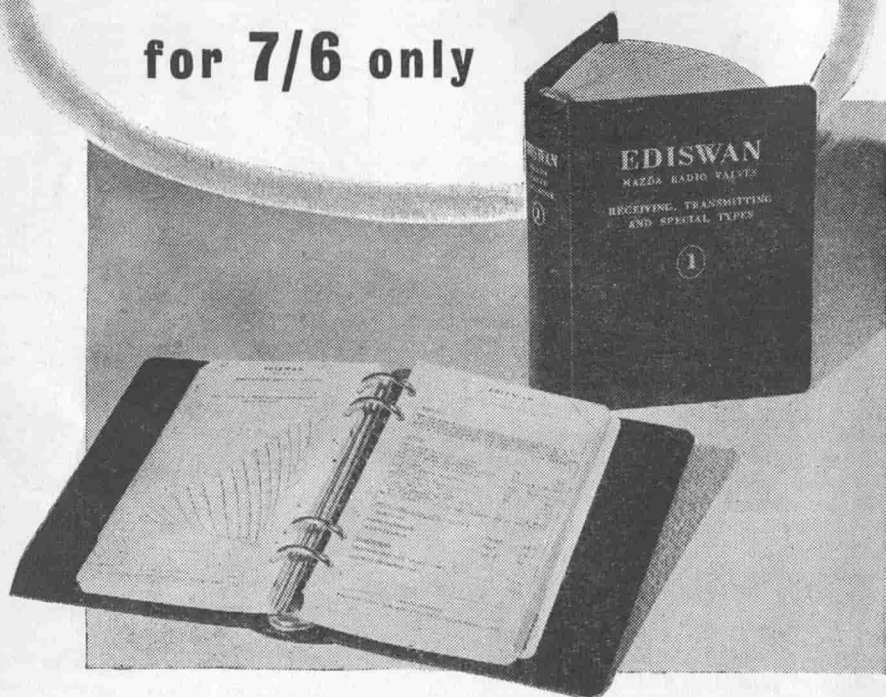
*Fully descriptive pamphlet
available on application*

D.C. VOLTAGE	D.C. CURRENT	A.C. VOLTAGE	A.C. CURRENT	RESISTANCE
2.5V.	50μA.	2.5V.	100mA.	First indication 0.5Ω
10V.	250μA.	10V.	1A.	Maximum indication 20MΩ.
25V.	1mA.	25V.	2.5A.	0—2,000Ω
100V.	10mA.	100V.	10A.	0—200,000Ω
250V.	100mA.	250V.	—	0—20MΩ
1,000V.	1A.	1,000V.	—	using internal batteries
2,500V.	10A.	2,500V.	—	
				using external batteries

Sole Proprietors and Manufacturers:

THE AUTOMATIC COIL WINDER & ELECTRICAL EQUIPMENT CO., LTD.
AVOCET HOUSE · 92-96 VAUXHALL BRIDGE ROAD · LONDON · SW-1 Victoria 3404/9

**THIS COMPLETE
TECHNICAL HANDBOOK
for 7/6 only**



For only 7/6d. you can obtain this comprehensive technical handbook containing data and characteristic curves on all valves and cathode ray tubes made by Ediswan.

The handbook is in two volumes, in loose-leaf form. As new valves are introduced additional sheets can be inserted. Ediswan make no charge for this additional service.

To obtain this valuable handbook and free follow-up service send a Postal Order for 7/6d. with your name and address to:—

EDISWAN

TECHNICAL PUBLICATIONS DEPARTMENT

THE EDISON SWAN ELECTRIC COMPANY LIMITED

155 Charing Cross Road, London, W.C.2

Member of the A.E.I. Group Companies

The Services' Textbook of Radio
Volume 3

ELECTRONICS

By

J. THOMSON, M.A., D.Sc., F.Inst.P., M.I.E.E.
Royal Naval Scientific Service

Edited by the Technical Staff of
'WIRELESS WORLD'

LONDON
HER MAJESTY'S STATIONERY OFFICE
1955

Crown Copyright Reserved

PUBLISHED BY HER MAJESTY'S STATIONERY OFFICE

To be purchased from

York House, Kingsway, LONDON, W.C.2 423 Oxford Street, LONDON, W.1

P.O. Box 569, LONDON, S.E.1

13a Castle Street, EDINBURGH, 2	109 St. Mary Street, CARDIFF
39 King Street, MANCHESTER, 2	Tower Lane, BRISTOL, 1
2 Edmund Street, BIRMINGHAM, 3	80 Chichester Street, BELFAST

or from any Bookseller

1955

Price 12s. 6d. net

The Government accepts no responsibility for any of the statements in the advertisements appearing in this publication, and the inclusion of any particular advertisement is no guarantee that the goods advertised therein have received official approval.

The Services' Textbook of Radio

VOLUME 1: ELECTRICAL FUNDAMENTALS

Electricity, magnetism and a.c. theory with a simple introduction to rotating electrical machinery and illumination.

VOLUME 2: TELECOMMUNICATION TECHNIQUES

The basis of communication theory, the principles of telegraphy, nature of speech, the principles of sound recording and reproducing, the basic principles of telephony, introduction to radio-frequency communications.

VOLUME 3: ELECTRONICS

Valves of all types, semi-conductors, photo-electric devices and cathode-ray tubes.

VOLUME 4: CIRCUITRY

The basic circuits used in radio communication and radar including amplifiers, oscillators, frequency-changers, filters, modulators and pulse generation; measuring and test instruments and power-supply circuits.

VOLUME 5: TRANSMISSION AND PROPAGATION

Properties of electrical lines, audio-frequency and radio-frequency considerations, waveguides, radiation and propagation, aerial systems.

VOLUME 6: WIRELESS COMMUNICATION

Characteristics of wireless systems, transmitters and receivers including v.h.f. and u.h.f. types, choice of systems for particular uses including installation considerations, the principles of facsimile and television.

VOLUME 7: RADIOLOCATION TECHNIQUES

(Including Direction Finding)

The principles of direction finding and the various systems in use, characteristics of radar equipment and the factors affecting performance, radar displays and typical radar systems, interference.

By Command of their Lordships

J. H. Lang

By Command of the Army Council

C. W. Finney

By Order of the Air Council

J. H. Barwell

Contents

<i>Chapter</i>	<i>Page</i>
Foreword	vi
Author's Preface	vii
List of Symbols and Abbreviations of Units	viii
1 Electrons in Metals	1
2 Thermionic Emission	16
3 Diodes and Triodes	32
4 Multi-Grid Valves	50
5 The Valve as a Distorter	67
6 Mixing and Frequency-Changing Valves	88
7 Modulation and Demodulation	100
8 Multi-Unit Valves	115
9 High-Power Valves	119
10 High-Frequency Effects	124
11 Velocity-Modulation Valves and Magnetrons	141
12 Noise	168
13 Gas-Filled Valves	179
14 Semi-Conductors	198
15 Electron Optics	219
16 Cathode-Ray Tubes	229
17 Photo-Electric Devices	247
Index	256

Foreword

THE SERVICES' TEXTBOOK OF RADIO is a series of seven volumes which has been written primarily to meet the needs of the three Fighting Services for a comprehensive and authoritative work on the theory and technique of radio, including in this term radar and direction-finding as well as telecommunications. Its usefulness, however, is by no means confined to the Fighting Services, for theory and technique are of universal application; it is, therefore, also well adapted for civilian needs and its availability has accordingly not been restricted to the Services.

The authors are all recognized authorities in their own particular spheres. The volumes have been written under the guidance of an Inter-Services Panel and have been edited by the Technical Staff of *Wireless World*.

The seven volumes of the series are listed, with brief synopses, on page iii. Together they cover the whole range of radio theory from the elementary principles of electricity and magnetism to the advanced techniques of radar and multichannel communication systems. Volume 1, which covers electricity and magnetism, is also Volume 1 of a companion series entitled *The Services' Textbook of Electrical Engineering*. Volume 2 includes descriptions of a number of telecommunications techniques on which information is not usually readily available, while Volumes 3, 4 and 5 cover the basic theory used in radio. Volume 6 contains information valuable to those responsible for the planning and operation of wireless communication systems. In Volume 7 will be found the principles of radar and direction-finding systems.

In conformity with modern practice, the rationalized m.k.s. system of units is employed throughout. In order to help readers who are more familiar with the c.g.s. system, however, the inter-relation between the two and the method of converting from one to the other are fully explained in Volume 1.

The book is intended to meet the needs of three distinct classes of reader. It is for the beginner, who has no prior knowledge of the subject; it is also for the technician and, finally, for the more advanced reader. The attempt to meet these three requirements in one book is an ambitious one and it has resulted in an unusual form of presentation which deserves some explanation.

The text, taken as a continuous whole, provides the full textbook for the technician. Parts of it, however, are too complex to be fully understood by the beginner and these parts are indicated by black triangles, one being placed at the beginning and one at the end of each such part. In cases where a marked part is continued from a previous page, a black triangle is placed by the page number. If the text is read without these marked sections it forms a continuous whole at a level suited to the beginner. In this way, two complete books are obtained at two different levels for two distinct classes of reader. Appendixes to individual chapters include those proofs and analyses which are highly mathematical, and are intended to supplement the main body of the text and assist those with mathematical ability. An extremely detailed index makes the work suitable for use as a reference book and, in this connection also, the appendixes will be found very useful.

Author's Preface

THIS VOLUME of the Services' Texbook of Radio is concerned with the electronic devices which have made possible radio as we know it to day. The majority of these devices are classified as 'valves', and of the latter there is a large and ever increasing number of types. The subject also includes the study of devices which are not normally classified as valves. For example, cathode-ray tubes, television-camera tubes, transistors and photo-cells are related and important tools used by the radio operator and radio engineer. Moreover, there are certain physical phenomena of importance in radio which are most conveniently discussed in a book on electronics. An outstanding example is the electronic 'noise' which is a fundamental limitation on the detecting power of all radio apparatus. All the above-mentioned subjects find a place in the present volume.

In a book which is primarily written for the beginner and for the technician it would be wrong to present the subject from the academic standpoint. No attempt has therefore been made to develop valve theory in the general sense, although attention has been directed in appropriate places to the valve as a non-linear circuit element. It is hoped, however, that general principles have been sufficiently stressed to enable the more advanced reader to appreciate the fundamental unity of all electronic devices, and to realize that all depend ultimately upon a small number of relatively simple physical principles. Mathematics has been used sparingly, bearing in mind that the only contribution which mathematics can make is arithmetical, if physical principles are well understood.

Electronics is a rapidly developing subject. It is difficult to know what to include and what to omit. In this volume the criterion has been the present or future applicability of the particular device to radio. No two authors would make exactly similar estimates of the relative importance of the devices discussed but it is hoped that nothing of immediate value has been neglected, and that at least a mention has been made of all known devices which may be used soon. In this connection a word should be said regarding transistors. The year 1951 has seen remarkable developments in solid physics which may well lead to far-reaching changes in the devices employed in many radio equipments. Although transistors manufactured today have very limited applications, the 'crystal valves' of tomorrow may compete with and, indeed, displace valves and their associated circuits. It may be, therefore, that Chapter 14 of this volume describes the embryo of a new applied science of first importance.

It is a pleasure to record the assistance which has been received from colleagues and friends in writing this volume. Professor Nicholson of the Royal Naval College, Greenwich, has always been willing to give me the results of his knowledge and experience in this field. The book bears marks in many places of his helpful suggestions, and I am grateful to him. Mr. Cocking of *Wireless World* has been a tower of strength. He has unerringly placed his finger on many weaknesses in the manuscript, and has done far more than his share as an Editor to see that the volume is suitable for its anticipated readers. To him and to my wife for her help with the manuscript my best thanks are due.

April 1952

J. THOMSON

List of Symbols

WHERE a symbol is used with different meanings in different chapters, the relevant chapter numbers are included in brackets in the following list. The symbols are arranged in alphabetical order, capital letters preceding small letters and Roman letters preceding Greek. Minor variations of meaning brought about by the addition of subscripts are not necessarily included.

<i>A</i>	Universal constant = 1.204×10^8 A/sq.m/degree ² (2) Arbitrary constant, (3, 10) Amplification	<i>V</i>	Potential difference Volume (1)
<i>B</i>	Frequency range, bandwidth (2) Flux density (11, 15)	<i>V_a</i>	Anode potential relative to cathode
<i>C</i>	Capacitance	<i>V_{GB}</i>	Negative bias (relative to cathode) on control grid
<i>D</i>	Distance	<i>V_g</i> or <i>V_{g1}</i>	Control-grid potential relative to cathode
<i>E</i>	Electromotive force Electric field strength (3, 16)	<i>V_{g2}</i> or <i>V_{SG}</i>	Screen-grid potential relative to cathode
<i>F</i>	Factor (12)	<i>V_{g3}</i>	Suppressor-grid potential relative to cathode
<i>G</i>	Conductance (10)	<i>V_{HT}</i>	Constant h.t. voltage applied to valve and circuit
<i>I</i>	Electric current	<i>V_k</i>	Cathode voltage relative to negative h.t.
<i>I_a</i>	Anode current	<i>V_s</i>	Signal voltage
<i>I_f</i>	Component of current at frequency <i>f</i>	<i>W</i>	Energy
<i>I_g</i>	Grid current	<i>Y</i>	Admittance
<i>I_k</i>	Cathode current	<i>Z</i>	Impedance
<i>I_s</i>	Signal current	<i>a</i>	Radius of cylindrical electrode (3)
<i>I_{sg}</i>	Screen-grid current	<i>c</i>	Velocity of light 3×10^8 m/sec
<i>L</i>	Inductance Loss (14)	<i>d</i>	Cathode-anode distance (10) Length (16)
<i>M</i>	Mass of molecule (1)	<i>e</i>	Electron charge = 1.6×10^{-19} coulomb
<i>N</i>	Number (1, 2) Noise factor (12, 14)	<i>f</i>	Frequency
<i>P</i>	Power	<i>f_c</i>	Carrier frequency
<i>Q</i>	Quantity of electricity (3,10) Ratio of (energy stored)/(energy dissipated per cycle) (10)	<i>f_i</i>	Intermediate frequency
<i>R</i>	Resistance Gas constant (1)	<i>f_m</i>	Modulation frequency
<i>S_b</i>	Bunching distance	<i>f_o</i>	Local-oscillator frequency
<i>T</i>	Temperature Transit time (10,11)	<i>f_r</i>	Recurrence frequency
<i>T_o</i>	1/ <i>f</i>	<i>f_s</i>	Signal frequency
		<i>g</i>	Internal valve conductance
		<i>g_c</i>	Conversion conductance

LIST OF SYMBOLS—continued

g_m	Mutual conductance	y	Length (in Cartesian co-ordinates) (16)
h	Planck's constant = 6.62×10^{-34} joule/cycle	α	Number (13) Acceleration (16)
j	Vector operator = $\sqrt{-1}$	β	Number (13)
k	Boltzmann's constant = 1.38×10^{-23} joule/degree Kelvin	γ	Propagation constant
l	Length	δ	Small increment
m	Mass Coefficient of modulation (7)	ε	Base of natural logarithms = 2.71828
n	Number	θ	Angle = $2\pi ft = \omega t$ (5) = $2\pi fT_g$ (10)
p	Pressure (1) Time duration (9)	κ	Permittivity
r	Internal valve resistance Duty cycle (9) Length (in polar co-ordinates) (16)	κ_0	Permittivity of free space = 8.854×10^{-12} F/m
s	Distance	λ	Wavelength
t	Time Noise temperature (14)	μ	Amplification factor
u	Mean-square speed Velocity (15)	π	Ratio circumference/diameter of circle = 3.14159
v	Velocity	ρ	Density (1) Charge density (3)
v_g	Signal voltage on control grid	ϕ	Net thermionic work function (2, 14) Conduction angle (5) Phase angle (7)
x	Distance Length (in Cartesian co-ordinates) (16)	ω	Angular frequency = $2\pi f$

Abbreviations of Units

A	Ampere	g	Gramme
Å	Angstrom unit = 10^{-10} m	k	Kilo = 10^3 (prefix)
F	Farad	m	metre milli = 10^{-3} (prefix)
H	Henry	p	10^{-12} (prefix)
M	Mega = 10^6 (prefix)	s or sec	Second
V	Volt	°	Degree, of angular measure of temperature, °C = degree Centigrade °F = degree Fahrenheit °K = degree Kelvin
W	Watt		
Ω	Ohm		
c	Cycle		
c/s	Cycles per second		
db	Decibel		
eV	Electron-volt		

CHAPTER 1

Electrons in Metals

THERMIONIC EMISSION may be defined as the process whereby electrons are freed from a solid or liquid when the latter absorbs energy in the form of heat. There are other similar processes whereby electrons may be freed, and it assists in the understanding of thermionic effects if the more general physical picture of all types of emission is first painted. It is the purpose of this chapter, therefore, to describe the nature of the metallic elements, to show how electrons behave in such elements, and to deduce under what circumstances electrons may be liberated from them. The four principal types of electron emission, viz., thermionic, photo-electric, secondary and field emission, will then be considered as special cases.

Electrons, Protons and Neutrons; Atomic Structure

To understand how valves and cathode-ray tubes work, it is necessary to know something about the electrons which give these devices their distinctive qualities; and to understand something of the properties of electrons it is necessary to know a little about the atomic structure of matter. It may appear that this is taking the reader rather deeply into the realm of pure physics, but the knowledge so acquired will stand him in good stead when he comes to study the application of physics to electronic devices.

All matter consists of atoms which contain electrons, protons and neutrons in a particularly intimate association. There are many things we do not know about electrons, protons and neutrons, but certain properties are quite definite. The electron has a negative charge of 1.60×10^{-19} coulomb. Its mass can be measured quite accurately, but varies with its speed (see Appendix 1.1). When it is travelling at the speeds usually encountered in electronics, its mass is 9.1×10^{-31} kg. The proton has a positive charge equal to the negative charge carried by the electron, but its mass is about 1840 times as large, being 1.66×10^{-27} kg. The neutron, as its name implies, has no charge—it is electrically neutral—and its mass is roughly equal to that of the proton.

Atoms are made up of these three 'particles' in a very simple manner. Each element possesses a certain number of protons and an equal number of electrons. For example, the hydrogen atom has one proton and one electron; the sodium atom has eleven protons and eleven electrons; the chlorine atom has seventeen protons and seventeen electrons. The number of electrons must always equal the number of protons to keep the atom electrically neutral, and each element is characterized by its number of electrons, or, as it is called, its atomic number.

Where then, it may be asked, do the neutrons come into this? Their role is much less important from one point of view than that of the protons and electrons. If an atom is weighed, it will be found that it contains more mass than can be accounted for by the protons and electrons. For example, sodium,

containing 11 protons and 11 electrons, weighs as much as 23 protons and 11 electrons. The remainder of the mass ($23 - 11$) is accounted for by the neutrons. So the atom of sodium contains 11 electrons, 11 protons and 12 neutrons. This picture of the atom accounts admirably for all its known properties. Hydrogen is the only element which has an atom devoid of neutrons. The commonest hydrogen atom is simply one proton and one electron. Helium contains 2 protons, 2 neutrons and 2 electrons.

Suppose now we write down all the numbers in order from 1 to 100 and let these be atomic numbers. Can we assign to each number an element? The answer is "Yes". Table 1.1 has been constructed in this way. Two points should be noted. First, there are no blank spaces. Every number has been accounted for up to 96. Secondly, there are far more atoms than there are numbers, for often one element possesses many atoms, each possessing a different number of neutrons. These atoms are called isotopes of the element. They have the same atomic number—they are identical chemically—but they differ in mass, or in the number of neutrons they possess.

Atoms do not simply consist of a jumble of electrons, neutrons and protons. They are arranged in a very beautiful manner. The protons and neutrons, comprising almost all the mass of the atom, are confined to an extremely small volume and together form what we call the nucleus. The electrons, on the other hand, are spread out in space round the nucleus in shells which have been compared to the orbits of the planets round the sun. The electrons can be stripped off the atom rather easily, leaving it positively charged, of course, but it is much more difficult to break up the nucleus. In fact this has only been done intentionally during the last twenty years.

The extra-nuclear electrons as they are called, are arranged in orderly shells. The first shell may contain one or two electrons, but not more than two. This is called the K-shell. Hydrogen and helium when in the normal state have their electrons in the K-shell. The second shell, further away from the nucleus, may contain anything up to 8 electrons. Clearly we have a full K and L shell in the case of the atom of atomic number 10; i.e., neon. The next shell, the M shell, also contains 8 electrons when complete and argon has complete K, L, M shells. After this there is a certain complication. The N shell can contain up to 18 electrons and so our next complete shell structure is at atomic number $2 + 8 + 8 + 18 = 36$; i.e., at the gas krypton. The O shell likewise contains 18 electrons and is complete at the gas xenon while the P shell, the last complete one, contains 18 electrons and finishes at the gas radon.

The interesting part about this classification is that it is based on the idea of complete electron shells and the atoms with these complete shells are helium, neon, argon, krypton, xenon and radon. These elements are all gases and are all completely inert. They will not combine with any other element. It was this fact which suggested the idea of electron shells. Conversely, atoms containing one electron more than is required to form a complete shell should be extremely active; and this is exactly true, for the atoms following the inert gases are lithium, sodium, potassium, rubidium, caesium and francium, all alkali metals and all extremely active chemically.

Following this idea up, the so-called Periodic Table of the elements has been constructed, Table 1.2, in which, arranging the elements in columns according to their chemical activity and likeness to each other, we find that

TABLE 1.1
Table of the Elements and their known Isotopes (1948)

Atomic Number	Name of Element	Number of known Isotopes	Atomic Number	Name of Element	Number of known Isotopes
1	HYDROGEN	3	49	INDIUM	10
2	HELIUM	3	50	TIN	22
3	LITHIUM	3	51	ANTIMONY	18
4	BERYLLIUM	3	52	TELLURIUM	17
5	BORON	3	53	IODINE	14
6	CARBON	5	54	XENON	21
7	NITROGEN	5	55	CAESIUM	16
8	OXYGEN	5	56	BARIUM	17
9	FLUORINE	4	57	LANTHANUM	9
10	NEON	5	58	CERIUM	9
11	SODIUM	5	59	PRASEODYMIUM	7
12	MAGNESIUM	5	60	NEODYMIUM	13
13	ALUMINIUM	4	61	ILLINIUM	10
14	SILICON	5	62	SAMARIUM	12
15	PHOSPHORUS	5	63	EUROPIUM	10
16	SULPHUR	7	64	GADOLINIUM	10
17	CHLORINE	6	65	TERBIUM	2
18	ARGON	6	66	DYSPROSIUM	8
19	POTASSIUM	6	67	HOLMIUM	3
20	CALCIUM	10	68	ERBIUM	9
21	SCANDIUM	8	69	THULIUM	3
22	TITANIUM	7	70	YTTERBIUM	10
23	VANADIUM	5	71	LUTECIUM	3
24	CHROMIUM	7	72	HAFNIUM	8
25	MANGANESE	5	73	TANTALUM	3
26	IRON	8	74	TUNGSTEN	10
27	COBALT	7	75	RHENIUM	7
28	NICKEL	9	76	OSMIUM	10
29	COPPER	9	77	IRIDIUM	5
30	ZINC	10	78	PLATINUM	10
31	GALLIUM	11	79	GOLD	10
32	GERMANIUM	10	80	MERCURY	11
33	ARSENIC	8	81	THALLIUM	12
34	SELENIUM	13	82	LEAD	12
35	BROMINE	9	83	BISMUTH	9
36	KRYPTON	18	84	POLONIUM	11
37	RUBIDIUM	14	85	ASTATINE	4
38	STRONTIUM	12	86	RADON	3
39	YTTRIUM	11	87	FRANCIUM	2
40	ZIRCONIUM	12	88	RADIUM	5
41	COLUMBIUM	11	89	ACTINIUM	3
42	MOLYBDENUM	12	90	THORIUM	8
43	TECHNETIUM	12	91	PROTACTINIUM	3
44	RUTHENIUM	14	92	URANIUM	6
45	RHODIUM	10	93	NEPTUNIUM	6
46	PALLADIUM	11	94	PLUTONIUM	3
47	SILVER	10	95	AMERICIUM	2
48	CADMIUM	41	96	CURIUM	2

the K, L, M, . . . shells fit perfectly into the picture. For example all the halogens—fluorine, chlorine, bromine, iodine, astatine—turn out to be elements with one electron less than a complete shell. They are, indeed, exactly opposed to the alkali metals, all of which have one electron more than a complete shell. This explains why these elements combine so readily to form such well-known compounds as sodium chloride, potassium bromide, etc, for in the combination one electron leaves the alkali and tacks itself on to the halogen, thus giving two complete shells. Chemical valency is also explained by the Periodic Table.

The alkalis have one electron too many (so to speak) and are monovalent, while the alkaline earths, beryllium, magnesium, calcium, etc, have two electrons in excess. They are therefore divalent. Both groups are electro-positive (i.e., their atoms easily give up an electron), again as one would expect. At the other side of the table, the halogens are monovalent, but electro-negative, the elements oxygen, sulphur, etc, are divalent and combine readily with the electro-positive elements while carbon and silicon are the most interesting cases of all. Poised equally between two complete shells, each may behave as an electro-positive (CCl_4) or as an electro-negative (CH_4) element. The valency, as might be expected, is 4.

Naturally, the picture painted above is too simple to explain all chemical and physical phenomena, but it is sufficient to allow the reader to understand many of the salient points in electronic action.

Explanation of Table 1.2

The number in the top left-hand corner of each rectangle is the *atomic number* of the element, the number of electrons it possesses. The letter in the top right-hand corner is its internationally-agreed symbol.

The number across the bottom of the rectangle is its atomic weight. This is the *average* weight of an atom of the element on a scale of mass in which oxygen is 16.0000. The atomic weight depends, among other things, on the number of isotopes possessed by the element. For example, chlorine is a mixture of two isotopes, one of which contains 18 neutrons and the other 20. In natural chlorine the isotope containing 18 neutrons clearly predominates, giving the natural gas an atomic weight nearer to 35 than to 37. Each electron-shell is composed of sub-shells, and these are not filled up in exact rotation, but according to certain quantum rules. The chemical nature of the element depends not only on the number of electrons in the shells but on their distribution. For example, the shell structures of Fe, Co, Ni, Cu, Zn are as follows:—

		K	L	M	N
Iron	Fe	2	2 + 6	2 + 6	6 + 2
Cobalt	Co	2	2 + 6	2 + 6	7 + 2
Nickel	Ni	2	2 + 6	2 + 6	8 + 2
Copper	Cu	2	2 + 6	2 + 6	10 + 1
Zinc	Zn	2	2 + 6	2 + 6	10 + 2

showing that Fe, Co, Ni are very much alike chemically, having the same number of electrons in the outermost shell, and that they are fundamentally divalent.

TABLE 1·2
Periodic Table of the Elements.

	GROUP I	GROUP II	GROUP III	GROUP IV	GROUP V	GROUP VI	GROUP VII	GROUP VIII	GROUP O.
K	1 H. HYDROGEN. 1·008								2 He. HELIUM 4·003
L	3 Li. LITHIUM 6·940	4 Be. BERYLLIUM 9·02	5 B. BORON 10·82	6 C. CARBON 12·010	7 N. NITROGEN 14·008	8 O. OXYGEN 16·0000	9 F. FLUORINE 19·000		10 Ne. NEON 20·183
M	11 Na. SODIUM 22·997	12 Mg. MAGNESIUM 24·32	13 Al. ALUMINIUM 26·97	14 Si. SILICON 28·06	15 P. PHOSPHORUS 30·98	16 S. SULPHUR 32·06	17 Cl. CHLORINE 35·457		18 Ar. ARGON 39·944
N	19 K POTASSIUM 39·096	20 Ca. CALCIUM 40·08	21 Sc. SCANDIUM 45·10	22 Ti. TITANIUM 47·90	23 V. VANADIUM 50·95	24 Cr. CHROMIUM 52·01	25 Mn. MANGANESE 54·93	26 Fe. IRON 55·84	27 Co. COBALT 58·69
N	29 Cu. COPPER 63·57	30 Zn. ZINC 65·38	31 Ga. GALLIUM 69·72	32 Ge. GERMANIUM 72·60	33 As. ARSENIC 74·91	34 Se. SELENIUM 78·96	35 Br. BROMINE 79·916		36 Kr. KRYPTON 83·7
O	37 Rb. RUBIDIUM 85·48	38 Sr. STRONTIUM 87·63	39 Y. YTTORIUM 88·92	40 Zr. ZIRCONIUM 91·22	41 Nb. NIOBIUM 92·91	42 Mo. MOLYBDENUM 95·95	43 Tc. TECHNETIUM 98·91	44 Ru. RUTHENIUM 101·7	45 Rh. RHODIUM 102·91
O	47 Ag. SILVER 107·88	48 Cd. CADMIUM 112·41	49 In. INDIUM 114·76	50 Sn. TIN 118·70	51 Sb. ANTIMONY 121·76	52 Te. TELLURIUM 127·61	53 I. IODINE 126·92		54 Xe. XENON 131·3
P	55 Cs. CAESIUM 132·91	56 Ba. BARIUM 137·36	57-71. RARE EARTHS	72 Hf. HAFNIUM 178·6	73 Ta. TANTALUM 180·88	74 W. TUNGSTEN 183·92	75 Re. RHENIUM 186·31	76 Os. OSMIUM 190·2	77 Ir. IRIDIUM 193·1
P	79 Au. GOLD 197·2	80 Hg. MERCURY 200·61	81 Tl. THALLIUM 204·39	82 Pb. LEAD 207·21	83 Bi. BISMUTH 209·00	84 Po. POLONIUM 210	85 At. ASTATINE		86 Rn. RADON 222
Q	87 Fr. FRANCIUM	88 Ra. RADIUM 226·05	89 Ac. ACTINIUM 227·05	90 Th. THORIUM 232·12	91 Pa. PROTACTINIUM 231	92 U. URANIUM 238·07	93 Np. NEPTUNIUM 237	94 Pu. PLUTONIUM 239	95 Am. AMERICIUM 241
								96 Cm. CURIUM 242	

Cu, on the other hand, is monovalent, and similar to the alkalis. If an element could exist with the structure $2 + 2 + 6 + 2 + 6 + 10$, it would be a rare gas. Quantum rules forbid this.

The Metallic Elements

Electrically, substances may be classified according to their conductivity, and when this is done, it is found that the best conductors are metals, silver and copper heading the list. Chemically, a metal is an electro-positive element; that is, it is an element which readily forms compounds in which it behaves as if each atom were positively charged. Add to this that the atomic physicist identifies the most electro-positive elements with atoms possessing one or more electrons which are rather loosely bound to the remainder of the electronic structure, and a fairly definite picture of a metallic conductor begins to emerge. We may imagine a metal to consist of a number of almost fixed atoms of the substance with one electron missing from each (such atoms are called positive ions) permeated with an equal number of electrons which are free to move about in the metal. Fig 1.1 gives some idea of the metal structure. The squares, marked MI, are the metal ions, heavy and more or less fixed in position, while the circles, marked E, are the electrons, drifting around from one point to another under the action of random electric forces.

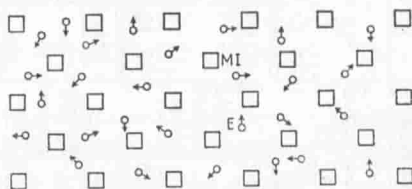


Fig. 1.1. Pictorial representation of a metal. The squares are the heavy positive ions and the small circles are the electrons. The arrows indicate that the electronic movement is random.

This is why a metal easily conducts electricity. An electric force, no matter how small it is, set up in the metal by means of a potential difference between its ends, causes the electrons to drift more in one direction than in the other. This constitutes an electric current. The electrons cannot easily leave the metal, for they are negatively charged and their loss would cause an excess positive charge in the metal which would pull them back. But some *do* escape, fortuitously, if their energy is large enough, just as molecules evaporate from a water surface even at room temperature. The kinetic energy of an electron can be expressed by the usual formula $W = \frac{1}{2}mv^2$, where W is the energy, m is the mass of the electron and v is its speed. In metals at ordinary temperature W is not large enough to allow an appreciable fraction of the electrons to escape. It is inconvenient in electronic calculations to express W in terms of mass and speed. Instead, a new unit of energy is used which is so common that the reader should master it immediately. It is the electron-volt. If an electron is accelerated from rest through a potential rise of V volts, then it acquires the energy $W = \frac{1}{2}mv^2 = Ve$ where e is the electronic charge. We speak of such an electron, after it has been accelerated, as possessing V electron-volts of energy. Clearly 1 electron-volt = 1.6×10^{-19} joule, since $e = 1.6 \times 10^{-19}$ coulomb.

We can measure the energy which an electron must have to leave a metal. This energy varies from one substance to another and is known as the "work function". The work function of an alkali, as might be expected, is small. The lowest is for caesium and is 1.75 electron-volts. Tungsten, on the other hand, which is much used as a thermionic emitter, has a work function of 4.55 electron-volts. As has been said, this is the energy which an electron must have to escape from the metal tungsten against the surface forces.

Electron Gas

▶ To begin with it was assumed that the electrons in a metal were free in exactly the same sense as the molecules of a gas. For that reason the term "electron gas" became well established as a short description of the conduction electrons. The characteristic property of a perfect gas is that its molecules are subject to quite negligible inter-molecular forces. When this idea is expressed in mathematical form, it is possible to calculate how the speeds of the molecules are distributed at any given instant. It is found, as one might expect, that the speeds range from zero to the speed of light, but that the majority of the molecules have speeds not far removed from the average. This average speed depends only upon the temperature. Indeed, the kinetic energy of the gas molecules is the heat content of the gas, and since the heat varies directly as the temperature, so does the kinetic energy and therefore the mean square of the speed of the molecules. Fig. 1.2 exhibits the distribution of velocities in a perfect gas, the ordinate expressing the fraction, $\delta N/N$, of the molecules having speeds between v and $v + \delta v$, where δv is a small increment, and the abscissa measuring v . Curve A shows the distribution for a gas at room temperature and curve B for a gas at 2,000°C. The mathematical expression corresponding to these curves and others is given in Appendix 1.2. It was first derived by Clerk Maxwell and is often referred to as Maxwell's distribution.

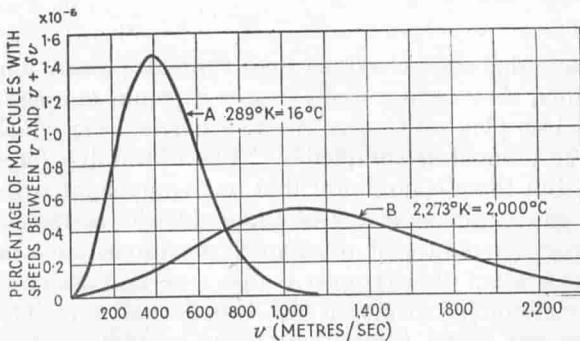


Fig. 1.2. Maxwell distributions of molecular velocities in oxygen for temperatures of 16°C and 2,000°C.

However, if the mean energy of an electron is calculated from the Maxwell formula, and if it is assumed that all the valency electrons in a metal are "free", a serious difficulty arises. There are so many atoms per cubic metre in a metal—about 6×10^{28} in tungsten—that the same number of electrons crowded into the same space would repel each other strongly, thus destroying the hypothesis of freedom of movement, and so denying the postulate on which the Maxwell

formula is built. It was Fermi and Dirac who solved this problem, applying the quantum theory properly to the possible states of energy of the electrons and thus deriving a velocity distribution law of a different nature. Fig. 1.3 exhibits the Fermi-Dirac distribution in the same way as Fig. 1.2 shows the Maxwell distribution, curves A and B referring once more to temperatures of 15°C and $2,000^{\circ}\text{C}$ respectively.

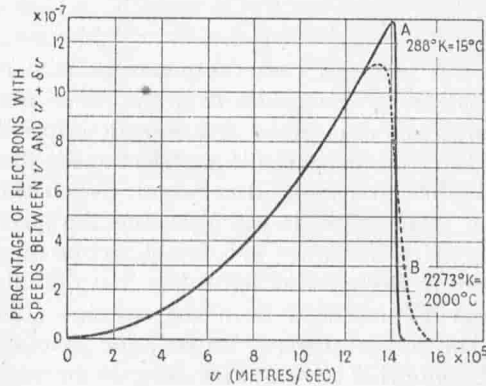


Fig. 1.3. Fermi-Dirac distribution of electronic velocities in tungsten for temperatures of 15°C and $2,000^{\circ}\text{C}$.

This is very different from the Maxwell distribution, and it leads to much higher average speeds. Only at exceedingly high temperatures do the two formulae correspond. At room temperatures the average electron velocity in tungsten is about 1.1×10^6 metres/sec. For the same temperature, using the Maxwell distribution, it would be about 10^5 metres/sec.

Surface Forces

Despite the fact that the electrons in a metal possess on the average considerable energy, they cannot easily escape at room temperature. Surface forces are called into play which give rise to a barrier over which the electrons must climb to free themselves completely. The restraining forces are not yet fully understood, but there is no doubt that one component is the electrostatic force due to the mirror image of the escaping electron in the surface. Whatever may be the exact nature of the forces, the important matter from the stand-point of the student of electronics is that a certain amount of work must be done by an electron to surmount the surface barrier. This work varies considerably from one metal to another, being smaller for the alkali metals (lithium, sodium, potassium, rubidium and caesium) and larger for such metals as silver and platinum. Unless the electrons in the metal have sufficient energy to perform the work of surmounting this barrier, escape is impossible. This fact is of major importance in all forms of electron emission.

The net quantity of work which must be done to free an electron from a metal is usually given in electron-volts and is called the 'net work function' of the metal. There are other work functions associated with a metal, but it is unnecessary to consider them in the meantime. The net work functions of the more common metals are listed in Table 1.3.

TABLE 1.3.

Metal		Net Work Function (electron-volts)
Caesium	Cs	1.75
Rubidium	Rb	1.80
Potassium	K	1.85
Sodium	Na	1.95
Lithium	Li	2.2
Copper	Cu	4.2
Molybdenum	Mo	4.25
Mercury	Hg	4.5
Tungsten	W	4.55
Silver	Ag	4.65
Platinum	Pt	6.15

The existence of this potential barrier at the surface of a metal explains many puzzling effects. For example, if a complete metallic circuit is made by joining two dissimilar metals together at both ends, the contact difference of potential at the two junctions can be explained as the difference between the work functions. The electrons escape more readily from the metal of lower work function, and do so until the flow has built up a sufficiently large opposing potential difference to bring the charge into a state of dynamic equilibrium.

Thermo-electric effects can be explained in the same way.

Electron Emission

At room temperatures very few electrons attain sufficiently high speeds to escape from the metal, but any agency which augments the electronic energy can produce electron emission. There are four principal ways in which this may occur:—

1. Thermionic emission (in which heat is given to the metal and to the contained electrons),
2. Photo-electric emission (in which electromagnetic radiation is absorbed by the electron gas),
3. Secondary emission (in which the energy of electrons or ions projected against the metallic surface is absorbed by the metallic electrons),
4. Auto-electronic or field emission (which is in a category by itself in that the additional energy is given to the electron gas by the action of an external electric field in overcoming the surface barrier).

Use is made of all four mechanisms in electronic devices, although thermionic emission is by far the most important, and so far very little use has been made of field emission. In many devices two or more types of emission occur simultaneously. Modern developments are mainly along the lines of improving our knowledge of these mechanisms, for it is only in the simplest cases that we have a thorough understanding of them. The full exploitation of the unexplored possibilities of electronic emission will require many years yet of patient study.

Thermionic Emission

When a metal is heated to a sufficiently high temperature, a considerable fraction of the free electrons will possess sufficient energy to overcome the surface forces and to escape entirely. This is thermionic emission. It is usually studied and employed with the emitting metal in an evacuated container so that the action of gas molecules will not complicate the process, although in a relatively small number of devices the additional action of the gas molecules and ions is put to a good use. If the emitting substance is tungsten at a temperature of $2,000^{\circ}\text{C}$, the *average* electron speed within the metal is about 1.2×10^6 metres/sec. This corresponds roughly to an energy of 4 electron-volts. But the work function of tungsten is only 4.55 electron-volts. Therefore many of the electrons can escape, and tungsten at this temperature should be a reasonably good emitter. In fact its total emission is about 1 mA/sq.cm.

Photo-Electric Emission

When electromagnetic radiation falls on a metal surface, some of the electrons can absorb the energy, transforming it into the kinetic form. It has been demonstrated by innumerable experiments of the most diverse type that, when an electron absorbs radiation of frequency f , it does so in packets or *quanta* of magnitude hf , where h is a universal constant having the value 6.60×10^{-34} joule/cycle, and known after its discoverer as Planck's constant. If the quantum hf is large enough, the kinetic energy of the electron which has absorbed it will be sufficient to surmount the surface barrier. This is photo-electric emission. To obtain some idea of the frequency of the radiation and therefore of its position in the electromagnetic spectrum, the metal tungsten may again be cited. In this case the work function is 4.55 electron-volts or $4.55 \times 1.6 \times 10^{-19}$ joule. To have a reasonable chance to escape, therefore, the electron must absorb a quantum of this value, and therefore $hf = 4.55 \times 1.6 \times 10^{-19}$. Substituting for h in this equation we find $f = 1.10 \times 10^{15}$. The wavelength, λ , and the frequency, f , of electromagnetic radiation in free space are connected by the equation $\lambda f = c$, where c is the velocity of light, 3×10^8 metres/sec. Therefore $\lambda = 3 \times 10^8 / 1.10 \times 10^{15}$ metre or 2.72×10^{-7} metre. An Angstrom unit ($\text{\AA}$) is 10^{-10} metre. Therefore $\lambda = 2,720 \text{\AA}$, and the required radiation is in the ultra-violet region. Hence, although visible radiation will not produce electron emission from tungsten, ultra-violet radiation (or X-radiation) of wavelength less than $2720 \text{\AA}$ will be effective. For comparison the calculation may be done for caesium for which the work function is 1.75 electron-volts. The required wavelength in this case is $7,070 \text{\AA}$ which is in the red region of the spectrum. Caesium will therefore emit electrons when almost any type of visible radiation falls on it.

Clearly X-rays, for which the frequency is much higher than for ultra-violet or visible radiation, will cause emission of electrons from any metal. Conversely, radiation in the far infra-red part of the spectrum cannot possibly produce a photo-electric emission from any pure metal.

Secondary Emission

When a stream of electrons bombards a metal surface, other 'secondary' electrons are emitted, particularly when the speed of the bombarding electrons is high. The effects produced are more complex than in the case of thermionic

and photo-electric emission for, as will be seen, there are two distinct mechanisms whereby the secondaries may be produced.

When the bombarding primary electrons fall on the surface at right angles, it is difficult to imagine how they can cause the emission of electrons which were originally free in the metal, since collisions with free electrons cannot give rise to the necessary momentum in the direction away from the metal. It is thought, therefore, that the secondaries emitted in this case were originally bound to atoms and were freed by the primary energy. The electron gas would only be involved as a means of transferring part of the energy inside the metal.

When the bombarding primaries fall on the surface at a small angle, however, the transfer of momentum to the direction away from the metal is a possibility. It is thought, therefore, that secondary emission in this case may contain electrons which were originally free as well as electrons which were originally bound. If this view of the matter is correct, it is to be expected that secondary emission will be greater at small than at large angles of incidence of the primaries. This has been verified experimentally.

One primary can produce more than one secondary. The primary energy may be absorbed by more than one atom or by more than one free electron, liberating several electrons with speeds great enough to surmount the surface barrier. *On the average*, however, and even under the most favourable conditions, a primary falling normally on a pure metal surface does not produce more than $1\frac{1}{2}$ secondaries. This figure increases in a pure metal to about 2 when the angle of incidence of the primary is small, and to 10 or more for the surfaces of certain metallic compounds.

The maximum number of secondaries is produced at primary energies of the order of 200–1,000 electron-volts, falling off slowly as the energy is increased beyond this point, and falling off sharply as the energy is decreased. Very few secondaries are observed when the primary energy is less than 10 volts, and this is in line with what might be expected from the mechanism sketched above. The decrease in emission for high energies is also what might be expected. It is due to the greater depth of penetration into the metal of the high speed electrons. The secondaries produced have then less chance to find their way to the surface and to escape.

Secondary emission is important in two ways. First, its effects in ordinary thermionic valves are highly undesirable and steps are taken to eliminate them. Secondly, the phenomenon is gaining in usefulness every day in devices where exceedingly small currents have to be directly amplified. Such devices employ emitters for which the ratio of secondaries to primaries is high and then cause the primary emission to be proportional to the current to be amplified. As the techniques of surface control improve, secondary emission will become more and more important. It is the adventitious variability of the effects which has made them useless in the past.

Field Emission

There is one means of producing increased electronic emission from a metal which has so far found little application, but which may be important in the future. If a metal surface is placed in close proximity to another metal surface and if the potential difference between the two is gradually increased,

there will come a time at which an electric discharge will occur. If, however, the space between the surfaces is thoroughly exhausted so that the gas pressure is exceedingly small (of the order of 10^{-7} mm of mercury), electronic emission from the metal surface which is negatively charged will occur. Since there is very little gas present, there is no means of sustaining an electric discharge. This emission is caused by the intense electric field near the negative electrode (cathode). It is supposed that the external field penetrates some little distance into the cathode metal, thus neutralizing to some extent the potential barrier at the surface. This augments the electron evaporation. The effect has been studied both theoretically and experimentally; the results indicate that the emission is almost independent of the temperature, but is largely dependent upon the work function of the metal. If we tentatively imagine the field which is necessary to produce a potential variation of the same order as the work function in about the same distance, its value must be about one hundred million volts per metre. This can readily be produced if required in the laboratory, but is excessive for a practical device at the present time.

Velocities of Emitted Electrons

- The electrons freed from the metal by the action of heat, radiation or any other form of energy, possess a velocity distribution determined by their history. Therefore, to a first approximation, their velocity distribution will correspond to that part of the Fermi-Dirac distribution applicable to high energy electrons travelling in a given direction. This is shown for two different temperatures by the curves of Fig. 1.4. If the energy of an electron after activation is V_E electron-volts, and it has overcome a barrier V_B , its residual energy will be $V_E - V_B$. It may be shown that the appropriate distribution is almost according to Maxwell's formula, when the latter is suitably modified to apply to one direction only.

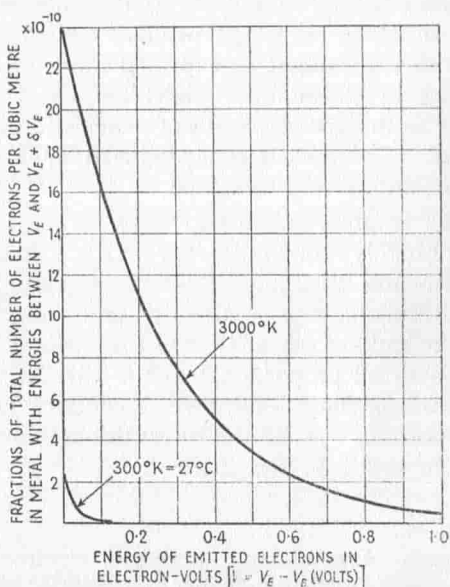


Fig. 1.4. Distribution of energies among electrons which can escape from tungsten.

We therefore find (i) that the mean speed of the emitted electrons depends upon (a) the temperature of the metal and (b) the total energy given to it in the form of heat, radiation, etc, and (ii) that in all types of emission the speeds of the electrons after they have left the metal are distributed approximately according to Maxwell's Law although, of course, the total number emitted is not what would be expected from an initial Maxwell distribution. The mean speed in the case of the electrons emitted by tungsten at $3,000^{\circ}\text{K}$ is about 3×10^5 metres/sec, corresponding to an energy of about 0.2 electron-volt. ◀

APPENDIX 1.1

Mass and Energy

It is one of the consequences of the theory of relativity that the total energy of a mass m moving with a speed v is mc^2 , where c is the speed of light. The quantity mc^2 is not a constant, for the mass m is given by

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

and therefore as v increases, m also increases. Where v/c is less than unity, the expression $(1 - v^2/c^2)^{-1/2}$ can be expanded by the Binomial Theorem in the form

$$1 + \frac{1}{2}v^2/c^2 + 3v^4/8c^4 + 5v^6/16c^6 + \dots$$

and if v/c is so small that v^4/c^4 and all higher terms may be neglected, then

$$\begin{aligned} mc^2 &= m_0c^2(1 + \frac{1}{2}v^2/c^2) \\ &= m_0c^2 + \frac{1}{2}m_0v^2 \\ &= \text{rest energy} + \text{classical kinetic energy.} \end{aligned}$$

This suggests that $(m - m_0)c^2$ is the true kinetic energy of the particle and m_0c^2 is the energy equivalent to a mass m_0 at rest.

The value of m/m_0 in Table 1.4 is given to four significant figures for various values of v/c . The results show that the classical assumption of the invariability of mass is correct to better than one part in 200 for all speeds less than one-tenth of that of light. Despite this it is necessary in certain electronic calculations to use the more exact formula.

TABLE 1.4

v/c	1/100	1/10	1/5	1/4	1/3	1/2
m/m_0	1.000	1.005	1.021	1.033	1.061	1.153

APPENDIX 1.2

Energy Distribution in a Perfect Gas

It can be shown theoretically and it has been verified experimentally that the distribution of velocities in a 'perfect' gas is according to the well-known Maxwellian formula

$$\delta N = 4\pi N \left(\frac{K}{\pi}\right)^{3/2} v^2 e^{-Kv^2} \delta v \quad (1.1)$$

where N is the total number of molecules, and δN have speeds between v and $v + \delta v$. K is a constant which can be evaluated in terms of the mean-square speed u , for from equation (1.1)

$$u^2 = 3/(2K) \quad (1.2)$$

It may also be shown theoretically that the pressure in a 'perfect' gas is given by the expression

$$p = \frac{1}{3}\rho u^2$$

where ρ is the density. The gas equation is, of course,

$$pV = RT$$

Let V refer to 1 metre³; then the gas constant R refers to 1 metre³ and

$$p = \frac{1}{3}\rho u^2 = RT$$

But $\rho = NM$, where N is the number of molecules per (metre)³ and M is the mass of a molecule.

Therefore,

$$\begin{aligned} \frac{1}{3}NMu^2 &= RT \\ \text{or } \frac{1}{3}Mu^2 &= \frac{R}{N}T = kT \end{aligned} \quad (1.3)$$

and $k = R/N$ is the gas constant/molecule, named Boltzmann's constant, and having the value 1.38×10^{-23} joule/degree Kelvin.

From equations (1.2) and (1.3) it follows that

$$K = \frac{M}{2kT} \quad (1.4)$$

Substituting in equation (1.1), the distribution of molecular energies is given as

$$\delta N = \frac{2N}{\pi^{1/2}(kT)^{3/2}} W^{1/2} e^{-W/kT} \delta W \quad (1.5)$$

where δN is the number having energies between W and $W + \delta W$

Energy Distribution in Electron Gas in Metal

Applying quantum mechanics to the states of the electrons in a metal, it may be shown that the formula corresponding to Maxwell's for the energy distribution is

$$\delta N = \frac{3N}{2W_0^{3/2}} \frac{W^{1/2} \delta W}{e^{(W - W_0)/kT} + 1} \quad (1.6)$$

where W_0 is given in terms of N and universal constants by the expression

$$W_0 = \frac{h^2}{2m} \left[\frac{3N}{8\pi} \right]^{2/3}$$

Here h is Planck's constant, 6.62×10^{-34} joule/cycle, and m is the electronic mass. N , the number of electrons/metre³ may be taken equal to the number of atoms in the metal in the same volume. This is not a constant, but a slowly varying function of the temperature, T . However, for all temperatures at which the metal is solid, the variation is negligible. Equation (1.6) is the so-called Fermi-Dirac distribution.

Velocity Distribution in Given Direction

Equation (1.1) gives the number of molecules having speeds between v and $v + \delta v$ in any direction. It is derived from the distribution of component speeds, and for any one direction the corresponding equation is

$$\delta N_x = N \left[\frac{K}{\pi} \right]^{1/2} e^{-Kv^2} \delta v \quad (1.7)$$

If this is the velocity distribution in the x -direction, the number of molecules having speeds between v and $v + \delta v$ which will strike unit surface perpendicular to the x -axis in unit time is

$$v \delta N_x = N \left[\frac{K}{\pi} \right]^{1/2} v e^{-Kv^2} \delta v$$

Writing this in terms of W instead of v , and remembering that $K = M/(2kT)$ (equation 1.4),

$$v \delta N_x = \frac{N}{(2\pi MkT)^{1/2}} e^{-W/kT} \delta W \quad (1.8)$$

The Fermi-Dirac equation corresponding to (1.7) is

$$\delta N_x = \frac{4\pi m^2 kT}{h^3} \log_e \left[1 + e^{(W_0 - W)/kT} \right] \delta v \quad (1.9)$$

and that corresponding to (1.8) is

$$v \delta N_x = \frac{4\pi m kT}{h^3} \log_e \left[1 + e^{(W_0 - W)/kT} \right] \delta W \quad (1.10)$$

When the exponential term is small compared with unity, equation (1.9) may be written

$$\delta N_x = \frac{4\pi m^2 kT}{h^3} e^{W_0/kT} e^{-W/kT} \delta v \quad (1.11)$$

This equation has the same form as equation (1.7). Therefore the two hypotheses, Fermi-Dirac and Maxwell-Boltzmann, lead to the same distribution for high energies. The mean speeds and the mean energies of the emitted electrons are the same in both cases. The difference lies in the total fraction of the electrons involved.

CHAPTER 2

Thermionic Emission

IF A LENGTH of tungsten wire is heated in a vacuum by passing an electric current through it, and if all the electrons emitted by the tungsten are collected on a positively-charged plate situated near the wire, then it is found that the thermionic current increases very rapidly indeed with the temperature of the tungsten wire. The curve in Figure 2.1 shows just how rapid the increase is. Valves employ some form of hot metal or compound as the thermionic emitter, and it is the purpose of this chapter to explain the types of emitters used in practice, and to show how these are arranged along with other metal parts to form the various thermionic devices we use today.

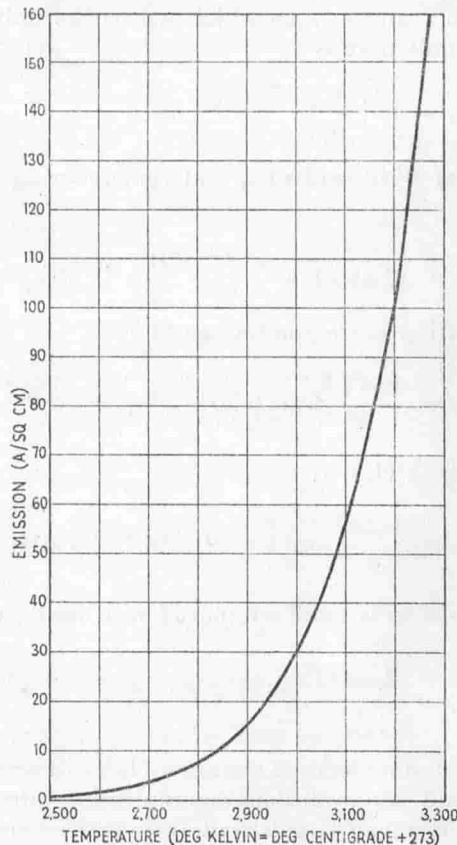


Fig. 2.1. Emission per unit area of pure tungsten at different temperatures.

The subject is a complex one and the more advanced student will want to know more about the detailed processes used to obtain good emitters. For this reason much of this chapter has been marked for a second reading. The less advanced student may proceed directly to the paragraph on "Types of Thermionic Cathode".

Thermionic Work Function

It was shown in Chapter 1 that an electron must possess a certain energy before it can escape from a metal and that this energy may be expected to vary from one metal to another being, on the average, less for the metals of Group 1 and greater for the metals of Group 8 (see Table 1.3). It should be noted that the absolute value of the escape energy is indeterminate; in fact it has not so far been explained how the values given in Table 1.3 may be obtained and therefore what exactly they signify.

Suppose that the average electronic energy in a metal at temperature T is W_T . Suppose also that the total potential barrier at the surface of the metal requires that an electron should have energy W_B to surmount it. Then the electron emission from the metal will depend very directly upon $W_T - W_B$ and only indirectly, if at all, upon the absolute values of W_T and W_B . The *net* work function given in Table 1.3 is so called because it depends upon $W_T - W_B$. Two other work functions, the inner and outer work functions, may be derived from the values of W_T and W_B . In fact, to a first approximation W_T and W_B are the inner and outer work functions. But the important quantity from the point of view of the electronic engineer is the net function which is generally represented by ϕ electron-volts.

It is shown in Appendix 2.1 that the thermionic current I emitted by unit area of a metal at temperature T is given by

$$I = AT^2 e^{-e\phi/kT} \quad (2.1)$$

where I is in A/sq metre, T is in degrees Kelvin, e is the electronic charge and k is Boltzmann's constant. A is theoretically a universal constant, equal to 1.204×10^6 A/sq. metre/degree². This may be termed the emission equation. It was first enunciated in this form by S. Dushman but a similar equation, based on the Maxwell distribution of velocities, was given much earlier by O. W. Richardson.

The thermionic current I varies very rapidly indeed with the temperature. Writing equation (2.1) in a more useful form

$$I = AT^2 e^{-1.16 \times 10^4 \phi / T}$$

where ϕ is in electron-volts. For tungsten $\phi = 4.5$ and therefore

$$I = 1.2 \times 10^6 T^2 e^{-5.22 \times 10^4 / T}$$

Consider temperatures in the neighbourhood of 3000°K.

$$\begin{aligned} \text{Then} \quad I &= 1.2 \times 10^6 \times 9 \times 10^6 e^{-17.4} \text{ A/sq. metre} \\ &= 2.997 \times 10^5 \text{ A/sq. metre} \\ &= 30 \text{ A/sq. cm} \end{aligned}$$

The graph of Fig. 2.1 shows the mode of variation of I with T in the case of tungsten from 2,500°K to 3,300°K. It will be noted that an increase of 100 degrees in the temperature in this region almost doubles the emission. In principle the thermionic work function ϕ and the constant A may be determined by an experimental examination of the dependence of I upon T . The emission varies so rapidly, however, that neither can be estimated very exactly. Moreover, as will be discussed later, ϕ depends upon the purity of the metal to a marked degree. The values quoted in Table 1.3 are averages of the best modern results for clean surfaces. It is usual to exhibit graphically the variation of I with T by writing equation (2.1) in the form

$$y = \log_{10} I - 2 \log_{10} T = \log_{10} A - 1.16 \times 10^4 \phi / 2.303T$$

and then to graph $\log_{10} I - 2 \log_{10} T$ against $1/T$. This gives a straight line the intercept of which on the y -axis is $\log_{10} A$. The experimental value of the constant A for tungsten is about one-half of that given above from theory. The explanation of the discrepancy is still somewhat conjectural but it is probably due to an increase with temperature of the work function.

Practical Thermionic Emitters

Very few substances can be termed good thermionic emitters and the practical problem of discovering and manufacturing suitable surfaces is as important today as it ever was. At present only three types of emitter are in use (a) pure metals (b) atomic films and (c) alkaline earth oxides.

In the case of emission from pure metals a practical criterion of suitability is the calculated emission at the melting-point, for the two desirable properties of a low work function and a high melting-point must both be taken into account. Thus tungsten, for which $\phi = 4.5$ and which has a melting-point of 3,380°C, has a calculated melting-point emission of 450 A/sq.cm, while tantalum, for which $\phi = 4.1$ and the melting-point of which is 3,030°C, has a corresponding emission of 360 A/sq.cm. Thorium, with a more desirable work function ($\phi = 3.4$) has a melting-point of 1,830°C and an emission at that temperature of 2.9 A/sq.cm. So the lower ϕ of thorium is more than off-set by its lower melting-point. Indeed, no pure metals other than tungsten and tantalum have been extensively used.

Large transmitting valves use tungsten emitters; and as a special case, where the control of the valve demands a variation in the temperature of the emitter, tungsten is again preferred. An example of this of which more will be said later is the so-called 'noise diode'. Tantalum is more easily worked than tungsten. While tungsten can only be readily employed in the form of wire which has to be hammered to the dimension required, tantalum can be formed into sheets. As will be seen later, this makes it useful in certain special cases, but it is easily oxidised and the emission of the oxide is negligible.

Very shortly after thermionic emitters became important it was discovered by Langmuir and Rogers that a tungsten wire containing a small amount of an oxide of thorium (ThO_2) could be made to give a considerably enhanced emission at temperatures between 1,230°C and 1,700°C. In fact, the emission at 1,700°C was about equal to that of pure tungsten at 2,500°C. This is an example of what has been called above an atomic-film emitter, and the general explanation of the phenomenon is as follows:—

A filament of tungsten which has been impregnated with the oxide, thorium, is called a thoriated-tungsten emitter. When this is heated in a vacuum to a temperature of about $2,500^{\circ}\text{C}$, the thorium is reduced to the metal. But the metallic thorium, the melting-point of which is $1,830^{\circ}\text{C}$ tends to evaporate if it is on the surface of the tungsten. Indeed, at $2,500^{\circ}\text{C}$, the emission of the thoriated tungsten is in no way different from the emission of tungsten. If now, the thorium having been reduced, the temperature is lowered to about the melting-point of thorium, the metal gradually diffuses over the surface of the tungsten in a very thin film. It is possible at this temperature to have about 80% of the surface of the tungsten covered by such a film. Thereafter, if the temperature is still further lowered, to prevent the too rapid diffusion of thorium to the surface, the atomic film is dynamically stable and gives the enhanced emission referred to above. However, when the emission is studied, it is found that the work function of the thoriated tungsten is about 3.0 electron-volts, the exact value depending upon the fraction of the surface covered by the atomic film. This work function is not only less than that of tungsten, it is less than that of thorium, showing that the thoriated emitter is not merely a device for holding the thorium to the surface, but one which offers an intrinsically low barrier to high speed electrons seeking to escape.

In using an atomic-film emitter a balance has to be struck between operating at too high a temperature—when the thorium diffuses too rapidly to the surface and gives rise to an increase in the work function—and operating at too low a temperature—when the emission is decreased simply for want of thermal energy. Thoriated-tungsten emitters can be run at temperatures lower than their rated values without loss of life. One merely sacrifices emission. They cannot be run at temperatures higher than their rated value without at least temporarily destroying the monatomic layer of thorium and the associated small work function.

Appreciable evaporation of thorium occurs at the operating temperature, but the rate is very considerably reduced if the wire is heated in a hydrocarbon vapour before mounting in the valve. This forms a surface layer of tungsten carbide, the process being known as 'carbonizing'. All modern thoriated-tungsten emitters are carbonized, although the process results in a more brittle wire of much higher conductivity.

By far the best practical emitter known at present and, curiously enough, one of the first to be used, is the oxide emitter. The physicist Wehnelt discovered in 1903 that the alkaline earth oxides emit electrons freely at temperatures in the neighbourhood of 700° – 800°C , and this was the genesis of an investigation into oxide emitters which has continued with increased momentum until the present day. The oxide emitter is employed in almost all small thermionic valves, and it is now replacing thoriated tungsten and pure tungsten in many of the larger valves.

Modern oxide emitters almost invariably take the form of a coating of a mixture of the oxides of barium and strontium on the surface of one of the more stable metals, nickel being the most popular although not the most efficient electrically. The combined emitter—core metal and oxide coating—may be raised to the operating temperature directly by passing a current through the core metal, or may be heated indirectly by passing a current through a separate tungsten filament placed inside the emitter but separate

from it. As in the case of thoriated tungsten, oxide-coated emitters require 'activation'. Indeed the process of making and activating an oxide-coated emitter is quite complicated. Without entering into too much detail it may be explained as follows:—

First the core metal, usually nickel, is formed to the required shape and is thoroughly cleaned, preferably by heating in an atmosphere of hydrogen at a temperature above 900°C . The coating is then prepared. This is usually a well-mixed suspension of the carbonates of barium and strontium in an organic liquid. The smaller the particle size in the suspension, the better the emission. This suspension is then applied to the required surface of the metal, spraying being a popular method. The material should then form a fine uniform coating from which the binder may be driven off by heat treatment. The emitter is then ready to be mounted in the valve, but until this can be done, care is taken to avoid contamination of the coating. When the valve is being 'pumped out' and 'processed', the last step before 'sealing off' is the activation of the emitter. This is first of all heated to about $1,200^{\circ}\text{C}$ until the carbonates all decompose, leaving the oxides behind. Carbon dioxide is, of course, driven off. Thereafter the temperature is lowered to about 850°C and a small current is taken from the emitter. During this phase the current generally builds up from something quite negligible to a stable value characteristic of the completed valve. Finally, a check is made to see that this current does not fall more than 10% when the emitter is run at the rated temperature, which is about 700°C .

An emitter made and activated in this way has a work function of the order of 1.0 electron-volt and a constant equivalent to A in the equation (2.1) of about 200 A/sq. metre/degree², giving an emission of the order of 200 mA/sq. cm at 700°C . This value is, however, only a rough guide. Very much better emitters have been made in the laboratory and there is little reason to doubt that emissions of the order of 2 A/sq. cm at the same temperature will be achieved in the future.

No attempt will be made to consider the theory of oxide-coated emitters. These form a rather complex example of a class of materials known as semi-conductors. The mechanism of emission is not fully understood yet, although recent work is clearing away many of the difficulties. Suffice it to say that oxide-coatings are very easily disturbed by contamination. A very small amount of any active gas, particularly oxygen or chlorine, reduces the emission to negligible proportions. Any attempt to run the emitter at a temperature lower than the specified value causes the oxide surface to become extremely susceptible to 'poisoning'. In this respect oxide emitters are the opposite of thoriated-tungsten emitters. The latter must not be run at too high a temperature. The former fail at too low a temperature. Consequently, where total emission is the first consideration, it is permissible to over-run an oxide cathode, but it must be remembered that this will shorten the life of the valve by increasing the rate of evaporation of the free barium and strontium which must exist on the surface for emission to occur at all.

Types of Thermionic Cathode

It is now necessary to describe in more detail how emitters are used in thermionic devices, and to do this some understanding of the various classes

into which such devices fall is required. A very simple valve classification is according to the number of electrodes employed. A diode is a valve with two electrodes, a cathode and an anode. Figure 2.2 is a very simple diagram of a theoretical diode, representing two plates A and K placed opposite and parallel to one another. The electrode K is the thermionic emitter, called the cathode and electrode A is called the anode. It must be supposed that these two electrodes have been sealed into a glass envelope which has been subsequently evacuated, so that the gas pressure between the plates is very small indeed.

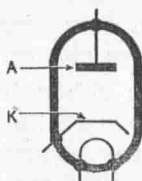


Fig. 2.2. Diagrammatic representation of a diode.

The diode gave the name 'valve' to such thermionic devices, for, as a moment's thought will show, electrons can flow from the cathode to the anode when the anode is positive, but no current can pass when the anode is negative. This unidirectional flow of current is analagous to the flow of a liquid through a valve.

The triode contains three electrodes. As before there is an electronic emitter, the cathode, and a collecting electrode, the anode, but in this case between these two, as shown diagrammatically in Fig. 2.3, there is a wire mesh or grid which to some extent controls the flow of electric current.

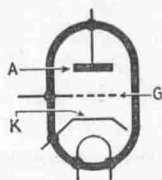


Fig. 2.3. Diagrammatic representation of a triode.

In practice diodes and triodes are hardly ever made up of plane parallel electrodes. A much more usual arrangement is a series of concentric cylinders, the anode being external and the cathode at the axis. Tetrodes, pentodes, hexodes, heptodes and octodes are valves containing respectively four, five, six, seven and eight electrodes. All possess a cathode, control grid and anode. The purpose of the other electrodes is studied in Chapters 4, 5 and 6.

A thermionic cathode made of pure tungsten or thoriated tungsten must take the form of a wire or tape which can be heated to the requisite temperature by passing through it an electric current. In this respect the filament of an electric lamp and a tungsten cathode are similar. The pure tungsten cathode operates at a temperature about $3,000^{\circ}\text{K}$, while a thoriated tungsten cathode operates at about $2,000^{\circ}\text{K}$. Consequently the loss of heat by radiation is about 5 times greater for pure than for thoriated tungsten (radiation loss is proportional

to T^4), and therefore the power required to heat the cathode is very much greater also. This is one reason why pure tungsten cathodes are not more used. They are uneconomical in heating power and, as a consequence, tend to heat up the remainder of the valve to an unwanted degree. Another difference between the two is that the resistivity of a pure tungsten filament is about sixteen times that of a 'carbonized' thoriated filament. Hence, for the same power input, the voltage across the thoriated filament is only a quarter of that across the pure tungsten filament, and this is important for certain uses of the valve. Both types of cathode suffer from the disability that the emitting surface is a comparatively narrow, long strip. This is very inconvenient at high frequencies or where a uniform electric field is required.

An oxide-coated cathode may be either directly heated in which case the core metal must again take the form of a wire or a tape, or it may be indirectly heated in which case the coating may be attached to a surface of almost any shape. The directly-heated oxide-coated cathode differs only in its operating temperature from a corresponding thoriated-tungsten cathode. The power required to run it is smaller than that required for the thoriated cathode, since the operating temperature is about $1,000^\circ\text{K}$. The voltage across the cathode need not be as large as in the case of the other two types. The most interesting cathode is the indirectly-heated cathode. In this case the temperature of the emitting surface is raised to the required value by means of a separate heating wire, usually enclosed in the emitting surface. A typical indirectly-heated cathode is shown pictorially in Fig. 2.4. The heater wire is made up in the form of a bifilar spiral which fits into the hollow cathode and is insulated from it by a coating of alumina.

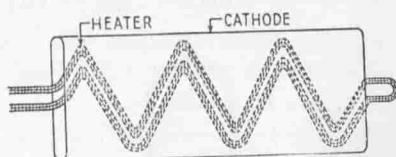


Fig. 2.4. Showing the construction of an indirectly-heated cathode with a non-magnetic heater.

This form of cathode possesses many advantages. First, the cathode can be given the best shape for its purpose. Secondly, the cathode can be maintained at a given potential relative to earth; it no longer has a potential gradient along its length due to the heating current. Thirdly, since the insulation provided by the alumina is very good, the heater can be maintained at a different mean potential from the cathode and, by winding the heater non-inductively, magnetic and electromagnetic effects due to using a heating current at power frequencies can be avoided.

In cathode-ray tubes and valves for ultra high frequencies, disc-shaped cathodes must be employed. This can readily be done, using indirect cathode heating. A diagram of a cathode-ray tube cathode is shown in Fig. 2.5, the heater being indicated by the spiral inside the 'pill-box' cathode. The latter is coated only on one circular surface, but a back plate is attached to prevent radiation and so raise the efficiency of the heating element.

Heating the Cathode

One of the best ways of specifying the efficiency of a given cathode is in terms of its emission per unit area for a given heating power input. The absolute value of this efficiency depends very largely upon what precautions have been taken in the valve to prevent loss of heat from the cathode along the leads, and so the absolute value is not particularly important. But when pure tungsten, thoriated-tungsten and oxide-coated cathodes are compared under equivalent conditions of heat loss, the figures give a valuable indication of the varying efficiencies. Typical figures are:—

Pure tungsten wire	..	..	..	..	..	5 mA/sq.cm/watt
Thoriated tungsten wire	..	..	..	..	..	50 mA/sq.cm/watt
Directly-heated oxide-coated wire	..	..	..	..	..	500 mA/sq.cm/watt
Indirectly-heated oxide-coated plate	..	..	..	..	..	100 mA/sq.cm/watt

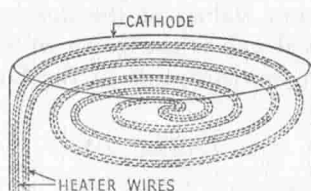


Fig. 2.5. Showing the construction of a circular cathode, as used in a cathode-ray tube. The heater is wound so as to be non-magnetic.

The lower efficiency of the indirectly-heated cathode relative to the corresponding directly-heated cathode is not hard to understand. Losses due to radiation are much greater for indirectly-heated cathodes.

It has always been a dream of the engineer to have standard fittings so that replacement is made simpler. There is, therefore, a great need to standardize the voltage required to heat a cathode to the required temperature. For many reasons, some of which will appear hereafter, complete standardization is not feasible, but so far as most Service valves are concerned 6·3 volts has been taken as the most desirable figure. Clearly this value has been suggested by the use of secondary batteries—accumulators—for cathode heating. It is, however, quite a convenient value when the heating power is taken from an a.c. supply via a transformer. Probably about 90% of the valves now made operate on a rated heater voltage of 6·3, and it is fortunate that the valve designer has been able to construct a cathode which is not particularly sensitive to heater voltage. It is rare for a small valve to be seriously affected as to emission by a change of voltage of $\pm 5\%$. The means employed to obtain this comparatively wide tolerance will appear when the variation of the current in a valve with the heater temperature is considered in the next chapter.

▶ One method of heating a cathode which is occasionally used in special circumstances is to bombard it with electrons. For example, if a plate of tantalum (melting point $2,800^{\circ}\text{C}$) is placed near, but insulated from, a tungsten filament, the electrons emitted by the latter may be accelerated by a potential difference between the two so as to give up their kinetic energy as heat to the tantalum plate. The latter, being thus raised to a high enough temperature, emits electrons copiously. This double-emission phenomenon makes it possible to

employ a sheet of pure metal as the cathode and to get the advantages of indirect heating usually associated with oxide coatings. The method is expensive in heating power and requires a complicated electrode structure, but it has been used in some experimental valves designed to generate ultra-high frequency currents.

An interesting example of indirect heating of a cathode is to be found in certain magnetrons. These valves are described in some detail in Chapter 11. Here the action of the valve is such that when it is operating efficiently a certain fraction of the emitted electrons is returned to the cathode with considerable energy. These returning electrons tend to over-heat the cathode to such an extent that it has been found possible in some cases to cut off the primary cathode-heating current altogether after the valve has begun to operate. The magnetron in this case may be regarded as supplying its own cathode heating as a by-product of the action.

In order to fix ideas some values of the average power required to heat the cathodes of some typical valves are given in Table 2.1. All of these are oxide-coated emitters with the exception of the large transmitting valves.

TABLE 2.1

Valve	Heater Power (watts)	Valve	Heater Power (watts)
Receiver	1.5-2	Large transmitter	100
Sub-miniature	0.06	Reflex klystron	3
Cathode-ray tube	3-4	Magnetron	20
Small transmitter	10		

General Construction of Valves

The majority of valves can be classified without any understanding of their function or operation according to the number of electrodes they possess. This has already been mentioned. It is now necessary to enquire more closely into their construction. By far the most popular design is one in which the electrodes are arranged as coaxial closed surfaces round the cathode, the simplest example being a series of coaxial cylinders. A section across the electrodes of a triode usually takes the form shown in Fig. 2.6(a), where K is the cathode, G the grid and A the anode.

This form has many advantages. First, the cathode emits over most of its surface, so that the maximum emission is obtained for a given heater power. Secondly, since the cathode and grid are completely surrounded by the anode, all electronic action is confined to the tubular spaces between the electrodes. This avoids any interference due to electrical charges on the glass walls of the envelope. Turning to Fig. 2.6(b), which shows the same structure pictorially, there are leads coming from the electrodes to the outside of the envelope to connect the valve to the external circuit. It is good practice, where the frequency of operation is not too high, and where the very best glass insulation is not required, to bring all the leads from one end so that they may be welded to the wires of a simple 'pinch'. This is an assembly of wires of a suitable composition sealed into a glass cup in such a way that the latter may itself be

sealed to the remainder of the glass envelope directly. Fig. 2.7(a) is a picture of a valve pinch with eight wires sealed through it. An alternative arrangement, which has found more favour recently for high-frequency work is the 'button' valve base illustrated by Fig. 2.7(b).

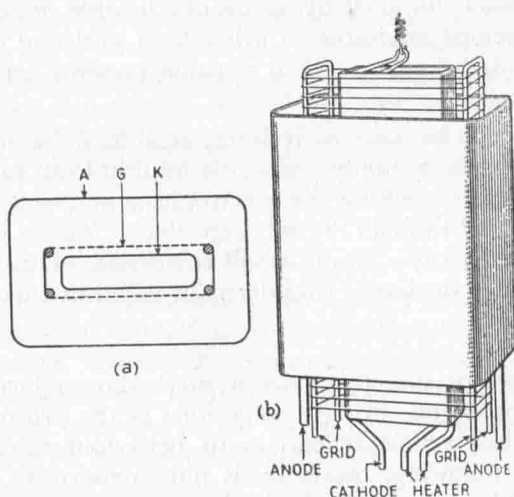


Fig. 2.6. Sectional view (a) of a typical triode and a general picture (b) of its electrode structure.

The type of construction which brings all the leads to a valve base is almost universal. It is cheap and convenient, lends itself to mass production, and enables a valve to be 'plugged into' a circuit by means of a suitable valveholder.

More complex valves may contain several gridded electrodes between the cathode and the anode or may, in the case of multi-unit valves, contain two cathodes and two complete sets of other electrodes. In all cases the attempt has been made to bring all the leads out through the base and, as a consequence, the valveholder has acquired considerable importance as a circuit element.

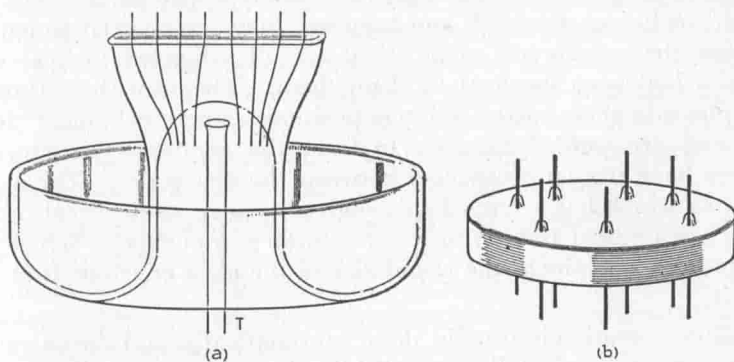


Fig. 2.7. A valve pinch (a) showing how the wires are brought through the glass. The tube T is used to evacuate the valve when the bulb has been sealed to the pinch. Another common type of valve base (b) is made of pressed or powdered glass.

For certain purposes, even at comparatively low frequencies of operation, it is necessary to separate one electrode (usually the anode) as completely as possible from the remainder of the valve structure. It is then common practice to bring the particular electrode lead out through the top of the envelope (top cap anode or grid) by means of a flexible connection. Occasionally it has been deemed advisable to bring both grid and anode connections out near the top of the valve. Such a valve presents certain difficulties in manufacture.

Although more will be said of it later, it should be noted at this stage that where the electrodes are only connected by thin leads to the outside of the valve, all the heat produced by the electronic action and the action of the heater must be radiated through the valve envelope. This is a serious limitation on the power handling capacities of small structures. How it is overcome in the case of high power devices is considered in detail in Chapter 7.

Valve Manufacture

Although it does not directly concern the radio engineer or operator, it is of operational importance to have some idea of the procedure in making a valve, for the finished article reflects in its behaviour the care taken in its construction. The following, therefore, is not intended to be an exhaustive description of what is a most complex technical process. It is intended as a general picture of the sequence of the principal events.

Almost all simple valve parts are made of nickel. The reasons are many. Nickel is easily worked and is relatively stable. It does not readily form the oxide and only slightly more readily forms the sulphide. Its melting point is $1,450^{\circ}\text{C}$, which is high enough for most purposes. It is a reasonably good conductor and is only slightly magnetic. In the following description it may be assumed that, unless otherwise stated, the metal parts are nickel. In some cases nickel cannot be used or must be used in a modified form. This is regrettable from the manufacturer's stand-point.

When a valve is to be made in large numbers it is, like any other complex mechanism, broken down into many small parts which can be made separately. In the case of a simple valve there may be as many as 100 parts. For example, the cathode, its heater, the leads and supports, may comprise six to ten different parts. After these parts are made, they are all subjected to heat treatment (usually in a hydrogen furnace) to clean them. They are then assembled on the valve pinch in their correct relative positions using mechanical aids known as 'jigs' and are welded together to form the composite structure. Most small valves have insulating spacers between the electrodes. These are often made of mica which has a remarkably high resistance to electrical breakdown, and which can be heat treated to remove surface impurities. When the parts are assembled on the pinch, the remainder of the glass envelope is ready to be sealed to it.

The sealing operation is usually done automatically and the valve proceeds directly to a pumping installation. While the pumps are still working, the valve is 'baked' for a considerable period in an oven at a temperature just slightly below the softening temperature of the glass envelope. This removes much of the absorbed water-vapour from the glass surface. Thereafter, the valve still being pumped, the electrodes are heated by induction to remove

the last traces of gas from the surfaces. This is particularly important in the case of the valve anode. Then, while pumping continues, the cathode is activated according to one of the methods described earlier in the chapter, and the electronic emission is tested. Finally, the 'getter' is 'fired'. A getter is a small pellet of an alkali metal contained in a gauze of a metal of higher melting-point. To 'fire' the getter, the pellet is heated by induction until the alkali evaporates and deposits itself as a thin film on the wall of the tube. This thin film, being extremely active chemically, is capable of combining with any residual gas which appears during the operation. This preserves, and in fact enhances, the vacuum after the valve is made. Immediately after firing the getter the tube is sealed off by allowing the narrow tube connecting it to the pumping system to collapse under local heating. The completed valve is then ready to be based and tested.

It will be seen that the manufacture of a valve entails a sequence of carefully controlled operations, each designed to improve the life and performance of the electronic device. There is no reason to suppose that the rapid improvements in valve manufacturing techniques which have taken place in the last twenty years will not continue in the future, until the electronic tube is as stable and reliable as its mechanical counterpart.

Reliability

As the number of valves used by the Services increases, it becomes more important that they should be reliable. If the chance of a particular valve failing in a given period is one in a hundred, it would generally be felt that the device was fairly reliable. If, however, a piece of equipment contains one hundred valves (and this is not at all uncommon) it is clear that the chance of the equipment failing in the same time due to valve faults is very high indeed.

Since the end of World War II valve manufacturers have been giving much attention to improving the reliability of their products, with the result that the Services can now expect very much better valves, judged in terms of average life and resistance to shock and vibration. This is particularly the case with miniature and sub-miniature valves.

In general terms reliability depends upon good electrical and mechanical design followed by sound methods of manufacture. To quote simple examples, a valve using a long filamentary cathode, supported only at its ends, tends to be unreliable because the filament resonates at a particular frequency of vibration. Short stout leads and accurately cut mica insulators are obviously desirable; welds should be made in such a manner that faults are unlikely; and the cathode coating must be made to adhere firmly to its metallic base.

Attention to details of this kind demands greater expense in valve production. Moreover, it has been found that many faults which appear in small valves can be traced to the accidental inclusion of tiny particles of foreign matter in the valve envelope. This suggests that care must be taken in manufacture to keep all valve assemblies spotlessly clean.

Reliable Receiver Valves

Much of the fundamental work on improving valve reliability was done on miniature types (using the B7G base). This development has established test specifications which ensure a high quality product. Commercially, such valves

are in considerable demand by the various air lines and their applicability to Service needs is obvious. Many of the well-known miniature valves are now available in "reliable" form with unchanged electrical characteristics. Service specifications exist defining "reliable" valves. These specifications are more stringent than the corresponding commercial specifications and are subject to gradual modification from year to year as reliability improves. The progress of development, therefore, is a continuous one and it is to be expected that there will be no end to it.

Similar work is being done to improve sub-miniature valves. In general terms the electrical characteristics of a typical sub-miniature valve cannot be as good as those of a corresponding miniature, but the smaller valve stands a better chance of being made resistant to shock. For this reason reliable sub-miniature diodes, triodes and pentodes tend to be used where the equipment may be subjected to mechanical ill-usage.

Flying-Lead Valves

A certain proportion of valve failures have always been due to bad connections to the external circuit. Some of these faults have been directly attributable to the valve pins themselves. For example, oxide coating on valve pins can introduce an undesirable resistance between a valve electrode and the circuit. Some of the faults have been due to the valve sockets. For example, early B7G holders, made of ceramic, were very apt to distort or break the valve pins; much better holders, using P.T.F.E. (polytetrafluorethylene) as the insulator, were subsequently made.

There has recently been a tendency, particularly where the valve is otherwise reliable, to solder the valve directly to its associated passive components. To do this without fear of damage to the valve and with reasonable convenience, valves have been made with flexible copper (or copper-plated) wires as leads and with no pins. These are known as flying-lead valves. Where reliability is a serious need it is argued that the elimination of the doubtful plug-in contact is essential and all sub-miniature valves and many miniatures are now available with flying leads.

Reliability in Special Valves

It has been argued in the past that there is little advantage to be gained in making conventional receiver-type valves very reliable, if magnetrons and other unconventional types are to remain unreliable. Although there is something in this argument, it must be recognised that *every* improvement in reliability is valuable even if it only focuses attention on the unreliable components.

Attempts to improve the reliability of special valves, such as micro-wave oscillators, meet much greater difficulties than are found in the case of receiver valves. For example, the alignment of the electron beam in a klystron or cathode-ray tube is difficult to maintain during vibration. The longer electron paths characteristic of many modern valves create a special problem. Moreover, where special glass-to-metal seals are employed these require extensive testing to ensure that they are robust enough to withstand vibration and shock, particularly where temperature variations are considerable. Nevertheless constant attention is now being paid to reliability in magnetrons, klystrons and cathode-ray tubes. Test specifications appropriate to the valves can be written.

An even higher degree of reliability, and a reliability of a slightly different kind, is required for valves in guided weapons. In this case it is economic to have special production lines arranged to give the most reliable product possible, irrespective of expense. New designs of valves are continually becoming available for this work and these valves are now being made by quite unconventional methods.

Shot Noise

Before leaving the general subject of thermionic emission, and before considering specific devices in detail, the reader must be introduced to one of the limiting factors in the use of all electronic devices, namely shot noise.

The emission of the electron stream from the cathode of a valve is a random effect. If the electrons came off at a constant rate, the emission current would be quite constant in the absence of any intentional variation. But since the thermionic effect itself depends upon the random variations in velocity of the free electrons in the metal, the emission current varies in a random manner about its mean value, and the greater the emission, the greater the associated variations. Consequently, if a valve is being used to amplify or to modify in any way a small current (or voltage), it must always be assumed that the random effect described above will be superimposed on the current (or voltage). When the latter (which we may call the signal) is of the same order as the random effect (called noise) the signal is submerged in the noise and the valve cannot be made to detect it. This constitutes a natural limitation on the effectiveness of the valve for the detection of small signals. Here the term 'detect' is used in its normal and not in its technical (radio) sense.

This noise effect is not confined to thermionic devices. If the attempt is made to detect a small signal (voltage or current) by any means whatever, the sensitivity of the detecting apparatus being continually increased, a point will be reached where the thermal agitation of the free electrons in the lead to the detector and the electronic agitation in the detector itself will together mask the signal. Thereafter, any increase in the sensitivity of the detector will only produce an increase in the noise detected. Thermal noise is the ultimate limiting factor in all detection and measurement.

In the subsequent discussions of the various types of valve, the noise associated with each will be considered in detail. At the moment it should be noted only that the small-signal limitation on any device is the existence of random electronic currents and their associated voltages.

It may be shown by the methods of quantum mechanics that a resistance at a temperature T degrees Kelvin is capable of delivering to a detector a noise power P which varies directly, as one might expect, with T and which also varies directly with the range of frequencies over which the detector is sensitive. Let the range be denoted by B .

The quantitative expression is

$$P = kTB \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

where k is, as usual, Boltzmann's constant.

In the case of a valve, if the emission current is I , the noise power (known in this case as *shot noise*) available in a resistance R through which the current flows is

$$P = 2eIBR \quad (2.3)$$

where e is, as usual, the electronic charge.

Equation (2.3) is only true if the valve is so arranged that the current I is the maximum current of which the emitter is capable. If for any reason the current taken from the cathode is less than the total emission I then the available shot noise is less, and may be as small as one-tenth of that given by equation (2.3). It can never, however, be zero.

Clearly, shot noise is only important if its magnitude is larger than that produced by the resistance through which the current flows. A comparison may easily be made.

$$\text{The quantity } x = \frac{2eIBR}{kTB}$$

gives the ratio of the shot noise to the thermal noise in the resistor R , when the valve current I passes through it.

$$x = \frac{2 \times 1.6 \times 10^{-19} IR}{1.38 \times 10^{-23} \times 300} \quad \text{where } I \text{ is in amperes, } R \text{ is in}$$

ohms and T is taken equal to 300° K .

$$\text{Therefore } x \approx 80V$$

where V is the voltage across the resistor due to current I . Taking $I = 1 \text{ mA}$, R must be less than 12.5 ohms to make the shot noise less than the thermal noise. As will be seen later I is usually greater than 1 mA and R is generally very much larger than 12.5 ohms . So shot noise is a serious factor.

APPENDIX 2.1

Derivation of the Richardson-Dushman Equation

Equation (1.10), giving the number of electrons striking unit area of the surface at right angles in one second with energies between W and $W + \delta W$ may be written

$$v\delta N_x = \frac{4\pi mkT}{h^3} e^{W_0/kT} e^{-W/kT} \delta W$$

In order to escape, an electron must have an energy greater than W . Therefore the number escaping per second per unit area is

$$\begin{aligned} \int_{E=W_0}^{\infty} v\delta N_x &= \frac{4\pi mkT}{h^3} e^{W_0/kT} \int_{W_0}^{\infty} e^{-W/kT} dW \\ &= \frac{4\pi mk^2 T^2}{h^3} e^{-(W_0 - W_0)/kT} \end{aligned}$$

Therefore the emission current is I , where

$$I = e \int_{W=W_B}^{\infty} v \delta N_x = \frac{4\pi m k^2 e}{h^3} T^2 e^{-(W_B - W_0)/kT}$$

Let $W_B - W_0 = \phi e$, where ϕ is the thermionic work function. Then

$$I = AT^2 e^{-\phi e/kT} \quad (2.1)$$

and

$$A = \frac{4\pi m k^2 e}{h^3} = 1.204 \times 10^6 \text{ A/sq.metre/C degree}^2$$

CHAPTER 3

Diodes and Triodes

THE SIMPLEST electronic device and one having a large number of applications is the diode. As has already been mentioned, it was the diode which suggested the name 'valve', for its most obvious characteristic is that it will allow electric current to pass in one direction but not in the other. Although all the uses to which the diode may be put can be explained in terms of this valve action, these uses are apparently so diverse that it is worth noting some of them at this stage. It may be employed (1) to transform alternating current to direct current, as a rectifier; (2) to produce from a mixture of two voltages of different frequencies a voltage which has a frequency equal to the sum or difference of the original frequencies, as a frequency changer; (3) to separate out the low-frequency (audio) component from a high-frequency alternating voltage the amplitude of which is varying at the low frequency, as a detector; (4) to change the waveform of a varying voltage by cutting off part of it, as a shaper; (5) to change a frequency or phase variation to a corresponding amplitude variation, as with a discriminator; (6) to control the action of other electronic devices by varying their condition of operation. In fact, just as there is no end to the uses to which electronic devices may be put, so there is no end to the uses of the diode.

Temperature Limitation

It was established in Chapter 2 that a cathode emits electrons at a rate depending upon its temperature. It might be imagined, therefore, that the current flowing in a valve always depended directly upon the cathode temperature. This is not so. It is far more usual for the valve current to be almost independent of the physical condition of the emitting surface, and it is necessary now to enquire into how this comes about. Let us consider a simplified version of the diode, consisting of a plane cathode, K, opposite and parallel to a plane anode, A, the distance between the two being D . (Fig. 3.1). As usual, let it be

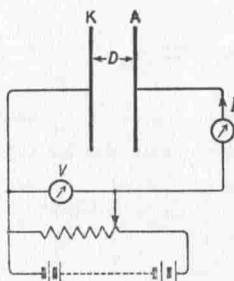


Fig. 3.1. Circuit for measuring the current-voltage characteristic of a diode.

assumed that the two electrodes are inside an evacuated container. Then the following experiment may be performed. A current-measuring instrument, such as a milliammeter, may be placed in one lead, and a potential difference V which can be varied, may be set up between K and A, K being negative.

Then the current I , corresponding to the voltage V may be measured for various values of the latter. The results of such an experiment are shown in the graph of Fig. 3.2. As V is increased from zero, the current I increases, at first slowly, and then more rapidly, and then slowly again, tending finally to a certain value I_T , which is almost independent of V . This current I_T is obviously the maximum which can be supplied by the cathode, and its value is given by equation (2.1). When V is greater than V_T , all the electrons emitted by the cathode are carried across to the anode. The diode is then said to be temperature limited, meaning that I_T depends only upon the temperature of the cathode and not upon V .

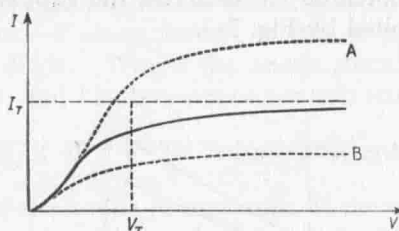


Fig. 3.2. Diode current-voltage characteristics.

Raising the temperature of the cathode (particularly in the case of a bright emitter) will raise the value of I_T . The corresponding curves for higher and lower temperature are shown by the dotted curves A and B in the figure. Changing the value of the distance D between the electrodes changes the value of V at which the current becomes approximately constant (or saturated, as it is called), but it does not change I_T . That depends only on the cathode surface and its temperature. As has been said, valves are seldom used in the temperature-limited condition. Indeed, valves with oxide-coated cathodes cannot operate in this way without deterioration.

The dynamic equilibrium of the oxide cathode, which gives rise to the enhanced emission, depends for its existence on there being a reservoir of electrons between the electrodes. If an oxide cathode is run in the temperature-limited condition, the emission will soon fall off to a small fraction of its original value.

Space-Charge Limitation

Returning to Fig. 3.2, let us consider what is happening in the diode when the value of V is less than V_T . If $V = 0$, there is no electric force between K and A to drive the emitted electrons towards A. Some will, however, find their way to A, for it must be remembered that these electrons have energies distributed about a mean value equivalent to about 0.2 volt. Those which do not find their way to A will distribute themselves throughout the space between K and A, forming a cloud of negative electricity which tends to force back to the cathode all electrons emitted with small energies. The charge cloud itself is a stable state, but its component electrons are continually chang-

ing and moving away, either to the cathode or to the anode. It is usually called the space charge. If V is now made slightly positive, tending to pull the electrons towards A, the current will increase, but the existence of the negative charge in the space will to some extent mask the effect of the electric field, particularly near the cathode. As a consequence, many of the electrons will still be turned back, and the current will still be much less than its maximum possible value. At the same time, as V increases, the electron cloud will move closer to the cathode, since the electrons nearest to the anode will be the first to be dragged away. Finally, when V is large (greater than V_T), the space charge, if it exists at all, will have an appreciable value only immediately in contact with the cathode. The state of affairs when V is less than V_T is that the current in the diode is limited by the space charge: the temperature of the cathode, except in so far as it produces the space charge, is of secondary importance. The situation is more easily understood by considering the manner in which the potential varies across the gap, starting from zero at the cathode. This is exhibited by Fig. 3.3.

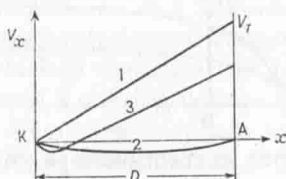


Fig. 3.3. Potential variation across a diode for different values of the anode voltage.

The different curves are drawn for different values of V . First, when $V = V_T$ all the electrons find their way to A and there is no space charge. This is curve 1. Hence the space between the electrodes behaves as if no electrons were present at all, and the electric field is uniform throughout, neglecting, of course, edge effects. The potential relative to the cathode at a point distant x from it is therefore $V_x = xV_T/D$. Another extreme case occurs when the anode and cathode are at the same potential. Owing to the existence of the cloud of electrons in the space between K and A, the potential at x will be lower than that of both anode and cathode. There will, in fact, be a potential minimum somewhere about the middle of the gap. This is shown by curve 2. An intermediate stage, the one of greatest practical importance, is represented by curve 3. Here the flow of electrons from the cathode has produced a cloud near its surface, but the electrons are drawn off from this cloud in appreciable numbers. The cloud of electrons forms a virtual cathode. There is little static charge anywhere in the space between the electrodes except near the cathode, and so the curve is almost a straight line over the greater part of its length. The potential again falls to a value below that of the cathode in the vicinity of the latter, giving rise to a potential minimum. This negative potential is very small, being of the order of 0.1 volt. Clearly, if it were greater, few of the electrons could surmount the barrier which it represents, since their original energies would not be large enough.

of V greater than 0. To be strictly accurate the current is not exactly zero at $V = 0$, since the finite velocity of emission of the electrons produces a small but rapidly diminishing current as V becomes negative. The graph is drawn for $D = 5$ mm, and gives the current in mA/sq.cm.

The resistance of the diode is the ratio of the voltage across it to the current which flows in it. Clearly, the diode does not obey Ohm's law. From the three-halves law $V/I =$ the static resistance $r_s = 1/C\sqrt{V}$ where C is a constant. Therefore, when V tends to zero, r_s tends to infinity, and as V increases, the resistance decreases. It is this non-linear characteristic which gives the diode its value for certain purposes in radio. As will be seen in Chapter 5, when a diode is introduced into a circuit, it distorts the waveform in a manner depending upon which part of the characteristic is used. In many cases the property of the diode of greatest interest is not the equivalent static resistance r_s , but what is termed the dynamic or a.c. resistance of the valve; that is, the ratio of a small change of anode voltage to the resulting change of anode current. This somewhat involved statement is explained by reference to Fig. 3.4. The dynamic resistance at voltage V is the ratio $\delta V/\delta I$, where the symbol δ is used to mean a small change in the quantity to which it is attached. This ratio we may term r_d . The inverse of r_d , the dynamic conductance of the valve, g , is used just as frequently.

- Using the V - I characteristic given by $I = CV^{3/2}$, the values of the resistance r_d and the conductance g are

$$r_d = 1/g = 2/(3C\sqrt{V}) = 2r_s/3 \quad (3.4)$$

For modern diodes with close spacing between anode and cathode the current-voltage characteristic is almost a straight line over a large part of its useful range. It is, therefore, permissible in many cases to assume that g and r_d are constants.

To exemplify the action of the diode as described above, let us consider the circuit shown in Fig. 3.5, where an alternating potential $V_0 \sin \phi$ is applied to the resistance R and the diode in series.

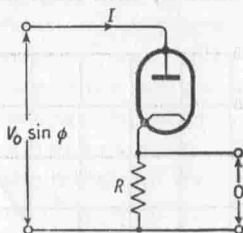


Fig. 3.5. Simple circuit illustrating the action of the diode.

Let $V_0 = 25$ volts, $R = 250$ ohms and the diode characteristic, when conducting, be given by $I = (V/25)^{3/2}$. These figures have been chosen to make the arithmetic simple. Then the potential difference $V_0 \sin \phi$ is made up of the voltage across the resistance RI and the voltage across the diode. This latter must be $V_0 \sin \phi - RI$, and therefore

$$I = \{(V_0 \sin \phi - RI)/25\}^{3/2}$$

or, writing it another way and inserting the values of V_o and R ,

$$I^2 = (\sin \phi - 10I)^3$$

This equation may be solved for various values of ϕ . Remembering that the valve does not conduct when the cathode is positive with respect to the anode, the curves of Fig. 3-6 are obtained. Curve A is $V_o \sin \phi$, curve B is the potential difference across the resistance, and curve C is the potential difference across the valve. Curve B is almost a half sine wave, and so, if the output from the circuit is taken across the terminals marked 0 in Fig. 3-5, the potential difference is in one direction only, and the current which may be taken is a rectified current.

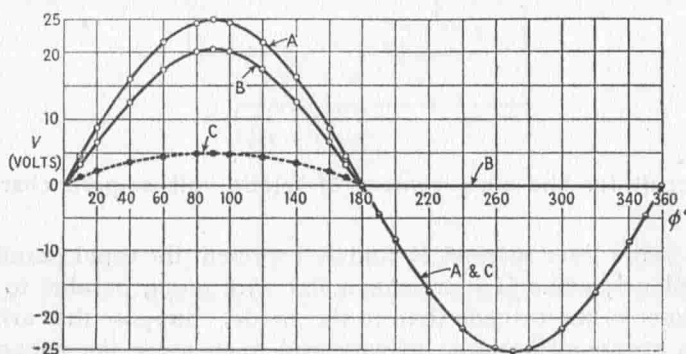


Fig. 3.6. Curves showing how the voltages vary with the phase of the input signal in different parts of the circuit of Fig. 3.5.

The simple circuit of Fig. 3-5 is the basic form of rectifier used in power packs to obtain direct current from alternating current. By storing the charge in a capacitor and inserting chokes in the output leads, the output voltage may be made as constant as we please, although its value will, of course, depend upon the load resistance to which it is applied. If the circuit is used at high frequency, however, the diode capacitance must be taken into account. In this case the simple theory given no longer holds, and the waveform obtained from the resistance R may be very different (see Volume 4).

Points of importance to be noted from Fig. 3-6 are that the peak current in the valve is 81.3 mA, although the average current is only 5.1 mA, and that the peak inverse voltage (that is, the peak voltage when the diode is not conducting) across the diode is 25 volts, although the peak direct voltage is only 4.7 volts. Diode rectifiers, therefore, must be capable of giving a very much greater peak emission than one might expect at first sight, and they have to withstand at least the maximum amplitude of the supply voltage in the non-conducting state. Where the charge is stored in a capacitor they have to withstand twice this voltage.

The Triode Valve

Radio would not have advanced very far, nor would the modern science of electronics have come into being, had we been limited to the use of the diode. The introduction of the third electrode, commonly attributed to Lee de Forest, and transforming the simple diode into a device which can control power

over an enormous range of frequencies, was an advance of the greatest importance. The operation of the valve is so greatly modified by the action of the 'grid', as the third electrode is called, that the mechanism requires a very careful study.

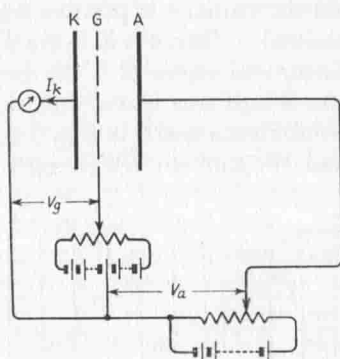


Fig. 3.7. Circuit for the measurement of triode volt-ampere characteristics.

In Fig. 3.7 the lines marked K and A represent the usual parallel cathode and anode planes, while G represents a flat wire mesh, parallel to A and K, but much closer to the cathode than to the anode. Suppose that arrangements are made by means of batteries to vary and to measure the voltages applied between K and A (V_a) and between K and G (V_g) and that the current through the valve from the cathode can be observed (I_k). Curves may then be obtained showing how this simple triode behaves. Generally the curves are plotted maintaining one of the variables V_a , V_g , or I_k at a constant value. For example Fig. 3.8 exhibits the variation of I_k with V_g , V_a being kept constant. Curves are drawn for various values of V_a .

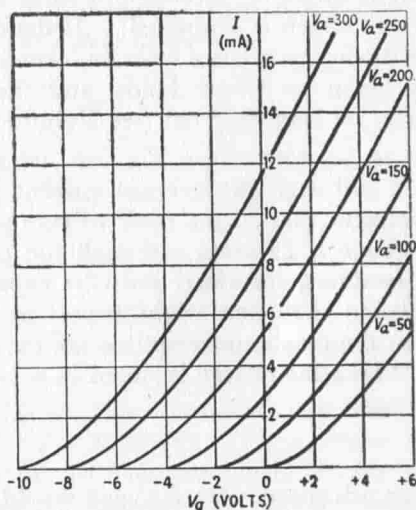


Fig. 3.8. Curves showing the variation of cathode current with grid voltage for various values of the anode voltage.

A casual inspection of these curves brings several interesting features to light. First, the anode potential influences the current through the valve, but secondly, the grid potential has obviously much more influence. Thirdly, current can flow through the valve even when the grid is negative, but fourthly, to every anode potential there corresponds a grid potential for which the current is just cut off. Indeed, if the current falls to zero at a grid potential of V_{go} , when the anode potential is V_{ao} , it is found that to a very good degree of approximation the ratio V_{ao}/V_{go} is a constant for the valve. These are the more obvious qualitative results which require explanation and understanding.

Since planar triodes are not much used, it is perhaps as well to think in terms of a much more common form, namely, the cylindrical construction. A cross-section of such a cylindrical triode is shown diagrammatically in Fig. 3.9.

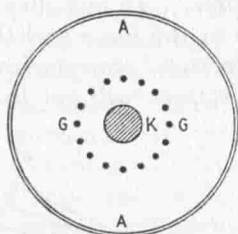


Fig. 3.9. A cross-section of a simple cylindrical triode valve.

The cathode is the inner cylinder, K; the grid is the series of wires stretched parallel to the axis round the cathode; and the anode, or, as it is sometimes called, the plate, is the outer coaxial cylinder. Using this as a model, let us try to explain the results of Fig. 3.8. As usual, defining all voltages with reference to the cathode, let V_g be made negative and V_a positive. Then no matter how positive V_a may be, there must be a negative potential V_g which will just prevent the flow of electrons from the cathode. As has been mentioned above, the ratio of V_a to this cut-off grid voltage is very nearly a valve constant. This must be so, since it is a non-dimensional quantity, depending only on the structural geometry of the valve. If V_a is doubled, V_g must also be doubled. The presence of space charge in the vicinity of the cathode modifies this simple picture slightly, but proportionality will still be approximately maintained. In practice it can usually be assumed.

Fig. 3.10 is a magnified version of a part of Fig. 3.9, showing the lines of electric force in the valve when the negative value of V_g is greater than that

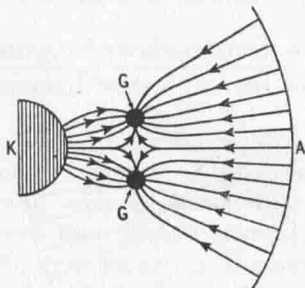


Fig. 3.10. The lines of electric force in a cylindrical triode when the negative potential on the grid is greater than that required to cut off the current.

required to cut off the current. The arrows showing the direction of the force are, of course, away from the anode since they conventionally show the direction in which positive charge would move. Electrons, being negative move against the arrows. It will be noted that no lines of force join the cathode to the anode. An electron leaving the cathode with zero velocity is an impossibility, since all the lines of force at the cathode surface tend to turn electrons back. The diagram requires slight modification to take account of space charge.

Consider now the case where V_g is still negative, but equal to one-half of the cut-off voltage. Then Fig. 3.11 shows the lines of force to be expected in this case. Some of the electrons leaving the cathode with zero velocity may now be accelerated away from it, and so some will reach the anode. It must be noted that our figure does not prove this, for the electron paths do not usually coincide with lines of force. In fact they only do so when the electron starts off in the same direction as the force and the line is straight. However, a careful examination of the electron 'trajectories' does prove that a considerable number of the emitted electrons will get through to the anode.

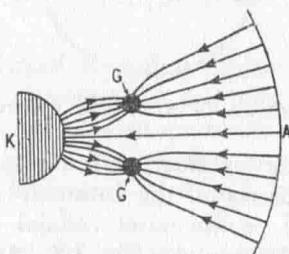


Fig. 3.11. The lines of electric force in a cylindrical triode when the grid is negative, but current can flow to the anode.

It is hardly necessary to draw another picture for $V_g = 0$. If we do not take account of space charge, there are no lines of force between the cathode and the grid wires, but the net effect is simply to increase the number of electrons arriving at the anode. If allowance is made for space charge, there are lines of force between the grid wires and the charge cloud in front of the cathode, for it will be remembered that there is a potential minimum between two electrodes at the same potential, when one of them is emitting electrons. This only modifies the current slightly.

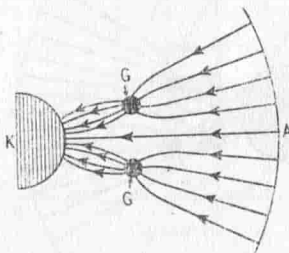


Fig. 3.12. The lines of electric force in a cylindrical triode when the grid is positive with respect to the cathode.

Finally, it is as well to consider the case where the grid is positive and at the potential which would exist at that part of the valve if the latter were a diode. This is sometimes called the 'natural potential' of the grid. This is illustrated by Fig. 3.12.

Under these circumstances the grid will absorb a fraction of the electron current depending upon the effective area which the wires present to the electron stream. If the wires are very fine, this current will be very small. It is as if the grid were transparent. The lines of force in this case are approximately the same as they would be in the absence of the grid.

Returning now to Fig. 3.8, the greater effect of the grid voltage in controlling the valve current is due to the fact that the grid is much closer to the cathode than is the anode. The anode current which flows when the grid is negative is due to the penetration of the lines of force from the anode to the cathode between the grid wires. It is to be expected that all the current will flow to the anode so long as the grid is negative, but when the grid is positive the current will divide itself between grid and anode. Even then, until the grid is at some potential above its 'natural' potential, the grid current will be small. Its value compared with the anode current will depend upon the diameter of the wires constituting the grid and upon their spacing from each other. Lastly, although the model used has been a cylindrical one, the results will hold qualitatively for any type of construction.

Amplification Factor of a Triode

In the diode valve the relation between the current and the anode potential is of the form $I = CV^{3/2}$. It may be shown both theoretically and practically that the current in the triode is given by a somewhat similar relationship

$$I = C(V_g + V_a/\mu)^{3/2} \quad (3.5)$$

where μ is a constant. We have already met this constant in the above explanation of the action of the triode, for it is obviously the ratio of the anode voltage to the cut-off grid voltage. This is one definition (but not the most usual one) of the amplification factor of a triode valve. A more usual definition of the factor is given by the equation $\mu = -\delta V_a/\delta V_g$ where δV_a is the change in anode voltage required to produce a small change in anode current, and δV_g is the change in grid voltage required to bring the anode current back to its original value. A third alternative is to define the factor by the equation

$$\mu = + \frac{\text{rate of change of anode current with grid voltage}}{\text{rate of change of anode current with anode voltage}}$$

All three definitions amount to the same thing so long as the grid voltage is small compared with the anode voltage. The straight lines of Fig. 3.13 are drawn from equation (3.5) and are representative of the V_a - V_g characteristics of a typical triode valve, where the cathode current has been maintained at a constant value. The amplification factor, μ , is given by the ratio $-\delta V_a/\delta V_g$, which is, of course, the same for all the lines. The values of μ met in practice range from 4 for a valve handling appreciable power to about 100 for an amplifier dealing with the minute signals entering the aerial of a radio receiver.

Mutual Conductance

Probably the most important characteristic of the triode is the mode of variation of the anode current with the grid voltage, the anode voltage being kept constant. These are the characteristics exhibited qualitatively by the curves of Fig. 3·8. Using the same equation as was used for Fig. 3·13, the I_a - V_g characteristics are shown quantitatively in Fig. 3·14. For the values of V_a and V_g chosen, $I_a = I_k$, the total cathode current. The grid current is zero.

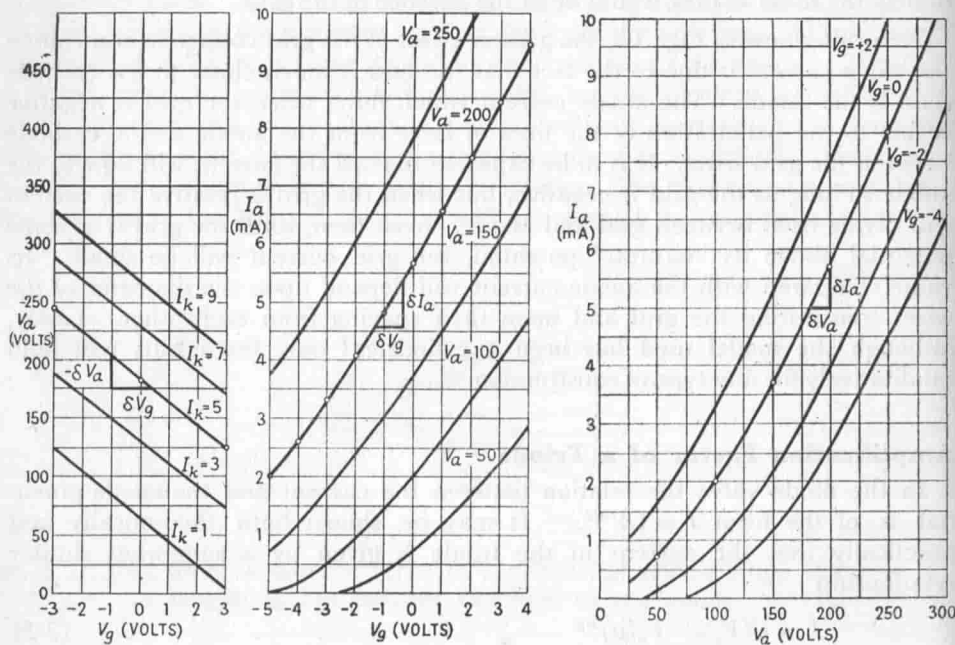


Fig. 3.13-3.15. Typical triode characteristics;

Fig. 3.13, $V_a - V_g$; Fig. 3.14, $I_a - V_g$; Fig. 3.15, $I_a - V_a$.

In this figure a great part of each curve is substantially straight, and the various curves are almost parallel over considerable ranges. This suggests the use of another valve 'constant' which has been named the mutual conductance or transconductance. The mutual conductance g_m is the ratio of the change in the anode current δI_a to the change in the grid voltage δV_g producing it, the anode voltage being kept constant. This is the gradient of any of the curves of Fig. 3·14. Clearly, g_m is not a true constant. For $V_a = 200$ volts, it varies as shown in Table 3·1. The mutual conductance should therefore be specified for a given value of the anode current. For example, if we take $I_a = 5$ mA in Fig. 3·14, $g_m = 0.815$ mA/V for any combination of V_a and V_g .

TABLE 3·1

V_g	-10	-5	-2.5	0	2.5	V
g_m	0	0.61	0.74	0.86	0.97	mA/V

Mutual conductance is measured in mA/V or in $\mu\text{A/V}$. These ratios have, of course, the names millimho and micromho, but it is quite usual to speak of mA/V directly. The mutual conductance of a triode can have any value between a few hundred and a few thousand micromhos. A typical value would be 1,000.

Anode Resistance

In the discussion of the diode it was pointed out that a valuable parameter was the rate of change of anode voltage with anode current, or its inverse, the rate of change of anode current with anode voltage. In the case of the triode the dynamic plate or anode resistance is similarly defined as the ratio of the change, δV_a , in the anode potential to the resulting change, δI_a , in the anode current, V_g remaining constant. The curves of Fig. 3.15 show the variation of I_a with V_a , V_g being kept constant. They are drawn from the same equation as was used for Fig. 3.13 and 3.14. The anode resistance, $r_a = \delta V_a / \delta I_a$ is like the mutual conductance in that it varies with V_a and V_g . But like the mutual conductance, the dynamic resistance is approximately constant for a given value of the anode current. In the figure its value at $I_a = 5 \text{ mA}$ is 24,600 ohms. In the case of a large transmitting triode r_a may be as low as a few hundred ohms. In receiving triodes it may be as large as 60,000 ohms.

Relation between the Three Triode Parameters

The three triode parameters, dynamic resistance, mutual conductance, and amplification factor are not independent, as may be seen by inspection. If a valve is set up so that V_a and V_g may be varied and I_a measured, then a cycle of variations can be gone through as follows:—

Let V_g be increased by δV_g , and let the corresponding change in I_a be δI_a . Then let V_a be decreased by an amount δV_a to bring I_a back to its original value. Then by definition,

$$g_m = \frac{\delta I_a}{\delta V_g}; \quad \mu = + \frac{\delta V_a}{\delta V_g}; \quad \text{and} \quad r_a = + \frac{\delta V_a}{\delta I_a}$$

and therefore $r_a g_m = \mu$ (3.6)

Returning to our previous characteristics, the values of g_m , r_a , and μ , measured at $I = 5 \text{ mA}$ were 0.815 mA/V, 24,600 Ω and 20 respectively. It may easily be verified that these are related as given by equation (3.6).

Importance of Mutual Conductance

The outstanding feature of the triode valve is the fact that it will allow a considerable current to pass when the grid voltage is negative and when the grid is therefore not taking any current itself, but that this anode current is closely controlled by the negative grid voltage. The triode valve is therefore essentially a device whereby the power drain from the supply to the anode can be controlled by a voltage swing at the grid. Neglecting features of the triode which we have not so far considered, this implies that by means of it a voltage can be used to vary a power. Theoretically, the grid voltage can be supplied using a negligible power, and so we can tap the power of the anode supply by means of a very minute triggering power. The mutual conductance

measures the efficiency of this triggering action, and so it may be regarded as the most important of the valve 'constants'. A simple example of the action of the triode will now be given.

Small Signal Operation of the Triode

Let us now examine a very simple form of triode amplifier in which a small 'signal' voltage δV_g is applied to the grid, the amplified 'signal' δV_a being taken from a resistance between the anode of the valve and the positive lead of the supply. Fig. 3.16 is a representation of the circuit.

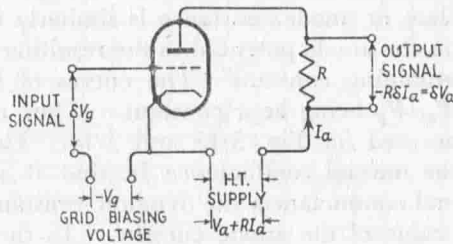


Fig. 3.16. A simple form of triode amplifier, illustrating the action of the valve.

Let the grid be biased negatively with regard to the cathode by means of a battery (as shown), and let this negative bias V_g be chosen so that the anode current I_a has a suitable value, and the grid current is negligible. Then, when the signal δV_g is applied to the grid, the change in the anode current would be $g_m \delta V_g$ if the anode voltage remained constant. But the presence of the resistance R causes the anode voltage to fall as the current rises, and so in examining the operation of the amplifier it must be assumed that V_g , I_a , and V_a all change. The overall change in the anode current is the change due to the increase in V_g minus the change due to the decrease in V_a . If these changes are small enough, we may write

$$\delta I_a = \delta I_a (V_a \text{ being kept constant}) + \delta I_a (V_g \text{ being kept constant})$$

$$\text{but } \delta I_a (V_a \text{ being constant}) = g_m \delta V_g$$

$$\text{and } \delta I_a (V_g \text{ being constant}) = \delta V_a / r_a$$

$$\text{whence} \quad \delta I_a = g_m \delta V_g + \delta V_a / r_a$$

But $\delta V_a = -R \delta I_a$, since the change in the anode voltage is due to the drop across the resistance R .

$$\text{Therefore} \quad \delta I_a (1 + R/r_a) = g_m \delta V_g$$

$$\text{or,} \quad R \delta I_a = R g_m \delta V_g / (1 + R/r_a) \quad (3.7)$$

$\delta V_a = -R \delta I_a$ is the change in the voltage appearing across R and available in a high impedance voltmeter, so we may say that the available signal voltage at the output is δV_a . The voltage amplification of the valve and circuit is therefore given by

$$A = -R g_m / (1 + R/r_a) = -\mu / (1 + r_a/R) \quad (3.8)$$

and as R becomes greater the amplification tends to μ , the amplification factor.

Finally, it should be noted that when the grid signal voltage is positive, the anode signal voltage is negative. As the grid voltage rises, the anode voltage falls; the output is in opposition to the input.

Now, if a generator of e.m.f. E and internal resistance r produces a voltage V across an external resistance R , carrying a current I ,

$$E = I(R + r) \text{ and } V = IR,$$

therefore

$$V = E/(1 + r/R).$$

Comparing this with equation (3.7), re-written,

$$R \delta I_a = -\mu \delta V_g / (1 + r_a/R),$$

if $V = R \delta I_a$, $E = -\mu \delta V_g$ and $r = r_a$. It may therefore be said that the valve behaves exactly like a generator of internal resistance r_a and e.m.f. $\mu \delta V_g$. This constitutes the simplest equivalent generator for the triode valve. Fig. 3.16 may therefore be replaced by the 'equivalent circuit' shown in Fig. 3.17, and small signal problems involving the triode amplifier may be solved by using the equivalent generator.

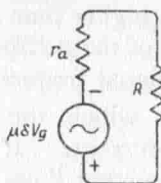


Fig. 3.17. The equivalent circuit of a small-signal triode amplifier.

To complete the discussion something must be said of what is meant by a 'small signal'. The analysis given above is exact, if it may be assumed that the valve parameters remain constant over the ranges of δV_g , δV_a and δI_a . As can be determined from the curves of Fig. 3.13—3.15, a change of 1 volt at the grid need not produce a change of more than 5% in g_m . A small signal can therefore be roughly defined as a signal less than 1 volt so far as the valve represented by the curves is concerned. To have some idea of the order of magnitude in practical cases, it may be stated that the signal in a radio receiver varies from about 1 microvolt to about 10 volts as it progresses from the aerial to the detector.

Large Signal Operation of the Triode

Triodes can also be used as a.c. power amplifiers with outputs varying from a few watts to many kilowatts, and in such cases the operation is very different from that described above. In the first place, the exciting signal is so large that the valve parameters can no longer be regarded as even approximately constant. Secondly, unless certain precautions are taken, the grid voltage tends to become so positive that an excessive part of the cathode current flows to the grid. This flow of grid current would, if it were allowed to occur, destroy the very high resistance between grid and cathode which is characteristic of small signal operation, and would give rise to serious distortion in certain applications of the amplifier. For these reasons, and for others connected with

the circuits used, the signal exciting the grid is applied in a different manner when the valve is required to act as a power amplifier. The mean grid voltage is made more negative; and care is taken to see that the supply voltage to the anode is high enough to allow the output signal voltage to swing to its peak negative value without producing excessive grid current. Power amplifiers can be made which function approximately according to the small signal theory given above, but they are very inefficient as generators. To increase the efficiency it is necessary to increase the mean negative grid bias.

This has led to a classification of amplifiers as A, B, and C which should be mentioned at this point.

A class A amplifier is one in which the grid bias and the signal voltage are such that the anode current never falls to zero. The amplifier discussed in the last section is an example of this class, as are all small-signal amplifiers. A class B amplifier is one in which the grid bias is adjusted to approximately the cut-off value. An alternating signal voltage applied to the grid therefore causes anode current to flow for one-half of a cycle only. This class of amplifier is often used for power amplification and is especially valuable where two valves are used in such a way that they allow current to flow in alternate cycles. This arrangement may be made to give distortionless amplification. The efficiency of a class B amplifier is higher than that of a class A amplifier. It is possible to adjust it so that 50% of the d.c. power absorbed from the supply is transformed into power at the signal frequency.

A class C amplifier is one in which the grid bias is greater than that required to cut off the anode current. If an alternating signal voltage is applied to the grid, the anode current flows for less than half a cycle. This type of amplifier is commonly used to amplify power at high frequency. Its efficiency is higher than that of class B, and can be made as great as 90% in favourable circumstances. Clearly, a class C amplifier compares favourably with any other generator so far as the conversion of anode d.c. supply power to signal power is concerned.

The theory of amplifiers in general will be given in Volume 4. Enough has been said here to indicate how the valve action changes as we pass from small to large signals.

Triode Inter-Electrode Capacitances

Our typical triode consists of three cylindrical structures, the grid inside and coaxial with the anode, and the cathode inside and coaxial with the grid. Such a structure forms a system of capacitors of small but not always negligible capacitances. Indeed, as will be seen in Volume 4, the inter-electrode capacitances play a major part in triode operation at high frequencies. It is therefore worth while at this juncture to obtain some idea of the order of magnitude of the capacitive effects and to relate these to the valve constants μ , g_m and r_a already defined. We shall assume throughout that the cathode is our reference conductor at earth potential. This is the basis of a large proportion of valve theory, although in certain circuits the grid or the anode may be the earthed electrode. First there is the grid-cathode capacitance. As shown in Volume 1, the capacitance per metre of concentric cylindrical conductors of radii a_1 and a_2 is

$$C = 2\pi\kappa/\log_e(a_2/a_1) \text{ farads}$$

If, for a moment, we regard the grid-cathode capacitance as being given only by the physical disposition of the parts, then for $a_1 = 1$ mm, $a_2 = 1.5$ mm, and $\kappa = \kappa_o = 8.85 \times 10^{-12}$ farad/metre, $C = 218$ pF/metre, or, for a valve 2 cm long, $C = 4.4$ pF. This value is almost certainly too small for the configuration chosen, since no account has been taken of space charge or of edge effects, but it gives the order of magnitude. The grid-anode capacitance could be calculated in the same way, and in this case the value would be much more nearly correct, since the electric field in the grid-anode space is more nearly electrostatic. The second point to notice is that there must exist in the triode valve a certain capacitance between cathode and anode. If this were not so—if the grid completely shielded the cathode from the anode—no electrons could pass through the valve at all. Therefore, a direct capacitance must be assumed, and the electrostatic arrangement in the triode must correspond to that shown in Fig. 3.18.

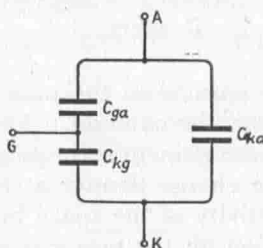


Fig. 3.18. Representation of the inter-electrode capacitances in the triode valve.

The charge on the cathode surface, Q_k , will be given by

$$\begin{aligned} -Q_k &= V_g C_{kg} + V_a C_{ka} \\ &= C_{kg} [V_g + V_a C_{ka}/C_{kg}] \end{aligned}$$

If we write $-Q_k = VC_{kg}$, it gives us the effective voltage V so far as the cathode is concerned. But we have already seen that

$$V = V_g + V_a/\mu$$

Therefore,

$$\mu = C_{kg}/C_{ka} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.9)$$

If, therefore, the amplification factor is high, the anode-cathode capacitance is small compared with C_{kg} . This merely indicates that the screening of the anode by the grid is good. Moreover, if, as in our typical valve, we take $\mu = 20$, and the electrostatic value of C_{kg} found above (i.e., $C_{kg} = 4.4$ pF) then C_{ka} is only about 1/5 pF. For some valves this figure is a good approximation, but for others the direct capacitance, cathode to plate, is greater than 1 pF. As will be seen later this has important repercussions on the use of the triode as an amplifier at very high frequencies.

It should, however, be noted that a valve is almost always placed in a valveholder which is itself connected in a circuit. The effective capacitance of the valve leads, the holder and the connections may therefore mask the effect of the true inter-electrode capacitances.

APPENDIX 3.1

Derivation of the Langmuir-Child Law

Gauss's theorem states that the total normal induction over a closed surface is equal to the charge inside the surface. Let the theorem be applied to the element of unit cross-sectional area and thickness δx shown in Fig. 3.19.

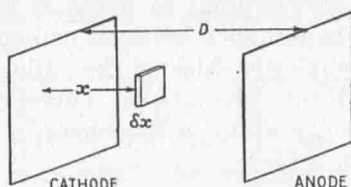


Fig. 3.19. Diagram of a planar diode, the parallel electrodes being distance D apart.

This element has its faces parallel to the plane parallel electrodes of the diode, and is at distance x from the cathode. Then, neglecting edge effects, the only electric force acting at the element is $\mathbf{E}_x$, perpendicular to the electrodes. $\mathbf{E}_x$ varies with x . Let ρ be the charge density at the element. Then ρ varies with x . Let κ be the permittivity of the space between cathode and anode. Then the total normal induction on the face x is $\kappa \mathbf{E}_x$ since the area of cross-section is unity. Similarly, the total normal induction at $x + \delta x$ is $\kappa(\mathbf{E}_x + \delta x d\mathbf{E}_x/dx)$. Therefore the normal induction over the closed surface is

$$\kappa \left(\mathbf{E}_x + \frac{d\mathbf{E}_x}{dx} \delta x \right) - \kappa \mathbf{E}_x = \kappa \frac{d\mathbf{E}_x}{dx} \delta x$$

By Gauss's theorem this is $\rho \delta x$.

Therefore $\kappa \frac{d\mathbf{E}_x}{dx} = \rho$, and since $\mathbf{E}_x = -dV_x/dx$,

where V_x is the potential at x ,

$$\frac{d^2 V_x}{dx^2} = -\frac{\rho}{\kappa} \quad . \quad . \quad . \quad . \quad . \quad (3.10)$$

This is the unidimensional form of Poisson's equation.

The current I is caused by moving electrons. Let the average velocity of the electrons at the element be v_x .

Then $I = \rho v_x$ and $\frac{1}{2} m v_x^2 = V_x e \quad . \quad . \quad . \quad . \quad . \quad (3.11)$

where m is the mass and e is the charge on each electron.

Eliminate v_x and ρ between the last three equations.

Then $\frac{d^2 V_x}{dx^2} = \frac{1}{\kappa} \left(\frac{m}{2e} \right)^{\frac{1}{2}} I V_x^{-\frac{1}{2}}$

and $\left(\frac{dV_x}{dx} \right)^2 = \frac{4}{\kappa} \left(\frac{m}{2e} \right)^{\frac{1}{2}} I V_x + \text{Constant}$

Integrating, and assuming that $V_x = dV_x/dx = 0$ at $x = 0$

$$\frac{4}{3} V_x^{3/4} = \frac{2}{\kappa} \left(\frac{m}{2e} \right)^{1/4} I^{1/2} x + \text{Constant}$$

But $V_x = 0$, when $x = 0$. Therefore, when $x = D$, and $V_x = V$,

$$I = \frac{4\kappa}{9} \left(\frac{2e}{m} \right)^{1/2} \frac{V^{3/2}}{D^2} \quad (3.12)$$

Let $\kappa = \kappa_0 = 8.85 \times 10^{-12}$ farad/metre, $e = 1.60 \times 10^{-19}$ coulomb and $m = 9.11 \times 10^{-31}$ kg, then

$$I = 2.34 \times 10^{-6} \frac{V^{3/2}}{D^2} \text{ A/unit area} \quad (3.2)$$

APPENDIX 3.2

The Current-Voltage Characteristic of the Conducting Diode

A particular condition of the diode valve which is of considerable importance is that in which a battery or other source of constant e.m.f. in series with a smaller voltage (frequently alternating) is connected across the valve. The fundamental circuit for this condition is shown in Figure 3.20.

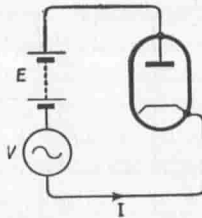


Fig. 3.20. A diode valve in series with a battery of e.m.f. E and a signal of voltage V .

It is often of value to be able to use a formula for the current flowing through the valve which does not involve the three-halves power, and for the case in question this can be done. Let the battery e.m.f. be E and the smaller voltage be V . Then

$$I = (CE + V)^{3/2}$$

But if V is smaller than E , we may write

$$I = CE^{3/2} (1 + V/E)^{3/2}, \text{ and by the binomial theorem} \\ = CE^{3/2} (1 + 3V/2E + 3V^2/8E^2 + \dots)$$

Hence $I = CE^{3/2} (1 + bV + cV^2 + dV^3 + \dots)$ where $b, c, d, \dots$ are appropriate constants.

But $CE^{3/2}$ is the standing current due to the battery. Therefore,

$$I - I(V = 0) = bV + cV^2 + dV^3 + \dots$$

or, the current due to the voltage V may be expressed as a power series in V . This result will be used later in the volume in connection with the actions of mixer valves and demodulators.

CHAPTER 4

Multi-Grid Valves

Difficulties of the Triode Construction

AT THE END OF Chapter 3 it was pointed out that the triode valve contains three separable capacitors, two in series, C_{kg} and C_{ga} , with another in parallel, C_{ka} . Of these three C_{kg} and C_{ga} are the most important, and the grid-anode capacitance in particular influences very considerably the action of the valve at high frequencies. In fact, although all three capacitances may often be neglected at audio frequencies, the effect of C_{kg} and C_{ga} begins to be felt at about 10 kc/s, first of all reducing the input impedance of the triode, and finally destroying the action of the valve as a simple amplifier. So important is this that it is worth while at this point considering the typical amplifier in more detail, as an introduction to the use of the multi-electrode valve. The theory about to be given is illustrative only. A more exhaustive treatment of the problem is given in Volume 4.

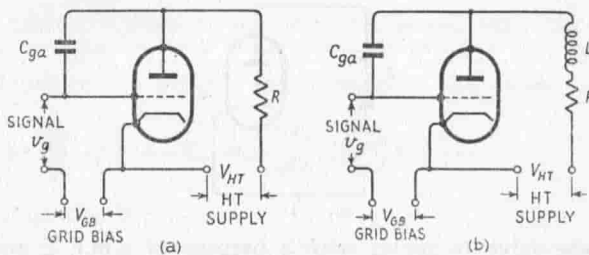


Fig. 4.1. Amplifier circuits showing the grid-anode capacitance of the valve (a) with resistive anode load and (b) with inductive reactance in the anode load.

Let Fig. 4.1(a) represent a simple amplifier with a resistive load R but, in this case, let the grid-anode capacitance be taken into account. Then, when the anode swings positively under the action of a negative grid swing, the voltage across C_{ga} is the sum of both the grid and anode voltage changes. For instance, if at the instant the grid swings 1 V negatively the anode swings 20 V positively, the voltage across the capacitance has changed by 21 V from its static value. The current taken by the capacitance is, therefore, 21 times as great as if the grid voltage only were acting on it. This current flows from the source of grid voltage and is the same as if C_{ga} were absent but a capacitance 21 times as great were connected between grid and cathode.

If the grid and anode voltages are exactly in opposite phase the effect of the grid-to-anode capacitance is exactly the same as if the grid-to-cathode capacitance were increased from its real value C_{kg} to

$$C_i = C_{kg} + C_{ga} (1 + A)$$

where A is the voltage amplification of the stage.

For a valve in which the geometrical values of C_{kg} and C_{ga} are 10 and 2 pF respectively, and which is operating at a voltage amplification of 20,

$$C_i = 10 + 42 = 52 \text{ pF}$$

a value much in excess of the geometrical capacitance.

In practice, the grid and anode voltages are rarely exactly in opposite phase. Even if C_{ga} is the only reactance present the current through it flows through R and modifies the amplitude and phase of the anode voltage. When the anode load is not a pure resistance R but contains also a reactive element, and it nearly always does in practice, the grid and anode voltages may depart appreciably from phase opposition. The current flowing through C_{ga} will not then lead V_g by 90° as it does in the simple case of grid and anode voltages in phase opposition.

The foregoing simple relation is, therefore, usually only an approximation, but it is nearly always sufficiently accurate for practical purposes over the audio-frequency range, say up to 10 kc/s. At higher frequencies it is necessary to take into account the effect of reactance in the anode load. This affects the phase of the anode voltage and, if the valve is giving any reasonable amount of amplification, the anode voltage is much larger than the grid voltage. The current in C_{ga} is, therefore, dictated mainly by the anode voltage and is very nearly 90° out of phase with it.

The current drawn from the source and flowing in C_{ga} is thus at a phase angle to V_g by an amount which depends on the phase angle of the anode voltage and in turn this depends on the nature of the anode load. Instead of leading V_g by 90° , the current leads by a small amount less or more than 90° . Consequently, the effect of C_{ga} can no longer be represented by a pure capacitance in shunt with C_{gk} ; it must have a resistance associated with it to make the phase angle of the equivalent circuit the same as that of the real one.

This resistance may be either positive or negative. Physically, this means that the anode circuit absorbs power from or supplies it to the input circuit. With a capacitive load the resistance is positive and power is absorbed from the input. With an inductive load, Fig. 4.1(b), the input resistance is negative and power is supplied to the input. This is true for a certain range of values of the resistance and reactance forming the load and is a very serious limitation of the triode at high frequencies. Under suitable conditions it may well happen that the power delivered from the anode circuit to the grid circuit exceeds the losses in the latter and the grid voltage of the valve is provided from the anode circuit. The system then generates self-sustained oscillations, and no input signal is needed to obtain an output. This self-oscillating action occurs very readily at high frequencies, for then the reactance of C_{ga} becomes fairly small.

Because of the effects of C_{ga} , the simple triode circuit is almost useless at high frequencies. To employ it effectively at such frequencies it is necessary to adopt special and more complex circuits. Even then the large input current flowing into C_{ga} may result in a serious dissipation of power in any resistance in series with it, and here even the resistance of the valve leads may be important.

Some of the effects of the grid-anode capacitance may be eliminated by special neutralizing circuits. At one time such circuits were widely used in

receivers for frequencies up to about 1 Mc/s. They are difficult to adjust, especially when the equipment must operate over a wide range of frequencies, and become almost impossible to operate when large amplification is needed. They are now only occasionally used in receivers, but are still common in transmitters, especially in high-power fixed-frequency apparatus.

In receivers the triode has been very largely superseded by a multi-electrode valve in which capacitance between the grid and anode is relatively very small. Instead of being about 1—3 pF it is commonly only 0.005—0.01 pF and values as low as 0.001 pF are sometimes found. The most commonly used valve of this type is the pentode but, historically speaking, the screen-grid tetrode was the first attempt to eliminate the grid-anode capacitance.

Tetrode Construction

Consider a triode valve to which a fourth electrode, a grid or mesh, has been added between the control grid and the anode, as in Figure 4.2.

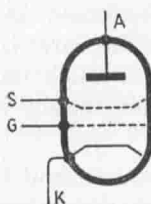


Fig. 4.2. Conventional representation of a screen-grid tetrode.

Let this second grid, S, called the screen, be connected to a d.c. supply at a voltage about two-thirds of that of the anode. Then the cathode, the control grid and the screen grid will act in much the same way as the cathode, grid and anode of a triode. That is, the electric field due to the potential of the screen will penetrate through the control grid to the cathode and will draw off some electrons, this current being actively controlled by the potential of the grid. But the electrons flowing towards the screen will not, in general, be absorbed there. Owing to the mesh structure of the screen, and to the higher potential of the anode, most of the electrons will flow past the screen to the anode, so that the anode current will be controlled, just as the anode current in the triode is controlled, by the potential of the grid. The new construction, therefore, gives a new type of valve in which (i) the current to the anode is controlled by a grid as in the triode, and (ii) the anode itself is screened from the control grid by another electrode.

The importance of the screen-grid lies in the use which can be made of it to reduce very considerably the undesired grid-anode capacitance which, as we have just seen, makes the triode difficult to use at high frequencies. The effect can best be understood by considering the example of a simple screen-grid valve amplifier with a resistive load, as shown in Figure 4.3.

Consider the screen-grid valve in which the screen S is maintained at a potential about two-thirds of that of + h.t. by means of the resistance R_s in which there is a certain potential drop due to the electron current flowing to the screen. The capacitor C_s is chosen to be of such a value that it is a virtual

short-circuit (zero impedance) at the signal-voltage frequencies. Therefore, although the voltage swing at the control grid varies the current flowing through the valve to both S and A, the capacitance C_s and the resistance R_s together maintain S at a constant potential. In fact, so far as the frequencies in the signal voltage are concerned, S is at earth potential. We have therefore interposed what is virtually an earthed electrode between the control grid and the anode without interfering too much with the current variations at the latter. This screens the anode effectively from the grid, and vice versa, reducing the capacitance C_{ga} to a very small value. Another way of looking at it is that we have substituted for the usual triode capacitances those shown in Fig. 4.4.

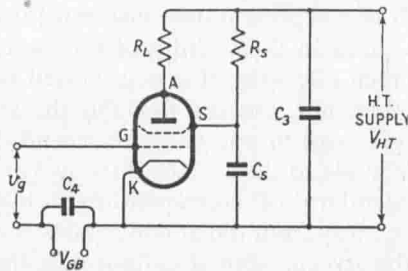


Fig. 4.3. Simple amplifier circuit using a screen-grid tetrode.

The capacitances C_{kg} , C_{gs} , C_{sa} , have the usual values, given approximately by the valve geometry and of the order of 10 pF each. The capacitor C_s , introduced externally, may have any value we choose; it can always be made so large that the fraction of the current from A to S which flows through G is negligibly small.

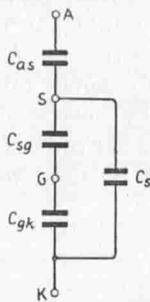


Fig. 4.4. Inter-electrode capacitances in a screen-grid tetrode, with an external capacitance between screen and cathode.

It has been stated that the grid-anode capacitance in the tetrode is very small. It is not always negligible, however. At very high frequencies it is often necessary to use circuits to neutralize the feedback even in the case of tetrodes. Moreover, theoretically, at least, the grid-anode capacitance sets a limit to the amplification which can be achieved by means of a tetrode at frequencies as low as a megacycle. The value of the capacitance is generally given in the manufacturer's valve data. It varies from 0.005 to 0.001 pF.

Secondary Emission Effects

The tetrode, or 'screen-grid valve' would be an ideal electronic device to use for amplification and for many other purposes but for the effect of secondary emission. In Chapter 1 it was pointed out that when electrons bombard a metal surface, other secondary electrons are produced, particularly when the speed of the bombarding electrons is high. The facts of secondary emission were also discussed in some detail. These secondaries may have speeds not much less than those of the primary electrons which produce them, and in the tetrode, with its rather complex potential distribution, the effects of secondary emission are very marked and may be very inconvenient. To examine these effects it is necessary to draw the tetrode characteristics, but first it is useful to trace the path of a representative electron through the valve. Pulled away from the charge cloud in the vicinity of the cathode by the field due to the potentials of the screen and grid, it is accelerated through the mesh of the grid and, more often than not, escapes through the screen also to strike the anode. Here, it may give rise to one or more secondaries. Since it impinges on the anode at a large angle to the surface if its own energy is large, it usually produces only one secondary of correspondingly high energy. This new electron can easily move away from the anode against the electric field and may, indeed, pass through the screen. But it cannot pass through the control grid, for its energy must be less than that of the primary which produced it. It returns, therefore, towards the anode, and either falls into the latter or is absorbed by the screen. If it is absorbed by the screen, the primary electron which caused it made no contribution to the anode current. It is therefore possible for many electrons to strike the anode without producing any net anode current.

Screen-Grid Tetrode Characteristics

The simplest characteristics of the tetrode are obtained by varying the grid voltage V_g and measuring either the current flowing from the cathode I_k or the current flowing to the anode I_a , the screen potential V_{sg} being maintained equal to or less than the anode potential V_a . The curves so obtained are of practical value, since they correspond to the usual operation of the valve. Fig. 4.6 exhibits a typical set of curves, showing I_a and I_k as functions of V_g . The circuit which could be used to obtain these curves is shown in Fig. 4.5.

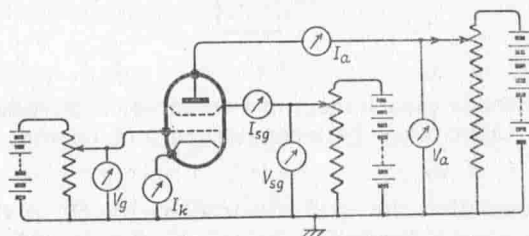


Fig. 4.5. Circuit for measuring the characteristics of a screen-grid tetrode.

The curves are very similar to those obtained from a triode. The mutual conductance, measured at $I_a = 15$ mA, is about 2 mA/V. The dotted lines represent I_k under the static conditions employed, and the difference between

a full line and an adjacent dotted line represents the net current to the screen. For any of the plotted curves the screen current is an almost constant fraction of the total current, but this is not true for other conditions.

Using the circuit of Fig. 4-5 once more, the currents I_a , I_{sg} , and I_k may be plotted as functions of V_a for special values of V_{sg} and V_g . The result is shown for a typical screen-grid valve in Fig. 4-6(b), where V_{sg} has been held at +100 V throughout and V_g is slightly negative. The curves have been slightly simplified to bring out the important features.

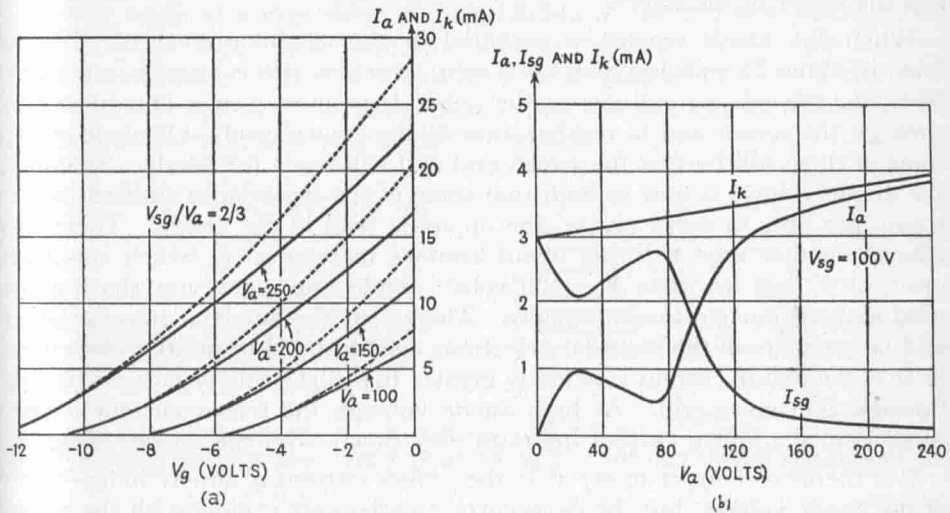


Fig. 4.6. Screen-grid tetrode characteristics; (a) anode current (full line) and cathode current (dotted) versus control-grid voltage; (b) anode, cathode and screen-grid currents versus anode voltage.

The general explanation of the variations in I_a and I_{sg} is not difficult. First it should be noted that the cathode current I_k is always equal to the sum of the currents I_a and I_{sg} , as might be expected. This is because the current must go either to the anode or to the screen when the control grid is negative. As V_a increases from zero, the number of electrons collected by the anode increases rapidly. Again, this is to be expected. In fact, one would expect a *very* rapid increase in the primary current to the anode, since all the electrons emitted by the space charge in front of the cathode have been accelerated between the cathode and the screen to a velocity corresponding to 100 volts, and are therefore sufficiently energetic to move against the opposing field between the screen and the anode, and to arrive at the latter. We should therefore expect that when the anode potential is only a few volts positive all the electrons which succeed in avoiding the screen in their first transit through the valve would reach the anode.

However, as the anode potential increases, the primary electrons striking it begin to be of sufficient energy to produce other secondary electrons; and these, finding themselves in an electric field pulling them away from the anode, tend to find their way to the most positive electrode, namely, the screen. According to Fig. 4-6, this secondary emission must start at about 10 volts, and

by the time that V_a has reached 25 volts the rate of increase of secondary emission has become equal to the rate of increase of primary emission. The anode current, therefore, becomes momentarily constant. As V_a is still further increased, the additional primary current (if any) is more than neutralized by the loss of secondaries, so that I_a decreases. This must mean that the number of secondaries produced per primary is still increasing, and is in accord with the theory given in Chapter 1. At the same time it must be remembered that secondary electrons are also being produced at the screen. These are not usually able to reach the anode when the potential of the latter is so much less than that of the screen.

When the anode reaches a potential in the neighbourhood of 75 volts (that is, about 25 volts less than the screen), two new effects become important. First, the secondaries emitted by the anode have now enough energy to pass through the screen and to reach almost to the control grid. On their return some of these will by-pass the screen-grid and will reach the anode. Secondly, the anode voltage is now so high that some of the secondaries emitted by the screen are able to move against the opposing field to the anode. These two effects together tend to bring about another increase in I_a which continues until, at V_a just less than V_{sg} (100 volts), anode and screen are sharing the total cathode current almost equally. Thereafter, the anode acquires a larger and larger share of the secondary electrons emitted by both positive electrodes, so that the anode current is actually greater than the primary current passing through the screen-grid. At high anode voltages the screen current is very small, and the anode current increases very slowly with the anode voltage.

It is therefore correct to say that the *cathode* current is almost independent of the anode voltage, but the *anode* current varies very rapidly with the anode voltage at certain intermediate values of the latter. This curious anode characteristic, typical of the screen-grid tetrode, leads to considerable difficulty in the use of the valve, so that it is now almost obsolete. It has been described in some detail above in order that its modern counterparts, the beam tetrode and the pentode, may be more easily understood. The simple screen-grid valve is, however, quite satisfactory when used as a small signal amplifier, where the anode potential does not vary very much. It then behaves like a triode of very high anode impedance, and with a negligible grid-anode capacitance. The screen electrode has, in this respect, done what was expected of it.

The Beam Tetrode

The effects of secondary emission are made very obvious by the characteristics of the screen-grid valve. If some measure could be adopted which would eliminate secondary emission, or render its effects nugatory, a much more useful electronic device would result. The beam tetrode is such a valve. It was developed several years after the pentode, which also eliminates the unwanted effects of secondary emission by another means. The beam tetrode is logically prior to the pentode, however, and for that reason it will be considered first. The important points in the construction of the beam tetrode are illustrated by Fig. 4.7(a) and 4.7(b).

The control grid G and the screen grid S consist of wires of equal pitch which are so aligned that the electrons escaping through the (negative) control

grid are formed into beams which do not strike the screen grid. This beaming action is so complete that the screen current is almost a negligible fraction of the total cathode current. The beaming action is improved by the introduction of the end plates B_1 and B_2 which are connected to the cathode and which keep the electrons from spreading outwards. Also, the anode is formed to a shape which, taken in conjunction with the end plates, causes the electrons to strike it at right angles to the surface. As may be remembered (see Chapter 1) this prevents one type of secondary emission. The beam tetrode is, in fact, a much more carefully designed valve than those discussed previously. Its inception came at a time when the mechanism of electron flow had received much consideration, and so care was taken to constrain the electrons to move along certain paths. Indeed, the valve may be considered to be made up of a number of small, carefully designed units, each acting as a combination of a cathode, a control system, and a collecting anode.

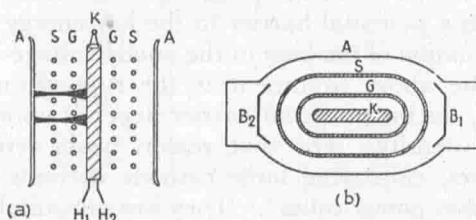


Fig. 4.7. (a) Axial section of a beam tetrode showing 'beams' of electrons; (b) cross-section of a beam tetrode showing the shaping of the anode and the positioning of the beaming electrodes.

The constructional details mentioned do not, however, explain the action of the valve. The most careful design cannot prevent the emission of secondary electrons from the anode and, unless some heretofore neglected action comes into play, these secondaries will be collected by the screen, when the anode potential is low, giving the characteristic curves of the screen-grid valve. The screen of the beam tetrode does *not* collect secondary electrons. This is due to the action of the increased space charge in the beam. A simple discussion of the mechanism will bring out the importance of this effect on the characteristics.

Reverting to Figure 4.7, let the screen be positive (say, 250 volts) the grid negative (-5 volts) and let the anode potential be varied, starting from cathode potential (zero volts). If the beams of electrons have been prevented from spreading by the action of the aligned grids and end plates, then the current approaching the anode must have roughly the same density as it had at the cathode. This space charge will produce a potential minimum between the screen and the anode at a small negative voltage. Such a space charge is the equivalent of a cathode in many of its effects and is, in fact, known as a 'virtual cathode'. As the anode potential is raised from zero, the space charge will persist, but the potential minimum will be at a positive voltage and will move towards the anode. There will no longer be a virtual cathode, since electrons will reach the anode with a finite speed. It is the slowing down of the electrons between the very positive screen and the not so positive anode which will produce a potential gradient greater than the average near the

screen, and a potential minimum near the anode. Electrons will be slowed down more than they would be in the absence of space charge, and so will be accelerated again towards the anode.

This potential minimum between screen and anode, considered in conjunction with the fact that almost all the secondaries will be of low energy, prevents electrons from reaching the screen from the anode. Indeed, the secondary electrons tend to increase the space charge and so to increase the repellant action on other secondaries. The emission of low energy secondaries is characteristic of electrons striking a metal surface at right angles, as has been described in Chapter 1. It is therefore important to place the anode surface perpendicular to the electron trajectories, and this is a feature of the design of a good beam tetrode.

The net effect of beaming the electrons is to produce high electron densities between screen and anode when the potential of the latter is the lower. The charge cloud presents a potential barrier to the low energy secondary electrons and prevents the formation of the kink in the anode voltage-current characteristic. Implicit in the above argument is the supposition that the current density is fairly high, for the potential barrier depends upon the absolute value of the space-charge density. For that reason beam tetrodes are generally power-handling valves, employing large cathode currents. They are known in the U.S.A. as 'beam power tubes'. They are not suitable for use as small-signal amplifiers, for they would be very inefficient in the sense that their power consumption would be large. More important still, as was indicated in Chapter 2, valve noise depends directly on cathode current; so, if the latter is large, the valve would cause a deterioration in the signal-to-noise ratio in the early stages of a receiver.

Beam Tetrode Characteristics

There is little difference in general shape between the anode-current-grid-voltage characteristics of a beam tetrode and those of any other amplifier valve. The important characteristic is that in which the anode current is exhibited as a function of the anode voltage at a constant screen potential. This is drawn for an average tetrode of anode dissipation 10 watts in Fig. 4.8.

It will be noted that there is no evidence of secondary emission. The extremely rapid rise of anode current at low anode voltages is followed by a sharp change of gradient, the remainder of the curve being almost a straight line. The screening effect of the second grid is signalled by the comparatively slow variation of current with voltage in the rectilinear part, the anode voltage having very little effect on the anode current. The corresponding dynamic resistance is about 8,000 ohms. Comparing the beam tetrode with the triode, the advantages which accrue from the introduction of the second grid are immediately obvious. The constants of the valve of Fig. 4.8 are amplification factor = 48, mutual conductance = 6 mA/V and, as has been stated, dynamic resistance = 8,000 ohms. The grid-anode capacitance would be about 0.2 pF. For a triode of equal anode dissipation, but with much inferior constants, the grid-anode capacitance would be about 3 pF. So far as power valves are concerned, therefore, the beam tetrode can be said to replace the triode with advantage.

The Pentode

In the beam tetrode the secondary electrons from the anode are prevented from reaching the screen by the repellant action of the space charge between screen and anode. This action can only occur when the anode is at a lower potential than the screen, but it is only under these circumstances that it is required. When the anode is at a higher potential than the screen, it attracts its secondaries back to itself; and in the beam tetrode the primary screen current is so small that the secondaries from the screen finding their way to the anode may be neglected.

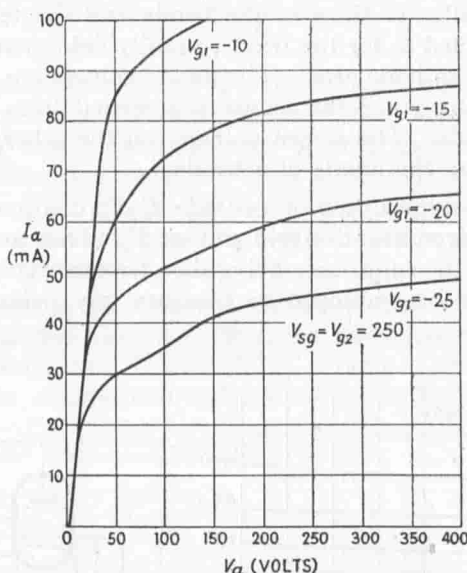


Fig. 4.8. Anode current versus anode voltage characteristics for a beam tetrode.

The pentode is (as its name implies) a five-electrode valve. It comprises a cathode, a control grid, a screen grid, and a coarse mesh grid between the screen and the anode, called the 'suppressor'. The function of the suppressor is to prevent the interchange of secondary electrons between the screen and the anode. Its action is very similar to that of the space charge in a beam tetrode, for it is maintained at cathode potential, repelling electrons coming towards it from either side. It is a conducting grid, depressing the potential between screen and anode.

The essential action of the pentode valve does not demand a meticulous attention to the prevention of primary screen or suppressor current. Alignment of the control grid and screen grid wires is not essential. But the pentode valve is so extensively used nowadays, and in so many different ways, that an enormous amount of effort has been directed towards improving its characteristics; and modern pentodes are designed to eliminate, as far as possible, primary current to all electrodes except the anode. There is another reason for doing this which has nothing to do with secondary emission: the unwanted noise produced by a valve amplifying a small signal depends to some extent on the division of current between the various valve electrodes; a good pentode

amplifier for small signals should therefore have the minimum screen and suppressor current. Care is also taken in the design to cause the anode current to increase very rapidly as the anode voltage is increased from zero. The rapid increase ends almost abruptly at a comparatively low anode voltage. Thereafter, the current only increases very slowly with the anode voltage. In short, the beam-tetrode characteristics of Fig. 4.8 are exaggerated in a good low-current pentode, with an even sharper shoulder in the I_a - V_a curve.

Pentode Characteristics

As is to be expected, the anode-current-grid-voltage characteristics of the pentode are similar to those of the triode and the tetrode. The mutual conductance g_m defined as for the triode, usually lies between 1 and 10 mA/V, depending upon the cathode area. The anode voltage has practically no effect on the anode current, since the anode is screened from the cathode by no fewer than three grids. The screen voltage, on the other hand, has an effect comparable to that of the anode of a triode.

Fig. 4.9 shows a typical set of pentode I_a - V_a characteristics, each curve corresponding to a given negative grid voltage V_g . They are drawn for a screen voltage of 250 volts, the suppressor being directly connected to the cathode. A circuit which might be employed to measure any pentode characteristic is shown in Fig. 4.10.

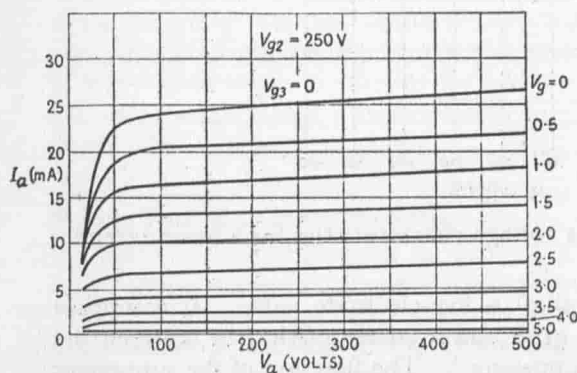


Fig. 4.9. Anode current versus anode voltage characteristics for a signal-amplifier pentode

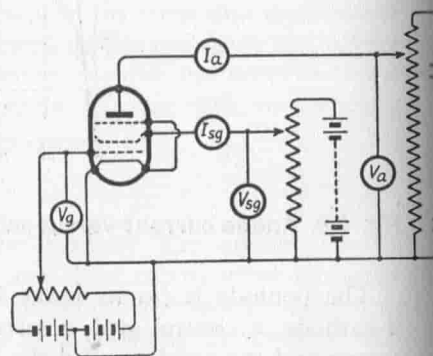


Fig. 4.10. Circuit for measurement of pentode characteristics.

The fact that the anode current is almost independent of the anode voltage gives the pentode one of its important properties as an amplifier; it causes the dynamic resistance of the anode to be extremely high. This is rarely less than 100,000 ohms and may be as high as 1.5 megohms. This property makes for an improved amplification, and μ being of the order of 1,000 is about ten times what can be expected from a high- μ triode.

Pentode Operation at Reduced Screen Voltage

The primary effect in a triode valve of reducing the anode potential is to reduce the total current flowing through the valve. The change is equivalent to increasing the negative voltage on the grid. Only to a much smaller degree is the mutual conductance changed.

In a pentode the screen grid takes the place of the anode in a triode so far as control of the cathode current is concerned. Decreasing the screen-grid voltage is roughly equivalent to increasing the negative bias on the grid, and does not decrease g_m to any important extent. This is a valuable property of the pentode which is made use of in certain resistance-coupled amplifiers.

Pentode and Triode Amplifiers.

In Chapter 3 it was shown that the voltage amplification of a triode is given by the formula

$$A = \frac{\mu}{(1 + r_a/R_L)} \quad (4.1)$$

Where μ is the amplification factor, r_a is the dynamic resistance, and R_L is the load resistance. The formula is only true for 'small signals', where the valve 'constants' remain really constant over all the variations of current and voltage. The amplification obtained from a pentode is given by the same formula (as a moment's thought will show) with appropriate values for μ , r_a and R_L . The pentode also behaves like a generator of open-circuit e.m.f. $e = -\mu v_g$ and of internal resistance r_a .

It is now necessary to enquire a little more closely into the application of equation (4.1) to real circuits. First, we note that the output voltage is $A v_g$ across the load resistance R_L (Fig. 4.11). Here v_g and $A v_g$ are considered to be the amplitudes of sinusoidal voltage swings on the grid and anode.

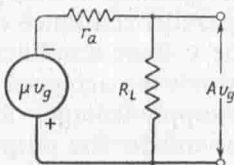


Fig. 4.11. Equivalent circuit of small-signal triode, tetrode or pentode amplifier with resistive anode load.

The a.c. power dissipated in the load is therefore given by P , where

$$P = \frac{A^2 v_g^2}{2R_L} = \frac{\mu^2 v_g^2}{2R_L(1 + r_a/R_L)^2} \quad (4.2)$$

remembering that the r.m.s. value of the voltage of amplitude v_g is $v_g/\sqrt{2}$. The power available in the load varies with the latter. When R_L is very small, $P = \mu^2 v_g^2 R_L / 2(R_L + r_a)^2$ will be small also. It will, in fact, tend to $\mu^2 v_g^2 R_L / 2r_a^2$. When R_L is very large, $P = \mu^2 v_g^2 / 2R_L(1 + r_a/R_L)^2$ will again be very small. In fact, P will tend to $\mu^2 v_g^2 / 2R_L$. Therefore the power available must have a maximum value for some intermediate value of R_L . This maximum value can be determined most easily by considering, instead, where the *minimum* value of $1/P$ occurs.

$$1/P = 2R_L(1 + r_a/R_L)^2 / \mu^2 v_g^2$$

Rewrite this as follows

$$\frac{1}{P} = \frac{2}{\mu^2 v_g^2} \left[\frac{(R_L + r_a)^2}{R_L} \right] = \frac{2}{\mu^2 v_g^2} \left[\frac{(R_L - r_a)^2}{R_L} + \frac{4r_a R_L}{R_L} \right] = \frac{2}{\mu^2 v_g^2} \left[\frac{(R_L - r_a)^2}{R_L} + 4r_a \right]$$

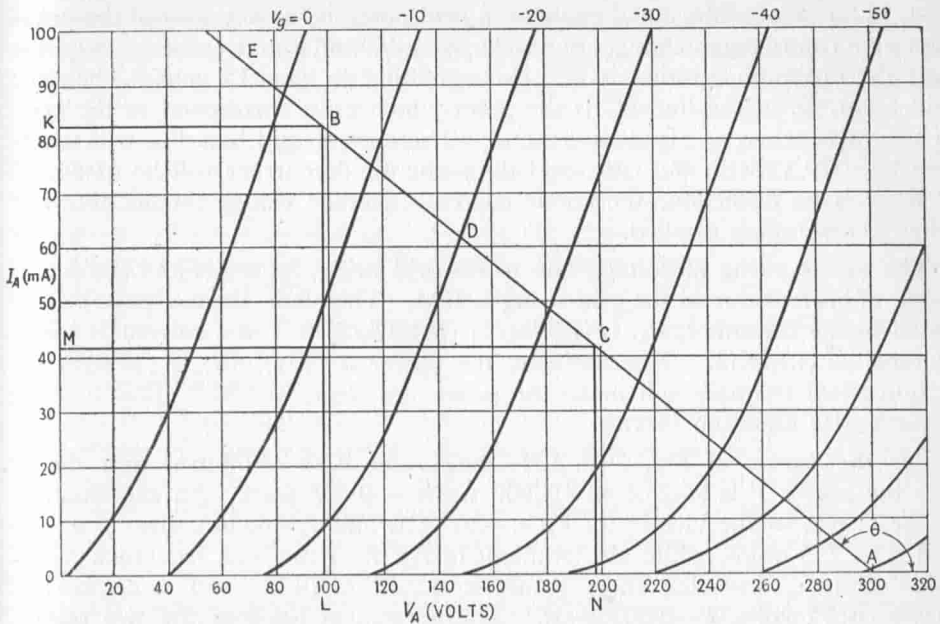


Fig. 4.12. Anode-current versus anode-voltage characteristics of a triode valve with superimposed load line of anode resistor.

question is that shown in Fig. 4.13, where the constant voltage applied to the load resistance, R_L , and to the valve, in series, is V_{HT} .

Let the negative grid bias be $-V_{GB}$, and the input signal be an alternating voltage of amplitude V_g . Then from the curves of Fig. 4.12 the negative grid bias V_{GB} which gives zero anode current I_a at an anode voltage V_{HT} may be read off. This will be a possible instantaneous state of the amplifier of Fig. 4.13. Let this point be A in Fig. 4.12. For any valve current I_a the anode voltage V_a will be given by $V_a = V_{HT} - R_L I_a$, and a straight line AE given by $I_a = (V_{HT} - V_a)/R_L$ may therefore be drawn across the characteristic curves, determining the possible simultaneous values of V_a and I_a with the valve in the circuit. Such a line does not depend only upon the valve but upon the load resistance, R_L ; it is called a 'load line'. The gradient of the load line is given by $\tan \theta = -1/R_L$, so, as R_L increases, θ becomes more nearly equal to 180° .

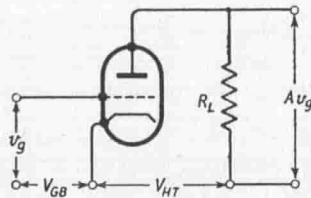


Fig. 4.13. Amplifier circuit showing the signal voltage V_g in series with the standing grid-bias voltage V_{GB} .

If the amplitude of the grid swing is v_g and minimum distortion is to be achieved, an intercept of grid voltage $2v_g$ must be chosen on the line AE so

that, as far as possible, equal changes in grid potential at any part of the intercept give rise to equal changes in anode potential and anode current. Suppose that the optimum position is BC, corresponding to $v_g = 15$ volts. Then the mid-point, D, of the line BC is the point which must correspond to the state of the valve when no signal is present. The correct grid bias V_{GB} will therefore be -20 volts in this case, and the static anode current will be 61.5 mA. If there is no distortion, the anode current (average value) should not vary when the signal is applied.

The anode swing of voltage due to the grid swing, $2v_g$ is LN and the anode swing of current due to the grid swing is KM. Therefore, the a.c. power dissipated in the resistance, R_L , is $(LN/2\sqrt{2})(MK/2\sqrt{2}) = \text{r.m.s. current} \times \text{r.m.s. voltage} = LN.MK/8$. Alternatively, the power is $(LN)^2/8R_L$ or $(MK)^2R_L/8$. A numerical example will make the procedure clear, and will allow a comparison with equation (4.2).

For the curves of Fig. 4.13, $LN=99.2$ volts, $KM=40.8$ mA, and therefore the power P is $99.2 \times 40.8/8,000$ watts $= 0.506$ watt. An examination of the curves in the vicinity of $V_b = -20$ volts and $I_a = 60$ mA gives $\mu = 4.3$ and $r_a = 735$ ohms, while the gradient of AE fixes the load resistance R_L as $2,444$ ohms. Inserting these values in equation (4.2), and remembering that v_g is 15 volts, $P=0.503$ watt. The agreement between the two results shows that the methods are equivalent. The load-line investigation gives more detailed information and should be used if possible.

Turning now to the pentode, a similar form of analysis may be applied in order to determine the anode current and potential swings due to a given grid excitation. The characteristics of the pentode look very different from those of the triode, as has already been established (see Fig. 4.9), but the technique of using the load line is the same. In Fig. 4.14 the I_a-V_a characteristics are shown for the CV109*, a representative receiving pentode. The corresponding values of g_m and r_a are 7 mA/V and 250 k Ω respectively. From the shape of these curves it may be deduced that there is considerable

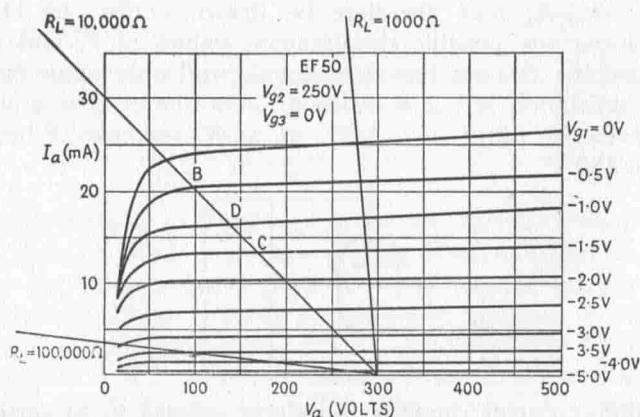


Fig. 4.14. Anode-current versus anode-voltage characteristics of a pentode with superimposed load lines for three different anode resistors.

* Mullard EF 50

latitude in choosing the value of the load resistance. Load lines for a total h.t. voltage of 300 are drawn for values of R_L ranging from 1,000 to 100,000 ohms. Choosing $R_L = r_a$ is obviously bad practice in this case, since the corresponding value of V_{GB} required (about -4 volts) takes us into the part of the I_a-V_g characteristic where g_m is varying rapidly. With a proper choice of R_L , however (say 10,000 ohms), the valve will give virtually undistorted amplification of any grid signal of swing less than 1.0 volt. Under these circumstances the power output is defined by the swing BDC about the mean grid voltage of -1.0 , and the power output is 60 mW. Using equation (4.2), and inserting $v_g = 0.5$ volt, $\mu = 1,750$, and $r_a = 250$ k Ω , $P = 57$ mW. This again agrees with the graphical method, but again the load line method is more illuminating.

Variable-Mu Pentode

There is a particular type of pentode valve which is much used in the early stages of a receiver and which deserves special mention. It is called a 'variable-mu' or 'remote cut-off' pentode. The valve is one in which the anode current diminishes more and more slowly as the grid is made negative, so that the values of g_m and μ vary considerably at low anode currents. The effect is one which is present to some extent in all valves, but in the variable-mu pentode it is purposely exaggerated. The I_a-V_g characteristics of such a valve are shown in Fig. 4.15, the dotted curves representing the more normal 'sharp cut-off' pentode.

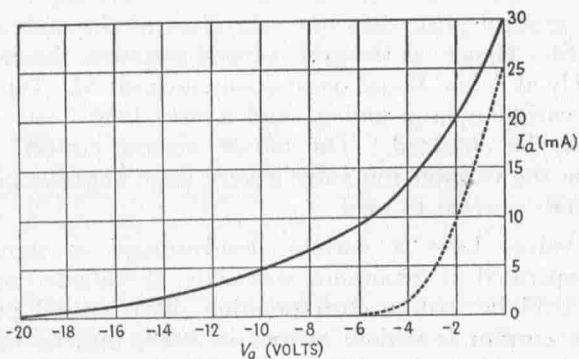


Fig. 4.15. The anode-current versus grid-voltage characteristic of a variable-mu pentode, showing the long 'tail' and the slow variation of g_m , with (dotted) the curve of an ordinary pentode for comparison.

The reason for making such a valve is that radio receivers have to handle without distortion signals of very different magnitude. At one moment a receiver may be required to deal with an input signal of 1 microvolt. Shortly afterwards the same receiver may be required to produce an equal audio or video response from an input signal of 100 millivolts. Moreover, the signal being received may fade or increase in an unpredictable manner, and it is not convenient to keep adjusting the gain control to give the same output. To get over this difficulty the variable-mu pentode was produced. When the input signal is small, the valve operates with a small negative grid bias and gives a large gain: when the signal is large, part of the rectified output from the detector is used to bias the grid of the valve to a more negative value. This auto-

matically reduces μ and the receiver gain, so that the signal amplifier settles down at a different average negative bias. 'Automatic gain control', as the device is called, is an important feature of all modern receivers, and is especially important in Service applications, where the input signals may vary over a range of a million to one.

The method of achieving the slow variation of μ with the grid voltage is ingenious. The control grid of the valve is wound so that the wires are close together near the ends of the electrode structure, but are much further apart at the middle. Fig. 4.16 represents an axial section of a cylindrical valve, showing the relative positions of the cross-sections of the wires forming the

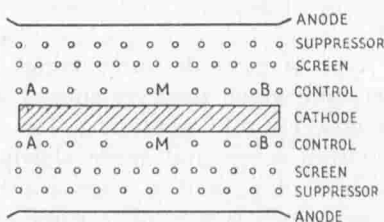


Fig. 4.16. Axial section of a variable-mu pentode, showing the variable pitch in the control grid.

electrodes. The control grid wires are very close at the ends A and B, but are far apart at M. Hence, as the grid is made negative, the cathode current shuts down quickly at A and B and more slowly towards M. With good design a very uniform variation in μ and g_m and a very long 'tail' to the I_a-V_g characteristic may be obtained. The name 'remote cut-off' is descriptive of the fact that in the variable-mu valve a very large negative bias is required to reduce the anode current to zero.

Variable-mu valves have a certain disadvantage as signal amplifiers. When they are operated at maximum sensitivity, the anode current is about twice what it would be with a corresponding sharp cut-off pentode. This additional anode current is seldom serious in itself, but, as will be seen, in Chapter 12, the 'noise' produced by a valve is almost directly proportional to the current I_a and so variable-mu valves are 'noisy' compared with their sharp cut-off counterparts. For this reason, the variable-mu valve is not often used in the first stage of a receiver, where its noise would have most effect on the receiver performance.

The Valve as a Distorter

We are now in a position to consider valves both as distorters and as amplifiers, and it must immediately be stated that distortion of the input signal is not necessarily an evil. The electronic engineer makes as much use of the distorting action of the valve as he does of the amplifying action. To mention important examples, it is because the valve is not a perfect amplifier that it makes an ideal device for the generation of audio- and radio-frequency currents; it is because the valve distorts the input signal that it can be used to detect the modulation in a radio-frequency wave; and it is the distorting action of the valve which enables us to use it to change the frequency of the incoming signal in a superheterodyne receiver.

Distortion due to the Curvature of the I_a-V_{g1} Characteristics

Consider the I_a - V_{g1} characteristics of a triode, beam tetrode or pentode as represented by curve A of Fig. 5-1. The curvature shown is typical of the average valve. In Chapters 3 and 4 it was assumed that, when the standing potential of the grid (the grid bias) was properly adjusted, a small signal δV_g would produce a proportional change in the anode current δI_a . This would be the case for *all* values of δV_g , if I_a were given by the equation

$$I_a = a + bV_g \quad (5.1)$$

This is the equation representing a straight line, and so it is usual to speak of it as a linear equation. The anode current, according to (5.1), varies linearly with the grid voltage. But we know from Fig. 5.1 that this is not actually true. In fact curve A may be represented much more accurately by the equation

$$I_a = C(V_g + V_a/\mu)^{3/2} \quad (3.5)$$

already given in Chapter 3.

Curve B of Fig. 6.1 is the tangent to curve A at $V_g = 0$. It is an example of a linear relationship of the form

$$I_a = a + bV_g$$

and is a first approximation to the true curve A.

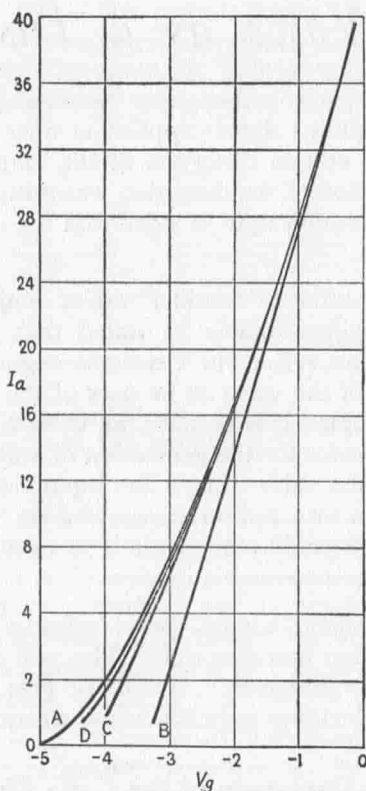


Fig. 5.1. Anode-current-grid-voltage characteristic of a valve and curves of simple mathematical types approximating to it.

Curve C is derived from curve B by adding another term, cV_g^2 , so that curve C is

$$I_a = a + bV_g + cV_g^2$$

Curve D which is almost indistinguishable from curve A for all values of V_g greater than -3 volts is

$$I_a = a + bV_g + cV_g^2 + dV_g^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.2)$$

By adding further terms in V_g^4 , V_g^5 , etc, even better approximations to curve A may be obtained. Equation (5.2), however, is sufficient for many purposes. This algebraic method of approximating to an actual valve characteristic is very useful. Examples will be found in what follows, and it will be seen that distortions can be classified as second order, third order, etc, according to the index of V_g in equation (5.2).

► For those who can use the binomial theorem, the derivation of equation (5.2) may be demonstrated as follows:—

The current flowing from the cathode of a triode is given by I_k , where $I_k = C(V_g + V_a/\mu)^{3/2}$, and if V_a is large and constant, and V_g is negative, then $I_k = I_a$, and the current may be regarded as a function of V_g only.

$$\begin{aligned} \text{Then } I_a &= C(V_a/\mu)^{3/2} (1 + \mu V_g/V_a)^{3/2} \\ &= C \left(\frac{V_a}{\mu}\right)^{3/2} \left[1 + \frac{3\mu V_g}{2V_a} + \frac{3}{8} \left(\frac{\mu V_g}{V_a}\right)^2 - \frac{1}{16} \left(\frac{\mu V_g}{V_a}\right)^3 + \dots \right] \\ &= a + bV_g + cV_g^2 + dV_g^3 \dots \dots \dots (5.3) \end{aligned}$$

Curve B of Fig. 5.1 is represented by the terms 1 and 2 on the right-hand side of the equation, curve C by terms 1, 2 and 3, and curve D by 1, 2, 3 and 4. It will be noted that

$$g_m = \frac{\delta I_a}{\delta V_g} = C \left(\frac{V_a}{\mu}\right)^{3/2} \left[\frac{3\mu}{2V_a} + \frac{3}{4} \frac{\mu^2 V_g}{V_a^2} - \frac{3}{16} \frac{\mu^3 V_g^2}{V_a^3} - \dots \right]$$

and, to a first approximation

$$g_m = \frac{3}{2} C \left(\frac{V_a}{\mu}\right)^{1/2} \dots \dots \dots (5.4) \blacktriangleleft$$

Effect of Valve Curvature on a Sine-wave Signal

It is now necessary to examine the effect produced by the curvature of the I_a - V_g characteristic on certain signals applied to the grid of a valve, and of these by far the most important signal is an alternating voltage. Suppose that the valve characteristic is represented by the curve A of Fig. 5.2 and that a steady voltage $-V_{GB}$, usually called the *standing bias*, is applied to the grid. Then, if the alternating signal voltage is represented by the sine wave FGHIJK, the corresponding current can be found from the characteristic for every value of the signal voltage. In the absence of any impedance in the anode circuit the current variation is a distorted version of the original signal-voltage variation, and it is not difficult to see how the distortion occurs. The curvature of the characteristic has produced a smaller change of current for the voltage peaks HT and JV than for the peaks SG and UI, and so the signal wave variation has been flattened at N and Q and sharpened at M and P.

The most convenient means of expressing this distortion, a method which is essential to an understanding of the action of a valve on many circuits, is to imagine the curve LMNPQR as the sum of a number of sine or cosine waves of different frequencies and amplitudes. For example, if the voltage wave at the grid is of frequency f (i.e., it is $\sin 2\pi ft$) then the current wave may be regarded as the sum of component sine or cosine waves of frequencies f , $2f$, $3f$, etc. The first of these, the wave of frequency f , is called the *fundamental* or *first harmonic*. Were there no distortion, only the fundamental would appear. The wave of frequency $2f$ is called the *second harmonic*, that of frequency $3f$ the *third harmonic*, and so on. The resultant wave, LMNPQR, has then been resolved into its component harmonics. This is called *harmonic analysis*. It may be carried out graphically or algebraically.

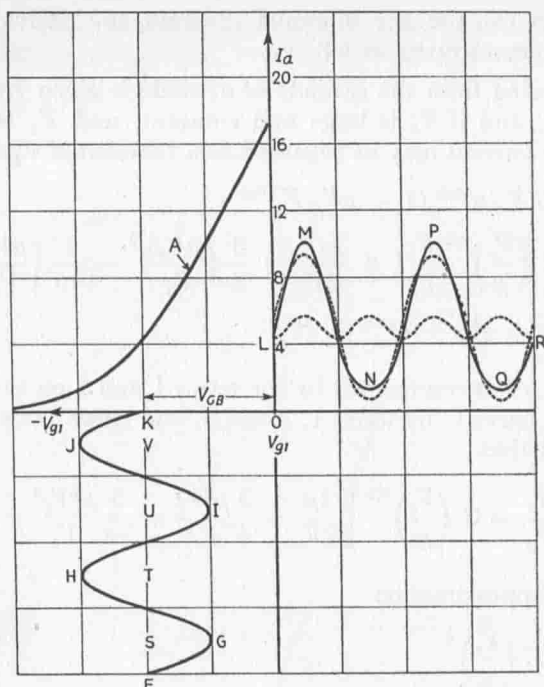


Fig. 5.2. Illustrating the production of second-harmonic distortion by the curvature of the $I_a - V_g$ characteristic.

In Fig. 5.2 a very good approximation to the resultant wave may be obtained by imagining it to be the sum of two waves only, one of the original frequency f and the other, of smaller amplitude, of frequency $2f$. These are shown by the dotted curves. The ratio of the amplitude of the second-harmonic component to the amplitude of the first-harmonic, or fundamental, component is a measure of the *second-harmonic distortion* due to the valve characteristic.

A more accurate examination of the resultant wave shows that it contains also components at frequencies $3f$, $4f$, etc, but that the amplitudes of these harmonics decrease rapidly as the order of harmonic increases, and they can often be neglected entirely. In short, the second-harmonic distortion is greater than the third which is itself greater than the fourth, and so on.

The last point to be noted is very important. If a valve $I_a - V_g$ characteristic is given by

$$I_a = a + bV_g + cV_g^2 + dV_g^3 + \dots \quad (5.2)$$

the second-harmonic distortion is mainly due to the V_g^2 term, the third-harmonic to V_g^3 , etc. To take a hypothetical case, if a valve could be constructed with a characteristic represented by $I_a = a + cV_g^2$, and the signal applied to its grid was at frequency f , the anode current would contain signals at frequencies f and $2f$ only. In practice it is impossible to obtain such a characteristic.

Summing up, the effect of the curvature of the $I_a - V_g$ characteristic is to produce in the anode current components at the grid signal frequency and at all the harmonics of the signal frequency. Since a is greater than b , and b is

greater than c , and c is greater than d , etc, the components decrease rapidly in amplitude as the index of the harmonic increases, but the harmonics are all there and can be emphasized or diminished by the use of particular circuits.

► For those who understand a little trigonometry the following treatment of the subject will probably be more satisfactory. Suppose that a signal $v_g = A \sin 2\pi ft$, (i.e., an alternating voltage of frequency f), is introduced at the grid of the valve. If the anode potential does not vary with the anode current, and if the grid bias is V_{GB} , then the anode current, as shown by equation (5.2), will be given by

$$I_a = a + b(V_{GB} + A \sin \theta) + c(V_{GB} + A \sin \theta)^2 + d(V_{GB} + A \sin \theta)^3, \\ \text{where } \theta = 2\pi ft,$$

or, rewriting,

$$I_a = a + bV_{GB} + cV_{GB}^2 + dV_{GB}^3 \text{ (a constant, or d.c. component)} \\ + bA \sin \theta + 2cAV_{GB} \sin \theta + 3dAV_{GB}^2 \sin \theta \text{ (the amplified component)} \\ + cA^2 \sin^2 \theta + 3dA^2 V_{GB} \sin^2 \theta \\ + dA^3 \sin^3 \theta \quad \left. \vphantom{I_a} \right\} \text{ (a series of distortion terms)}$$

But, $\sin^2 \theta = \frac{1}{2} (1 - \cos 2\theta)$, and $\sin^3 \theta = \frac{3}{4} \sin \theta - \frac{1}{4} \sin 3\theta$.

Therefore,

$$I_a = a + bV_{GB} + cV_{GB}^2 + dV_{GB}^3 + \frac{1}{2}cA^2 + \frac{3}{4}dA^2 V_{GB} \quad \text{(d.c.)} \\ + A \sin \theta (b + 2cV_{GB} + 3dV_{GB}^2 + \frac{3}{4}dA^2) \quad \text{(amplified signal)} \\ - \frac{1}{2}A^2 \cos 2\theta (C + 3dV_{GB}) \quad \text{(2nd harmonic)} \\ - \frac{1}{4}dA^3 \sin 3\theta \quad \text{(3rd harmonic)} \quad \left. \vphantom{I_a} \right\} (5.5)$$

More Exact Theory of Triode Amplifier

To illustrate the effect of the distortion introduced by the curvature of the valve characteristics, consider the case of an amplifier with a resistive load handling a reasonably large signal. Fig. 5.3 shows the amplifier circuit. The signal is an alternating potential of amplitude 30 volts and frequency f applied to a triode valve, the negative grid bias of which is 30 volts. The anode load is 5,000 ohms and the anode supply voltage is 300. Fig. 5.4 is a set of V_g - I_a characteristics for the valve, the curves being drawn for anode voltages of 300, 250, 200, 150, 100 and 50 volts.

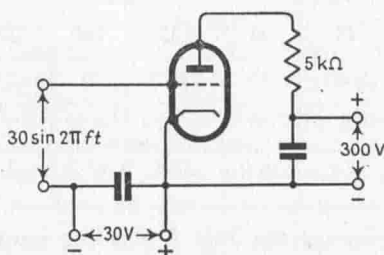


Fig. 5.3. Simple triode amplifier circuit.

To find the output signal the best method is to employ what is known as the *dynamic characteristic* of the valve and the anode load taken together. This is the curve which gives the anode current to be expected at each value of the voltage, drawn across the V_g - I_a characteristics, and it is found in the case under discussion as follows:—

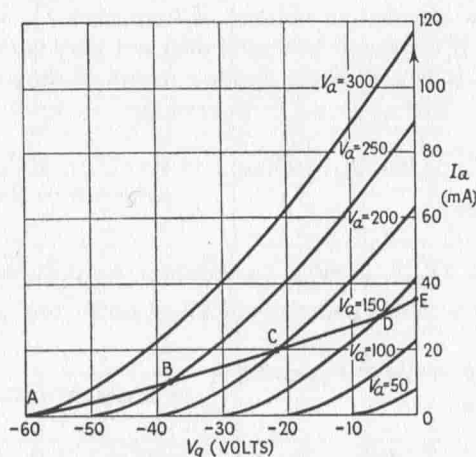


Fig. 5.4. Static and dynamic characteristics of a triode.

The valve current will be zero at an anode potential of 300 volts and a grid potential of -60 volts. Since the current is zero there is no drop in the anode resistor and therefore point A in Fig. 5.4 is a point on the dynamic characteristic. Similarly, at an anode potential of 250 volts, the potential drop in the anode resistor must be 50 volts, corresponding to a current of $(50/5000)$ A or 10 mA. Therefore, the point B ($V_a = 250$ V, $I_a = 10$ mA) must lie on the dynamic characteristic. Similarly, points C, D, E can be found. The dynamic characteristic is not a straight line. It would only be straight if the valve static characteristics were straight.

The output signal can now be deduced from the input signal by a graphical method. Let $\theta = 2\pi ft$, then $V_g = -30 + 30 \sin \theta$. Therefore V_g may be written down for representative values of θ . From Fig. 5.4 the corresponding values of I_a can then be read off, since curve ABCDE shows the current flowing in the anode load for all values of V_g . This has been done in Table 5.1.

TABLE 5.1

θ°	0	45	90	135	180	225	270	315	360
$30 \sin \theta$	0	21.21	30	21.21	0	-21.21	-30	-21.21	0
$-30 + 30 \sin \theta$	-30	-8.79	0	-8.79	-30	-51.21	-60	-51.21	0
I_a (mA) from Fig 5.4	15.2	27.6	34.0	27.6	15.2	3.5	0	3.5	15.2

The full line of the graph in Fig. 5.5 is the anode current. It will be noted that there is marked distortion. The positive half cycle is 'peaky' and the negative half cycle is 'flat' compared with a sine wave.

Neglecting third and higher harmonic components in the anode current I_a we may write

$$I_a = \text{d.c. component} + I_f \sin 2\pi ft - I_{2f} \cos 4\pi ft \quad . \quad . \quad (5.6)$$

where I_f and I_{2f} are the amplitudes of the fundamental and second harmonic, or, using θ for $2\pi ft$.

$$I_a = \text{d.c. component} + I_f \sin \theta - I_{2f} \cos 2\theta \quad . \quad . \quad (5.7)$$

In Fig. 5.6 the variable parts of the equation (5.7) have been graphed for the arbitrary value of $I_{2f} = I_f/4$.

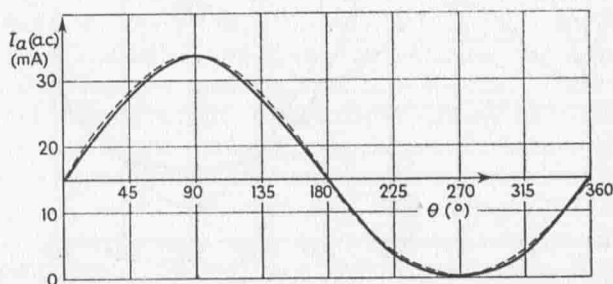


Fig. 5.5. A.C. component of anode current, showing distortion. The dotted curve is a theoretical one with second-harmonic distortion only.

Curve A is $I_f \sin \theta$ and curve B is $-(I_f/4) \cos 2\theta$. Adding these we have the distorted signal (curve C) and it will be noted that this curve is similar to the full line curve of Fig. 5.5 which was found from the dynamic characteristic. Indeed, as the reader may easily verify, the curve of Fig. 5.5 is almost exactly given by

$$I_a = 16.10(\text{d.c.}) + 17.0 \sin \theta - 0.9 \cos 2\theta \quad . \quad . \quad (5.8)$$

This equation is represented by the dotted curve of Fig. 5.5. Hence the distortion of the output signal is accounted for very well by the presence of the 2nd harmonic of the signal frequency, and if we define the distortion by the ratio of the amplitude of the 2nd-harmonic component to the amplitude of the fundamental component, the distortion is $0.9 \times 100/17.5 = 5.2\%$. To obtain the *exact* curve of Fig. 5.5 account would have to be taken of the 3rd and other harmonics. In certain cases 3rd-harmonic distortion can be considerable.

In the amplifier valve just considered the amplitude of the input signal was taken to be one-half the negative grid voltage necessary to cut off the anode current at the given value of the h.t. In Chapter 3 it was shown that if the h.t. voltage is V_{HT} and the amplification factor of the valve is μ , the cut-off voltage for the grid is V_{HT}/μ . Therefore, in the case under discussion, the signal amplitude is $V_{HT}/2\mu$. Under these circumstances when the signal reaches its maximum negative value, the anode current is zero, and when the signal reaches its maximum positive value, the grid voltage is zero. If the grid is allowed to become positive with respect to the cathode, grid current will flow. This produces distortion of a different kind which will be discussed later. If the grid becomes more negative than the cut-off value ($-V_{HT}/\mu$), the waveform of the output signal will obviously be very seriously distorted. It follows,

therefore, that in the amplifier discussed the signal is the greatest which can be handled by a single valve if relatively distortionless amplification of a signal of complex waveform is required. It is the extreme case of a Class A amplifier, as that was defined in Chapter 3. Even in this case, as we have seen, a 5% second-harmonic distortion is present for the particular valve considered.

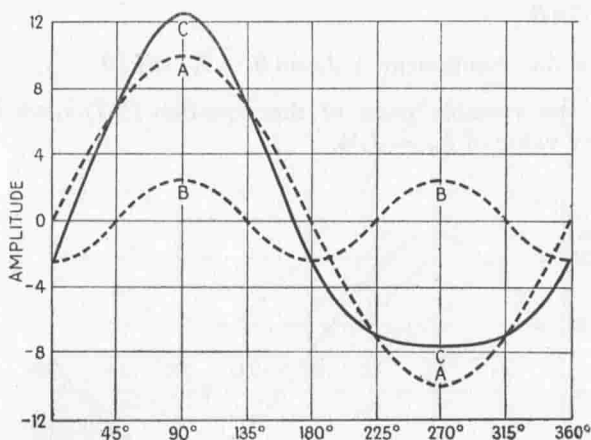


Fig. 5.6. Curves showing resolution of a distorted wave (C) into its fundamental (A) and second-harmonic (B) components.

It must not be imagined, however, that valves cannot handle larger signals. This can be done in two ways. First, we may use two valves working together in 'push-pull'. Such a combination greatly reduces second-harmonic distortion. Secondly, where the signal is sinusoidal (i.e., contains only one frequency) or consists of a group of signals of almost equal frequencies, advantage can be taken of this fact to nullify the distortion which the valve introduces by using suitable circuits. The first of these cases will now be considered. The second is a circuit matter and will be discussed in Volume 4.

Push-Pull Amplifier

Imagine two identical valves connected as shown in Fig. 5.7 so that the signal to be amplified can be introduced across $G_a G_b$. The anodes are connected together through two equal resistors R_1 .

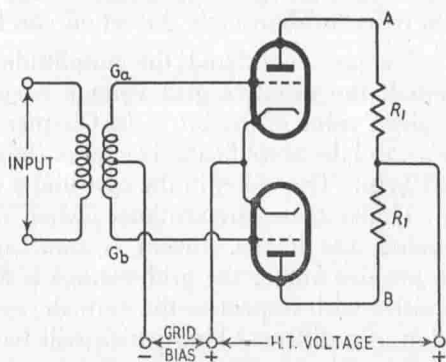


Fig. 5.7. Simple push-pull amplifier.

The two valve cathodes are connected directly together; the two grids have a common standing bias applied at the centre-tapping of the coil $G_a G_b$; the anodes have a common h.t. supplied at the mid-point of the resistance $2R_1$. Such a combination of two valves is known as a push-pull amplifier although, in practice, the anode load is seldom a resistor. It is either a tuned circuit (for radio-frequency amplification) or a transformer (for audio- or video-frequency amplification). The appropriate circuits and their effects are discussed in Volume 4. The class A push-pull amplifier greatly reduces second-harmonic distortion and, in the limit, can amplify signals of almost double the amplitude which could be handled by each valve separately. This cannot be done by using the same two valves in any other way.

To understand the operation of the class A push-pull amplifier first imagine that, as in the single-valve amplifier just discussed, the signal amplitude is $V_{HT}/2\mu$ and the common standing bias is $-V_{HT}/2\mu$. Then each valve will give a distorted output, but the voltage across AB will be virtually undistorted. The reason is that in the anode circuit the second-harmonic components of current flow in opposition to one another in the two resistances, while the fundamental components are in phase. The voltages developed across the resistances by these currents each comprises a fundamental and a second-harmonic component. Viewed across both resistances together, however, between the points AB, the fundamental components of voltage add, whereas the second-harmonic components cancel. A simple analysis shows this immediately.

▶ We have seen that the anode current of one valve is given by

$$I_a = a + bV_g + cV_g^2 + dV_g^3 + \dots \quad (5.2)$$

and that $I_a = a + bV_g + cV_g^2$ is a good approximation.

Therefore
$$I_a = a' + I_f \sin \theta - I_{2f} \cos 2\theta \quad (5.7)$$

where $\sin \theta = \sin 2\pi f t$, and I_f and I_{2f} are the amplitudes of the fundamental and second-harmonic components in the anode current respectively.

Then in the push-pull amplifier, if I_a = current in valve A and I'_a = current in valve B

$$I_a = a' + I_f \sin \theta - I_{2f} \cos 2\theta$$

and
$$I'_a = a' + I_f \sin (\theta + \pi) - I_{2f} \cos 2(\theta + \pi)$$

We write $\theta + \pi$ in the second equation since the valves act in the same way at opposite phases of the cycle.

Therefore, the potential at A is

$$V_a = V_{HT} - R_1(a' + I_f \sin \theta - I_{2f} \cos 2\theta)$$

and the potential at B is

$$V'_a = V_{HT} - R_1(a' - I_f \sin \theta - I_{2f} \cos 2\theta)$$

and
$$V'_a - V_a = 2R_1 I_f \sin \theta \quad (5.8)$$

or the potential difference across the two resistances is an undistorted amplified version of the signal. The second-harmonic terms have cancelled out.

The current taken from the h.t. supply is

$$I_a + I'_a = 2a' - 2I_{2f} \cos 2\theta \quad (5.9)$$

This contains the component at frequency $2f$. ◀

If an exact analysis of a perfect push-pull amplifier is made, it emerges that all the even-harmonic terms cancel, but that the odd harmonics remain. The push-pull amplifier, therefore, theoretically eliminates *even* harmonic distortion, but leaves the *odd* harmonics unaffected. Since the curvature of the valve characteristic gives rise mainly to second-harmonic distortion, the push-pull amplifier is of considerable value.

It should be mentioned in passing that the class B push-pull amplifier is the limiting case of this type. It was stated in Chapter 3 that the class B amplifier is one in which the standing grid bias is approximately equal to the grid cut-off voltage. In Fig. 5.8, two identical I_a - V_g characteristics have been drawn together in such a way as to represent a class B push-pull amplifier circuit.

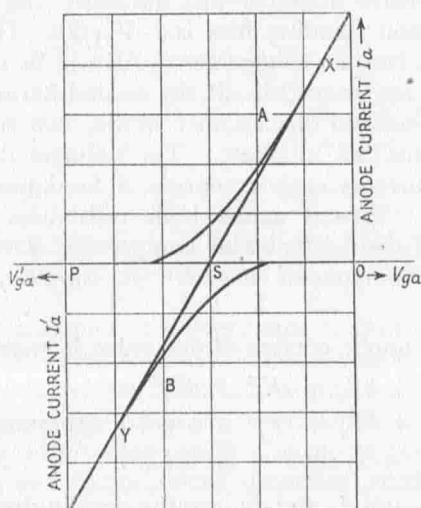


Fig. 5.8. Anode-current-gridvoltage characteristics of a push-pull amplifier, showing how distortion is reduced by this arrangement.

The anode current taken by valve A is I_a and that taken by valve B is I'_a . Since these flow in opposite directions in the anode circuit, the axes have been drawn in opposition. When the signal in the grid circuit is positive with respect to one valve, it is negative with respect to the other, and so the V_g axes have also been drawn in opposition. Let curves A and B represent static characteristics for each valve taken separately, and let the common standing bias be $OS = SP$. Then if a transformer or parallel resonant circuit is substituted for the resistances R_1 , of Fig. 5.7, a dynamic characteristic which is almost a straight line is obtained. It derives from the almost straight line XSY, obtained by adding algebraically I_a and I'_a . This class B amplifier can handle a signal of amplitude SO, without serious distortion. As in the case of the class A push-pull amplifier the second-harmonic distortion has been eliminated by the differential action of the two valves. A complete analysis of the class B amplifier demands a knowledge of the circuits attached to the valves. This will be given in Volume 4. At the moment it is sufficient to note that this is an extreme case of push-pull action.

The arrangement of two valves according to the circuit of Fig. 5.7 has many different uses. Such an arrangement can be used as a balanced modulator, a balanced mixer or a class C amplifier. It is, therefore, important to appreciate the fundamental action whereby the even harmonics, introduced by each valve, are cancelled by the combination.

Distortion due to Grid Current

So far it has been assumed that all the electrons leaving the valve cathode have found their way to the anode, and that the grid could be regarded simply as a control electrode to which or from which there was no conduction current. This is very nearly true if the anode is positive and the grid is negative, both with respect to the cathode. It is not absolutely true even then, but the grid current is certainly small enough under these circumstances to be neglected.

If, however, the grid becomes positive relative to the cathode for any interval of time, however short, a certain electron current will flow from the cathode to it, and this grid current may cause severe distortion of the amplified signal. There are two causes for this distortion. First, since current flows to the grid, the anode current is less than is to be expected from the triode equation

$$I_a = C(V_g + V_a/\mu)^{3/2}$$

Indeed, if the grid current is I_g and the current from the cathode is I_k , the equation is more properly written

$$I_k = I_a + I_g = C(V_g + V_a/\mu)^{3/2}$$

This flow of current to the grid distorts the shape of the I_a - V_g characteristic when V_g is positive, but the effect on the amplifier is not usually serious. Secondly, the flow of current to the grid decreases the input resistance across the grid-cathode terminals, so that the valve is now a device which requires a finite power to actuate it. This variation in the input impedance can cause serious distortion, unless it is corrected. A valve which has effectively an infinite input resistance for negative grid voltages may have an input resistance of only a few thousand ohms at a small positive grid voltage. This second cause of distortion can be overcome to some extent by the use of low impedance input circuits, and good use can then be made of the flow of grid current to provide automatic grid bias.

The grid current flowing in a valve is determined, as we might expect, by two parameters, V_g and V_a . The curves of Fig. 5.9 exhibit typical characteristics. Fig. 5.9(a) shows how the current varies with the grid voltage for three given anode voltages. The curves are not of the form $I_g = A(V_a + B)^{3/2}$ because of secondary emission. The primary grid current in the region delineated is greater than the net current shown, since most of the secondary electrons liberated when V_g is greater than 10 volts are attracted to the anode.

Fig. 5.9(b) shows how the current varies with the anode potential, for certain positive grid voltages. The form of these curves is important, for it means that the fraction of the total current going to the grid increases as the anode voltage decreases. This is a secondary cause of distortion when the anode voltage swing is large. The grid circuit is more heavily damped as the anode voltage swings about its smallest value.

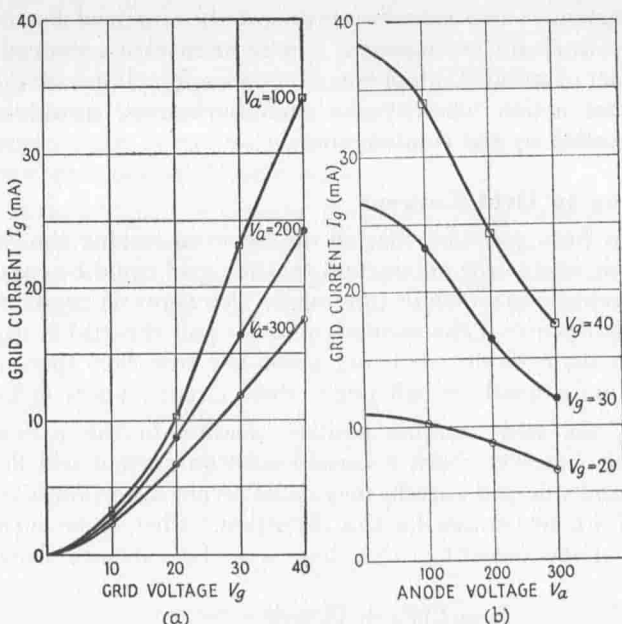


Fig. 5.9. Grid-current characteristics; (a) grid current versus grid voltage and (b) grid current versus anode voltage.

Automatic Grid Bias

Since automatic bias is universally used in receivers and is used in all but the largest transmitters, and since it depends in one form on the utilization of grid current, it is worth while at this point explaining the possibilities.

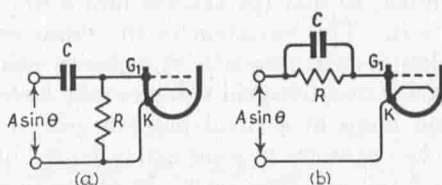


Fig. 5.10. Two methods of obtaining automatic grid bias by utilizing grid current.

Imagine a valve of any kind in which a signal $A \sin \theta$ may be applied via a large capacitor C to the control grid G_1 , as shown in Fig. 5.10(a). If there is a resistor R connected directly between grid and cathode, then in the absence of a signal, but with the h.t. voltage applied, electrons will flow in the valve from the cathode to the grid. This current flowing through R will tend to bias the grid negatively, but it obviously cannot bias it beyond the point of grid-current cut-off, for the grid current would then be zero and the potential difference across R would vanish. Therefore, in the absence of a signal G_1 will adjust itself automatically to about the same potential as the cathode.

When the signal is applied, the potential of G_1 will swing positive and negative. During the positive half-cycle the grid will take current and during

the negative half-cycle it will not. If the capacitor C were very small (which it is not), the grid would then vary in potential as shown in Fig. 5.11(b), where (a) is the corresponding representation of $\sin \theta$.

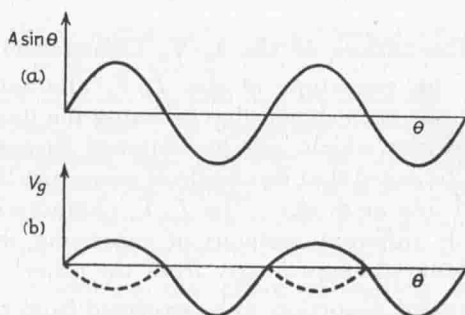


Fig. 5.11. Showing how the potential of the grid would vary in the circuit of Fig. 5.10 (a), if a signal $A \sin \theta$ were applied to it and C were negligibly small.

The distortion at the peak of the positive half cycle is due to the negative voltage produced by the grid current and shown by the dotted curve. The changing impedance of the circuit might produce further distortion. Increasing the capacitance changes this picture considerably. First, the flow of current to the grid no longer flows directly through R . It must also charge the capacitor C . Also, after the flow of current to G_1 has ceased, the capacitor has still to discharge through the resistor. In fact the capacitor smooths the potential at G_1 due to the grid current, and by suitably choosing the time constant of the CR circuit, the curves of Fig. 5.11 can be transformed to those of Fig. 5.12.

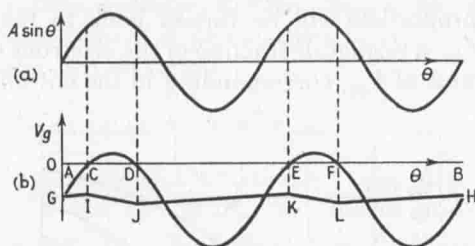


Fig. 5.12. When the time constant CR of Fig. 5.10 has a suitable value the mean grid potential takes up a negative value such as that shown by the curve $G I J K L$.

Here curve (a) is again the reference sine wave. In curve (b) the line AB represents zero grid voltage. The grid conducts during the intervals CD or EF , but the effect of the capacitor is to level off the voltage due to grid current to the approximate curve $G I J K L$. During the interval IJ , the capacitor is being charged. During the period JK it is being discharged. The time constant depends upon the product CR as shown in Volume 1. There is still a certain distortion of the waveform, but this can be made almost negligible by a proper choice of C and R . This form of automatic bias is used for many purposes, particularly where signals of varying amplitude have to be handled.

Within limits the valve is self-adjusting. An increase in signal amplitude produces a corresponding increase in the average negative grid bias. The circuit of Fig. 5.10(b) is an alternative to that of 5.10(a) which is commonly used.

Distortion due to Curvature of the I_a - V_a Characteristic

Distortion due to the curvature of the I_a - V_g characteristic has already been discussed, and it has been shown that by using the dynamic characteristic of the valve the distortion which will be obtained in specific cases may be calculated. It should be noted that this mode of procedure is only an alternative to the use of the load line method. The I_a - V_a characteristics and the I_a - V_g characteristics are only different methods of expressing the same thing. In fact one set can be deduced immediately from the other.

So there is no *additional* distortion to be expected from the curvature of the I_a - V_a characteristics in the simple amplifier, but all the effects due to the non-linear variation of I_a with V_g are mirrored in the non-linear variation of I_a with V_a . One additional effect deserves notice, however. At low anode voltages the grid is at its most positive value. Should the positive grid voltage exceed the minimum anode voltage, very severe distortion is to be expected. It cannot be calculated except in conjunction with the parameters of the external circuit and therefore will not be further considered here.

Distortion due to the Use of Two Control Grids

Let us consider the pentode valve, but on this occasion let the potential of the suppressor grid V_{g3} be deliberately varied. Starting from $V_{g3} = 0$ relative to the cathode, as V_{g3} is made negative fewer electrons will reach the anode. Having passed through the screen grid at high speed, they will lose their energy as they approach the suppressor grid and, as V_{g3} becomes more negative, a larger proportion will be turned back to the screen grid. At a particular value of V_{g3} , a negligible fraction of the electrons will reach the anode. This is the cut-off value of V_{g3} , corresponding to the cut-off value of V_{g1} .

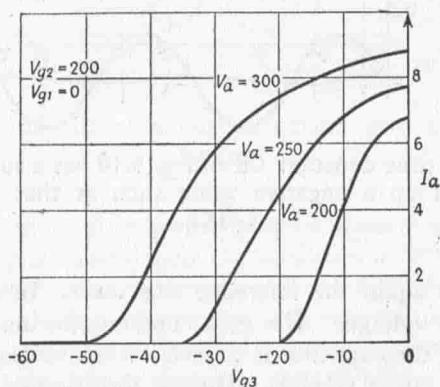


Fig. 5.13. Suppressor-grid characteristics in a pentode. Anode current versus suppressor-grid voltage.

A typical set of suppressor-grid characteristics is shown in Fig. 5.13 for various anode potentials. As might be expected from what has just been

said, the current is zero for a given value of V_{g3} and increases fairly rapidly as V_{g3} tends to zero, but levels off as V_{g3} becomes positive, since at this voltage all the electrons passing through the screen-grid G_2 are able to reach the anode.

The effect on the control-grid characteristics of making V_{g3} negative follows from the discussion just given. The cathode current I_k flowing through the valve is hardly affected by V_{g3} , and so if we draw the I_k - V_{g1} characteristics, they too are unaffected. But I_a is considerably reduced, since much of the current goes to the screen, and therefore the I_a - V_{g1} characteristics reflect the change in V_{g3} . The curves remain in the same relation to each other as before, but the anode currents are all reduced by a common factor. This is equivalent to a reduction in the mutual conductance g_m .

Another way of describing the effect of varying V_{g3} which is of great value in the applications to be given in later chapters is as follows:— First suppose that the I_a - V_{g3} characteristics are straight lines, as indeed they may be considered to be over a certain range. We know that I_k is approximately constant. Therefore

$$I_a/I_k = a' + b' V_{g3}$$

where a' and b' are arbitrary constants.

Suppose, too, that the I_k - V_{g1} characteristics are straight lines. This again is true over a considerable range of values. Then

$$I_k = a + bV_{g1}$$

The first equation expresses the fact that only a certain proportion of the cathode current arrives at the anode and that the proportion depends upon V_{g3} . The second equation expresses the fact that, if the screen-grid potential is constant, the cathode current depends upon V_{g1} .

Substitute for I_k in the first equation.

$$\text{Therefore } I_a = (a + bV_{g1}) (a' + b' V_{g3}) \quad . \quad . \quad . \quad (5.12)$$

and this equation expresses mathematically to a first approximation the simultaneous actions of the two grids, the control grid and the suppressor grid. In fact the equation has a more general significance, for it may be shown that it applies to any two grids in a valve, although the ranges over which it holds may be quite different. It shows that if a signal is applied to one grid and another signal to a second, the anode current will give a distorted reproduction of both, for

$$I_a = aa' + a'bV_{g1} + ab'V_{g3} + bb'V_{g1}V_{g3},$$

and of the terms on the right-hand side,

$$aa' = \text{d.c.}$$

$$a'bV_{g1} = \text{a reproduction of the signal at } G_1$$

$$ab'V_{g3} = \text{a reproduction of the signal at } G_3$$

$$\text{and } bb'V_{g1}V_{g3} = \text{a distorted signal due to both.}$$

The equation (5.12) is worth examining in more detail. As an example, let us take the case of the valve CV1091. Its I_a - V_{g1} characteristic is shown in Fig. 5.14.

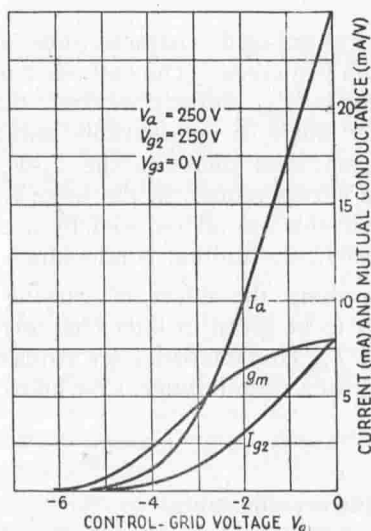


Fig. 5.14. Anode current versus control-grid voltage in CV1091.

Over the straight part (from $V_{g1} = -2$ to $V_{g1} = 0$) the current I_a is given by

$$I_a = 25 + 7.5 V_{g1} \quad (5.13)$$

with $V_a = V_{g2} = 250$ V and $V_{g3} =$ cathode potential. I_a is in mA and V_{g1} is in volts.

The I_a - V_{g3} characteristics of the valve are shown in Fig. 5.15.

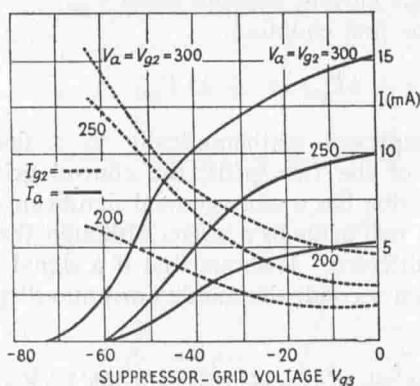


Fig. 5.15. Anode current versus suppressor-grid voltage in CV1091. for $V_{g1} = -2.0$ V.

Over the straight part of the characteristic for $V_a = V_{g2} = 250$ V and $V_{g1} = -2$ V (i.e., from $V_{g3} = -60$ V to $V_{g3} = -30$ V),

$$I_a = 15 + 0.25 V_{g3} \quad (5.14)$$

with I_a again in mA and V_{g3} in volts.

Combining these

$$I_a = 25(1 + 0.3V_{g1})(1 + 0.016V_{g3}) \quad (5.15)$$

This is the equation showing how the anode current will vary for small changes in V_{g1} about $V_{g1} = -1$ volt and for changes in V_{g3} about $V_{g3} = -45$ volts.

Another way of looking at the matter is that changes in the suppressor grid voltage produce changes in the slope of the I_a - V_{g1} curve; i.e., changes in g_m . From equation (5.15)

$$g_m = 25 \times 0.3V_{g1}(1 + 0.016V_{g3})$$

i.e., for $V_{g3} = -30$ V, $g_m = 3.8$ and for $V_{g3} = -60$ V, $g_m = 0$, g_m being expressed in mA/V. As has been mentioned, although equation (5.12) has been developed specifically for a pentode in which both control-grid and suppressor-grid voltages are varied, it applies equally to any valve in which small changes are made in the potentials of two grids.

The general equation, taking account of curvature in the characteristic, is

$$I = (a + bV_{g1} + cV_{g1}^2 + dV_{g1}^3 + \dots)(a' + b'V_{g3} + c'V_{g3}^2 + d'V_{g3}^3 + \dots) \quad (5.16)$$

The importance of the effect, as will be seen in the following chapters, lies in its application to frequency-changers, modulators and electronic switches.

An interesting point about the variation of the suppressor-grid voltage in the pentode is that a *decrease* in V_{g3} is accompanied by an *increase* in I_{g2} as shown by the dotted curves of Fig. 5.15. If, therefore, G_2 is connected to + h.t. via a resistance or a resistive impedance, and the screen-to-cathode by-passing capacitor is omitted, the potential of G_2 will follow the potential of G_3 or, putting it another way, the triode comprising cathode, suppressor grid and screen grid forms an amplifier in which the output is in phase with the input. This is of importance in the action of the transitron oscillator, discussed in Volume 4.

Transmitting Pentodes

Finally, before leaving the subject of suppressor-grid action, the effect of raising the suppressor grid to a positive voltage with respect to the cathode should be mentioned.

In a transmitting pentode, where the current is large and the space charge near the suppressor grid is large in consequence, the bend in the I_a - V_a characteristic for positive values of V_{g1} is not as sharp as is desirable. This is shown by the full line of Fig. 5.16, drawn for $V_{g1} = V_{g2}/4$.

The reason for the slow increase in I_a is that the electrons crowd together near the suppressor grid, where their velocities are small, and all but the most energetic are deflected in the direction of the screen until V_a is larger than V_{g2} . If, however, the potential of the suppressor grid is raised to about $V_{g2}/10$, the virtual cathode near the suppressor disappears, and a considerably greater fraction of the electrons can reach the anode at low potentials of the latter.

This is shown qualitatively by the dotted curve of Fig. 5.16, drawn for a suppressor-grid voltage of $V_{g2}/10$. Hence in transmitting pentodes it is usual to maintain the suppressor grid at a small positive potential relative to the cathode, thus extending the flat part of the I_a - V_a characteristic.

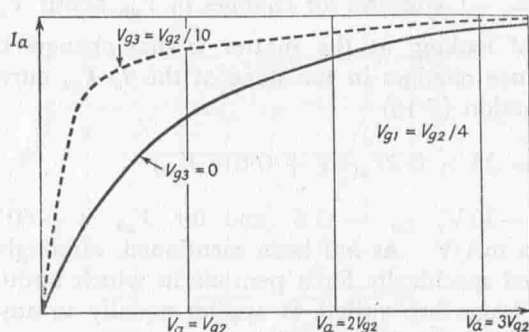


Fig. 5.16. Anode current versus anode voltage in a power pentode for a positive voltage on the control grid. Two curves are drawn for two values of suppressor-grid voltage.

The Class C Amplifier

Before leaving the subject of valve distortion, a most important case of a distorter-amplifier must be mentioned. The full discussion of it must be postponed to Volume 4, since it involves the consideration of circuits.

It has already been mentioned that amplifiers can be classified as A, B, C, according to the standing bias on the grid. Classes A and B have already been discussed. The class C amplifier, used to obtain radio-frequency power in all but the smallest transmitters, and for many other purposes, is one in which the grid is maintained at an average potential more negative than that required to cut off the anode current. Therefore, when a sinusoidal signal is applied to the grid, current flows in the valve for less than half of each cycle. This state of affairs is exemplified by Fig. 5.17, where the anode current is shown as a function of the angle θ , assuming the signal to be $A \sin \theta$, and the anode of the valve to be connected directly to the d.c. supply.

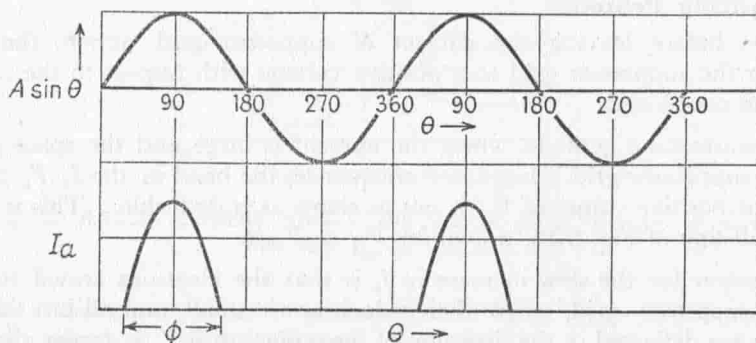


Fig. 5.17. Showing how the anode current varies with the phase of the signal voltage on the grid for a class C amplifier, the anode potential being kept constant.

If a circuit such as that shown in Fig. 5.18 is attached to the valve, the pulses of current to the anode excite the parallel circuit, which is tuned to the signal frequency, and the voltage fluctuations across it are almost sine waves in series with the constant supply voltage. Extending Fig. 5.17 to give the picture in this case, in Fig. 5.19 V_a is the anode voltage, I_a is the anode current and $A \sin \theta$ is the exciting voltage. All are drawn as functions of θ .

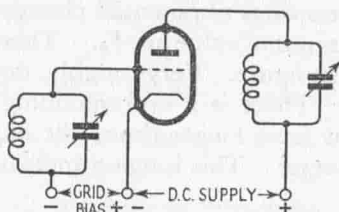


Fig. 5.18. Arrangement of circuits in a class C radio-frequency amplifier.

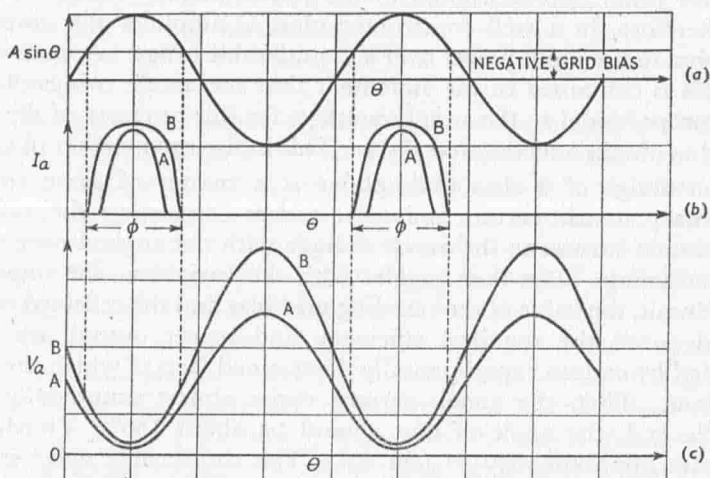


Fig. 5.19. Currents and voltages in a class C amplifier. (a) Signal voltage applied to grid, (b) pulses of anode current A and B for two different anode-supply voltages, and (c) the corresponding anode voltages for two different supply voltages.

It will be observed that when the anode current is a maximum the anode voltage is a minimum, and when the anode voltage is a maximum the anode current is zero. This effect reduces considerably the amount of heat generated at the anode, and very much increases the efficiency of the valve and its circuit as a machine for the generation of radio-frequency power. In fact, as may be shown by calculation, anode efficiencies of 90% are feasible, so that only 10% of the d.c. power entering the valve is dissipated as heat at the anode.

To be efficient a class C amplifier requires a considerable grid swing. For the amplifier to operate properly the grid must become positive over a small part of each cycle, and grid current then flows. The grid must not,

however, become more positive than the anode at any time for this would lead to excessive grid current and a kink in the anode-current curve. Since the anode is at its lowest potential when the grid is at its highest, this condition imposes a fundamental limitation on both swings. A very rough approximation, which is nevertheless useful, is that both grid and anode swing to one-tenth of the supply voltage on the anode.

Referring again to Fig. 5.19(c), if the anode supply voltage is changed from OA to OB, the anode swings of potential change to correspond, but there is little change in the minimum value of V_a . This is indicated by the two curves A and B drawn in the figure. Very roughly, doubling the supply voltage doubles the anode swing. There is a corresponding change in the pulses of anode current. These flow for a longer time—the angle of flow ϕ is greater—and their amplitudes are larger. This is again indicated in Fig. 5.19(b) by the two curves A and B.

Consequently, doubling the supply voltage roughly doubles the voltage swing on the anode and also doubles the mean anode current. But the input power (d.c.) is the (mean anode voltage) $\times$ (mean anode current), and the output power (a.c.) is proportional to the square of the voltage swing at the anode. Therefore, in a well constructed class C amplifier the output power is proportional to the input power over a considerable range. A more important aspect of this is contained in the statement that the anode swings of potential are nearly proportional to the supply voltage, for this property of the amplifier is employed to obtain a modulated wave. This is discussed further in Chapter 7.

The exact design of a class C amplifier is a matter of some complexity. It is customary to use certain universal curves connecting the ratio of the maximum anode current to the anode voltage with the angle ϕ over which the valve is conducting. One then juggles with the variables—the impedance of the anode circuit, the value of the standing grid bias and the reflected impedance of the load—until the required efficiency and power output are obtained. One is guided by certain experimentally determined facts of which the following are important. First, the anode current varies almost sinusoidally with the angle ϕ . Second, the angle of flow should be about 150° . Third, the grid current varies approximately as $(\sin \phi)^2$. The choice of a valve is made in terms of the required power output and the frequency of operation. The class C amplifier is considered in more detail in Volume 4.

Suppose that a valve has to be chosen to give a continuous output of 60 watts. Assuming an efficiency of 75% (which is very good in practice), the d.c. input will be 80 watts, or $60 \times 4/3$, and the valve anode must be capable of dissipating 20 watts. Such a valve would require a supply voltage in the neighbourhood of 500, giving a mean anode current at full load of $80 \times 1000/500 = 160$ mA. What is required, therefore, is a valve capable of giving 160 mA at an h.t. voltage of 500 with an anode capable of dissipating 20 watts. Allowing a reasonable factor of safety to ensure long life in the valve, the designer would probably look for a valve capable of giving 200 mA at 500 volts and with an anode which would dissipate 30 watts. It should be noted that a bigger valve would be required if the output had to be modulated.

Importance of the Distortion Formulae

In this chapter two valve equations have been derived which are of such far-reaching importance in radio that the reader is advised to make himself

thoroughly familiar with them and with their explanation. The first expresses the static anode current as a function of the potential of a control grid. It is equation (5.2) which is repeated here for reference

$$I_a = a + bV_g + cV_g^2 + dV_g^3 + \dots \quad (5.2)$$

The second equally important equation expresses the anode-current in a multi-electrode valve as a function of the potentials of two grids, the grids chosen in the derivation being G_1 and G_3 . This is equation (5.12) and it is also repeated here for reference.

$$I_a = (a + bV_{g1})(a' + b'V_{g3}) \dots \quad (5.12)$$

This second equation is actually an approximation to equation (5.16), but for very many purposes it is exact enough.

There is no new electronic mechanism introduced in the next two chapters (6 and 7). Both are concerned with important practical applications of the valve distortions characterized by the two equations just given. The equations may be regarded as short-hand expressions of the effects described in this chapter, and the equations will be used to exhibit the phenomena of frequency-changing, modulation and demodulation.

Oscillators

Before proceeding to Chapter 6, the reader should have some idea of what an oscillator is, for the term is frequently used in this and later chapters. An oscillator is a device which produces a regularly-recurrent waveform of voltage or current and it comprises two basic elements—an amplifier and a frequency-determining circuit. The types of oscillator and the kinds of waveform which they produce vary enormously and, as they are matters of circuitry rather than of electronics, they are dealt with in Volume 4.

The most common oscillator, however, produces an output waveform which approaches the sinusoidal very closely and comprises a valve and a tuned circuit. An amplifier valve with a parallel tuned circuit in its anode lead will give an amplification A at the resonance frequency of the tuned circuit and, at this frequency, the anode voltage will be in phase opposition to the grid voltage. If a fraction $1/A$ of the anode voltage is taken, reversed in phase, and fed to the grid, the grid voltage is obtained from the anode voltage. The attenuation and phase reversal can be obtained by including a small coil in the grid circuit and coupling it to the anode coil.

With such a circuit, oscillation starts up at a small amplitude and builds up until the non-linearities of the valve limit its growth. These non-linearities play an essential part and one of the more important of them comes into action through the use of an automatic grid-bias circuit of the type shown in Fig. 5.10 (p. 78). As the amplitude of oscillation increases, so does the grid voltage and hence the grid bias. The increasing bias reduces the mutual conductance of the valve and so the amplification, and an equilibrium condition is at length reached at which a steady amplitude of oscillation is generated.

All types of valve can be used as oscillators, but the triode is most commonly employed, although in some circumstances a pentode is advantageous.

CHAPTER 6

Mixing and Frequency-Changing Valves

ONE OF THE most important applications of the thermionic valve as a distorter or non-linear circuit element is in the transformation of a signal at one frequency to a similar signal at a different frequency. This process, commonly called frequency changing, is widely employed in telecommunications and is particularly important in most modern receivers. The latter are known as supersonic-heterodyne receivers or, more shortly, as 'superheterodynes'. A superheterodyne receiver is one in which the sequence of events as the signal proceeds through the receiver is (a) amplification, (b) transformation of the signal to a fixed 'intermediate frequency' which is usually lower than that of the incoming signal, (c) further considerable amplification at the intermediate frequency, (d) demodulation and (e) further amplification of the resultant audio or video signal. The reasons for this elaborate procedure and the detailed analysis of the circuits required will be given in Volume 4. It is the transformation of the incoming signal to a fixed intermediate frequency which is known as frequency-changing, and which must engage our special attention in this chapter. Special valves are made for the process, although, as will be seen, ordinary types can be used.

Frequency-changing implies mixing the signal voltage at frequency f_s with an oscillation of frequency f_o , generated in the receiver and known as the 'local oscillation', to obtain a new signal at the difference frequency, $f_o - f_s$. This difference frequency is what is known as the intermediate frequency. Frequency changing can only be performed by using a valve of one description or another. To obtain the new frequency conductors which obey Ohm's law cannot be employed. On the other hand, valves may be used in various ways which will be described in what follows.

Two-Electrode Mixing

As has been mentioned, valves may be used in different ways as frequency changers. One common method is to apply the signal and the local oscillation simultaneously to two electrodes of a valve. To understand the mixing action, imagine first a hypothetical valve in which there are two *control* grids between the cathode and the anode, as in Fig. 6.1. The diagram is not meant to represent a tetrode of the type described in Chapter 4, but a valve in which the potentials of both G_1 and G_2 can be deliberately varied. Clearly the electron current to A will depend on the potentials of both G_1 and G_2 , but will change more when a certain signal voltage is applied to G_1 than when the same voltage is applied to G_2 .

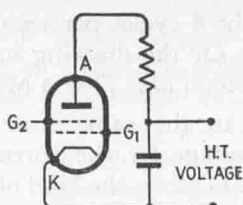


Fig. 6.1. A simple form of valve for two-electrode mixing.

If the signals are very small, then, to a first approximation, the effect at A will simply be one of superposition, and if A is connected to the cathode through a resistor and a large enough capacitor, small sine-wave voltages applied simultaneously to G_1 and G_2 will give amplified voltages at the same two frequencies across the anode resistor. But this is only an approximate theory. The current to A, as was shown in Chapter 5, varies with V_{g1} and V_{g2} according to the equation

$$I_a = (a_1 + b_1 V_{g1}) (a_2 + b_2 V_{g2}) \quad (6.1)$$

even when V_{g1} and V_{g2} are small, for the rate of change of I_a can be reduced to zero by applying a suitable negative voltage either to G_1 or to G_2 .

Therefore, if V_{g1} is a sine-wave voltage of frequency f_s and V_{g2} is a sine-wave voltage of frequency f_o , the resultant current contains components at frequency f_o and f_s , and a component due to the term $b_1 b_2 V_{g1} V_{g2}$ in equation (6.1); i.e., due to the product of the one voltage and the other. As was explained in Chapter 5, this is because the mutual conductance of the grid G_1 with respect to the anode depends upon the potential of G_2 .

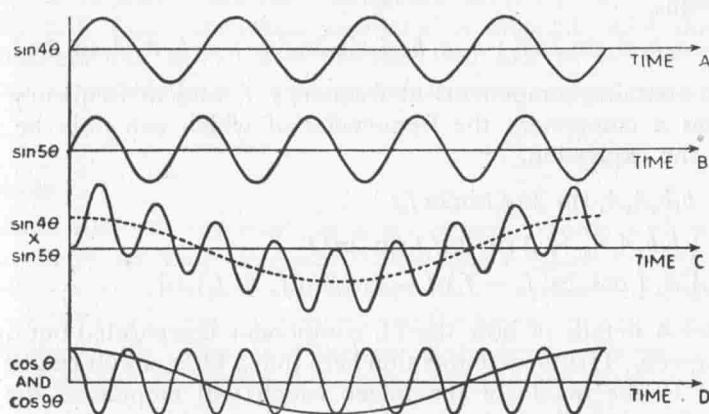


Fig. 6.2. Illustrating two-electrode mixing. A, the signal voltage on G_1 as a function of time; B, the local-oscillator voltage on G_2 ; C, the waveform produced by the product of the two voltages; D, the components of C drawn separately—the sum and difference terms.

Now consider Fig. 6.2. Curve A represents the variation with time of the signal voltage and curve B the variation with time of the local oscillation. Each drawing may be imagined to represent one second, in which case the

frequency of the signal would be 4 cycles per second, and the frequency of the oscillator 5 cycles per second. Or the drawing may be imagined to represent one-millionth of a second, in which case $f_s = 4 \text{ Mc/s}$ and $f_o = 5 \text{ Mc/s}$.

Curve C has been drawn to the same scale by multiplying the instantaneous value of the signal voltage by the corresponding value of the local oscillator voltage, so that curve C shows the kind of wave obtained from $V_{g1} V_{g2}$. It will be noted that curve C has two entirely different frequencies in it. One is high compared with f_s and f_o . In fact there are 9 cycles in the drawing. The other is low compared with f_s and f_o . There is only 1 cycle in the drawing. This is indicated by the dotted curve. If $f_s = 4$ and $f_o = 5$, the higher frequency in curve C is $f_o + f_s = 9$ and the lower, $f_o - f_s = 1$, showing that the two new frequencies produced by the valve action are the sum and difference frequencies. These are sometimes called the *heterodyne* or *beat* frequencies. In superheterodyne receivers $f_o - f_s$ is called the intermediate frequency (i.f.). The curves D are the components of curve C separated out. Curve C may be obtained from D by subtracting the instantaneous values of the one curve of D from the corresponding values of the other in D.

A final point should be noted. Since the intermediate frequency component is due to the term $b_1 b_2 V_{g1} V_{g2}$, its amplitude is proportional to the product of the amplitudes of the local-oscillator voltage and the signal voltage. Within limits, and subject to other modifying factors, the i.f. output varies directly with both the signal and local-oscillator voltages.

► The algebraic description of the valve action is as follows:—

If V_{g1} is a sine-wave of amplitude A_s and frequency f_s , and V_{g2} is a sine-wave of amplitude A_o and frequency f_o ,

$$I_a = (a_1 + b_1 A_s \sin 2\pi f_s t) (a_2 + b_2 A_o \sin 2\pi f_o t)$$

Expanding this,

$$I_a = a_1 a_2 + a_2 b_1 A_s \sin 2\pi f_s t + a_1 b_2 A_o \sin 2\pi f_o t + b_1 b_2 A_o A_s \sin 2\pi f_s t \sin 2\pi f_o t$$

I_a therefore contains components at frequency f_s and at frequency f_o , but it also contains a component the frequencies of which can only be found by simplifying the expression.

The term $b_1 b_2 A_o A_s \sin 2\pi f_s t \sin 2\pi f_o t$

is equal to $\frac{1}{2} b_1 b_2 A_s A_o \times 2 \sin 2\pi f_s t \sin 2\pi f_o t$

$$= \frac{1}{2} b_1 b_2 A_s A_o [\cos 2\pi(f_o - f_s) t - \cos 2\pi(f_o + f_s) t] \quad . \quad . \quad . \quad (6.2) \blacktriangleleft$$

In Volume 4 details of how the i.f. component is separated out and amplified will be given. It suffices to mention here that a filter circuit can be arranged in the lead to the anode of the mixer which will emphasize the required frequency. Our hypothetical valve theoretically would be a 'mixer' suitable for use in a superheterodyne receiver, or, in any circuit designed for frequency changing. Such a valve was, in fact, the forerunner of present-day mixers, but it suffers from several disadvantages which have rendered it obsolete.

The Hexode

The simplest practical form of mixer valve, but one which is not much used nowadays is the hexode, shown diagrammatically in Fig. 6.3. The valve has two control grids, G_1 and G_3 , but G_3 is shielded both from G_1 and

from the anode A by a pair of screen-grids G_2 and G_4 . These are connected in the usual manner, so that their common potential does not vary appreciably and is maintained at a positive value rather less than that of the anode. The voltage from the local oscillator is applied to G_1 , causing variations in the anode current at the oscillator frequency f_o . The signal to be transformed is applied to G_3 . This also varies the anode current. Owing to the combined effects of these two voltages, the total anode current contains a component at the frequency $f_o - f_s$. This is used to excite a parallel-tuned circuit resonant at the same frequency, and this intermediate frequency component is then amplified by succeeding stages in the receiver.

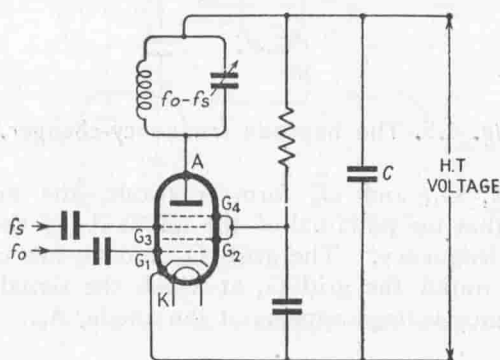


Fig. 6.3. A hexode mixer with a tuned anode circuit.

The presence of the almost complete screen round G_3 is intended to prevent any interaction (other than that transmitted by the electron current) between G_1 and G_3 . The local oscillator is isolated, and the presence of G_4 diminishes the capacitance between G_3 and A and so lessens the danger of an appreciable fraction of the anode voltage appearing at the signal grid.

The Heptode

The hexode may be regarded as a screen-grid tetrode with two control grids. An alternative form of mixer which may similarly be regarded as a pentode with two control grids is the heptode. The heptode is represented diagrammatically in Fig. 6.4. It differs from the hexode in having a suppressor grid next to the anode. This, as in the case of the pentode, is usually operated at cathode potential. The signal may be applied either at G_1 or at G_3 .

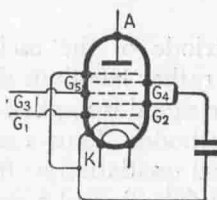


Fig. 6.4. The heptode mixer.

Heptode Frequency-Changer

A valve with seven electrodes and therefore called a heptode, but of very different construction from the heptode mixer, can be used simultaneously as the valve in the local-oscillator circuit and as the mixer. Such a combination valve we shall designate a frequency-changer. In this case the electrodes are connected as shown in Fig. 6.5, but it must be noted that the actual electrodes are very different from those used in the heptode mixer.

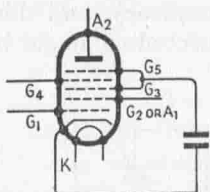


Fig. 6.5. The heptode frequency-changer.

The cathode, K, G_1 , and G_2 form a triode, the active part of the local oscillator, so that the potential of the anode A_1 of the triode varies at the local-oscillator frequency. The grids G_3 and G_5 are connected together and form a screen round the grid G_4 at which the signal is applied. The intermediate-frequency voltage appears at the anode, A_2 .

Octode Frequency-Changer

The heptode frequency-changer combines a triode and a screen-grid tetrode. This suggests a further modification, producing a combined triode and pentode, and this valve, having eight electrodes, is known as an octode. Fig. 6.6 shows an octode frequency-changer.

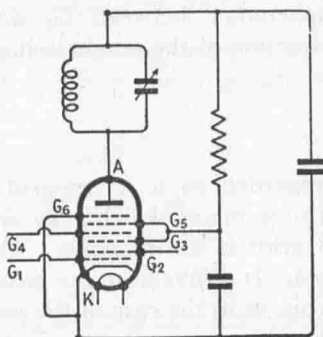


Fig. 6.6. The octode frequency-changer.

K, G_1 and G_2 form the triode of the oscillator. G_3 and G_5 are maintained at a constant potential rather less than the mean potential of the anode A. G_4 is the grid at which the signal is applied and G_6 , the 'suppressor grid', is connected directly to the cathode. Thus a single valve can effectively deal with the production of the local oscillation at frequency f_o , the mixing process to obtain a signal at frequency $f_o - f_s$, and a certain degree of amplification of this intermediate-frequency signal.

Triode-Hexode Frequency-Changer

Perhaps the most important of all frequency-changing valves is that known as the triode-hexode. This is a double-purpose valve carefully designed to give the best results as a frequency-changer. There are two forms, shown in Fig. 6.7(a) and 6.7(b), the more common being that of Fig. 6.7(a). In both valves the physical arrangement of the electrodes is complicated by the steps which have been taken to persuade the electron currents to flow in the required directions, but in both cases the triode has been divorced as far as possible from the mixer. In the valve of Fig. 6.7(a) the signal is applied to G_1 .

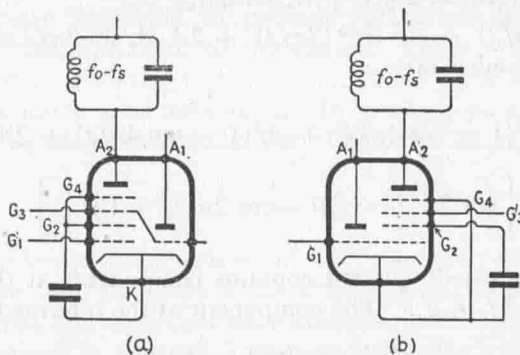


Fig. 6.7. Two forms of the triode-hexode frequency-changer.

This allows the maximum effect to be obtained from a small signal. In the valve of Fig. 6.7(b) the small signal is applied to G_3 , an electrode relatively close to the anode, and therefore having less control of the anode current. For this reason the design of Fig. 6.7(b), though simpler in manufacture, is not so sensitive to small signals.

Single-Electrode Mixing

A rather different mixing principle is employed when the signal frequency is very high. Hexodes, heptodes and octodes have certain disadvantages at frequencies above a certain value which may be arbitrarily stated to be about 30 Mc/s. These disadvantages will be discussed later in the chapter on valve noise. At the higher frequencies pentodes, tetrodes, triodes and diodes are used as mixers, but instead of applying the signal and local-oscillator voltages to two different electrodes, these are applied simultaneously to one electrode only. The theory of this type of mixing will now be discussed.

In Chapter 5 it was shown that, due to the curvature of the grid-voltage-anode-current characteristic, a variable voltage applied to the grid of a triode, tetrode or pentode results in an amplified but distorted voltage across the anode load resistor. It was also shown that the grid-voltage-anode-current characteristic could be represented by the equation

$$I_a = a + bV_g + cV_g^2 + dV_g^3 \quad (5.2)$$

with a fair degree of accuracy, and as exactly as required by adding the terms

$$eV_g^4 + fV_g^5 + \dots$$

The distortion of the output signal due to the term cV_g^2 is used to produce the intermediate frequency.

► Let the signal and local-oscillator voltages applied simultaneously to the control grid of a triode, tetrode or pentode be, as usual, $A_s \sin 2\pi f_s t$ and $A_o \sin 2\pi f_o t$. Then the anode current for a negligible anode load resistor is

$$I_a = a + b(A_s \sin 2\pi f_s t + A_o \sin 2\pi f_o t) + c(A_s \sin 2\pi f_s t + A_o \sin 2\pi f_o t)^2 + d(A_s \sin 2\pi f_s t + A_o \sin 2\pi f_o t)^3 + \dots \quad (6.3)$$

Consider the term $c(A_s \sin 2\pi f_s t + A_o \sin 2\pi f_o t)^2$

This is $c[A_s^2 \sin^2 (2\pi f_s t) + A_o^2 \sin^2 (2\pi f_o t) + 2A_s A_o \sin 2\pi f_s t \sin 2\pi f_o t]$ which may be expanded into

$$\frac{c}{2} \left[A_s^2 (1 - \cos 4\pi f_s t) + A_o^2 (1 - \cos 4\pi f_o t) + 2A_s A_o \left\{ \cos 2\pi(f_o - f_s)t - \cos 2\pi(f_o + f_s)t \right\} \right]$$

This shows that the anode current contains components at the two beat frequencies, $f_o - f_s$ and $f_o + f_s$. The component at the intermediate frequency is

$$cA_s A_o \cos 2\pi(f_o - f_s)t.$$

Fig. 6.8 may assist in explaining the action. The signals applied in series to the control grid are represented by the curves A and B, except that the amplitudes have been taken to be the same for simplicity. Curve C shows their combined

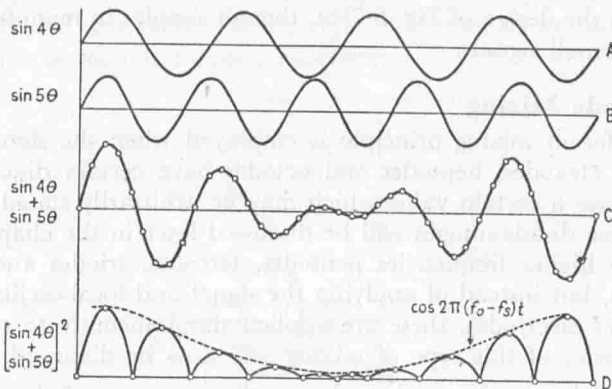


Fig. 6.8. Illustrating waveforms in single-electrode mixing. A, Signal voltage; B, local oscillator voltage; C, sum as applied to G_1 ; D, (sum of the voltages)². In D it will be noted that the envelope is the wave of the difference frequency, $f_o - f_s$.

effect at the grid. It is obtained by adding the ordinates of curves A and B. The valve current then contains a component proportional to curve C, but it also contains a component proportional to the square of the grid voltage. Curve D shows the result of squaring the ordinates of curve C and, as may be seen from the dotted curve, contains a component at the difference frequency $f_o - f_s$.

Therefore, by applying the two voltages to the control grid, and by selecting the component of frequency $f_o - f_s$ by a suitable circuit at the anode, a triode, tetrode or pentode can be made to perform as a mixer. In this case the distortion produced by the valve is an essential part of the process, and a large local-oscillator voltage is used. The output voltage at the intermediate frequency is, indeed, proportional to the product of the signal and local-oscillator amplitudes, as in the other types of mixing, and if it were not for other considerations which will be discussed later, the local-oscillator voltage could be increased indefinitely with advantage.

Special valves are employed as pentode and triode mixers in which the valve adds the minimum noise to the signal. These are, in effect, signal-amplifier pentodes and triodes which are suitable for use at high frequencies. Tetrodes are not much used nowadays. In general, pentodes are employed between 30 and 100 Mc/s, and triodes from 100 to 400 Mc/s. Above this diodes are used.

Diode Mixer

The current-voltage characteristic of the diode is not linear, and the curvature may be used to produce the beat frequencies by applying to one electrode the signal and local-oscillator voltages just as in all the other cases discussed. As we saw in Chapter 3, the diode current is given by $I = CV^{3/2}$, where V is the potential difference across the valve and C is a constant. It may be shown (see Appendix 3.2) that over any finite range of I the equation may be written

$$I = bV + cV^2 + dV^3 + \dots \quad (6.4)$$

where, of course, $b, c, d, \dots$ are appropriate constants. As in the case of the multi-electrode valve, the term cV^2 gives rise to the intermediate frequency component when

$$V = \text{signal voltage} + \text{local-oscillator voltage}.$$

Diode mixers are extremely inefficient, but are the only thermionic devices so far available for frequencies above 400 Mc/s. They can be made very small, and can be fitted to almost any circuit.

Crystal Mixers

A particular form of the diode valve in which current passes much more easily in one direction than in the other is a combination of certain crystals in contact, or a crystal in contact with the end of a metallic conductor. Such a device has a non-linear current-voltage characteristic, and therefore possesses the essential requisite for a mixer. The crystal and 'catwhisker' used in very early radio receivers functioned as a demodulator because of its non-linear characteristic. Nowadays valves are preferred for this operation, but a similar device is the best available mixer at frequencies above 400 Mc/s. The action of such crystal valves will be considered in detail in a later chapter. At the moment it suffices to note that the mixer crystal behaves very much like the diode.

Conversion Conductance

In a simple triode or pentode valve, acting as part of an amplifier, the mutual conductance g_m is defined by $g_m = (\text{change in anode current}) / (\text{change$

in grid voltage causing it) the anode voltage remaining constant. The mutual conductance is therefore a measure of the value of the valve in an amplifier circuit. As was shown in Chapter 3, equation (3.8), the total change in the anode current δI_a is given by

$$\delta I_a = g_m \delta V_g + \delta V_a / r_a$$

where δV_a is the change in anode voltage due to the change in current δI_a , the dynamic resistance of the valve being r_a . In any mixer valve it is possible to define a similar constant, the *conversion conductance* g_c so that $g_c = (\text{change in anode current at the intermediate frequency}) / (\text{change in grid voltage at the signal frequency causing the change in anode current})$, the anode voltage remaining constant. By analogy with g_m this quantity is a measure of the value of the valve as a mixer or converter, and in this case

$$\delta I_a = g_c \delta V_g + \delta V_a / r_a \quad (6.5)$$

where δI_a and δV_a refer to the frequency $f_o - f_s$ and δV_g refers to the frequency f_s . If the anode circuit is tuned to give a resistive impedance R then

$$R \delta I_a = R g_c \delta V_g + R \delta V_a / r_a$$

and

$$\delta V_a = -R \delta I_a, \text{ as in Chapter 3.}$$

Therefore

$$\begin{aligned} \delta V_a &= -R g_c \delta V_g / (1 + R/r_a) = -g_c \delta V_g R r_a / (R + r_a) \\ &= -g_c R_p \delta V_g \end{aligned} \quad (6.6)$$

where R_p is the resistance of r_a and R in parallel. The rules to be applied to the converter are exactly the same as those applicable to an amplifier, the conversion conductance g_c taking the place of the mutual conductance g_m .

The conversion conductance of a mixer valve varies with the local-oscillator voltage available. It has already been shown that in both types of mixer the component of the anode current at the intermediate frequency is proportional to $A_o A_s$, where A_o and A_s are the amplitudes of the local-oscillator voltage and the signal voltage respectively. According to this simple theory g_c must be proportional to $A_o A_s / A_s$; i.e., to A_o . But there is a definite upper limit to g_c . As A_o is increased, a time comes when the negative swings of voltage cut off the current in the valve altogether. Further increases in A_o are then inoperative in increasing g_c . Moreover, large positive swings of local-

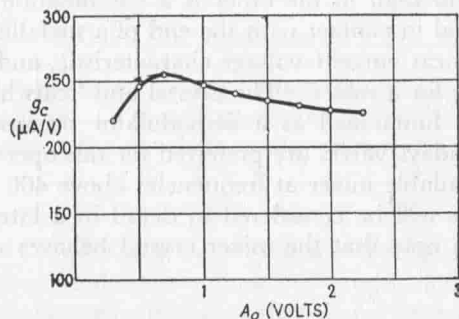


Fig. 6.9. Illustrating the variation of the conversion conductance with local-oscillator voltage.

oscillator voltage cause current to flow to the grid in question, limiting the positive swing and damping the source of oscillation. Therefore, although g_c varies almost linearly with A_o for small values of the latter, the curve flattens out, and indeed falls off slowly when A_o exceeds a certain value. The curve, taken from a representative heptode, and shown in Fig. 6.9, is typical of the variation of g_c with A_o .

Valve manufacturers generally exhibit the variation of g_c with the direct current flowing in the grid biasing resistor of the oscillator grid. This current is almost proportional to A_o , and so the general shape of the curve is the same. The g_c versus I_{g3} curve for a typical triode-hexode is shown in Fig. 5.10.

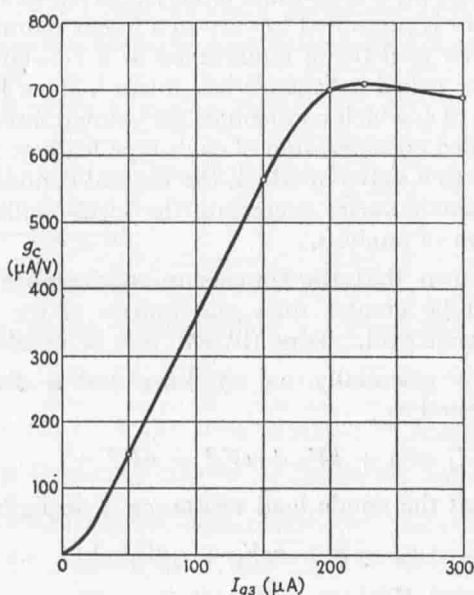


Fig. 6.10. Variation of conversion conductance with oscillator grid current.

APPENDIX 6.1

Relation between Conversion and Mutual Conductances

In Chapter 6 two apparently different types of mixer valve have been discussed. In type (a) the local-oscillator and the signal voltages are applied to two separate grids, each of which controls in some measure the electron stream flowing to the anode. In type (b) the two voltages are applied to a single electrode of a relatively simple valve which may even be a diode.

For the purpose of elementary analysis (a) and (b) were regarded as being of different types. Type (a) is sometimes loosely referred to as a 'linear mixer', since the anode current is presumed to vary in a linear manner with each of the grid voltages, the other grid being maintained at a constant potential. Type (b) might similarly be called a 'square-law mixer', since it is the V_g -squared term in the expansion of I_a which contributes the wanted intermediate-frequency signal. A more detailed consideration of each type leads to the conclusion that each can be regarded as a valve in which the mutual conductance of the signal grid relative to the anode varies according to the local-oscillator voltage. This is the more exact form of analysis.

It will now be shown that the conversion conductance of the signal grid in either case cannot be greater than one-quarter of the maximum mutual conductance for the same grid. Type (b) will first be considered.

Since any valve is essentially an amplifier and a distorter, the anode current may be expressed as

$$I_a = a + bV_g + cV_g^2 + dV_g^3 + \dots \quad (6.7)$$

assuming, as usual that the anode load resistance is negligible.

$$\text{Then } dI_a/dV_g = g_m = b + 2cV_g + 3dV_g^2 + \dots \quad (6.8)$$

Neglecting terms beyond V_g^2 in equation (6.7)

$$dI_a/dV_g = 0 \text{ when } b = -2cV_g, \text{ or } V_g = -b/2c.$$

Suppose that the local oscillator produces a voltage at the grid $A_o \sin 2\pi f_o t$, then g_m will just fall to zero at the negative peak of the oscillation if $-A_o - B_g = -b/2c$, where $-B_g$ is the standing negative bias applied to the grid; i.e., $g_m = 0$, when $b - 2c(A_o + B_g) = 0$ (6.9)

Under these circumstances the maximum g_m is reached at the positive peak of the oscillation;

$$\text{i.e., } g_m = \text{maximum, when } g_m = b + 2c(A_o - B_g) \quad (6.10)$$

$$\text{Substituting from (6.9) in (6.10), maximum } g_m = 4cA_o \quad (6.11)$$

But, as shown in Chapter 5, the intermediate-frequency current at the anode due to the local-oscillator voltage $A_o \sin 2\pi f_o t$ and the signal voltage $A_s \sin 2\pi f_s t$ is

$$\delta I_a = cA_s A_o \cos 2\pi(f_o - f_s)t$$

$$\text{whence } g_c = \frac{\delta I_a \text{ (intermediate frequency)}}{\delta V_g \text{ (signal frequency)}} = cA_o$$

Therefore, combining equations (6.11) and (6.12),

$$g_c = (g_m \text{ maximum})/4$$

or, putting it another way, the conversion conductance cannot be greater than one-quarter of the mutual conductance for the same grid.

In the case of type (a) mixing equation (6.7) must be replaced by

$$I_a = a + b_1 V_{g1} + b_2 V_{g2} + c_1 V_{g1}^2 + c_2 V_{g2}^2 + c_{12} V_{g1} V_{g2} + \quad (6.11)$$

and
$$\frac{\delta I_a}{\delta V_{g1}} = g_{m1} = b_1 + 2c_1 V_{g1} + c_{12} V_{g2} \quad (6.12)$$

Let G_1 be the electrode at which the signal voltage is applied.

Then the variation in V_{g1} will be small compared with the variation in V_{g2} due to the local oscillator, and $c_1 V_{g1}$ can be neglected in comparison with $c_{12} V_{g2}$.

Therefore, $\frac{\delta I_a}{\delta V_{g1}} = 0$, when $b_1 = -c_{12} V_{g2}$.

As before, assume that the local oscillator produces a voltage $A_o \sin 2\pi f_o t$ at G_2 . Then g_{m1} will just fall to zero if $-A_o - B_{g1} = -b_1/c_{12}$, where $-B_{g1}$ is the standing negative bias on the grid.

Therefore, $g_{m1} = 0$, when $b_1 - c_{12} (A_o + B_{g1}) = 0 \quad (6.15)$

Under these circumstances the maximum g_{m1} is reached when

$$g_{m1} = b_1 + c_{12} (A_o - B_{g1}) \quad (6.16)$$

Substituting from (6.15) in (6.16), maximum $g_{m1} = 2c_{12} A_o \quad (6.17)$

But the intermediate-frequency current at the anode due to the two voltages is

$$\delta I_a = \frac{1}{2} c_{12} A_o A_s \cos 2\pi(f_o - f_s)t$$

whence $g_{c1} = \frac{1}{2} c_{12} A_o$ and $g_{c1} = (g_{m1} \text{ maximum})/4$, as in type (b).

It must not be assumed from the above that the working g_c of a valve used as a mixer (for example, a pentode) is less than a quarter of the working g_m of the same valve used as an amplifier. In an amplifier the maximum g_m can only be used for vanishingly small signals. Usually, the working g_m is about three-quarters of the maximum value. Hence, a rough rule for the relation between the working values of g_c and g_m is that g_c is about $g_m/3$.

CHAPTER 7

Modulation and Demodulation

SO FAR IN THIS volume the voltages applied to valves have been considered to be constant or sinusoidal with respect to time. This simplification has been made to allow the various functions of valves to be brought to light without unnecessary difficulties to the reader. In practice it is unusual for valves to have to deal with such simple voltages, and now is the time to discuss the electronic effects of the more complex waveforms encountered in radio. It will be found that the valve actions already discussed are of great value in producing the particular waveforms required and that the theory already given only requires slight extension to explain all the electronic actions employed in practice.

Amplitude Modulation

In Volume 2 it is shown that a radio-frequency carrier wave can only be a vehicle for the transmission of intelligence which is superimposed upon it in the form of modulation, and that the simplest and most common form of modulation is that in which the amplitude of the carrier wave varies according to the amplitude of the modulating wave.

► It is also shown in Volume 2 that the algebraic expression for an amplitude-modulated wave, the modulation being itself sinusoidal, is

$$A = A_c \sin 2\pi f_c t [1 + m \sin 2\pi f_m t]$$

where m can have any value between zero (no modulation) and unity (100% modulation), f_c is the carrier frequency, and f_m is the frequency of the modulation.

It will now be shown how any valve may be made to give rise to an amplitude-modulated wave when it is excited by the carrier wave and the modulating wave simultaneously. There are two types of modulating action just as there are two types of mixer. Indeed modulation, demodulation and mixing are very similar in their principles of action.

First, let the two voltages be applied in series to the control grid of a triode, tetrode, or pentode. As we have seen in Chapter 5, the anode-current-grid-voltage characteristic of a valve is not a straight line, and so the valve distorts the combined input signal. Let the sum of the carrier voltage and the modulating voltage be represented by the curve A of Fig. 7.1, and let B be the dynamic I_a-V_g characteristic of the valve and its anode load. Then when the modulating voltage is most negative, as at D, the amplification is small and the radio-frequency anode current is a minimum, as at E of curve C. But when the modulating voltage is positive, as at F, the amplification is large, and the anode current swings are a maximum, as at G. Hence the amplitude of the radio-

frequency current at the valve anode rises and falls with the modulating voltage. This is the essential feature of an amplitude-modulated wave.

The modulation has been brought about by the curvature of the I_a-V_g characteristic or, as mentioned above, the valve action responsible for producing the beat frequency (frequency-changing) is also responsible for modulation.

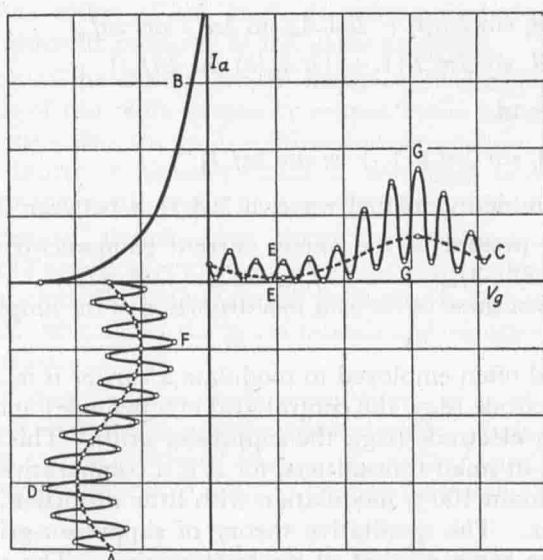


Fig. 7.1. Illustrating grid modulation. The sum of the carrier and modulating voltages at the grid produces the modulated wave at the anode. The component of the anode current which varies at the modulating frequency is shown by the dotted curve.

The anode current is not, of course, a pure amplitude-modulated wave. The latter is one component only; in addition to it there are other frequencies present, notably the original modulating frequency. But, as in the case of the mixer valve, it is possible to select the modulated carrier wave alone by the use of a suitable tuned circuit in the anode lead. This is exactly what is done in practice.

- This qualitative explanation of the valve as a modulator can be made quantitative by referring to the general expression for the anode current I_a in terms of the grid voltage V_g . This expression was derived in Chapter 5 [equation (5.3)] and is

$$I_a = a + bV_g + cV_g^2 + dV_g^3 + \dots \quad (5.3)$$

Let $V_g = A_c \sin 2\pi f_c t + A_m \sin 2\pi f_m t$, where the two terms are the carrier and modulating voltages. Then, substituting in equation (5.3).

$$\begin{aligned} I_a = & a + b (A_c \sin 2\pi f_c t + A_m \sin 2\pi f_m t) \\ & + c (A_c \sin 2\pi f_c t + A_m \sin 2\pi f_m t)^2 \\ & + d (A_c \sin 2\pi f_c t + A_m \sin 2\pi f_m t)^3 + \dots \end{aligned} \quad (6.1)$$

Expanding the right-hand side of this equation

$$I_a = a + bA_c \sin 2\pi f_c t + bA_m \sin 2\pi f_m t \\ + \frac{1}{2} cA_c^2 (1 - \cos 4\pi f_c t) + \frac{1}{2} cA_m^2 (1 - \cos 4\pi f_m t) \\ + 2cA_c A_m \sin 2\pi f_c t \sin 2\pi f_m t + \dots \quad (6.2)$$

Of all the terms, consider only

$$bA_c \sin 2\pi f_c t + 2cA_c A_m \sin 2\pi f_c t \sin 2\pi f_m t \\ = bA_c \sin 2\pi f_c t [1 + (2cA_m/b) \sin 2\pi f_m t]$$

This is equivalent to

$$A_c \sin 2\pi f_c t [1 + m \sin 2\pi f_m t]$$

i.e., it is an amplitude-modulated wave, if $2cA_m/b$ is between 0 and 1.

There are also present in the anode current components of frequency f_m , $2f_m$, $3f_m$, $2f_c$, $3f_c$, $(f_c \pm 2f_m)$, $(f_c \pm 3f_m)$, etc. As will be seen later some of these terms lead to a distortion of the amplitude-modulated wave.

Another method often employed to modulate a carrier is to apply the carrier voltage to one electrode (e.g., the control grid of a pentode) and the modulating voltage to another electrode (e.g., the suppressor grid). This form of modulation is very useful in small transmitters, for it is a comparatively simple matter by its means to obtain 100% modulation with little distortion, using very little modulating power. The qualitative theory of suppressor-grid modulation is almost exactly the same as that of pentode mixing. The variations in the modulating voltage increase or decrease the anode current, thus changing the effective g_m of the control grid, and so altering the net amplification of the carrier wave. This action of the suppressor grid was considered in Chapter 5. By causing the suppressor to become sufficiently negative, (and this can be done by using the appropriate negative bias and a sufficiently large modulating voltage) the anode current can be cut off entirely and, as may be seen from Figure 5.13, the slope of the I_a - V_{g3} characteristic is almost constant over a considerable range of I_a , so that the distortion is small.

► In the case of the pentode, to a good approximation the anode current is given by

$$I_a = (a + bV_{g1})(a' + b'V_{g3}) \quad (5.12)$$

so that, if $V_{g1} = A_c \sin 2\pi f_c t$ and $V_{g3} = A_m \sin 2\pi f_m t$,

$$I_a = a + a'bA_c \sin 2\pi f_c t + ab'A_m \sin 2\pi f_m t \\ + bb'A_c A_m \sin 2\pi f_c t \sin 2\pi f_m t \\ = a + ab'A_m \sin 2\pi f_m t \\ + a'bA_c \sin 2\pi f_c t [1 + (b'A_m/a') \sin 2\pi f_m t] \quad (6.3)$$

The third term on the right-hand side of equation (6.3) is an amplitude-modulated wave, the depth of modulation varying as $b'A_m/a$. The second term, consisting of a wave of frequency f_m , can be filtered out by an appropriate circuit. Suppressor-grid modulation is therefore a particularly simple method of obtaining a modulated wave relatively free from distortion.

In conclusion it should again be emphasized that to obtain an amplitude-modulated wave by either of the methods described above, use is made of the 'second-order' distortion term in the valve characteristic; i.e., either of the term cV_g^2 or of the term $bb'V_{g1}V_{g3}$.

Anode Modulation

There is an entirely different method of obtaining a modulated wave, used more generally than either of the methods described above, and depending upon an entirely different property of the valve amplifier.

In the discussion of the class C triode amplifier in Chapter 5 it was stated that the amplitude of the radio-frequency output varies almost directly as the anode voltage on the valve, the grid exciting voltage, or drive voltage, remaining constant. A similar relationship holds for the beam tetrode and pentode class C amplifiers if the anode and screen voltages are varied together.

All that is necessary, therefore, to obtain an amplitude-modulated wave from a class C amplifier is to vary the anode voltage at the modulating frequency, taking care to see that the amplitude of the latter is less than the constant supply voltage. In the cases of the beam tetrode and pentode both screen and anode must be varied.

▶ This method of modulation—anode modulation, as it is called—does not depend upon any curvature in a valve characteristic. It is a linear process. The anode voltage can be expressed as $A_o + A_m \sin 2\pi f_c t$ and, if the unmodulated radio-frequency voltage is $A_c \sin 2\pi f_c t$, then the combined output voltage is proportional to

$$(A_o + A_m \sin 2\pi f_m t) A_c \sin 2\pi f_c t.$$

Anode modulation may be obtained by means of a variety of circuits described in Volume 4. It should be noted that, unlike control-grid or suppressor-grid modulation, anode modulation takes a considerable amount of power from the modulating source. In fact the latter supplies half the total input power when the modulation is 100%.

Frequency and Phase Modulation

Frequency modulation and phase modulation are discussed in Volume 2, where it is stated that in a frequency-modulated wave the amplitude remains constant but the frequency varies according to the audio or video signal which it transmits, the changes in frequency being proportional to the amplitude of the modulating signal and occurring at the same frequency. Similarly, a phase modulated wave is one in which the amplitude is constant and the phase varies directly according to the modulation.

▶ Algebraically, if the modulation is given by $m \sin 2\pi f_m t$, the frequency of the modulated wave is given by

$$f = f_c + m \sin 2\pi f_m t$$

and the frequency-modulated wave itself by

$$\sin 2\pi[f_c t + (m/2\pi f_m) \sin 2\pi f_m t].$$

Similarly, if the phase ϕ is given by

$$\phi = 2\pi(f_c t + m \sin 2\pi f_m t)$$

the phase-modulated wave itself is given by

$$\sin 2\pi(f_c t + m \sin 2\pi f_m t)$$

There are two ways of phase or frequency modulating a carrier wave by means of valves. In the first the valve is used as a *variable reactance*. In the second a particular combination of two valves and their circuits known as a *balanced modulator* is used. Both the *reactance valve* and the balanced modulator have many applications in radio apart from their use to produce a frequency- or phase-modulated wave. They will now be discussed simply as electronic devices, and their application to frequency or phase modulation will be mentioned among other things in passing.

The Valve as a Variable Reactance

Imagine a pentode arranged in a circuit so that, as shown in Figure 7.2, a voltage $A_a \cos \theta$ is applied in series with the h.t. voltage V_{HT} to the anode, and a voltage 90° out of phase with this, $A_g \sin \theta$ is applied to the grid in series with the negative grid bias V_{GB} .

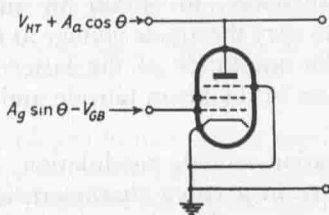


Fig. 7.2. A reactance valve.

The alternating current through the valve is chiefly determined by the grid potential, if $A_a \cos \theta$ is small, and is in phase with it. In fact, for a pentode in which the anode load is small compared with the resistance of the valve, I_a is approximately $g_m V_g$ which equals $g_m (A_g \sin \theta - V_{GB})$. The fluctuations in the anode voltage due to $A_a \cos \theta$ are 90° out of phase with this, and so the valve behaves like a reactance. Even if there is an appreciable resistance in the anode circuit, the reactive component is still present, and if the valve is connected across a tuned circuit, it behaves like a parallel reactance, the magnitude of which depends upon the grid voltage. If the tuned circuit and valve together constitute the frequency-determining element of an oscillator, the average frequency of which is the carrier frequency, the changes in the grid voltage are reflected as changes in the reactance of the valve and appear as changes in the frequency of the oscillatory circuit. The changes in reactance are due to the curvature of the grid-voltage-anode-current characteristic. Changes in the grid voltage are not exactly imaged in the changes in the anode current.

It can now be seen how the reactance valve is used to produce frequency modulation. The carrier voltage is applied to the grid and anode simultaneously, but the grid is caused to be 90° out of phase with the anode by suitable circuit arrangements. The modulating voltage is applied to the control or suppressor grid and, as it varies, the reactance presented by the valve to the carrier-frequency oscillatory circuit varies. So long as these changes are small, the frequency variations which result mirror faithfully the modulating-voltage variations.

Reactance valves can be employed for other purposes. Chief among these is automatic frequency control of a transmitter or receiver. In this case the

reactance of the valve is controlled by the signal to which the transmitter or receiver is to be tuned, and the reactance in turn controls the frequency of the transmitter or receiver. This action is of considerable importance in certain radar and communication systems.

The Balanced Modulator

In Chapter 5 the combination of two valves known as a push-pull amplifier was described and its action discussed. The arrangement of Figure 5.7 was shown to have certain important properties which are shared by the balanced modulator. Consider the pair of valves connected as in Figure 7.3, so that an audio- or video-frequency signal $A_m \sin 2\pi f_m t$ is applied between the two grids but, in addition, both grids are connected in parallel to a carrier-frequency

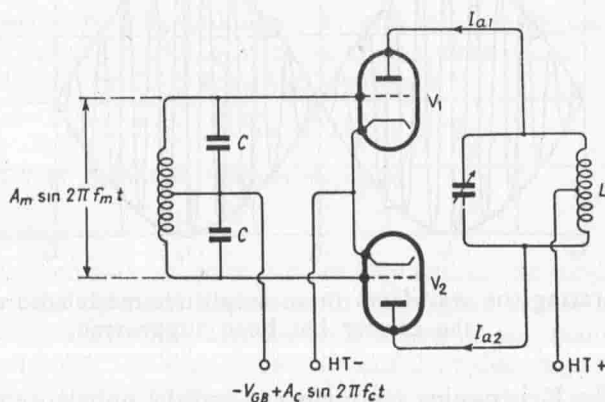


Fig. 7.3. A simple balanced modulator.

source, giving both a potential $A_c \sin 2\pi f_c t - V_{GB}$ relative to the cathode. V_{GB} is the negative standing grid bias. Without going into details the figure indicates how these potentials might be applied. The audio voltage is supplied by the secondary of a transformer, the carrier voltage is applied by means of capacitors C , and the constant potential $-V_{GB}$ is applied at the centre-tapping. Between the two anodes a parallel resonant circuit is connected, tuned to frequency f_c but of a low enough Q to accommodate without appreciable loss frequencies between $f_c - f_m$ and $f_c + f_m$. As in practice, let it be assumed that f_c is very great compared with f_m .

The action of the balanced modulator may be explained as follows:— Suppose first that only the carrier voltage is present. Then the potentials of both grids rise and fall together at the carrier frequency. Each time the valve grids become positive, each valve takes a greater current but, as the currents are in opposing directions in the coil L , no excitation of the resonant circuit occurs. The valves are balanced and suppress the carrier altogether. When the audio voltage is applied, the valves are only balanced instantaneously when $A_m \sin 2\pi f_m t$ is zero. At all other times, one grid is more positive than the other and consequently an amplified carrier-frequency voltage appears across L . This varies according to the positive half of $A_m \sin 2\pi f_m t$ but is zero and a maximum twice in each cycle of the modulating voltage. The action

is illustrated by Fig. 7.4, where the top graph is $2A_m \sin 2\pi f_m t$ as a function of time, and the second graph is the voltage V_L across L , drawn to the same time scale.

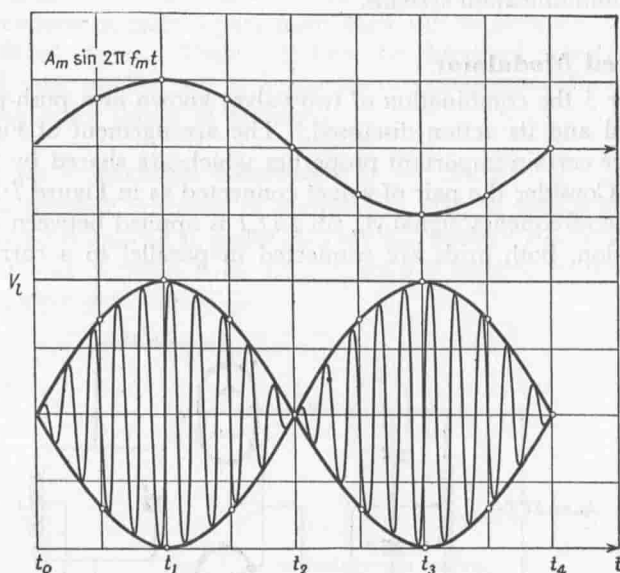


Fig. 7.4. Illustrating the waveform in an amplitude-modulated wave in which the carrier has been suppressed.

At time t_1 valve V_1 is passing more current and the unbalance gives a carrier-frequency output. At time t_3 valve V_2 is passing more current and exciting the carrier-frequency component again. Only at t_0, t_2, t_4 is the carrier-frequency voltage zero. Clearly the envelope of the modulated wave corresponds to full-wave rectification of the modulating wave. This is what is known as *suppressed-carrier modulation*.

► Algebraically the action of the valves in a balanced modulator can be described in much the same way as a push-pull amplifier.

Let I_{a1} and I_{a2} be the anode currents in valves V_1 and V_2 respectively. Then (equation 5.2)

$$I_{a1} = a + bV_{g1} + cV_{g1}^2 + dV_{g1}^3 + \dots$$

$$\text{and } I_{a2} = a + bV_{g2} + cV_{g2}^2 + dV_{g2}^3 + \dots$$

The two equations will only be identical if the two valves are exactly matched. This can be done approximately by the use of various circuit devices.

$$\text{But } V_{g1} = (A_m/2) \sin 2\pi f_m t + A_c \sin 2\pi f_c t - V_b$$

$$\text{and } V_{g2} = -(A_m/2) \sin 2\pi f_m t + A_c \sin 2\pi f_c t - V_b$$

$$\begin{aligned} \text{Therefore, } I_{a1} = & a - bV_b + (bA_m/2) \sin 2\pi f_m t + bA_c \sin 2\pi f_c t \\ & + c[(A_m/2) \sin 2\pi f_m t + A_c \sin 2\pi f_c t - V_b]^2 \\ & + d[\dots]^3 + \dots \end{aligned}$$

with a similar expression for I_{a2} , namely,

$$I_{a2} = a - bV_b - (bA_m/2) \sin 2\pi f_m t + bA_c \sin 2\pi f_c t \\ + c[-(A_m/2) \sin 2\pi f_m t + A_c \sin 2\pi f_c t - V_b]^2 + \dots$$

Assuming that only first- and second-order terms need be considered,

$$I_{a1} - I_{a2} = \text{current flowing in } L, \text{ neglecting the resonating voltage,} \\ = bA_m \sin 2\pi f_m t + 2cA_m A_c \sin 2\pi f_m t \sin 2\pi f_c t$$

But the circuit between the anodes is tuned to frequency f_c and is wide enough to accept frequencies between $f_c - f_m$ and $f_c + f_m$. Therefore the voltage across L is given by V_L , where, neglecting the sign

$$V_L = c'A_m A_c \sin 2\pi f_m t \sin 2\pi f_c t$$

This is suppressed-carrier modulation. It may be transformed to ordinary amplitude modulation by adding a carrier in the same phase. For example, if a carrier $(c'A_m A_c/m) \sin 2\pi f_c t$ is added, we obtain

$$(c'A_m A_c/m)[1 + m \sin 2\pi f_m t] \sin 2\pi f_c t,$$

the expression for an orthodox modulated wave.

Phase Modulation

A phase-modulated wave may be obtained by adding the suppressed-carrier modulation from the balanced modulator to the original carrier if the phase of the latter is changed by 90° . The methods belong to the realm of circuitry rather than electronics. It should only be noted at this stage that although the resultant wave is not a true phase-modulated wave, it is a sufficiently good approximation to it in certain cases to be satisfactory for all practical applications. The theory is given in Appendix 7.2.

Frequency Modulation

Again, unless a reactance modulator is employed, this is not an electronic action. As may easily be shown, any phase-modulated wave can be transformed to a frequency-modulated wave and vice-versa by the appropriate use of a resistor and capacitor.

Pulse Modulation

In pulse modulation the carrier wave is electronically switched on or off according to some prearranged code, the modulation determining the amplitude of the pulses, or their length or their frequency. There is no new electronic principle involved in the valve action in pulse modulation, although the action of a valve switch or electronic switch deserves a few words.

Electronic Switch

If an alternating voltage at some frequency f_c , the carrier frequency, is to be switched on and off many times per second, the advantage of the valve switch is that it can do this at an exceedingly high speed. Any mechanical switch suffers inevitably from the mechanical inertia of its parts. Its speed of action is limited. Even for very small mechanical switches it is difficult to perform the switching in less than 1 millisecond. If inductance is kept to a

minimum in the circuits associated with a valve, the electrical inertia is determined by the electron itself. This has a mass of about 10^{-30} kg. Consequently valve switches can be designed which perform the switching in one-hundredth of a microsecond. This allows pulses of length one-tenth of a microsecond to be used.

Demodulation

As has been explained in Volume 2, to transmit information by radio it is necessary at the transmitter to impress the information on the carrier in the form of modulation. Conversely, it is necessary at the receiver to recover the information from the modulated carrier. This process is known as demodulation.

Amplitude Demodulation.

To recover the amplitude modulation from a modulated carrier, a valve or some similar distorting device must be used. Passive circuits cannot by themselves effect the necessary transformation. Although there are many varieties of demodulators, or 'detectors' as they are often called, the basic principle of demodulation is the same in all cases. To illustrate it consider first *any* device which has the current-voltage characteristic shown in Fig. 7.5, where $I = 0$ at $V = -V_c$, and the current varies in a non-linear manner with the voltage.

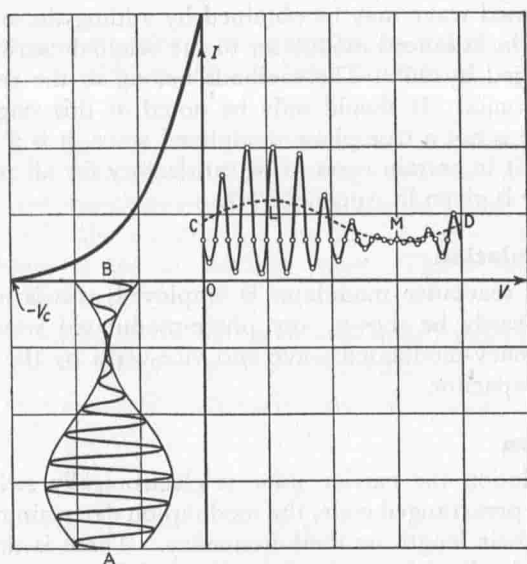


Fig. 7.5. Illustrating demodulation in any electronic device.

If the voltage applied across the device is at a radio frequency, and is modulated at an audio or video frequency, as shown, for example, by the wave AB, the current through the device does not mirror this faithfully, since the current-voltage characteristic is curved. The current emphasizes the positive excursions of voltage and compresses the negative excursions, giving rise to a current waveform such as is shown by CD. If the distorted output current is then applied to a circuit which suppresses variations at the carrier frequency, the

output voltage corresponds to the average value of the curve CD, that is, it corresponds to the dotted curve CLMD. This curve approximately reproduces the modulation impressed on the wave AB, and the device with its attached circuits acts as a demodulator.

Algebraically, the demodulation may be explained as follows:—

If the non-linear device has a characteristic represented by

$$I = a + bV + cV^2 + dV^3 + \dots$$

and if V , the voltage applied across it, is the amplitude modulated wave

$$V = A_c \sin 2\pi f_c t [1 + m \sin 2\pi f_m t],$$

$$\begin{aligned} \text{then } I = & a + bA_c \sin 2\pi f_c t [1 + m \sin 2\pi f_m t] \\ & + cA_c^2 (\sin 2\pi f_c t)^2 [1 + m \sin 2\pi f_m t]^2 \\ & + dA_c^3 (\sin 2\pi f_c t)^3 [1 + m \sin 2\pi f_m t]^3 + \dots \end{aligned}$$

The third term on the right-hand side of the equation may be expanded thus

$$\frac{cA_c^2}{2} [1 - \cos 4\pi f_c t] [1 + 2m \sin 2\pi f_m t + m^2 (\sin 2\pi f_m t)^2]$$

This contains the important middle term

$$\frac{cA_c^2}{2} \times 2m \sin 2\pi f_m t = cA_c^2 m \sin 2\pi f_m t \quad (A)$$

and, using the last term, we also have,

$$\frac{cA_c^2 m^2}{2} (\sin 2\pi f_m t)^2 = \frac{cA_c^2 m^2}{4} [1 - \cos 4\pi f_m t]$$

$$\text{which contains } (cA_c^2 m^2 / 4) \cos 4\pi f_m t \quad (B)$$

The term (A) is the required modulation term. Its appearance in the expansion shows that if suitable circuits are employed with the device, a voltage proportional to the modulation can be recovered from it. The term (B) is the chief second-harmonic distortion term. This is usually the most important distortion. Since the amplitude of (B) is $m/4$ times the amplitude of (A), the second-harmonic distortion increases with the modulation index. When m is unity (100% modulation), the second-harmonic amplitude may be as large as one-quarter of the required modulation amplitude. Steps may, however, be taken to reduce this by the use of suitable circuits. The maximum second-harmonic distortion at 100% modulation is, therefore, one sixteenth of the power in the modulation, or, expressing it in the usual manner, -12 db relative to the required signal.

The Square-Law Demodulator

Demodulators are sometimes classified as 'square-law' and 'linear'. The theory given applies strictly to the square-law type, for the output current (or voltage) is proportional to the square of the input carrier voltage. Any valve may be used as a square-law demodulator; diodes, triodes, and occasionally pentodes are so employed. The circuits and exact theory are discussed in Volume 4.

The Linear Demodulator

When the voltage available from the modulated wave is large, the output current or voltage is more nearly proportional to the input carrier voltage than to its square, and for this reason it is quite common to speak of such a large-signal demodulator as being 'linear'. Consider, for example, a triode valve in a demodulator circuit, the anode load being a resistor. Then, as was seen in Chapter 5 (see Fig. 5.4) the dynamic I_a-V_g characteristic of the valve and resistor is much more nearly a straight line than the I_a-V_g characteristic of the valve itself. Suppose that the curve AB of Fig. 7.6 represents such a dynamic characteristic. Then, if a standing bias = $-OA$ is given to the grid, and a large modulated wave is applied, the output current is given by the curve OK, and, since AB may be regarded as a straight line over the greater

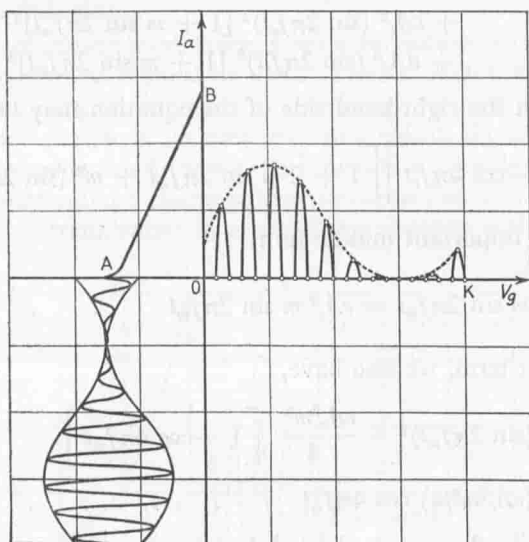


Fig. 7.6. Action of the 'linear' demodulator.

part of its length, the output current is proportional to the input voltage. One of the most important points about the linear demodulator is that the second-harmonic distortion is less than in the case of the square-law detector. In the case of the ideal linear demodulator harmonic distortion is zero. The ideal characteristic is as shown in Fig. 7.7 where two straight lines of different

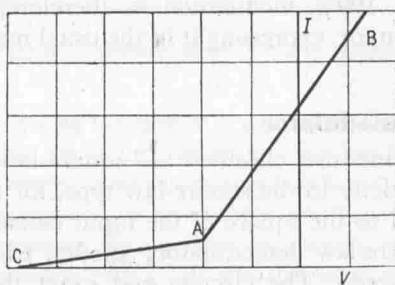


Fig. 7.7. Ideal characteristic of the 'linear' demodulator.

gradients meet at A. Such a complete discontinuity can never occur in practice. It should be noted, too, that despite the name 'linear demodulator', demodulation demands a non-linear device. The characteristics of Fig. 7.6 and 7.7 are not linear, since the straight part or parts are strictly limited in length. Diodes, triodes and multi-electrode valves can be used in linear demodulators. The circuits are discussed in Volume 4.

Frequency and Phase Demodulators

No new principle of valve action is involved in these units. In both cases it is customary to use the valve or valves as in the linear amplitude demodulator. The arrangements necessary to transform phase or frequency variations to amplitude variations are entirely a matter of circuits. The same is true of pulse demodulators where these operate as pulse-length demodulators or pulse-frequency demodulators. They are essentially similar in action to the two fundamental demodulators already considered. No new valve action is involved.

Cross-modulation

One of the most undesirable effects produced in a receiver by a non-linear device such as a valve is the phenomenon known as *cross-modulation*. This denotes the appearance of the modulation from one signal on the carrier frequency of another signal. Suppose, for example, that a receiver is tuned to a signal the carrier frequency of which is f_c , but that an interfering signal at a carrier frequency f'_c cannot be excluded. The first non-linear device in the receiver can then transfer the modulation from the carrier of frequency f'_c to the carrier of frequency f_c and vice-versa. The effect on the listener is that of a distorted chatter, when there is modulation on both carriers. In Service receivers cross-modulation may be a serious limitation particularly where a receiver has to operate in the vicinity of a strong transmitter on a neighbouring frequency.

Cross-modulation generally arises either in the first signal-amplifier valve of a receiver or in the mixer valve. It is unlikely to be produced in the i.f. amplifier, since the selectivity is then usually sufficient to reduce the unwanted signal to an extremely low level. Attempts have been made without success to design signal-amplifier valves to give the minimum of cross-modulation. Generally this results in obtaining the minimum of amplification as well, which does not help. The important feature of cross-modulation, as is shown in Appendix 7.2, is that the amplitude of the unwanted modulation is proportional to the square of the amplitude of the unwanted carrier. Hence any improvement in selectivity produces a much greater improvement in the signal to cross-modulation ratio. Greater selectivity appears to be the best answer. No electronic device has yet been discovered which reduces the effect in practice.

APPENDIX 7.1

Grid Modulation Distortion due to the Departure of the I_a - V_g Characteristic from Square-Law

The ideal modulator would be one in which the anode current varied as the square of the grid voltage. In fact, however,

$$I = a + bV + cV^2 + dV^3 + eV^4 + fV^5 + gV^6 + \quad (7.5)$$

Let the carrier and modulating voltages applied to the grid give

$$V = A_c \sin \omega_c t + A_m \sin \omega_m t, \text{ where } \omega = 2\pi f.$$

Then each term above bV in equation (7.5) contributes something to the modulation pattern. Expanding I in terms of the value of V , and considering only the terms involving the first powers of $\sin \omega_c t$, Table 7.1 is obtained, showing the amplitudes of the carrier, the fundamental component of the modulation and the various harmonics given by terms up to and including gV^6 .

TABLE 7.1
Amplitudes of Components

Carrier	Fundamental	2nd Harmonic	3rd Harmonic	4th Harmonic	5th Harmonic
bA_c	$2cA_cA_m$	$-3dA_cA_m^2/2$	$-eA_cA_m^3$	$5fA_cA_m/8$	$3gA_cA_m^5/8$
$3dA_c^3/4$	$3eA_cA_m^3$	$-5fA_cA_m^4/2$	$-15gA_cA_m^5/8$		
$5fA_c^5/8$	$15gA_cA_m^5/8$	$-15fA_c^3A_m^2/4$	$-15gA_c^3A_m^3/4$		
$3dA_cA_m^2$	$3eA_c^3A_m$				
$15fA_cA_m^4/8$	$45gA_c^3A_m^3/4$				
$15fA_c^3A_m^2/4$	$15gA_c^5A_m/4$				

Writing bA_c , the undistorted amplitude of the carrier wave as unity, and $2cA_cA_m$ as m times this amplitude, the other terms are seen to be small. The largest distortion is due to the 2nd-harmonic term $-3dA_cA_m^2/2$ which is proportional to m^2 , for the amplitude of the modulating voltage A_m must be many times that of the carrier voltage if a reasonable depth of modulation is to be obtained.

The distortion is, of course, greatly modified by the circuits attached to the valve.

APPENDIX 7.2

Cross-Modulation

To show how cross-modulation arises, consider a non-linear device simultaneously excited by two signals, the first being a modulated carrier, $A_c \sin \omega_c t (1 + m \sin \omega_m t)$ written A for short, and the other being an unmodulated carrier of different frequency $A_c' \sin \omega_c' t$, written A' for short. Then the applied voltage V is given by

$$V = A + A',$$

and, if the characteristic of the valve is represented as usual by

$$I = a + bV + cV^2 + dV^3 + \dots$$

the term dV^3 gives rise to

$$d(A^3 + 3A^2A' + 3AA'^2 + A'^3)$$

Consider the expansion of the term $3A^2A'$

$$\begin{aligned} \text{Then } 3dA^2A' &= 3dA_c^2 \sin^2 \omega_c t [A_c \sin \omega_c t (1 + m \sin \omega_m t)]^2 \\ &= \frac{3}{2} dA_c^2 A_c' \sin \omega_c' t (1 - \cos 2\omega_c t) (1 + 2m \sin \omega_m t + m^2 \sin^2 \omega_m t) \end{aligned}$$

Extract the term $3dmA_c^2 A_c' \sin \omega_c' t \sin \omega_m t$

and the appropriate carrier $\frac{3}{2} dA_c^2 A_c' \sin \omega_c' t$.

Then there appears the new modulated carrier given by

$$\frac{3}{2} dA_c^2 A_c' \sin \omega_c' t (1 + 2m \sin \omega_m t)$$

The modulation has been impressed upon the previously unmodulated carrier.

Two points are worthy of note. Cross-modulation is due to terms of the 3rd. order; that is, to terms of the order of dV^3 or $dV_1^2 V_2$ or $dV_1 V_2^2$. To be accurate cross-modulation is produced by all odd-order terms, V^3 , V^5 , V^7 , etc. But in practice only the 3rd-order terms are important. The second point is that phase or frequency modulation cannot be transferred to another carrier in the same manner. This is important since it has been shown experimentally that frequency-modulation systems are free from such direct cross-modulation effects.

Finally, it should be noted that the amplitude of the cross-modulation (on the 'wanted' carrier), is proportional to the square of the amplitude of the 'unwanted' carrier. The effect therefore increases very rapidly with the strength of the interfering signal.

APPENDIX 7.3

Transformation of Suppressed-Carrier Modulation to Phase Modulation

It has been shown that a balanced modulator produces the modulation sidebands without the carrier, and that the suppressed-carrier modulation may be represented by

$$y_1 = A_m \sin \omega_c t \sin \omega_m t,$$

where ω_c is the pulsance of the carrier and ω_m is the pulsance of the sinusoidal modulation.

Now suppose that a carrier is added to the wave in a linear network, the phase of the carrier being 90° different from the phase of the carrier component in the suppressed-carrier modulation. Then the sum is y_2 , where

$$y_2 = A_c \cos \omega_c t \pm A_m \sin \omega_c t \sin \omega_m t.$$

$$\text{Let } A_c = A \cos \theta, \quad \text{and } A_m \sin \omega_m t = A \sin \theta.$$

Then $y_2 = A \cos (\omega_c t + \theta)$, where

$$\tan \theta = A_m \sin \omega_m t / A_c \text{ and } A = A_c \sqrt{1 + (A_m \sin \omega_m t)^2 / A_c^2}.$$

If A_m/A_c is small, $\theta \approx \tan \theta$ and $A \approx A_c$, so that

$$y_2 \approx A_c \cos [\omega_c t + (A_m \sin \omega_m t) / A_c].$$

This is the expression for a phase-modulated wave.

It should be noted (i) that the output wave is really both amplitude and phase modulated, but that the amplitude modulation can be made vanishingly small compared with the phase modulation, since the latter depends upon A_m/A_c , while the former depends upon $(A_m/A_c)^2$. It should also be noted (ii) that the distortion of the phase modulation due to the non-equality of θ and $\tan \theta$ can be made as small as we please by decreasing the ratio A_m/A_c , and (iii) that the phase modulation may be multiplied almost indefinitely by the use of frequency multipliers, leaving the residual amplitude modulation unaffected.

CHAPTER 8

Multi-Unit Valves

TO SAVE SPACE and to economize in wiring valves have been developed which incorporate in one envelope two or more complete sets of electrodes, each set behaving as a simple valve. Other advantages have been discovered for this construction, and, as will be seen later, such 'multi-unit' valves are of considerable importance. No new electronic principles are involved in their action. They are used only because they are more economical in materials, or more efficient, or both.

Twin or Double Valves

The first class of multi-unit valve which requires attention is that in which two exactly similar valves are placed side by side in one container. The three most important are twin diodes, twin triodes and twin beam tetrodes. The twin diode is very widely used as a full-wave rectifier. For this application there is no difficulty in justifying its use to replace two single diodes. The complete valve may have two separate cathodes, or a common cathode, serving two anodes. The cathode or cathodes may be directly or indirectly heated. In almost all cases the heater pins are common to both sections of the valve. A typical circuit, using a twin diode as a full-wave rectifier, is shown in Fig. 8.1.

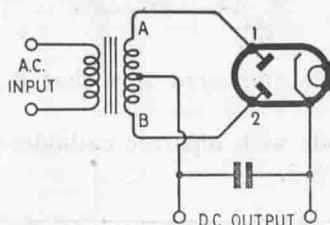


Fig. 8.1. Double diode used as h.t. supply rectifier.

The two diodes pass current in alternate half cycles. When A is positive, diode 1 is conducting; when B is positive, diode 2 is conducting. Such twin diodes are of almost universal application where the peak voltage across AB is not too large.

The twin triode is also a very popular valve. It consists of two identical triodes in one envelope. These may or may not have a common cathode. One important use of the twin triode is as a push-pull amplifier, where it has a considerable advantage over the use of two separate valves. It will be recalled that the theory of push-pull amplification demands that the two valve characteristics should be identical. This is impossible to arrange in practice, and to obtain even approximately 'matched' characteristics is not easy. Where the

two triode units are incorporated in the same envelope, and are therefore manufactured and processed together, there is a better chance of obtaining a good match. In particular, the cathode emission per unit area tends to be the same for both cathodes, and this has a marked effect on the characteristics, in spite of the fact that the valves are used in the space-charge limited condition.

When a double triode is used for push-pull amplification, its cathodes are usually at a fixed, common, potential, but for certain other uses it is convenient to have two triodes in one envelope but with separate cathodes. For example, in certain radio transmitters frequency multiplication is necessary. If a multiplication factor of 4, 6 or 9 is required, this can be conveniently done using a twin triode, each section multiplying by 2 or 3. But in this case, since the signal levels are different at the two grids, it is necessary to have the cathodes separate. Although both are 'earthy' so far as radio-frequency voltages are concerned, they should be at different average potentials to give the grids different bias voltages. Twin-triodes are, in fact, widely used to replace two separate triodes, wherever the functions of the two valves do not demand complete separation of the wiring. Clearly, the residual lead capacitances and mutual inductances inherent in twin valves are of lesser importance at the lower frequencies, and it is in this region of the radio spectrum that advantage can be taken of their lower cost and smaller bulk. A typical circuit, employing a twin triode as a class B amplifier for audio frequencies is shown in Fig. 8.2

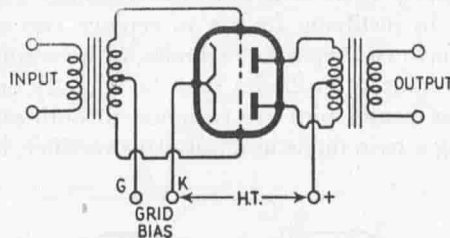


Fig. 8.2. Double triode employed as a class B push-pull amplifier.

and the use of a twin triode with separate cathodes as a two-stage frequency multiplier is shown in Fig. 8.3.

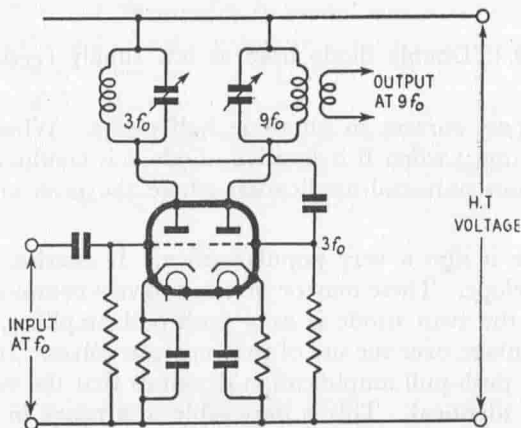


Fig. 8.3 Double triode functioning as a two-stage frequency multiplier.

The use of the twin beam tetrode has become very common for power amplification both at audio and radio frequencies. The construction in this case is more elaborate than merely placing two sets of electrodes in one envelope. Clearly there is no reason why the screen grids of the two sections, since they are maintained at the same constant potential, should not be joined together, and so twin tetrodes usually have a common cathode and a common screen connection. Frequently the two sides of an almost flat cathode are used for the two sections, so that the valve resembles a normal beam tetrode except that the grid and anode have been separated into two equal parts. Since 1945 some very elegant twin beam tetrodes have been developed. Such valves are capable of acting as power amplifiers at frequencies up to 400 Mc/s. In certain cases the capacitor, maintaining the screen at cathode potential with reference to the radio frequencies, is incorporated in the envelope to avoid the reactance of leads outside the valve. The capacitance used is small, but large enough to be effective at the higher frequencies. For lower frequency work it can be augmented by an external capacitor; in this case the reactance of the leads is no longer important. For such very high-frequency work the anodes are usually brought out through the top of the valve. More will be said about this in Chapter 10. A typical radio-frequency power amplifier circuit, using a twin beam tetrode, is shown in Fig. 8.4.

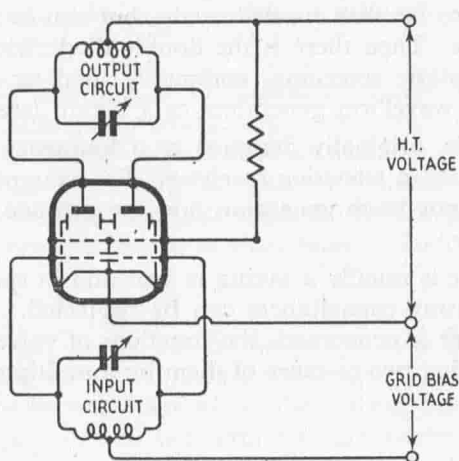


Fig. 8.4. Double beam tetrode operating as a push-pull r.f. amplifier.

Multi-Purpose Valves

A more elaborate type of multi-unit valve is one which includes two or more different units in the one envelope. Of such is the double-diode-triode which incorporates, as its name suggests, two diodes and a triode, all sections usually having a common cathode. This valve operates simultaneously in the demodulator circuit and in the first audio- or video-amplifier stage of a receiver, the double diode demodulating with one section and providing a rectified voltage for use in automatic gain control with the other. Circuits for this application are rather elaborate, and are discussed in Volume 4. Another similar valve, performing the same three functions of demodulation, rectifica-

tion and amplification is the double-diode-pentode. The conventional representation of these two valves is shown in Fig. 8.5, where (a) is the triode and (b) the pentode.

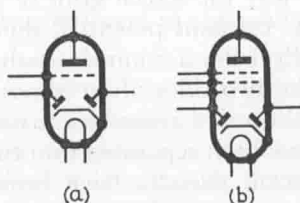


Fig. 8.5. Conventional symbols (a) for a double-diode-triode and (b) for a double-diode-pentode.

Valves incorporating a pentode and a single diode, both operating from the same cathode, are sometimes used as signal amplifiers. In this case the diode unit provides the d.c. signal for automatic gain control. The single-diode-triode is used as an audio amplifier, the diode section providing the automatic gain control voltage. A less common combination which is used at higher frequencies is the diode-triode frequency changer. Here the triode energizes the local-oscillator circuit, and the diode is the mixer. This valve performs the same functions as the more familiar triode-hexode, but can be used at frequencies as high as 100 Mc/s. Then there is the double diode with separate cathodes and internal electrostatic screening, commonly used as a demodulator and a.g.c. valve, and in waveform generators as a clamp (see Volume 4).

The triode-pentode, originally designed as a frequency changer, has found many uses, particularly in television receivers. For example, the triode section may be used in the saw-tooth generator, and the pentode in the output stage of a frame time base.

The multiple valve is usually a saving in cost and in space, but it can only be used where the stray capacitances can be neglected. For this reason, so far as the set designer is concerned, the functions of valves must be carefully studied before replacing two or more of them by a multiple type.

CHAPTER 9

High-Power Valves

IT HAS BEEN pointed out earlier in this volume that there is a fairly clear line of demarcation between 'signal' or voltage amplifiers on the one hand and power amplifiers on the other, and this despite the fact that fundamentally both amplify power.

When a valve is used to transform d.c. power to a.c. power, whether as an amplifier or as a self-sustained oscillator, it is of some importance that its efficiency of transformation should be reasonably high, for otherwise much power would be wasted. It has already been noted that the anode efficiency of a class C amplifier is about 70 to 80%; i.e., if the d.c. input power from the h.t. supply is 100 watts, the a.c. output power is 70 to 80 watts. Only 20 to 30 watts are wasted. Similarly, a class C oscillator should have about the same efficiency. A signal amplifier, on the other hand, may easily have an anode efficiency of less than 10%, but this is of little importance from an engineering stand-point, since the total h.t. input power is seldom more than two or three watts, and in general is less than one watt.

In this chapter it is proposed to discuss the special features possessed by valves capable of handling appreciable a.c. power. It will be found that these special features are all bound up with the fact that their operation is class B or class C. It will also be found that the general character of such valves is the same whether the power involved is 10 watts or 10 kilowatts. The chief difference between these two extremes is in their relative sizes.

Cathodes

Receiving valves and low-power transmitting valves generally have oxide-coated cathodes since, as was explained in Chapter 2, oxide-coated emitters are the most efficient known. But where the cathode current is large and the anode voltage is high, oxide-coated cathodes deteriorate rapidly. This is due to the bombardment of the cathode surface by positive ions. Although the gas pressure in a valve is extremely low, there are always enough gas molecules present to give rise to some positive ions. Where the cathode current and anode potential are large, these ions may be sufficient in number to affect the life of the cathode.

When oxide-coated cathodes can no longer be used because of ionic bombardment, thoriated tungsten cathodes replace them. Thoriated cathodes are not so sensitive to positive-ion bombardment as are oxide-coated cathodes, but even they cannot be used in the largest valves. Where the cathode current is greater than an ampere and the h.t. is greater than 10,000 volts, pure tungsten cathodes are preferred. These require many more watts to give the required emission, but the life of a cathode is proportionately increased.

A rough rule for the use of the various cathodes is that oxide can be used at voltages up to about 1,500 and total emission currents less than 500 mA. Thoriated cathodes are useful up to about 6,000 volts and 8 A. Above these figures pure tungsten is essential.

An apparent exception to these rules is where a cathode operates under 'pulsed' conditions; i.e., where the h.t. is applied periodically for a very short time, as, for example, in a radar transmitter. If the pulses, each of time duration p , recur f_r times per second, the 'duty cycle', r , of the valve, the ratio of the time it is working to the total time, is given by

$$r = pf_r$$

If r is less than $1/10$ and f_r is of the order of 1000 or more, two new cathodic phenomena emerge which are very important. First, the total emission obtainable from a cathode under such pulsed conditions is 50 to 100 times greater than under corresponding conditions with a constant h.t. voltage. Secondly, owing perhaps to the fact that positive ions accelerate very slowly compared with electrons and therefore may not reach the cathode while the h.t. is applied, a pulsed cathode does not deteriorate at high voltages. In fact, oxide-coated cathodes are used in pulsed valves with anode voltages up to 20,000. The same explanation applies to all valves in which the anode current is cut off for the greater part of the time. In a blocking oscillator, for example, anode current only flows for a small part of each cycle: over the greater part of the cycle the grid is so negative that no current can flow. This increases enormously the peak current which can be taken from the cathode, and enables a small receiver valve to cope with currents of the order of an ampere.

Power-valve cathodes may require up to 2,000 watts of heater power. The problem of dissipating this heat is by no means simple. Thick leads of good electrical conductivity are essential. The heater current may be as much as 100 A, and this current must not be allowed to heat the valve until it reaches the actual emitter. Even with small power valves, using 3 to 5 watts of heater power, it is important to see that the seals where the heater pins enter the envelope (usually glass) are kept cool. Bad contacts can be the cause of valve failure.

Anodes

The anode of a power valve may be called upon to dissipate as heat a power of 10 kW. Even in the case of a small power valve an anode dissipation of 10 W is quite common. Two types of anode may be distinguished. First, where the heat produced at the anode does not exceed 100 watts, it is usually possible to dissipate it inside the valve envelope by radiation alone. For the anode structure itself a metal of high melting point and good heat emissivity is employed. Typical metals, taken in order of their dissipation per unit surface, are nickel, carbonised nickel, graphite, tantalum and molybdenum. When the power to be dissipated is greater than 100 watts, it is usually necessary to connect the anode to an external radiator by means of a good conductor, and to cool the radiator either by air blast or by a stream of water. This leads to a very different type of valve construction in which the object is to conduct the heat away through the envelope as fast as possible. Copper is often used as the metal conducting the heat to the outside of the valve. Fortunately, copper can

be sealed to glass in a variety of ways. In large power valves the seal is often made on the surface, internal or external, of a cylindrical copper tube, the copper being very thin at the actual joint. These are called 'Housekeeper' seals after their inventor.

A power valve is often characterized by its anode dissipation, for it is this factor which determines, more than anything else, the amount of a.c. power available from the valve and its circuit. If a valve has an anode dissipation of 100 watts, it is likely that it will be capable of producing 300 watts of a.c. power as a class C amplifier, providing that the frequency is not too high. Valve manufacturers match the cathode emission to the anode dissipation, and it can usually be assumed that the cathode will be capable of emitting a current sufficient to produce a power output equal to 3 or 4 times the anode dissipation. This may not be the case, however, with valves used at ultra-high frequencies.

Grids

Power valves differ radically from voltage-amplifier valves in the average control-grid current which they take during operation. As we have seen, valves used in voltage amplifiers rarely operate under conditions which give rise to grid current, whereas it is impossible to achieve optimum efficiency in a class C amplifier without driving the grid positive with respect to the cathode.

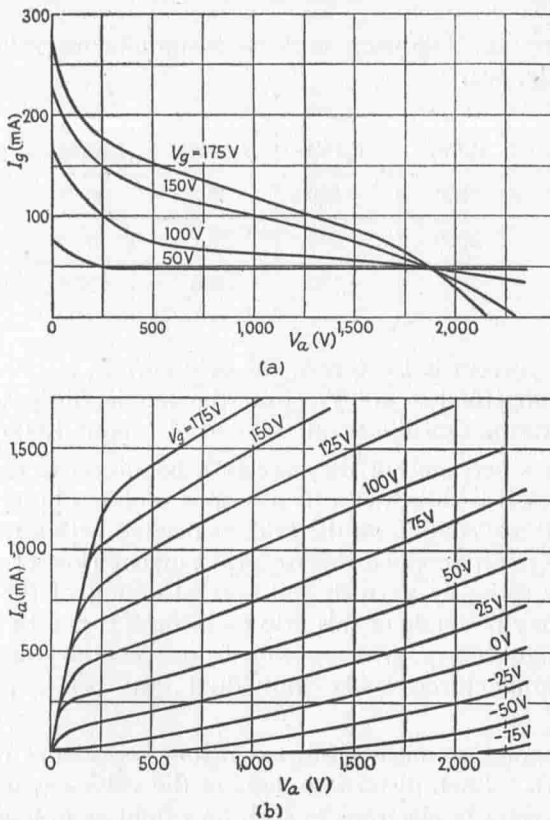


Fig. 9.1. Grid-current - anode-voltage (a) and anode-current - anode-voltage (b) curves of the Mullard TY2-125 transmitting triode.

As a result of this, more care must be taken with power valves to ensure that the control grid (and, for that matter, the screen grid of a beam tetrode) is able to dissipate the heat generated at its surface. Also it should be noted that the control grid becomes hot even in the absence of grid current, owing to its proximity to the high-temperature cathode. Grids are almost always made of wires of a metal of high melting point, such as molybdenum. These wires are wound into grooved supporting posts and are fixed in position by turning the metal of the support over the grooves. In a very few cases grids are made of copper to obtain high electrical and thermal conductivity, but this is a special construction only used in disc-seal triodes or tetrodes, valves used at ultra-high frequencies. To exemplify some of the facts just mentioned, the simple analysis of an actual valve under working conditions may be made.

A typical power triode is the Mullard TY2-125, rated to dissipate 135 W at the anode, and operating at voltages up to 2,500. Its characteristics are shown in Fig. 9.1.

When this valve is being used as a class C amplifier, typical operating conditions are shown as follows:—

$$V_a = 2,000 \text{ V}, I_a = 200 \text{ mA}, I_g = 40 \text{ mA}, V_g = -150 \text{ V (bias voltage)}$$

Amplitude of drive voltage = 300, giving an output power = 280 W at an efficiency of 70%.

From these figures the following table of instantaneous values may be constructed, using Fig. 9.1.

$V_a(\text{V})$	$V_g(\text{V})$	$I_a(\text{mA})$	$I_g(\text{mA})$
>2000	-450	0	0
2000	-150	0	0
250	+150	1000	150

When the anode current is 1,000 mA, the grid current is 150 mA. The input power is $200 \times 2000 / 1000 = 400 \text{ W}$, and the anode dissipates $400 - 280 = 120 \text{ W}$. At this dissipation the anode is red hot (about $1,000^\circ \text{ K}$).

When grid wires become hot they are in a condition to emit electrons and 'grid emission', as it is called, can be a serious problem in large power valves. A trace of the cathode coating on the grid, associated with a temperature of the order of $1,000^\circ \text{ K}$, can produce inverse grid current (loss of electrons by the grid) which may seriously perturb the normal action of the valve. On the other hand use may be made of this grid emission if it can be controlled. It is not easy to distinguish true grid emission from secondary emission in a large valve. Most manufacturers make individual tests before passing the valve for use.

A grid may become seriously distorted if its temperature rises much above that of its supports. Such distortion changes the valve capacitances and may even distort the beams of electrons to such an extent as to lower the efficiency. Special steps are taken in large valves to keep the grid wires stretched at all working temperatures.

Leads to Electrodes

It has already been pointed out that the heater leads must be sufficiently robust to carry without undue heating the current required to heat the cathode. A power valve operates with a large swing of potential both at the grid and at the anode. These swings give rise to considerable alternating currents in the anode and grid leads. At the higher frequencies, where 'skin effect' is important, the leads must have a good conducting surface to avoid resistive losses. Some metals used for sealing to glass have comparatively poor qualities as conductors. For example, tungsten pins tend to become oxidised and to dissipate power in this way. Care should therefore be exercised in examining and, if necessary, cleaning the valve leads wherever large powers are being handled.

To complete this description, Table 9.1 lists some facts regarding representative power valves. The values given are averages and do not refer to particular valves. It will be observed that beam tetrodes are preferred at anode dissipations up to 100 watts. Above this value triodes only are possible at the present time.

TABLE 9.1

Heater Power (W)	Max I_k (mA)	Max V_a (V)	Max I_{g1} (mA)	Anode Diss. (W)	g_m (mA/V)	Class C Output (W)	Valve Type
3.5	50	400	5	7.5	2	10	Beam Tetrode
6.0	150	600	5	25	6	80	Beam Tetrode
50	225	2,000	25	100	4	300	Beam Tetrode
100	750	4,000	100	500	6	1,500	Triode
250	1,400	5,000	150	1,500	6	4,000	Triode
1,000	1,600	12,000	350	4,000	5	10,000	Triode

CHAPTER 10

High-Frequency Effects

THE BEHAVIOUR of a valve at high frequencies is often quite different from its behaviour at low frequencies, and that for a variety of reasons. Were it not for the fact that a number of new effects begin to become important as the frequency of operation is increased, the radio-frequency spectrum would have extended to a wavelength of 1 mm shortly after valves were invented. As will be seen from what follows, when a valve of a certain size and structure is made, an upper limit to the frequency band over which it can work is automatically set. To obtain operation at a higher frequency a smaller valve, or one of a different construction, or one employing a different electronic principle, must be made. There are a number of reasons why valves fail in performance at high frequencies. These will be discussed in this chapter in their order of importance as the frequency is increased. The modifications which have been made to correct for this failure will also be described in so far as they apply to conventional valve operation. In the next chapter valves operating on different principles will be considered.

Dimensions of Valve Electrodes and their Leads

It was noted in Chapter 3 that the triode valve comprises a generator having a certain internal resistance, r_a , and certain internal capacitances due to the sizes and relative positions of the various electrodes. Indeed, as was described at the beginning of Chapter 4, the reason for discarding the triode valve as an amplifier in favour of the beam tetrode or pentode is the existence of the comparatively large grid-anode capacitance in the triode. As the frequency increases, the reactance of a capacitance decreases, and so, as the frequency of operation is increased, the valve inter-electrode capacitances become more important. For example, suppose that the input capacitance of a valve is 10 pF. Then its reactance at a frequency of 100 kc/s is about 160,000 ohms. If the frequency is increased to 300 Mc/s, the reactance falls to about 53 ohms, and such a low reactance can seldom be neglected. Certainly the impedance presented at the input terminals (grid to cathode) can no longer be regarded as very large.

Similarly, the leads to the valve electrodes cease to have negligible inductive reactances at the higher frequencies. A straight piece of wire, isolated from all other conductors, has an inductance which, for an inch of No. 20 s.w.g. copper is about 0.015 μ H. At a frequency of 100 kc/s this is a reactance of 0.009 ohm and may usually be neglected. At a frequency of 300 Mc/s, the reactance is about 28 ohms which is certainly not negligible and may be very important. To illustrate the importance consider the circuits shown in Fig. 10.1.

The impedance across the grid-cathode pins in (b) at 100 kc/s is capacitive and equal to 160,000 Ω . The impedance in (c) at 300 Mc/s is inductive and

equal to $3\ \Omega$. At a frequency of about 290 Mc/s, corresponding to resonance, the leads and the electrodes connected to G and K would have zero reactance. In the absence of resistance, the terminals would appear to be short-circuited.

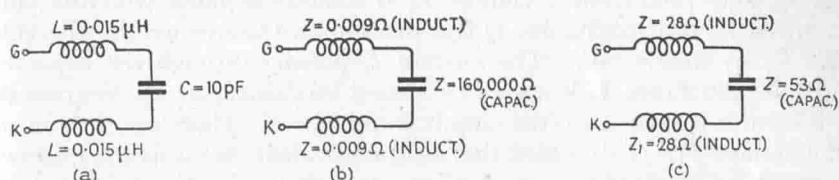


Fig. 10.1. Equivalent circuit of grid-cathode path of valve (a) and the reactances of the elements (b) at 100 kc/s and (c) at 300 Mc/s.

The matter is complicated still further by the resistances of the leads. As the frequency increases, the resistance of a given length of wire of given cross-sectional area increases owing to the decreasing depth of penetration of the electron current. This 'skin effect' causes the resistance of a length of No. 20 s.w.g. copper wire at a frequency of 300 Mc/s to be almost 60 times its d.c. resistance, or 1.5 ohms per metre. Although the resistance of an inch of wire is still small, therefore, at 300 Mc/s, it is not always negligible, particularly when the h.f. current carried by it is large.

Input Impedance of Valve at a High Frequency

The results of the preceding paragraph may now be used to investigate a practical case. Over an important range of frequencies (from 20–300 Mc/s, say) the following circuit elements inside the valve envelope may be neglected:— (a) the resistances of the leads and (b) the grid-anode capacitance. The lead resistances may be neglected, since, as shown in the last paragraph, they are small. The grid-anode capacitance is not of course negligible in itself, but the use of special circuits neutralizes its effect. In particular the 'earthed-grid amplifier' (see Volume 4) almost entirely eliminates the Miller effect.

Over the same range of frequencies the following circuit elements inside the valve envelope must be taken into account:— (a) the inductances of the leads, and particularly that of the cathode lead, (b) the grid-cathode capacitance.

Making these assumptions, the input impedance of an amplifier valve may now be studied. The valve may be represented by the circuit shown in Figure 10.2. K, G and A represent the valve pins connected to the cathode, grid and anode respectively. L_g , L_a and L_k are the lead inductances; C_{gk} is the grid-cathode capacitance. When a signal V_s at a frequency f is applied between G and K a current I_s flows via L_g , C_{gk} and L_k to K. This produces a voltage V_g between the actual grid and cathode. This signal V_g produces a current I_a in the anode circuit in the usual manner. So long as the frequency is not too high, the inductance L_a can be incorporated in the external circuit as part of an LC circuit, forming a resistive load. The inductance L_a can, therefore, be neglected. Similarly, since the anode current passes through L_k , and since for all frequencies at which the valve is useful I_a is large compared with I_s , the voltage across L_k is almost entirely due to I_a , and is large compared with that produced across L_g . It is assumed that the inductances L_k and L_g are of the same order of magnitude. The inductance L_g may also, therefore, be neglected.

With these assumptions it may easily be shown that the cathode inductance produces the effect of a resistance connected between grid and cathode, and thus reduces the input impedance of the valve. The anode current flowing through L_k and producing a voltage V_k is almost in phase with the voltage V_g for, when V_g is a maximum, I_a is a maximum. Consequently, the voltage V_k leads V_g by nearly 90° . The current I_s flowing through the capacitance C_{gk} and the inductance L_k leads V_g by almost 90° , since, at all frequencies at which the valve can be used, the circuit is capacitive. Therefore, I_s is in phase with the voltage V_k . This means that L_k is equivalent, so far as the grid circuit is concerned, to a resistance in series (or in parallel) with the capacitance C_{gk} .

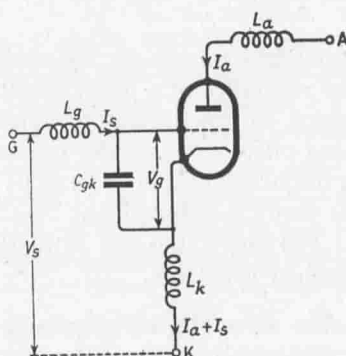


Fig. 10.2. Triode valve showing the lead inductances and the grid-cathode capacitance.

The effect, therefore, of the inductance in the cathode lead is to give rise to a resistive grid-cathode admittance. It is shown in Appendix 10.1 that the value of this admittance is $4\pi^2 f^2 g_m L_k C_{gk}$. Since, in fact, there is no ohmic resistance in the circuit, the power taken from the signal source is not lost; it appears in the output. But it means that the power required to produce a given V_g between grid and cathode increases as the square of the frequency. If $f = 300$ Mc/s, $g_m = 10$ mA/V, $L_k = 0.015$ μ H, $C_{gk} = 10$ pF, the power required to produce 1 volt between grid and cathode is about 5 milliwatts, using the formula which has just been given.

Reducing Valve Inductance and Capacitance

The obvious modification to reduce L_k and C_{gk} is to reduce the lengths of the leads and to make the valves generally smaller. This is sometimes useful and it has been done to some extent in miniature and sub-miniature valves, which operate satisfactorily at frequencies at which their larger counterparts are very inefficient. Miniature valves are not made for this reason, but they exemplify one method of increasing efficiency at very high frequencies.

This obvious modification does not take us very far. Two factors limit the use of very small valves. One is that they are inconveniently small to manufacture economically. Human hands and tools are too large. The other is that the amount of power which can be handled by a valve decreases as its size decreases, since the smaller cathode will not give the necessary electron

emission and the smaller anode becomes too hot. If, therefore, any considerable improvement in valve performance is required, some other method of reducing L_k and C_{gk} must be sought.

The second modification which is used is to decrease as far as possible the inductance L_k (and, incidentally, L_g) by using much heavier leads to the cathode (and grid). The inductance of a conductor decreases rapidly at high frequencies as its surface area increases, and so valves are made in which the cathode, grid and anode connections are either cylinders of considerable diameter or short plates of considerable surface area. By this method, and without seriously affecting the ease of manufacture, ordinary valves may be made to give much improved performance at the higher frequencies.

The limiting case of this technique is to be found in the so-called 'disc-seal' valves which are now much used at ultra-high frequencies. An axial section of a typical disc-seal triode is shown in Fig. 10.3.

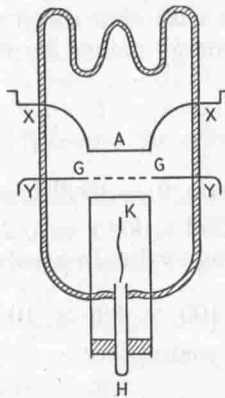


Fig. 10.3. Axial section of disc-seal triode.

Here the connection from the cathode to the external circuit is a cylinder and disc, and the connections to grid and anode are of similar nature. The active electrodes of a disc-seal valve usually form a planar structure, although this is not invariably the case. In the figure the disc cathode K faces a plane grid mesh G through which the electrons pass to the anode A. In the corresponding tetrode a screen is interposed between grid and anode and in this case cylindrical electrodes are most commonly used. The electrode connections are the most important feature of the valve. They are copper annular discs, usually about 10 thousandths of an inch thick, sealed through the glass as shown at X and Y in Fig. 10.3. The whole valve is symmetrical about an axis at right angles to the plane of the electrodes. Such valves are now made for frequencies between 300 and 3,000 Mc/s. They are used in special concentric-line circuits which will be considered later.

Time of Transit of an Electron in a Valve

In all valve actions so far studied it has been tacitly assumed that the electrons move from the cathode to the anode in a negligibly small time. For example, it has been assumed throughout that the anode current responds instantaneously to a change in the grid voltage, although the thoughtful reader

are virtually brought to rest by the space-charge. It is shown in Appendix 10.2 that the transit time in a diode where the current is space-charge limited is $3/2$ times the transit time in the same diode when the current is temperature limited. So the effect of transit time on valve action begins to be felt at a lower frequency than is suggested by the previous calculation.

Effects of Finite Time of Transit in a Valve

When the time taken by an electron to traverse a valve is no longer a negligible fraction of a period of the high-frequency alternating voltage, two major effects appear which combine to destroy the utility of conventional valve action.

First, the finite time of transit between the cathode and the control grid T_g gives rise to an additional conductance in parallel with C_{gk} and equal to $4\pi^2 A g_m f^2 T_g^2$, where A is a non-dimensional constant for a given valve, g_m is the mutual conductance and f is the frequency of the applied signal. This conductance still further lowers the grid-cathode input impedance, and dissipates the power available from the signal source. Secondly, if a valve is used in an oscillator at a frequency high enough to exhibit transit-time phenomena, the time T_a of transit between the grid and the anode causes the anode current to lag behind the grid voltage. This lowers the efficiency of the oscillator.

Grid Current Induced by an Electron Stream

To understand the effects mentioned in the last paragraph it is necessary to enquire a little more closely into what happens when an electron passes through a grid. The passage of the electron, as we shall see, induces a momentary or transient current in the grid which is of little importance at low frequencies but is of major importance at high frequencies. Indeed the action of the more recent valves introduced to operate at frequencies above 1,000 Mc/s depends very largely upon this momentary current.

Let Fig. 10.5 represent an aperture in the control grid of a valve, and let $P_1, P_2, P_3, \dots, P_8$ be successive positions of an electron passing through the aperture with a steadily increasing speed. This is the normal state of affairs in a negative-grid amplifier valve when a current is passing. As the electron approaches, it induces an opposite positive charge on the grid, just as the charge at a certain part of a conductor at a given potential is increased by bringing another oppositely charged conductor up to that part. Since the electron's speed is continuously increasing, the charge induced diminishes as the electron recedes more rapidly than it increased, and if the temporary charge is graphed against time, a curve similar to that of Fig. 10.6(a) is obtained.

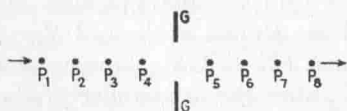


Fig. 10.5. Successive positions of an electron passing through a grid G are shown here.

The current I_g to the grid, producing the temporary charge Q is the rate at which Q increases or diminishes. I_g is therefore the gradient of the curve of

Fig. 10.6(a). This is shown in Fig. 10.6(b) on the same time scale. The average value of I_g is, of course, zero, for the electron is not captured by the grid.

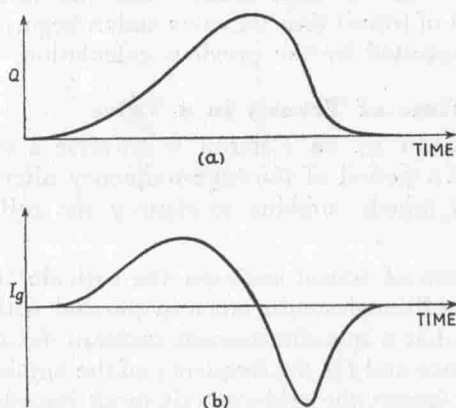


Fig. 10.6. Variation with time of the charge (a) induced on a grid by an electron passing through it and (b) the resulting current.

Fig. 10.6(b) represents a transient pulse of current produced at the grid by a single electron passing through it. If a continuous stream of electrons passes through the grid, each gives rise to a little pulse of the type shown, but *if the electron current is constant*, these pulses will cancel out on the average and give rise to nothing more than a small random current to the grid. If, however, as is the case in an amplifier valve, the grid voltage varies sinusoidally, producing roughly sinusoidal variations in the electron current, then the transient pulses when the current is large will be greater than when the current is small, so that a resultant transient having roughly the shape of the curve of Fig. 10.6 will occur once per cycle of the electron current. Moreover, the resultant pulse will occur exactly when the current is a maximum and so, as is shown in Fig. 10.7, a relation will subsist between the phase of the grid voltage and the resultant pulse.

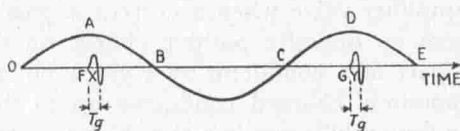


Fig. 10.7. The induced grid-current pulses FX and GY occur at the positive peaks of the input voltage OABCDE.

Curve OABCDE is the grid voltage graphed against time, and the resultant pulses at the grid are shown at X and Y. The time of transit T_g is represented by the distance $FX = GY$, showing that the maximum current reaches the grid at time T_g after the occurrence of the maximum grid-cathode accelerating voltage. Now consider the effect of these resultant pulses of grid current. Each of the diagrams of Fig. 10.8 shows pulses on the same time scale; (a) corresponds to a low-frequency variation of grid voltage, (b) to a high-frequency variation, corresponding to the onset of transit-time effects, and (c) to an ultra-high frequency variation, when the time of transit is a large fraction of a period of the grid voltage.

Clearly the transient in Fig. 10·8(a) is all over in such a small fraction of a period of oscillation that it can contribute nothing to the action of the valve and may be neglected. In diagram (b) the transient can no longer be neglected. It will give rise to a component of current at the signal frequency which will be small but not negligible. From the mode of repetition of the pulses, this component will have the phase and magnitude shown by the dotted curve. In diagram (c) the signal frequency has been raised to such a value that the transient fills a large fraction of a period and will therefore have a relatively large component at that frequency. This is again indicated by the dotted curve.

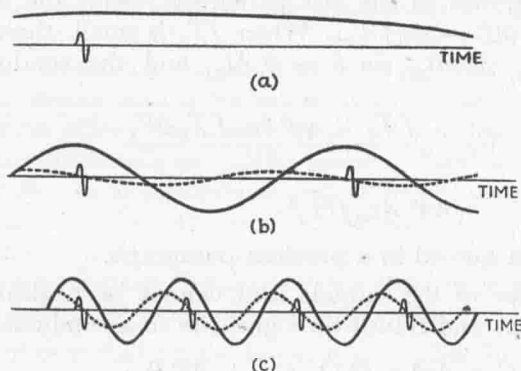


Fig. 10.8. At (a) the induced grid-current pulse occupies so small a fraction of the period of the signal voltage that its effect is negligible. At (b) it occupies a relatively larger period and is beginning to have appreciable effect, while at (c) because the time of one cycle of the v.h.f. signal is small it has a large effect. The dotted curves show the components of the induced current at the signal frequency.

The magnitude of the component I_{gf} of the induced grid current at the signal frequency f is proportional to T_g/T_o , where $T_o = 1/f$ is the period of the signal, or $I_{gf} \propto T_g f$. This is strongly suggested by the diagrams of Fig. 10·8 for, when T_g/T_o is negligible, the current I_{gf} is negligible, and as T_o decreases (f increases) I_{gf} shows a proportionate increase. The proportionate increase will, however, only hold when T_g/T_o is small. For the extreme case illustrated by Fig. 10·8(c), it would no longer be strictly true.

The phase of the component I_{gf} of the induced current may be deduced from Fig. 10·7. The zeros of the component at frequency f must occur at X and Y, so that zero current lags behind *maximum* grid voltage by an angle $\theta = 2\pi f T_g$. Therefore the component of current I_{gf} leads the grid voltage by an angle $90^\circ - \theta$. This is what might be expected. When θ is negligible, the extremely small I_{gf} is 90° out of phase with the grid voltage. As the frequency increases, the magnitude of I_{gf} increases, and it leads the grid voltage by an angle less than 90° , thus giving rise to a resistive component in the grid circuit,

Effect of Transit Time on the Input Impedance of a Valve

► The result of the preceding paragraph may now be applied to an amplifier, the grid voltage of which is changing at frequency f . The valve current is

proportional to the grid voltage or, as stated in previous chapters, $\delta I_a = g_m \delta V_g$, where δI_a is the change in anode current produced by a change in the grid voltage δV_g , g_m being the mutual conductance and changes in the anode voltage being neglected. The induced current is proportional to the change in anode (electron) current, or, writing δI_g for the induced grid current, δI_g varies directly as $g_m \delta V_g$. But, as has been shown in the preceding paragraph, the component δI_{gf} of δI_g at frequency f , varies directly as $f T_g$. Therefore $\delta I_{gf} \propto g_m f T_g \delta V_g$ or, using a numerical constant characteristic of the valve,

$$\delta I_{gf} = 4 \pi^2 A g_m f T_g \delta V_g$$

This current, as shown in the last paragraph, leads the voltage δV_g by an angle $90^\circ - \theta = 90^\circ - 2\pi f T_g$. When $f T_g$ is small, therefore, the resistive component of δI_{gf} is $\delta I_{gf} \sin \theta \approx \theta \delta I_{gf}$ and the conductance G_T due to transit-time is

$$\begin{aligned} G_T &= \frac{f T_g \times 4 \pi^2 A g_m f T_g \delta V_g}{\delta V_g} \\ &= 4 \pi^2 A g_m f^2 T_g^2 \end{aligned}$$

This is the formula quoted in a previous paragraph.

If the inductance of the cathode lead cannot be neglected, the combined effects of inductance and transit-time give rise to a conductance.

$$G = 4 \pi^2 g_m f^2 (L_k C_{gk} + A T_g^2) \quad (10.1)$$

Effect of Transit-Time on Output Impedance

In certain cases, notably where the valve is part of an oscillator, the transit-time lag of the anode current relative to the grid voltage may have important effects on the output circuit. For example, the maximum anode current in a power valve may not occur when the anode potential is a minimum. This causes greater heat dissipation in the valve and decreases the anode efficiency. The effect is only important in special cases. In a power amplifier the phase of the anode voltage adjusts itself relative to the current pulse so that the current maximum and voltage minimum occur simultaneously. In an oscillator this is not necessarily true. In fact if there is the usual 180° phase shift between the grid and anode voltages, the anode current must lag behind its low-frequency phase. For this reason valves function better as power amplifiers than they do as power oscillators at ultra-high frequencies.

Reducing Transit Time Effects

The only obvious modification to valve structure which might be expected to decrease transit time is to decrease the inter-electrode distances. This is done in ultra-high-frequency triodes and tetrodes, where the grid-cathode spacing is made as small as is practicable. The plane electrode construction of certain disc-seal valves, as illustrated in Fig. 10.3, materially assists the designer in keeping the grid-cathode distance as small as possible. Most valves of this type are now manufactured with cathodes which can be adjusted externally to their final positions after the envelope has been evacuated and sealed off. This allows the final cathode position to be fixed by electrical measurement, and this technique has certainly improved average valve performance. But the spacing between grid and cathode is limited by several factors such as the

radiation of heat from the cathode, the mechanical accuracy of tools and jigs and the thickness of electrodes. There is, therefore, a definite upper frequency limit for conventional valves. This limit appears to lie between 3,000 and 6,000 Mc/s. Beyond that triode or tetrode construction is useless and a new valve principle must be invoked.

Integration of Valve and Circuit

It will be clear from the foregoing that the effects of increasing frequency on conventional valve action appear in the following order:— First, the inductances of the leads become troublesome, particularly that of the cathode lead which in association with the grid-cathode capacitance simulates a conductance in parallel with the grid-cathode space. This absorbs power from the generator driving the valve. Secondly, the resistances of the leads become important. Any step taken to reduce lead inductances, however, reduces lead resistances. Thirdly, transit-time effects begin to appear. These first of all bring about an increase in the grid-cathode conductance. They also affect the phase angles of an oscillator so as seriously to reduce its efficiency. The steps taken in succession to overcome these effects are (a) to reduce lead inductances and to incorporate all valve inter-electrode capacitances in the external circuit and (b) to decrease inter-electrode distances to the limit of reasonable mechanical control, thus displacing the onset of transit-time effects to a higher frequency.

It will now be shown that by the use of a concentric line the valve may be made so much a part of the external circuit that it is impossible to identify the valve leads in any case. This technique has raised the upper limit of usefulness of the conventional valve from about 300 to about 3,000 Mc/s, and has made transit-time effects the only real limitation.

Consider the circuit elements shown in Fig. 10.9. The high- Q parallel-tuned circuit (a), consisting of an inductance, a capacitance and a small resistance, resonates at a frequency f given approximately by $1/f = 2\pi\sqrt{LC}$ and presents a high resistive impedance L/CR between the terminals. Such an element is

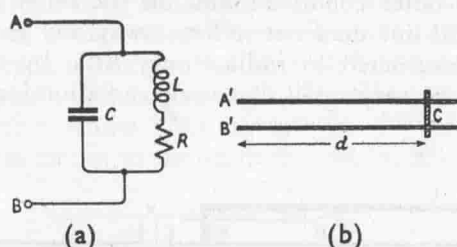


Fig. 10.9. Lumped LCR circuit (a) and its equivalent in the form of a quarter-wave line (b).

widely used in radio-frequency amplifiers, oscillators, frequency changers, etc. At very-high and ultra-high frequencies it is difficult to realize this circuit in practice, since L and C would be so small that they could not conveniently be made, and if they were made, their power handling capacity would be very small.

But, as is shown in Volume 5, a pair of conductors running parallel to each other [Fig. 10.9 (b)] and short-circuited at C by a connecting conductor behave very much like the circuit (a) if the distance A'C is approximately given by $d = \lambda/4 = c/4f$, where c is the velocity of electromagnetic

waves in air $= 3 \times 10^8$ m/s, λ is the wavelength in metres and f is the corresponding frequency. If a signal of frequency f is applied to the pair of conductors a standing wave is set up, having maximum voltage variation across A'B'. In fact such an 'open pair' of conductors acts as a resonator and presents a high impedance at the end A'B'.

A similar resonance effect occurs whenever the distance A'C equals an odd multiple of $\lambda/4$ (i.e., at $3\lambda/4$, $5\lambda/4$, etc.) and, as will appear later, there are some cases where it is convenient to use one of these longer lines.

The distance d is very large for ordinary radio frequencies. For example, if $f = 1$ Mc/s, $d = 75$ metres. But if $f = 100$ Mc/s, $d = 75$ cm, which is a length comparable with the linear dimensions of ordinary circuits of type (a) at lower frequencies. By using a 'line', therefore, it is possible to obtain electromagnetic resonators at ultra-high frequencies the dimensions of which are comparable with those of ordinary 'lumped' circuits at low frequencies.

Another form of 'line' which is even more valuable is composed of coaxial conductors, as shown in Fig. 10.10, where (a) is an axial section and (b) is a cross-section. Again, if $d = \lambda/4 = c/4f$, the line resonates at frequency f .

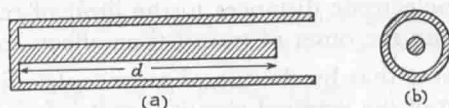


Fig. 10.10. Axial (a) and cross (b) sections of a coaxial resonator.

In both cases, the device is known as a quarter-wave line or quarter-wave resonator.

The coaxial or 'concentric-line' resonator possesses certain advantages over the open-line resonator. Since all the high-frequency currents flow on the inner surface of the outer conductor and on the outer surface of the inner conductor, the coaxial line does not radiate away any power. The open-line resonator can be constructed to radiate very little by making the distance between the conductors very small, but some radiation inevitably escapes.

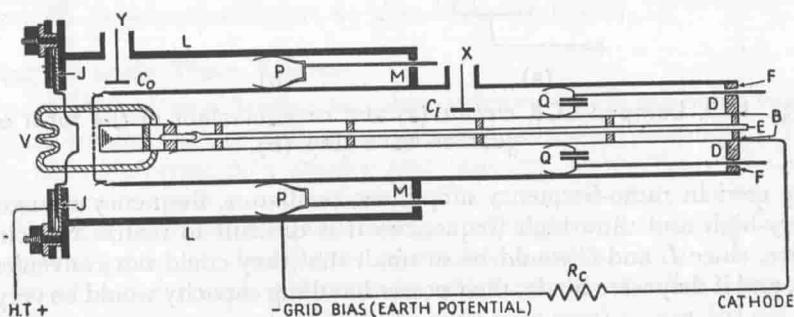


Fig. 10.11. Complete amplifier with disc-seal valve V. There are two coaxial resonators, one comprising the grid and anode cylinders and the other the grid and cathode cylinders.

Disc-seal valves are manufactured to be used specifically in coaxial lines; the connections to the cathode, grid and anode are suitable for direct fitting into a double coaxial-line circuit. Fig. 10.11 illustrates a complete amplifier, incorporating a disc-seal valve.

V is the valve, the anode, grid and cathode surfaces being opposite each other. The cathode lead is extended outside the valve as the smallest axial cylinder, emerging at B through an insulating annular disc D. The heater lead is brought down inside the cathode cylinder and emerges at E. The grid disc with its edge pressed over (see Fig. 10.3) is wedged firmly into a second cylinder, coaxial with the cathode cylinder, holding the insulating disc at F. The anode is clamped against a thin annular sheet of mica, J, which is set between the anode and the face of the outermost coaxial cylinder, L. This cylinder is braced to the grid cylinder at M. There are, therefore two coaxial-line circuits, one formed by the grid and anode cylinders and the other formed by the grid and cathode cylinders. The length of the grid-anode line is determined by the annular tuning plunger P which may be controlled by rods or any other suitable mechanical device. The length of the grid-cathode line is also determined by an annular plunger Q, but this second plunger is more complicated for the following reason. Although the plungers are intended to terminate the lines, and must therefore act virtually as short-circuits at the high frequency, the valve electrodes are at different mean voltages. The mica disc, J, allows the anode to be maintained at a high voltage, and the capacitance formed by the anode and the plate on the other side of the mica is sufficiently large to act as a virtual short-circuit at frequencies of 300 Mc/s and above. But the plunger Q must separate the grid and the cathode cylinders statically and yet short-circuit them at high frequency. It therefore contains a built-in capacitor (or several) and this is indicated on Fig. 10.11 by the parallel lines.

At X and Y flexible coaxial lines may be attached to inject power into the amplifier (X) and to extract the amplified power (Y). The coupling elements in each case are small variable capacitors, C_1 and C_0 .

The d.c. supplies are directly connected to the circuit, the heater connection, E, and the cathode, B, to the heater circuit, B to the negative of the h.t. line, the anode clamping ring to the positive of the h.t. line, and the external body of the amplifier (at grid potential) to earth. It will be noted that the negative h.t. is not then at earth potential. The alternative, which is preferred, is to adopt the usual cathode-bias circuit as shown in Fig. 10.12, with the grid at negative

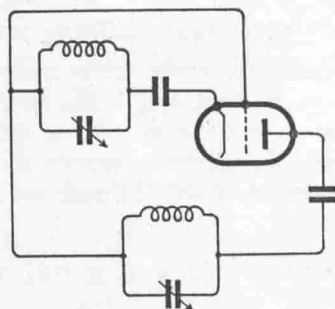


Fig. 10.12. Earthed-grid amplifier stage which is equivalent to the resonator form of Fig. 10.11.

h.t. (earth potential), and the cathode positively biased by means of the anode current flowing in the resistance R_c . Chokes are nowhere required so long as the capacitances due to J and Q are large enough.

The equivalent of the amplifier of Fig. 10·7 drawn for a lower frequency is shown in Fig. 10·12. This is the so-called earthed-grid (grounded-grid) amplifier which is universally used at ultra-high frequencies.

Attention may now be called to the integration of valve and circuit achieved by the amplifier of Fig. 10·11. The following points should be noted:—

- (a) The best possible use has been made of the small inductance disc seals.
- (b) The cathode lead forms an integral part of the 'external' circuit.
- (c) The large diameter conductors keep resistive losses to a minimum.
- (d) The valve inter-electrode capacitances have become terminating capacitances of the concentric lines.

It should also be noted that the length of the grid-cathode line is much greater than that of the grid-anode line of Fig. 10·11. This is not necessarily the case, but it is often done to make it easy to inject power into the grid circuit at the appropriate point. The grid-cathode circuit is usually equivalent to a three-quarter wave line and the grid-anode to a quarter-wave line. This allows the length of the former to be greater than that of the latter.

Finally, it should be noted that the cathode-grid and grid-anode circuits of Fig. 10·11 are not exactly quarter-wave coaxial lines; the valve capacitances, terminating the lines at their 'free' ends, cause the lengths to be less than a quarter of a wavelength. The reduction in length may be readily calculated from the constants of the line in each case. As the frequency is increased, the lengths of the lines decrease, but before the circuits have become unmanageably small, the valve has ceased to operate owing to transit-time loading. Transit-time effects form the ultimate limitation on the use of conventional valves.

APPENDIX 10.1

Input Impedance of a Valve with Inductance in the Cathode Lead

Fig. 10.2 may be reduced to the circuit shown in Fig. 10.13 so far as the input to the valve is concerned.

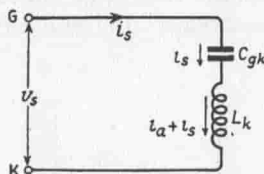


Fig. 10.13. Equivalent circuit of Fig. 10.2 as far as the input is concerned.

Using complex quantities, if v_s is the input voltage and i_s is the input current, and remembering that $i_s \ll i_a$, where i_a is the anode current,

$$v_s = v_g + j\omega L_k i_a$$

where v_g is the voltage between grid and cathode; i.e., across the capacitance C_{gk} .

But $i_a = g_m v_g$ and $v_g = i_s / j\omega C_{gk}$

Therefore $v_s = i_s / j\omega C_{gk} + j\omega L_k g_m i_s / j\omega C_{gk}$

$$= i_s \left[\frac{1}{j\omega C_{gk}} + L_k g_m / C_{gk} \right]$$

Therefore $Z_{\text{input}} = 1 / j\omega C_{gk} + L_k g_m / C_{gk}$ (10.2)

The input impedance contains the series resistance of value $L_k g_m / C_{gk}$.

This series resistance may be replaced by a resistance in parallel with C_{gk} by making the usual transformation. Assume that $\omega L_k g_m$ is small compared with unity.

Then $Z = (1 + jL_k g_m \omega) / j\omega C_{gk}$

and $\frac{1}{Z} = Y = j\omega C_{gk} (1 + j\omega L_k g_m)^{-1} \approx j\omega C_{gk} (1 - j\omega L_k g_m)$

or $Y \approx j\omega C_{gk} + \omega^2 L_k C_{gk} g_m$ (10.3)

and the conductance of the parallel circuit is $4\pi^2 f^2 L_k C_{gk} g_m$.

It is interesting to examine under what conditions $\omega L_k g_m$ is small compared with unity. Let $L_k = 0.01 \mu\text{H}$, $C_{gk} = 5 \text{ pF}$, $g_m = 10 \text{ mA/V}$.

Then $\omega L_k g_m = 10^{-8} \times 10 \times 10^{-3} \omega = 10^{-10} \omega$, whence the approximation made in deriving (10.3) is clearly good for all values of $\omega < 10^9$, or for all values of f , the frequency, less than 150 Mc/s. At this frequency, the conductance is:—

$$\frac{1}{R} = 10^{18} \times 10^{-8} \times 5 \times 10^{-12} \times 10^{-2}, \text{ giving } R = 2,000 \text{ ohms.}$$

This shows the importance of the effect. The source is shunted by a resistive impedance of about 2,000 ohms.

APPENDIX 10.2

Transit-Time in a Space-Charge Limited Diode consisting of Parallel-Plane Electrodes

Referring to Fig. 10.14, consider the electrical field at a point P, distance x from the cathode K. Let the charge density at P be ρ and the electric field strength be X . As is usual, assume that edge effects may be neglected and that X is perpendicular to the electrodes. Let the electrode separation be d .

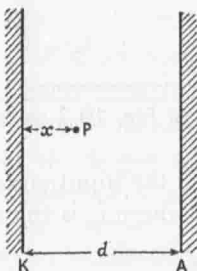


Fig. 10.14. Diagram used in calculating the electric field at P.

Then, assuming a unidimensional flow of electricity, Poisson's equation becomes:—

$$\kappa_0 \frac{\partial X}{\partial x} = \rho,$$

where κ_0 = permittivity of the space between K and A, approximately equal to $10^{-9}/36\pi$, and ρ is the charge density. The total current I_0 at P is the sum of the conduction current I_c and the displacement current I_d . $I_c = \rho u$, where u = speed of the average electron and $I_d = \kappa_0 \partial X / \partial t$ by definition.

$$\text{Therefore } I_0 = \kappa_0 \frac{dx}{dt} \frac{\partial X}{\partial x} + \kappa_0 \frac{\partial X}{\partial t} = \kappa_0 \left[\frac{\partial X}{\partial x} \frac{dx}{dt} + \frac{\partial X}{\partial t} \right] = \kappa_0 \frac{dX}{dt}$$

where dX/dt is the complete differential of X . But I_0 , the total current, does not vary with the time if the potential difference V_0 between K and A is constant. Therefore, if V_0 is constant, dX/dt is constant.

Consider now the equation of motion of an individual electron,

$$m \, d^2x/dt^2 = eX, \text{ where } m \text{ and } e \text{ have their usual significance.}$$

$$\text{Therefore } \frac{d^3x}{dt^3} = \frac{e dX}{m dt} = \frac{I_0 e}{\kappa_0 m} = \text{a constant, as has just been shown.}$$

$$\text{Whence } \frac{d^2x}{dt^2} = \frac{e I_0}{m \kappa_0} t, \quad \frac{dx}{dt} = \frac{1}{2} \frac{e I_0}{m \kappa_0} t^2, \quad x = \frac{1}{6} \frac{e I_0}{m \kappa_0} t^3 \quad . \quad . \quad . \quad (10.4)$$

assuming that $x = dx/dt = d^2x/dt^2 = 0$, when $t = 0$.

The approximation that $dx/dt = 0$ when $x = 0$ is justified so long as V_0 is greater than a few volts, since the average initial speed of an electron at the cathode corresponds to about 1/10th of a volt. When $t = T =$ transit-time of the electron between cathode and anode, $x = d$ and $dx/dt = \sqrt{(2V_0e/m)}$. Substituting in equations (10.4)

$$\sqrt{(2V_0e/m)} = \frac{\frac{1}{2} e I_0}{m \kappa_0} T^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (10.5)$$

$$\text{and} \quad d = \frac{\frac{1}{6} e I_0}{m \kappa_0} T^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (10.6)$$

Dividing (10.6) by (10.5),

$$T = \frac{3d}{\sqrt{(2V_0e/m)}}$$

This is the transit-time in a space-charge limited diode.

The transit-time in a temperature-limited diode is T' , where,

$$T' \frac{\sqrt{(2V_0e/m)}}{2} = d$$

$$\text{or} \quad T' = \frac{2d}{\sqrt{(2V_0e/m)}}$$

Therefore $T/T' = 3/2$. The transit-time in a space-charge limited diode is 1.5 times the transit-time in the same diode, temperature limited, the accelerating voltage being the same in the two cases.

APPENDIX 10.3

Length of the Line in a Disc-Seal Valve Concentric-Line Amplifier

In Appendix 1, Chapter 3, Volume 5, it is shown that the input impedance of a line of length l , terminated at the far end by a short-circuit, is $Z_o \tanh(\gamma l)$, where Z_o is the characteristic impedance of the line and γ is the propagation constant. Let Fig. 10.15 represent either the grid-cathode or grid-anode circuit of a concentric-line amplifier, terminated at the valve end by the inter-electrode capacitance C .

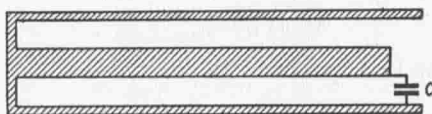


Fig. 10.15. Short-circuited coaxial line terminated in a capacitance C .

Then, if Z_i is the input impedance, across the capacitance C ,

$$\frac{1}{Z_i} = j\omega C + \frac{1}{Z_o \tanh(\gamma l)}$$

where $\omega = 2\pi f = 6 \times 10^8 \pi / \lambda$ and λ is the wavelength in metres. If the line losses are negligible.

$$\tanh \gamma l = j \tan \beta l = j \tan (2\pi l / \lambda)$$

Then

$$\frac{1}{Z_i} = j \left[\frac{6 \times 10^8 \pi C}{\lambda} - \frac{1}{Z_o \tan (2\pi l / \lambda)} \right]$$

When the line is resonant at frequency f and wavelength λ , Z_i is infinite, and

$$\frac{6\pi \times 10^8 C}{\lambda} = \frac{1}{Z_o \tan (2\pi l / \lambda)}$$

$$\text{or} \quad l = \frac{\lambda}{2\pi} \tan^{-1} \left(\frac{\lambda}{6 \times 10^8 \pi Z_o C} \right) \text{ metres.}$$

This length is shorter than that of the open quarter-wave line by an amount depending upon $Z_o C$. To take a specific example, if $C = 5$ pF, $Z_o = 100$ ohms and $\lambda = 30$ cm,

$$l = 2.23 \text{ cm.}$$

This is about the smallest length of line which could be accommodated on a disc-seal valve. At wavelengths smaller than this a three-quarter wave line would have to be used.

CHAPTER 11

Velocity-Modulation Valves and Magnetrons

As HAS BEEN shown in Chapter 10, conventional valves are limited by transit-time effects to operation at frequencies less than 3,000 Mc/s and, indeed, are difficult to manufacture for frequencies much above 1,000 Mc/s. By making use of transit-time effects valves working on different principles have been evolved. These new valves fall into several categories. First, there are valves employing what are called 'velocity-modulation' effects directly. Secondly, there are valves employing velocity-modulation effects in conjunction with the effects of crossed electric and magnetic fields. Valves in the first category are usually called velocity-modulation valves. Examples are the 'klystron', and the 'reflex klystron' or 'reflection klystron'. Valves in the second category are called magnetrons. Two entirely different types are the 'split-anode magnetron' working in a particular way and the 'multi-resonator magnetron'. A third type of valve which is still in the experimental stage is called a 'travelling-wave valve', and employs a stream of electrons moving with an electromagnetic wave and velocity-modulated by it. In principle valves of the third type have important advantages over klystrons and magnetrons. The promise of these advantages has yet to be fulfilled in practice.

It is the purpose of this chapter to set forth the principles on which these valves work, and to describe examples of them in detail. The properties of these more modern valves are in certain cases quite different from those of conventional types. By their use it has been found possible to extend the radio-frequency spectrum to frequencies beyond 100,000 Mc/s.

Parts of this chapter cannot be fully understood without some knowledge of cavity resonators. These are treated in Volume 5 and so this chapter should be read again (or, at least, parts of it) after Volume 5 has been studied.

Principle of Velocity Modulation

Suppose a circuit, resonant at frequency f , is connected to a pair of plane parallel grids G_1 and G_2 as shown in Fig. 11.1, and a high-frequency alternating voltage is maintained between G_1 and G_2 (the signal voltage). Suppose also that a uniform stream of electrons, all travelling initially with the same high speed, is passing through the two grids along the paths AB, CD, EF, etc. Then, if the distance between the grids is chosen so that the electrons traverse it in a small fraction of a period $T = 1/f$, each electron will be accelerated or retarded during its passage according to the phase of the alternating voltage at that particular time. For example, an electron crossing when G_2 is most positive with respect to G_1 will have its speed increased by a certain amount. An

electron crossing when G_2 is most negative with respect to G_1 will have its speed diminished by the same amount. An electron crossing when G_1 and G_2 are at the same potential will emerge with its speed unchanged.

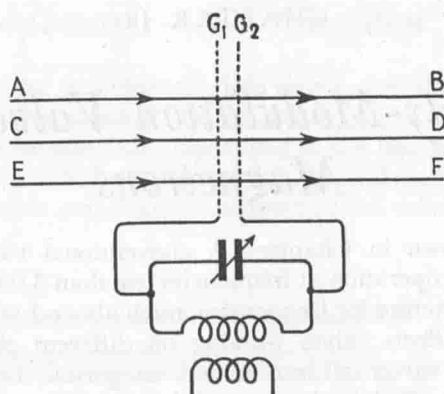


Fig. 11.1. The electrons move in paths such as AB, CD, and their velocity to the right of the grids is modified by the h.f. potentials of the grids.

Suppose now that the beam of electrons travels on undisturbed by any additional forces. The electrons are then said to move in a 'force-free space'. The faster electrons will gradually overtake the slower electrons and at a certain distance beyond G_2 bunches of electrons will tend to form where the fast overtake the slow. There will be corresponding places where the electric density will be very much less than it was in the original uniform beam. The alternating voltage between G_1 and G_2 has therefore modified the uniform electron stream, producing at a certain distance from the grids concentrations of electrons at regular time intervals. These concentrations are known as 'bunches' and the gap between G_1 and G_2 as the 'buncher' gap because of its action.

The best picture of the bunching process, which, be it noted, is due to 'velocity modulation' of the electron stream, is obtained by drawing a space-time diagram for representative electrons leaving the gap. In Fig. 11.2 the ordinate gives the position in space of representative electrons at times given by the abscissa. Time is measured from an arbitrary zero when the voltage across the grids is zero and is about to become negative. The variation of the voltage with time is shown below the space-time diagram with the axis on the same scale.

Suppose that an electron, leaving the grids at time $t = 0$, moves so that its position at any subsequent time is on the line OA at 45° to both axes. Then an electron leaving at time $t = T/2$ will have its position given by the line PB, an electron leaving at $t = T$ by the line QC and so on. All these are parallel, since the electron speed is the same for each case. But an electron leaving at time $t = T/4$ experiences the full retarding voltage due to the high-frequency signal. Its speed on leaving the grids is less and the gradient of the line on the space-time diagram is less. Let the line be UG. Similarly, the line for the electron leaving at $t = 5T/4$ is VH and that for the electron leaving at $t = 9T/4$ is WI. In the same way, electrons leaving at $t = 3T/4, 7T/4, 11T/4$ are

accelerated between the grids and continue on their way with an increased speed. Their paths are given by XJ, YK and ZL respectively. Similar constructions can be made for electrons leaving at intermediate times.

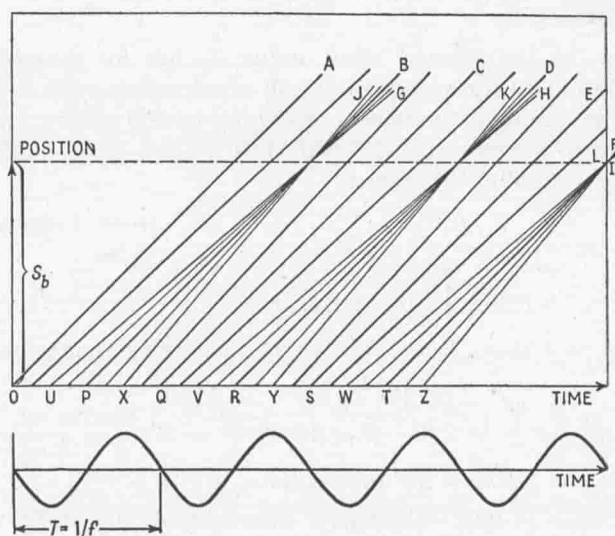


Fig. 11.2. The Applegate diagram indicates in graphical form the distance S_b at which the maximum bunching occurs.

The diagram (known as the Applegate diagram, after its inventor) clearly indicates that at distance S_b from the grids the electrons appear in bunches which are separated in time by the interval $T = 1/f$. The tendency to form bunches begins immediately the electrons leave the buncher gap, but maximum bunching occurs at distance S_b . This distance depends on two factors, (i) the initial speed of the electrons and (ii) the voltage between the grids G_1 and G_2 .

Calculation of the Bunching Distance (S_b)

Referring again to the Applegate diagram, let the speed of the electrons entering the buncher gap be v_0 . Let the speed of those gaining most energy in the buncher gap be $v_0 + v_1$ and the speed of those losing most energy be $v_0 - v_1$. Then the electrons leaving at time $T/4$ have speed $v_0 - v_1$. These electrons at time t have reached a point distant $(t - T/4)(v_0 - v_1)$ from the buncher. This corresponds to the line UG. The electrons leaving at $T/2$ have speed v_0 and at time t reach a distance $(t - T/2)v_0$. The electrons leaving at time $3T/4$ have speed $v_0 + v_1$ and at time t reach a distance $(t - 3T/4)(v_0 + v_1)$.

To a very rough approximation,

$$S_b = \left(t - \frac{T}{4}\right)(v_0 - v_1) = \left(t - \frac{T}{2}\right)v_0 = \left(t - \frac{3T}{4}\right)(v_0 + v_1)$$

Solve these equations for t and S_b . Then $S_b = v_0^2 T / 4 v_1$.

But $T = 1/f$, so, to this approximation $S_b = v_0^2 / 4 f v_1$.

If the calculation is done more accurately, it can be shown that although electrons travelling with these speeds do *not* in fact meet at one point in space and time, their maximum density (bunch) occurs when $S_b = 1.84v_0^3/2\pi f v_1$, or, very nearly, $S_b = v_0^3/3f v_1$. This may be taken as the bunching distance for most calculations.

It must now be determined what value S_b has for reasonable values of v_0 , v_1 and f . Let v_0 be the speed due to an accelerating potential difference of 1000 volts, and let the signal voltage across $G_1 G_2$ be 100 volts. Let the buncher circuit be tuned to a frequency of 3,000 Mc/s. Then $\frac{1}{2}mv_0^2 = V_0e$ where V_0 is the accelerating potential difference.

$$\begin{aligned}\text{Therefore } v_0 &= \sqrt{\left(\frac{2eV_0}{m}\right)} = \sqrt{\left(\frac{2 \times 1.6 \times 10^{-19} \times 1000}{9 \times 10^{-31}}\right)} \text{ metres/sec} \\ &= 1.885 \times 10^7 \text{ metres/sec.}\end{aligned}$$

Similarly, $\frac{1}{2}m(v_0 + v_1)^2 = (V_0 + V_1)e$ where V_1 is the signal voltage.

$$\begin{aligned}\text{Therefore } v_0 + v_1 &= \sqrt{\left(\frac{2 \times 1.6 \times 10^{-19} \times 1100}{9 \times 10^{-31}}\right)} \text{ metres/sec} \\ &= 1.976 \times 10^7 \text{ metres/sec.}\end{aligned}$$

$$\text{Whence } v_1 = (1.976 - 1.885)10^7 \text{ metres/sec} = 9.1 \times 10^5 \text{ metres/sec.}$$

$$\text{But } f = 3 \times 10^9 \text{ c/s. Therefore,}$$

$$\begin{aligned}S_b &= \frac{v_0^3}{3f v_1} = \frac{(1.885)^3 \times 10^{21}}{3 \times 3 \times 10^9 \times 9.1 \times 10^5} \text{ metres} \\ &= 4.34 \text{ cm.}\end{aligned}$$

The electrons will be bunched as tightly as possible in a distance 4.34 cm. This is a reasonable distance if a valve is to be made. The figure is not exact, for the self-repulsion of the electrons causes a certain debunching. The magnitude of this debunching effect depends upon the electron current. The value of S_b used above is only true for such small currents that 'space-charge debunching' may be neglected. However, for most practical currents the correction is not very large.

Current in the Drift Space

It has already been mentioned that the electrons after leaving the buncher are assumed to move freely in a force-free space. This space in a velocity-modulation valve is known as the drift space. We must now examine a little more closely the electron current as a function of time at various points in the drift space.

Reference may be made to Fig. 11.3. Diagram (a) represents the mode of variation of the current with time at or near the buncher gap. There is, in fact, no variation with time. The electron beam is uniform.

Diagrams (b), (c), (d), (e), (f) represent the variations of current at distances from the buncher increasing in that order. The bunches are beginning to form in (b). In (c) there are quite large bunches. In (d), corresponding to the optimum bunching distance S_b single narrow pulses of current of considerable amplitude occur once per cycle. In (e) the distance is greater than

S_b , and the faster electrons have gone through the bunch which is beginning to break into two. In (f) the process is continued. There are two distinct peaks in each bunch. A beam in which the current varies as in (f) is said to be 'over-bunched'.

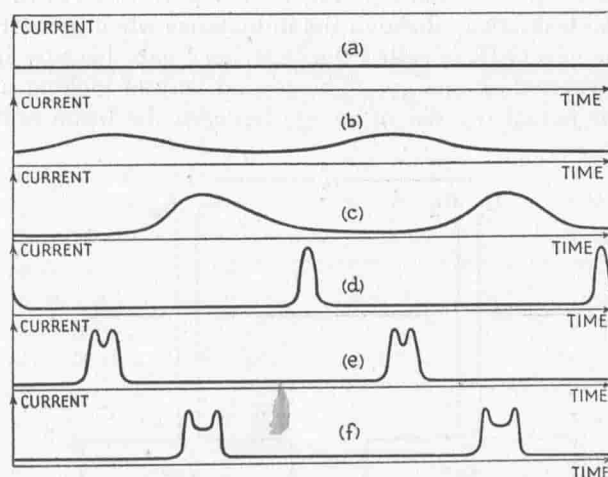


Fig. 11.3. Variation of current with time at various distances from the buncher gap; (a) is for a point close to the gap and the other diagrams for points at increasing distances from it. The optimum distance S_b is represented by diagram (d).

Klystron Amplifier Theory

Now suppose that a valve is made according to the drawing in Fig. 11-4. Let two pairs of grids, G_1G_2 and G_3G_4 be arranged in the path of a beam of electrons. It should be noted that the function of the grids is to give uniform electric fields through which the electron beam must pass. Ideally, they would be quite 'transparent', so that all the electrons would pass through the mesh. In practice some electrons are intercepted by the grids and are lost so far as the action of the amplifier is concerned. Let the first pair of grids, forming the buncher gap, be connected across a circuit L_1C_1 , resonant at frequency f . Let the distance between G_1G_2 and G_3G_4 be d . Then if G_3G_4 are connected across a circuit L_2C_2 which is also resonant at frequency f , amplification of a signal injected into L_1C_1 may be obtained due to the velocity modulation and subsequent bunching of the electron beam.

Suppose that the distance d is chosen so that optimum bunching occurs at the grids G_3G_4 for the signal injected into L_1C_1 having a given amplitude and a given frequency f . Then the bunches of electrons pass through the grids G_3G_4 regularly f times per second. Each bunch excites the resonant circuit L_2C_2 , and, since f is its resonant frequency, large oscillations are built up in L_2C_2 . When the amplitude of the injected signal is constant, and the steady state has been reached, the grid G_4 will always be negative with respect to G_3 when the bunch is passing through, for it is only under these circumstances that the circuit can take energy from the electron beam.

This can be looked at in two ways. First, it will be remembered that in the conventional valve amplifier the valve anode is always at its most negative when the current to it is a maximum. In fact the current to the anode produces the voltage fall at the anode through the resistance of the tuned circuit. Similarly, in the klystron amplifier a current flowing between G_3 and G_4 produces an immediate fall in the potential of G_4 . The energy thus stored in the capacitance of the circuit then leaks away through the inductance when the electron current diminishes. The gap G_3G_4 is called the 'catcher' gap, because it catches the bunches and extracts their energy. The second way of looking at the matter is to consider the actual transfer of energy between the beam and the circuit.

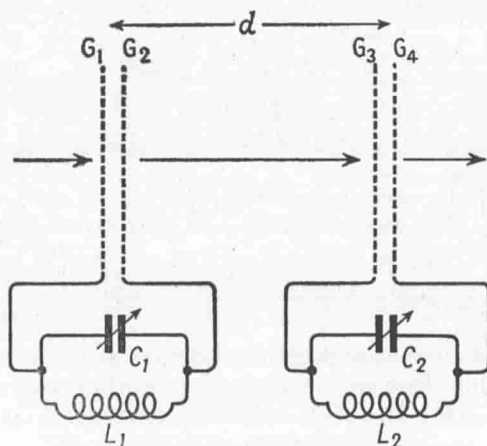


Fig. 11.4. General form of klystron with the buncher grids G_1G_2 connected to the input tuned circuit L_1C_1 and the catcher grids G_3G_4 joined to the output tuned circuit L_2C_2 .

If the electrons are accelerated in traversing the gap G_3G_4 , their energy is thereby increased and this energy must come from the circuit L_2C_2 . If, however, the electrons are decelerated, their energy is decreased and the d.c. energy in the beam is partly released to the circuit. Since the bunches carry the greater part of the electron current, the electrons in the bunches must be decelerated if the net exchange of energy is to be in favour of the circuit, and this must be the case if the valve is to amplify. Therefore, in a klystron amplifier the grid G_4 must be negative with reference to G_3 as each bunch passes through the gap.

The larger the potential swing produced across G_3G_4 by the bunches, the greater will be the transfer of energy to the circuit, until, if the swing is *too* large, some electrons may actually be turned back from G_4 to G_3 . This decreases the efficiency and constitutes a limitation on the power output of the amplifier. A corresponding limitation exists for the conventional power amplifier if the anode swings more negative than the most positive potential of the grid.

Klystron Amplifier Practice

In practice the circuits L_1C_1 and L_2C_2 cannot be conventional lumped circuits of the low-frequency type. The frequency range over which klystrons operate is so high that the circuits must be cavity resonators, taking the form shown in Fig. 11.5. The drawing (a) represents an axial section of a cavity

which has cylindrical symmetry about AB, while the drawing (b) is a pictorial representation of the cavity with the re-entrant cylinders shown by the dotted lines.

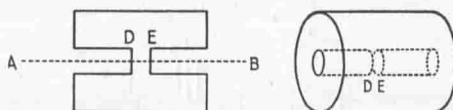


Fig. 11.5. Axial section (a) and perspective sketch (b) of a coaxial resonator suitable for a klystron.

The two grids corresponding to G_1 and G_2 or to G_3 and G_4 of Fig. 11.4 are represented by D and E. Such a cavity may be compared to a short piece of concentric line (see Chapter 10), broken by the gap between the grids. It possesses the property of resonating electrically at a frequency determined primarily by its size and by the nature of the gap between D and E. To understand its relation to the conventional LC circuit shown in Fig. 11.4, imagine a circuit such as is shown in Fig. 11.6, consisting of a copper conductor L, joining a pair of disc-shaped grids, G_1 and G_2 placed opposite each other.

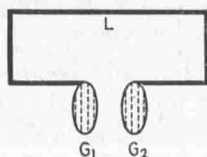


Fig. 11.6. The cavity resonator may be considered as made up of a large number of simple circuits such as this in which the wire L represents the inductance and the grids G_1G_2 the capacitance.

Clearly the conductor L and the grids G_1G_2 form a simple circuit in which L is the inductive part and the grids G_1G_2 form a capacitance. The circuit has a resonating frequency f determined, as usual, by the inductance of L and the capacitance of G_1G_2 . Now suppose that a large number of conductors identical with L are joined to the grids at intervals round their peripheries. These form a large number of inductances in parallel, placed across the pair of grids and the whole resembles the cavity resonator of Fig. 11.5 and possesses a resonant frequency much higher than that of the circuit of Fig. 11.6. All the high-frequency currents flow on the inside surface of the conductors, and very little energy is radiated. The Q of such a cavity resonator is extremely high if the conductor is good, for the resistance is extremely low. All resonant circuits at frequencies of the order of 3,000 Mc/s or more take the form of cavities. These are discussed in more detail in Volume 5.

A practical klystron amplifier contains four important elements. First, there is the resonator with the buncher gap incorporated. This is shown at B in Fig. 11.7 and is the input circuit of the amplifier. Then there is the resonator incorporating the catcher gap. This is the output circuit of the amplifier. The electron beam, produced by an electron 'gun' G flows down the tubes E, D and F, while D is the drift space in which the electron bunches are formed. The electrons which pass through all the grids are collected on the target

electrode T. The power is fed into the input cavity B by means of the concentric line L_1 terminated in a coupling loop. The amplified power is taken from the cavity C by means of the concentric line L_0 which is again loop-coupled to the circuit.

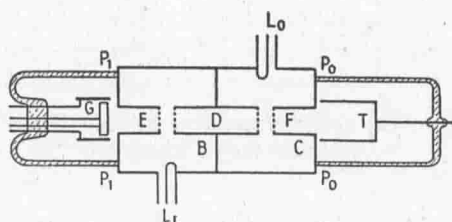


Fig. 11.7. Klystron valve showing the electron gun G, the buncher and catcher cavities B and C, the drift space D, and the target T.

The valve itself consists of a metal body, comprising the two cavities, joined at one end to a glass tube containing the 'gun' assembly and at the other to a similar tube through which the target lead emerges. Arrangements are made to tune each circuit cavity over a small range of frequencies. These arrangements take different forms in different valves. A common device is to make the end plates of the cavities, P_1 and P_0 , of thin diaphragms, so that they may be moved mechanically by screws outside the valve.

Klystron amplifiers have so far only been successfully used as power amplifiers. As receiver amplifiers for small signals they produce an undesirable noise power which causes a deterioration in the signal to noise ratio. As power amplifiers they are moderately useful. Klystrons are operated with the resonators (cavities) at or about earth potential. This allows connections to be made by means of concentric lines. The cathode is highly negative, producing the accelerating voltage for the electron beam. This voltage may be regarded as the h.t. for the valve. The target or collector, T, (Fig. 11.7) is generally operated at or near earth potential. In the majority of cases it is connected directly to the cavities. The electron gun is usually simple, consisting of a disc-shaped cathode surrounded, as is shown in Fig. 11.7, by a focusing cylinder. The latter is maintained either at cathode potential or negative with respect to the cathode. Its potential gives a fine control of the shape of the beam and is adjusted so that there is no undue spreading as the electrons travel along E, D and F. Electron guns are discussed in more detail in Chapters 15 and 16. Typical operating figures for a simple two-cavity klystron amplifier are as follows:—

$$\text{High tension} = 1,000 - 2,000 \text{ V}$$

$$\text{Total cathode current} = 20 - 40 \text{ mA}$$

$$\text{Maximum power gain (small signal output)} = 10 - 20 \text{ db}$$

$$\text{Efficiency} = 10 - 20\%$$

This would result in a maximum power output of about 15 watts, but at this level the gain would probably be less than 10.

More efficient amplifiers, giving considerably greater gain, can be made by introducing a third cavity into the beam. The three-cavity klystron so formed employs the first two cavities as a voltage amplifier. No power is taken

from the second cavity, but partial bunching of the beam produces large voltage swings in it. The second and third cavities together form a power amplifier and the power is extracted from the third. Power gains of the order of 30 db can be obtained from such valves and efficiency can be stepped up to almost 30%.

The principal objection to the use of klystron amplifiers, and one which is particularly troublesome with three-cavity types is that the tuning of each cavity is fairly critical. Indeed, the over-all pass-band of a three-cavity amplifier working at 3,000 Mc/s may be no more 1 Mc/s when it is working at maximum efficiency. This makes its operation difficult, particularly when the temperature variations in the metal of the cavities produce changes in length which are reflected in frequency drifts.*

The Klystron Oscillator

An electronic amplifier can always be made into a self-oscillator by feeding back a sufficient fraction of the output power to the input circuit in the proper phase. This is true of the klystron amplifier.

If the output cavity C_2 of Fig. 11.8 is connected to the input cavity C_1 by a concentric line L then the amplifier may become an oscillator, the power being abstracted by another concentric line, as at O .

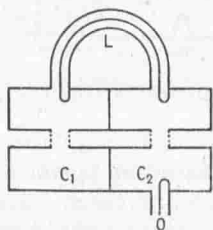


Fig. 11.8. In a klystron oscillator energy is fed from the output resonator C_2 to the input C_1 by the coaxial line L . Power is drawn out through the coupling loop O .

The action is as follows:—Suppose, first, that the coupling between the two resonators, C_1 and C_2 , is large enough to sustain oscillation, but so small that the two-circuit system is characterised by a single frequency when each is properly tuned (see Volume 1). Then to maintain oscillations the phase change round the path—buncher gap to catcher gap via the drift space plus catcher gap to buncher gap via the concentric line—must be N times 360° , where $N \times 1, 2, 3, \dots$. In short, the voltage fed back to the buncher gap must be in phase. Therefore the klystron will only oscillate when the time taken by the electrons to travel from the buncher gap to the catcher gap satisfies this relationship. Suppose that this time in the drift space is just right when the h.t. voltage which determines it is V_N . Then, if this corresponds to a phase difference of $2N\pi$, the klystron will also tend to oscillate when the voltage is V_{N+1} , corresponding to a phase difference of $2(N+1)\pi$. It will also tend to oscillate at V_{N+2} , corresponding to $2(N+2)\pi$, and so on. On the other hand, oscillations

*Note: During the last few years advances in klystron design have resulted in greatly improved performance, both as regards gain and bandwidth.

cannot possibly be maintained at values of V corresponding to a phase difference of $(2N - 1)\pi$, where $N = 1, 2, 3, \dots$, for then the feedback is entirely negative. In the klystron oscillator, therefore, there are definite *modes* of oscillation at a given frequency, occurring at certain values of the h.t. voltage.

However, if the h.t. is *nearly* correct for a phase difference of $2N\pi$, the circuits pull themselves into step by a slight adjustment of their common frequency, and so klystrons oscillate over certain finite ranges of h.t. voltage in the neighbourhood of the values quoted above.

It may be shown by a calculation of the phase changes in the different parts of the feedback loop that the critical values of V to maintain oscillation are given by:—

$$V_N = A/(N - \frac{1}{2})^2, \dots (N = 1, 2, 3, \dots) \quad (11.1)$$

where A is a constant for any given klystron tuned to any given frequency. On either side of these critical values of V , the power output of the oscillator falls off rapidly. Fig. 11.9 shows the power output versus h.t. voltage diagram for a typical klystron oscillator.

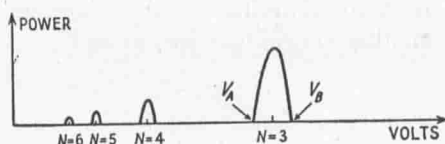


Fig. 11.9. Variation of output power with h.t. voltage in a klystron oscillator.

It will be noted that the maximum power obtained in any mode falls off rapidly as the mode number N increases. This is to be expected, since the corresponding decrease in V produces a rapid decrease in the d.c. input power. Finally, when N is greater than a certain value (6 in Fig. 11.9), the power fed back from the catcher is insufficient to supply the resistive losses in the circuits and oscillation can no longer be maintained. There are therefore no modes at $N = 7, 8, \dots$.

A second case of the klystron oscillator may be distinguished. If the coupling between the cavities is increased until the two-circuit system is characterized by two frequencies, (as explained in the theory of coupled circuits in Volume 1) the number of possible modes of oscillation is doubled. The equation corresponding to 11.1 in this case is

$$V_N = A/(N \pm \frac{1}{4})^2 \dots (N = 1, 2, 3, \dots) \quad (11.2)$$

and the two modes given by the $+$ and $-$ signs are at different frequencies. The difference can be as much as 60 Mc/s at 3,000 Mc/s. As before, the klystron fails to oscillate for values of N greater than a certain number and the power obtained increases as N diminishes.

Klystron Oscillator Performance

As in the case of conventional power valves, a klystron oscillator is designed to give a certain output with maximum efficiency. This usually entails designing the valve to work at a certain value of N —for example at $N = 3$.

For smaller values of N the power output may then decrease despite the increased d.c. input, for the time in the drift space may be too short to give optimum bunching of the electrons. In practice, for a power oscillator giving about 20–100 watts, N could be 1, 2 or even 3. The efficiency obtained from a klystron oscillator is about the same as for an amplifier—20%, and a power output of 100 watts, produced from a d.c. input of 500 watts can be obtained at about 4,000 volts. This means an effective beam current of 125 mA. Under these circumstances the voltage swings in the buncher and catcher are of the same order of magnitude, the Q of the buncher being many times that of the catcher which is supplying power to an external circuit and is therefore heavily damped.

The fact that the klystron oscillates over a finite range of h.t. voltage in one mode is important. For example, in Fig. 11.9 the $N = 3$ mode extends from $V = V_A$ to $V = V_B$. Over this range the frequency of oscillation adjusts itself to give the correct phase of the feedback voltage. The h.t. voltage is therefore a fine control of the klystron frequency and may be used as such in certain communication and radio-beacon circuits. Increasing the voltage increases the frequency of oscillation.

Bunching of a Reflected Beam of Electrons

One of the most important velocity-modulation valves is the so-called *reflex klystron* which is widely used as a receiver local oscillator at centimetric wavelengths. To understand its action it is necessary first to consider what will happen to a beam of electrons (Fig. 11.10) which, originally moving with a constant speed v_0 , is velocity modulated at a gap G and is then reflected in a retarding electric field between the gap G and a negatively charged plate R , before returning to the gap. The simplest way to examine the action is to draw

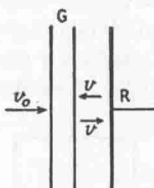


Fig. 11.10. Representation of reflex klystron with combined buncher and catcher gap G and reflector plate R .

a distance-time diagram of the type shown in Fig. 11.2 for these new circumstances. This is given in Fig. 11.11.

The alternating potential applied across the gap G is represented as a function of time by the curve LMN. The velocity of an electron leaving the gap at time $t = 0$ is v_0 . It moves with a constant deceleration towards the reflector, coming to rest at A and thereupon turns back to the gap, reaching it at the time represented by OB with the same speed v_0 . Similarly, electrons leaving half a period later and a full period later traverse paths represented by the parabolas PCD and QEF , where $OB = PD = QF$. But electrons leaving the gap a quarter of a period after $t = 0$ have experienced an accelerating force in the gap and have therefore an increased speed. This allows them to penetrate

further towards the reflector, coming to rest at G. They take a longer time to return to the gap than the unaccelerated electrons. They return at a time represented by OH.

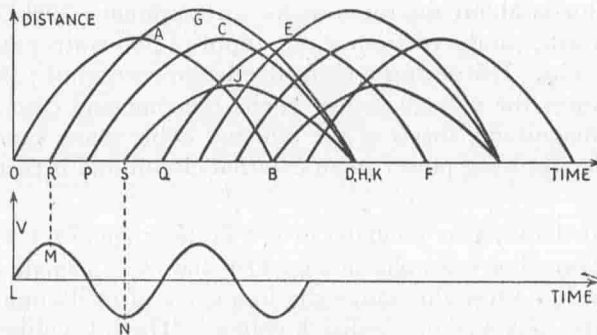


Fig. 11.11. Applegate diagram for reflex klystron.

On the other hand the electrons starting three-quarters of a period after $t = 0$ trace the path SJK, for they are unable to penetrate so far into the reflecting field and therefore return sooner to the gap. Electrons leaving at intermediate times take longer or shorter times to return according to the velocity modulation acquired.

Clearly, the velocity-modulated beam, reflected by the metal surface R, tends to form bunches which are similar to those produced in a force-free drift space. By suitably choosing the reflector voltage and its distance from the grids, the electrons can be made to form tight bunches on their return.

The Reflex Klystron Oscillator

Imagine a valve made so that an electron gun U causes a beam of electrons to traverse the space between two grids, G_1 and G_2 , forming the capacitive gap in a cavity resonator C (Fig. 11.12). Let the beam, velocity-modulated by a

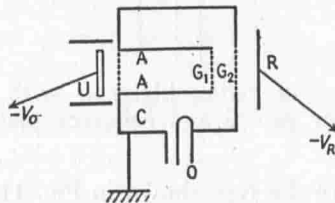


Fig. 11.12. Reflex klystron with electron gun U, resonator C, gap G_1G_2 and reflector R.

high-frequency signal in C, be reflected by the plate R which is maintained at a negative potential $-V_R$. The reflector R must be more negative than the cathode of the gun which we shall assume to be at potential $-V_0$. The h.t. of the valve is then V_0 , for the electrons entering the cavity (which is at earth potential), have been accelerated through this potential difference. The electrons of the beam after reflection again traverse the gap between the grids, this time in the opposite direction and in bunches. Each bunch, passing through

the cavity, induces a current in it at the signal frequency, and the cavity C may be regarded both as a buncher and as a catcher. The bunches, slowed down by their passage between the grids are finally absorbed on the walls of the cylinder at A. Power may be taken from the cavity at O, for the valve is in principle an oscillator employing electronic feedback.

To understand the action thoroughly consider the way in which the self oscillations will build up when the valve is switched on. There is always an infinitesimal signal in the cavity at the resonant frequency. The mere action of switching on the valve produces a 'ringing' of the circuit. This infinitesimal signal produces velocity modulation of the beam and this in turn produces a larger signal at the same frequency when the bunches return to the cavity, provided, always, that the currents induced by the bunches are in the correct phase relationship and that the circuit losses are not too high. This oscillatory current builds up until it is limited by some other factor such as the over-bunching or de-bunching of the beam. A steady state is usually reached where the valve is a generator of sinusoidal oscillations at or near the resonant frequency of the cavity. Such a valve is a reflex klystron oscillator.

As has been mentioned, two basic conditions must be fulfilled for stable oscillations to be maintained. First, the power fed back by the electron bunches must be large enough to supply the losses in the oscillator. Secondly, the time of return of the bunches must be properly related to the phase of the current oscillations in the cavity.

The phasing condition is easily examined. The centre of the bunch in 11.11 leaves the cavity gap when the voltage of G_1 is becoming positive relative to G_2 (Fig. 11.12). To give up energy to the h.f. field the electron bunch must return when G_1 is at or about its most positive value with respect to G_2 , for it is then that the bunch experiences its maximum deceleration and therefore its maximum loss of d.c. energy. Therefore the phase difference between the time of setting out of the electrons and the return of the bunch must be $2N\pi - \pi/2$ or $2\pi(N - \frac{1}{4})$. Calculation shows that the time in the drift space is proportional to $\sqrt{V_o/V_R}$, where V_o is the h.t. voltage and R is at the potential $-V_R$ relative to the cavity. Therefore,

$$V_R/\sqrt{(V_o)} = A/(N - \frac{1}{4}) \quad . \quad (N = 1, 2, 3 \quad . \quad .) \quad . \quad . \quad (11.3)$$

where A is a constant for a given valve. As in the case of the two-resonator klystron oscillator, the reflex klystron only oscillates at or about certain values of the operating voltages. In the case of the reflex oscillator there are two voltages which may be varied, although in practice V_o is usually maintained constant.

Modes of Oscillation

The reflection valve usually operates at a mode number N of 2, 3, or 4, although early valves worked at $N = 5$ or 6. The way to use the valve is illustrated by Fig. 11.13. The h.t. voltage between the cathode K and the resonator A is obtained by means of a power pack in the usual fashion, but A is earthed as it forms the 'circuit' of the valve and requires tuning. The cylinder F, which by its position and potential tends to direct the electron beam through the cavity, or, as we say, to 'focus' it, and the reflector R, are then usually operated from a small auxiliary power pack. Neither electrode

takes any current, but both require a very constant potential source. Sometimes the focusing cylinder is connected directly to the cathode. The reflector potential ($-V_R$ relative to earth and $V_o - V_R$ relative to the cathode) must be variable. The auxiliary power supply is therefore connected across a potentiometer.

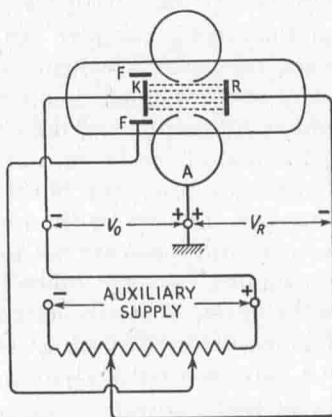


Fig. 11.13. Circuit of reflex-klystron oscillator. The cavity resonator is represented by A.

Usually the h.t. voltage V_o is kept constant and V_R is adjusted to give the correct phase relationship for oscillation. As with the two-resonator klystron oscillator, the valve oscillates over a certain range of V_R , the power output having a maximum at a particular value. The valve automatically adjusts the phase of the reflected bunches by a slight change of frequency. The curves of Fig. 11.14 give some idea of the variation of power and frequency for a typical reflex-klystron oscillator working at about 3,000 Mc/s. V_o in this case would be about 250 volts, the cathode current about 20 mA and the efficiency of the oscillator about 3% at its maximum value. The variation of frequency

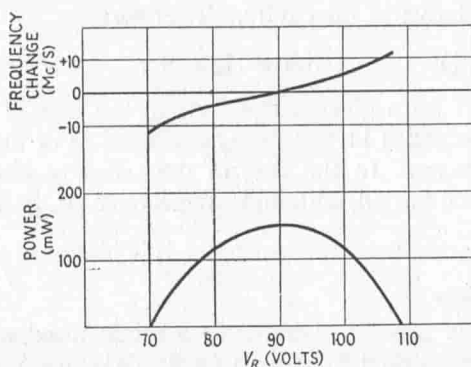


Fig. 11.14. Variation of output power and frequency with reflector voltage.

with reflector voltage is a valuable property of the klystron when it is used as a receiver local oscillator. It is used to good effect both in radar and communications.

Mechanical Tuning of the Reflex Klystron

The mechanical tuning of a klystron, as distinct from the electronic tuning just mentioned, may be done in two ways. Valves in which the cavity resonator is part of the envelope are usually tuned by distorting the shape of the cavity. Most American reflex klystrons employ this method, particularly at frequencies around 10,000 Mc/s. Mechanical tuning of this type must not be abused. The diaphragm which controls the dimensions of the cavity is made of thin metal, and a too violent distortion or too large a number of smaller movements may cause fracture, destroying the valve. For this reason mechanical tuning in a reflex klystron of this type should be employed sparingly and with care.

A second form of tuning, used extensively in valves for use around 3,000 Mc/s, is obtained by screwing one or more suitably sized plungers into the cavity. This can only be done when part of the cavity is outside the valve envelope, as it is in the usual British valves for this part of the spectrum. Screwing a plunger into the cavity reduces the effective inductance and so raises the frequency of oscillation, unless the end of the plunger is brought too close to the grids, when the effective capacitance may be increased and the frequency lowered. Fig. 11.15 is an axial section of a typical reflex-klystron resonator for use at a wavelength of 10 cm. As may be seen, the tuning plungers (of which there are three) are placed so that their insertion decreases the size of the cavity without seriously increasing the capacitance between the two grids.

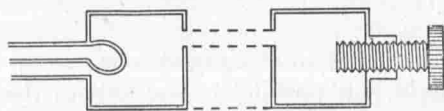


Fig. 11.15. In this diagram one of the three tuning screws is shown. These screws alter the effective size of the cavity and, hence, its natural frequency.

One plunger usually has a finer screw movement than the other two, and is so fitted that it gives smooth tuning with the valve working. The two rough plungers are used to set the valve approximately to the frequency required. They are then locked in position by nuts. The fine tuner is then used while the valve is running to set the frequency exactly. A variation of 50 kc/s can be made at a frequency of 3,000 Mc/s by a skilled operator, using the fine tuner.

The Magnetron Oscillator

The name *magnetron* was originally given to a cylindrical diode valve placed in a magnetic field acting parallel to its axis. Under certain circumstances when such a valve is connected to an external resonant circuit, it becomes an oscillator; that is, it generates high-frequency alternating current from the d.c. power supplied by the h.t. battery.

In this chapter only modern forms of the magnetron oscillator will be considered. There is no such thing as a practical magnetron amplifier. The valve is an oscillator only. Moreover, there are several ways in which a magnetron may oscillate. Terms such as 'dynatron' and 'negative resistance' have been used to describe a particular mechanism of oscillation. Another kind of oscillation has been called the 'travelling-wave' oscillation, and this differs from the 'cyclotron-frequency' oscillation. To make matters more com-

plicated, it has to be admitted that a quantitative theory of the modern 'multi-resonator' magnetron does not exist, although a qualitative explanation of its behaviour can be given. To simplify the matter and to avoid describing phenomena which are of little importance, attention will be focused on how the modern 'multi-resonator' valve behaves in terms of the magnetron effect and the *transit-time effect*. The explanation of the term 'multi-resonator' magnetron will be given in due course.

The Magnetron Effect

First let us consider a cylindrical diode represented by the cross-section in Fig. 11.16, where K is the cathode and A is the coaxial anode. Suppose that A

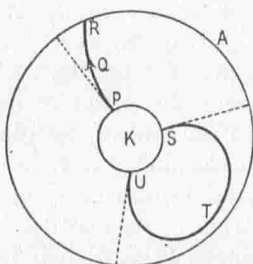


Fig. 11.16. Cross section of cylindrical diode forming the basis of the magnetron oscillator.

is maintained at a constant potential V_A with reference to the cathode and that a uniform magnetic field acts parallel to the axis of the valve; i.e., at right angles to the plane of the paper. Then an electron leaving the cathode with zero speed is accelerated towards the anode, its initial path being along a radius of the circle formed by A. But a charge e moving with speed v constitutes an element of electric current and, in a magnetic field of flux density B at right angles to its direction of motion it experiences a force Bev . This force acts in the plane of the paper and at right angles to the direction of motion of the electron.

The electron trajectory is therefore curved and there are several possibilities.

- (a) If the potential V_A is large enough and B is small enough, the trajectory is of the PQR type. The electron reaches the anode but is displaced from its original direction.
- (b) It is theoretically possible that the electron should circle about the cathode in a closed orbit. This would require a particular value of B and of V_A .
- (c) If V_A is small and B is large, the trajectory is of the STU type. The electron turns back to the cathode before reaching the anode.

This suggests that there is some critical value of B corresponding to any value of V_A at which the anode current I_A changes abruptly, and this is borne out by the I_A - V_A characteristics of the valve which are shown in Fig. 11.17, drawn for various values of B . It may in fact be shewn that the critical value of V_A , namely, V_c , for which the anode current abruptly changes is given by $V_c = KB^2$, where K is a constant for the valve, and B is the magnetic flux density. This is what has been termed the *magnetron effect*.

The Transit-Time Effect

Now suppose that the cylindrical anode is split into two semi-cylindrical segments, as shown in Fig. 11-18. The valve is then known as a split-anode magnetron. It is very little used at present, but it serves as a useful introduction to the more complex structures.

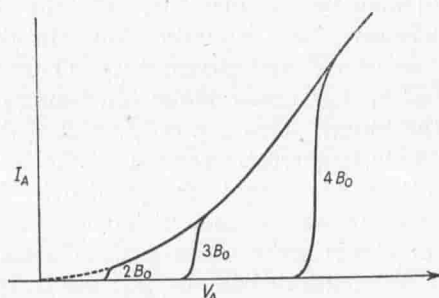


Fig. 11.17. Anode-current—anode-voltage curves for a magnetron with various values of magnetic flux density.

Let the magnetic field be such that, when the two segments are at potential V_A , the anode current is cut off, and the time taken by an electron to travel from the cathode towards the anode and back to the cathode again is $T = 1/f$. Suppose, also, that an alternating potential difference of frequency f is established between the two segments A_1 and A_2 , so that their potentials are $V_A + V \sin 2\pi ft$ and $V_A - V \sin 2\pi ft$, respectively. Then an electron emitted from the left-hand side of the cathode K at $t = 0$ finds itself in an electric field greater than the average, and reaches A_1 if V is large enough. But an electron emitted from the right-hand side of the cathode at $t = 0$ is not so rapidly accelerated and fails to reach A_2 , turning back towards A_1 as the alternating electric field

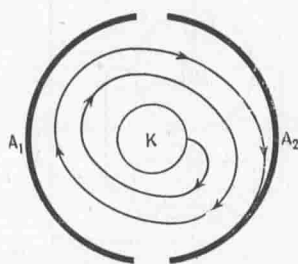


Fig. 11.18. Cross section of a split-anode magnetron.

is changing sign. As the electron crosses the gap towards A_1 it again finds itself in a retarding electric field which decreases its angular velocity about K . This considerably reduces the deflecting force due to the magnetic field, so that the electron moves towards A_1 . However, as its speed increases, it is again deflected, this time towards A_2 , crossing the gap between the two segments when the phase has again changed, so that it is again retarded. This once more reduces the deflection produced by the magnetic field, allowing it to approach A_1 . This process of alternate acceleration and retardation can

occur many times before the electron finally falls into the anode. Moreover, its speed at the anode may be small compared with what it would have acquired in falling through a potential difference V_A .

Since one electron can circulate many times round the cathode, losing energy each time it crosses from one segment to the other, the energy gained by the alternating field may be considerable. On the other hand, electrons accelerated by the alternating field and therefore absorbing energy from it, immediately cross to the anode and disappear. There may therefore be a considerable exchange of energy between the h.t. battery and the alternating field to the benefit of the latter. This can only occur if the transit-time of the electron round the cathode is correctly related to the frequency of the alternating field. The transit-time is roughly determined by V_A/B and the dimensions of the valve. In fact, it may be shown that for a cylindrical magnetron, $f \approx 2 \times 10^{10} B$ c/s, where B is given in webers/sq. metre and V_A , the cut-off value, is given by KB^2 , as mentioned in the last paragraph. In Fig. 11·18 a typical electron trajectory is drawn, showing the roughly elliptical spiral characteristic of an electron which transfers d.c. energy to the alternating field. It should be noted that here, as in the klystron, the electrons act as carriers of electric power, transferring it from the battery to the alternating field. Every deceleration of an electron by the h.f. field means a transfer of energy from the electron to the field. Similarly, every acceleration of the electron by the d.c. field means a transfer of energy to the electron from the field. The electron is to be regarded therefore as the active component coupling the battery or other d.c. source to the h.f. field.

The Split-Anode Magnetron Oscillator

Now imagine a parallel-tuned circuit connected as in Fig. 11·19 across the two anode segments of a magnetron. As stated in the last paragraph, the

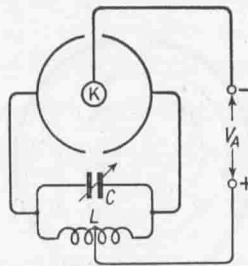


Fig. 11.19. Tuned circuit connected to a split-anode magnetron.

transit-time of the electrons in the absence of an alternating field can be made equal to $T = 2\pi\sqrt{LC}$ by adjustment of V_A and B within certain limits. Then a small impulse given to the tuned circuit, and causing it to oscillate, will tend to build up owing to the transfer by some electrons of d.c. energy to the alternating field. This build up will actually occur if the radiation and resistance losses of the valve and circuit are less than a certain value. The valve then becomes stable as a generator of alternating power at a frequency $f = 1/T$. A quantitative theory of this magnetron oscillator has not yet been achieved, but certain properties of the device are important and well-known. First, there is

a certain latitude in the settings of V_A and B for a given L and C . Conversely, for given values of V_A and B , the external circuit may be tuned over a narrow range of frequencies. The action of the valve, as might be expected, depends very definitely upon the space-charge surrounding the cathode K : a change in the space-charge alters the transit-time of the electrons, and for this reason it is usual to stabilize the voltage applied to the magnetron filament-cathode. Split-anode magnetrons have been used as oscillators at many frequencies up to 60,000 Mc/s, with special circuits attached at frequencies above 1,500 Mc/s. Efficiencies range from 20% at 2,000 Mc/s to a fraction of one per cent at 60,000 Mc/s. As receiver local oscillators they are 'noisy' and cannot compare with the reflex klystron when the latter is properly adjusted.

The Multi-Resonator Magnetron

The most important magnetron valve, used extensively as the generator of micro wave oscillations in pulsed radar and communication systems, is the multi-resonator magnetron, a valve which may be regarded as a modification of the split-anode magnetron, but which differs from it so radically that it requires very special consideration.

Imagine first a magnetron having, as usual, cylindrical symmetry about the axis of the cathode, and the cross-section shown in Fig. 11.20. The cathode is surrounded by a coaxial cylinder of copper out of which eight identical

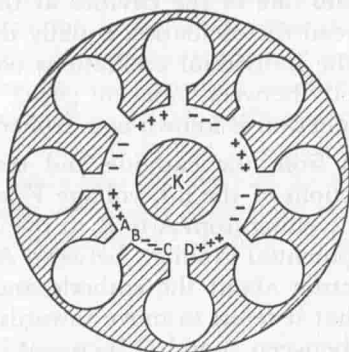


Fig. 11.20. Cross section of multi-resonator magnetron.

cavities have been cut. Each cavity is resonant at the same high frequency, the frequency being determined by its dimensions. The mode of oscillation used is such that the high alternating potential difference is developed across

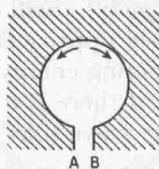


Fig. 11.21. Enlarged view of one cavity of the multi-resonator magnetron.

the neck AB (Fig. 11.21) and the h.f. current circulates round the cavity in the directions of the two arrows alternately.

The cavities are not entirely divorced electrically from the cathode which, by its proximity, has a considerable effect upon the capacitance of the equivalent circuit. The latter might be represented by the elements shown in Fig. 11.22, where C_1 is the direct capacitance between A and B, and C_2 is the

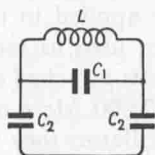


Fig. 11.22. Equivalent lumped-circuit diagram of the cavity of Fig. 11.21. C_1 represents the direct capacitance between A and B and C_2 the capacitance to cathode.

capacitance between the segment at A and the cathode. Now imagine that an alternating potential at the resonant frequency f of a cavity is established between A and B. Referring to Fig. 11.20, if A is positive, B is negative, C is negative and D is positive. Proceeding round the ring we find that each copper segment between the gaps is alternately positive or negative, and a symmetrical distribution of potential is achieved by the use of eight resonators. Indeed, if power is fed into one of the cavities at the correct frequency, the coupling capacitors C_2 spread the oscillation equally throughout all the cavities and keep the phases of the individual oscillations correct, so that there is a phase difference of π exactly between adjacent gaps. Such a mode of oscillation in a multi-cavity magnetron is known as a π -mode for that reason.

If an electron emerges from the cathode and travels towards the anode under the simultaneous actions of the h.t. voltage V_A and the magnetic field B , it may pass the gap AB, travelling from A to B, at the instant when A is positive and B is negative. The potential gradient between A and B then reduces the angular speed of the electron about the cathode and increases the radius of curvature of its path so that it tends to move towards the anode. This movement towards the anode between B and C increases its speed again, so that it again curves away from the anode at C. If by this time the electric field has changed across CD, making C positive and D negative, the electron again gives up some energy in traversing the gap. This action may repeat itself several times and, indeed, the electron may travel right round the cathode being alternately accelerated by the d.c. field and retarded by the h.f. field. For this to occur, the time of rotation of the electron about the cathode must be about four times the period of the h.f. oscillation, and in its circulation round the cathode, during which it slowly approaches the anode and finally reaches it, the electron is continually extracting energy from the h.t. battery and giving it to the alternating field. Thus, if there is a sufficient fraction of such useful electrons in the anode current, the valve may be a generator of h.f. power.

But there is another action which occurs in the multi-resonator magnetron. A close examination of the paths of the useful electrons—those which transfer energy to the h.f. field—shows that these tend to bunch as they pass the gaps AB, CD, etc., the faster electrons over-taking the slower. This bunching effect greatly increases the transfer of energy to the resonant circuit.

Enough has now been said to indicate that a valve of this type, used at suitable values of V_A and B , may become a self-oscillator, and since eight resonators are involved, the total power available may be considerable. The eight cavities, coupled together, behave as one rather high- Q resonant circuit from which power may be abstracted by a coupling loop inserted into any one of them.

Modes of Oscillation of a Magnetron

In the last paragraph the π -mode of oscillation was discussed, in which the phase difference between the alternating voltages across adjacent gaps is 180° . This is the most efficient mode of oscillation and the one which the magnetron designer tries to encourage. But the valve has other modes of oscillation, and these can complicate the operation to such an extent as to render the magnetron almost useless. The only condition which must be satisfied in the oscillator is that there shall be no phase discontinuity in the standing wave existing round the anode block or, stating it otherwise, the sum of the phase differences between all the resonators, taken in order, must be $2n\pi$ where $n = 1, 2, 3, \dots$. The phase-difference between adjacent gaps must therefore be $2n\pi/N$, where there are N resonators, or, for an eight-resonator magnetron, the phase difference between adjacent gaps may be $\pi/4$, $\pi/2$, $3\pi/4$ or π , and for a six-resonator magnetron, $\pi/3$, $2\pi/3$ or π . Each mode of the magnetron has a different frequency from the others, as might be expected from the fact that no two cavities will be identical in shape and size. In fact all modes except the π -mode have two separate frequencies, corresponding to two possible electric configurations for each value of n . Therefore there are in all $2N-1$ frequencies associated with a valve with N resonators, and of these only one is required in operation. A valve which jumps to another mode and therefore to another frequency is not satisfactory. Much of the work which has been done on the magnetron valve has been directed towards emphasizing the stability of the π -mode at the expense of the others.

Magnetron Strapping

Although the theory of it is not well understood, one of the best methods of emphasizing the stability of the π -mode, and incidentally of increasing the efficiency of the oscillator, is to 'strap' alternate segments of the anode block

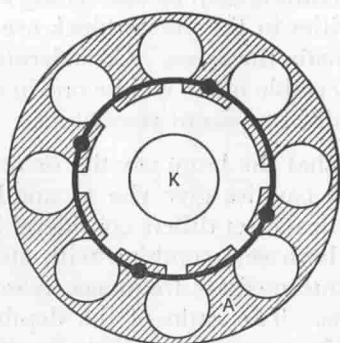


Fig. 11.23. Eight-cavity magnetron with ring strapping.

by means of wire bridges. A single ring strapping of an eight-cavity magnetron is illustrated in Fig. 11.23. The strap on one end of the block is shown as a ring, the points at which it is braced to the segments being shown by black circles. On the other end of the anode block is a similar ring, braced to the alternate segments. Each ring is recessed into the block as shown in Fig. 11.24.

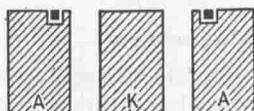


Fig. 11.24. Vertical section through a strapped magnetron showing how the strapping ring is recessed into the anode block.

Other forms of strapping are also used. The common feature is the metallic connection of alternate segments or bars by means of wires.

The first effect of strapping is to decrease the frequency of the π -mode of oscillation without seriously affecting the frequencies of the other modes. This separation of the mode frequencies is advantageous. It allows the valve to be used over a wider range of anode voltage and current without 'mode jumping'. The second effect which is not so well understood is to increase the efficiency of oscillation in the π -mode, so that about 60% of the d.c. power is transformed into h.f. power at 3,000 Mc/s. The greater efficiency is due in part to the fact that the suppression of other modes makes it possible to operate the magnetron at higher input currents.

There is no doubt that the primary effect of strapping is to tend to lock the π -mode, as was originally suggested by Randall and Sayers in 1941. It is probable that the a.c. path offered by the strap is capacitively coupled to the segments to which it is not directly connected, and that these additional capacitances add effectively to the resonator capacitances in the π -mode, but are not additive in the other modes. A more exact analysis allows of fairly accurate computation of the frequencies to be expected.

The Rising-Sun Magnetron

An alternative magnetron modification, equally satisfactory in separating out the π -mode, is the structure known as the 'rising sun'. In this, as shown in Fig. 11.25, adjacent cavities in the anode block are made of different sizes, but alternate cavities are made the same. Considerations of symmetry immediately suggest that the only stable mode will be one in which a phase difference of π exists between each pair of adjacent resonators.

It may in fact be shown that the large cavities or resonators give rise to one set of modes and the small cavities give rise to another set. These interact weakly with each other and each set differs considerably in frequency from the other. The π -modes of both sets combine with strong coupling to give a predominant mode at an intermediate frequency reasonably widely separated from each of the other sets. The ratio of the depths of the two cavities is generally about 2 to 1. The π -mode may be more widely separated from the others by changing this ratio, but the resulting oscillation is not so efficient.

Rising-sun magnetrons are universally used for the highest frequencies. Strapping is a practical procedure at 3,000 Mc/s, but at frequencies above this the straps distort the symmetry and absorb power. A larger number of cavities can be employed using the rising-sun construction; eighteen- and even twenty-four-cavity magnetrons are quite common at the highest frequencies.

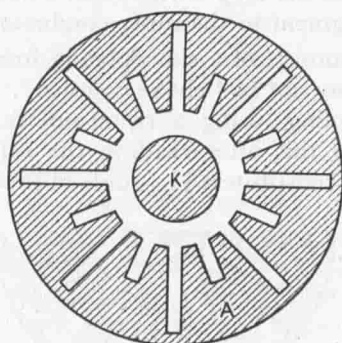


Fig. 11.25. Cross section of rising-sun magnetron in which adjacent cavities are of different sizes.

Output Coupling

The resonant cavities of the magnetron valve taken together are analogous to the resonant circuit of a conventional valve oscillator. A means must therefore be provided for coupling the resonators to the load which is usually the aerial of a radar or communication transmitter. The coupling unit may be either a loop of wire terminating a coaxial cable or a slot terminating a waveguide. The former is more common at frequencies below 4,000 Mc/s, while the latter is almost universally used at the higher frequencies. Fig. 11.26 shows a simple method of coupling the system of cavities to a coaxial cable.

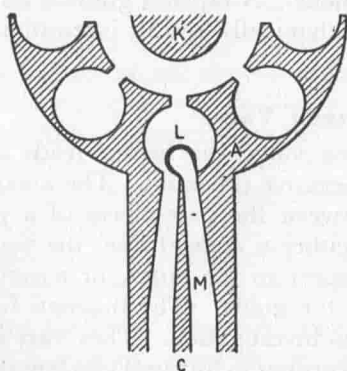


Fig. 11.26. A coaxial cable is coupled to a cavity magnetron by means of a coupling loop L placed in one of the cavities.

The loop of copper L placed centrally in one of the cavities, leads to a concentric line C via a matching section M. Although the loop appears to cause asymmetry in the magnetron block, the coupling between the cavities is large

enough to make this negligible, and the magnetic flux through the loop is sufficient to allow all the useful power to be transferred to the load, provided that the impedance of the latter is not too small.

Other types of coupling to a coaxial line are used. Sometimes the inner conductor is directly connected to a strap which picks up the energy from the resonators, and sometimes the loop is introduced at one end of the magnetron block, connecting one segment to the inner conductor.

Waveguide coupling consists of a slot opening from one of the resonators to a guide which may or may not be 'choke coupled'. Fig. 11.27 shows a simple form of waveguide coupling, using a rising-sun magnetron. The reader is advised to refer back to this after reading the relevant part of Volume 5. The slot S is arranged to constitute a H_{10} guide of low characteristic impedance,

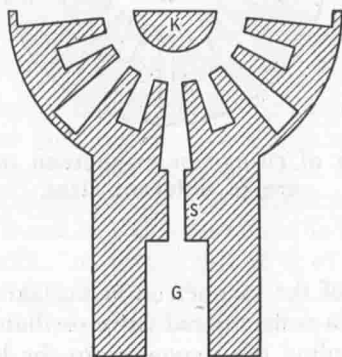


Fig. 11.27. Waveguide coupling to a rising-sun magnetron. S is a quarter-wave matching-section.

opening into another guide of higher characteristic impedance. The slot is a quarter-wave section, transforming from a low to a high impedance. The low impedance is required at the resonator end of the slot, since this is a current anti-node and a voltage node. A tapered guide is an alternative arrangement, but it is not so simple mechanically and its insensitivity to frequency variations is not usually important.

The Complete Magnetron Valve

In a typical magnetron valve the heater leads are brought in via glass-vacuum seals at a high negative potential. The anode block is at earth potential, and fits exactly between the pole-pieces of a permanent magnet. The output coupling unit is either a coaxial line, the vacuum seal positioning the inner conductor with respect to the outer, or a waveguide, the vacuum seal then being a window in the guide. The magnets for magnetron valves have been the subject of careful investigation. They vary in weight from less than a pound to about 300 lb according to flux and gap length. The magnetic material used is Alnico. Some magnetrons, the so-called 'packaged' type, comprise valve and magnet in one unit.

Typical operating figures for pulsed magnetrons are shown in Table 11.1. In this table it has been assumed that the duty cycle (see Chapter 9) is small enough to keep the heat dissipation within limits and to obtain the requisite pulsed current from the cathode.

TABLE 11.1

Frequency (Mc/s)	Anode Voltage (kV)	Magnetic Flux Density (weber/sq. metre)	Power Output (kW)
3,000	22	0.20	250
3,000	30	0.28	1,000
10,000	15	0.65	60
10,000	33	0.65	500

Frequency Pulling in Magnetrons

It should be noted that the magnetron oscillator, like all other oscillators, automatically adjusts its frequency to the reactance and resistance of the load across it. It is therefore possible to change the frequency of the magnetron to some extent by varying the load circuit. This must always be borne in mind when a magnetron is used in a transmitter. Tuning the valve for maximum output varies the frequency, and any adventitious change in the loading has the same effect.

The Travelling-Wave Valve

In any amplifier or oscillator the function of the electron stream is to convert d.c. energy from the h.t. supply to a.c. energy at the required frequency. In the multi-cavity magnetron a considerable fraction of the electrons leaving the cathode move gradually towards the anode alternately gaining energy from the h.t. supply and then losing this energy to the a.c. circuit. This they do many times before reaching the anode and this repetition of interchange of energy accounts for the relatively high efficiency of the device. The relatively low efficiency of the two-resonator klystron amplifier is fundamentally due to the fact that each electron has only one opportunity of transferring its energy to the circuit: the higher efficiency of the three-resonator klystron is due to the fact that the electrons have two opportunities of acting.

If a valve could be made in which a large fraction of the electrons continuously gave up their d.c. energy to the h.f. circuit over a large number of periods of the latter, a high-efficiency amplifier or oscillator would result, and this is to a certain extent successfully accomplished in the travelling-wave valve. The latter may be used either as an oscillator or as an amplifier. As an amplifier it is characterized by a high gain associated with an extremely large bandwidth. Its noise-factor, though not much if at all better than that of a silicon crystal, is considerably less than that of a klystron. It has therefore a certain value in some special communication and radar circuits.

Imagine a concentric line the inner conductor of which is a wire helix, as in Fig. 11.28. If a signal is applied across the terminals AB, it travels along

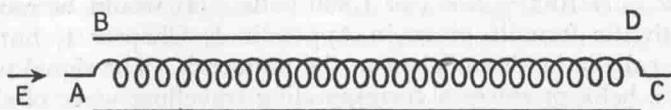


Fig. 11.28. Concentric line with the inner conductor in the form of a helix.

the inner conductor to the terminals CD at a speed which is determined by the total length of the wire forming the helix. That is, if the helix consists of N turns of diameter d , the total length is $N\pi d$ and, to a first approximation, the time taken by a signal to pass from A to C is $N\pi d/c$, where c is the velocity of light. If the shortest distance between A and C (along the axis of the helix) is l , the apparent speed of the signal along BC is $lc/N\pi d$. By giving N and d suitable values, this axial speed may be made small compared with c .

Now imagine that an electron gun at E produces a beam of electrons travelling along AC with speed v . Then, if $v = lc/N\pi d$, the electrons in the beam travel at approximately the same speed as the signal and a continuous interchange of energy between the electron stream and the circuit is feasible.

It was R. Kompfner who first stated the principle of the travelling-wave valve, associating the slowly-moving signal with the words 'travelling wave' in contradistinction to the 'standing wave' which exists in a klystron or a conventional valve amplifier. It should be noted, however, that the signal in a multi-resonator magnetron may be regarded similarly as a travelling wave, rotating round the anode.

The Travelling-Wave Amplifier

Suppose now we have a complete valve arranged as shown diagrammatically in Fig. 11.29 with an electron gun G and an electron collector C. The electrons are kept in the form of a narrow pencil as they traverse AB by the action of an axial magnetic field produced by the direct current in the external solenoid S. This action is explained in Chapter 16, when magnetic focusing of a beam of electrons is considered.

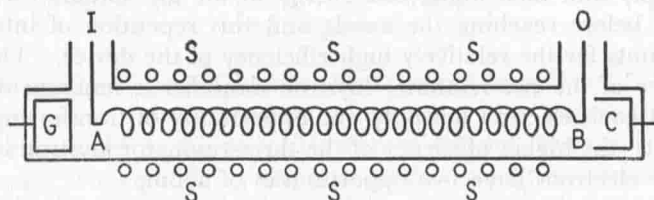


Fig. 11.29. Travelling-wave valve with electron gun G and collecting electrode C.

A small sinusoidal alternating voltage is introduced at the input terminals I and the steady accelerating voltage is varied to make the speed of the electrons equal to the axial speed of the travelling wave. In an experimental valve used by Kompfner the frequency of the signal voltage was 3,300 Mc/s, the length of the helix AB was 60 cm and its diameter 0.476 cm. There were 7 turns per cm of the length. Consequently the length of one turn of the helix was $\pi d = 1.5$ cm, or one-sixth of a wavelength, so that there were approximately 70 complete waves in the helix. The axial speed of the travelling wave was about $c/10.5$, equal to the electronic speed produced by an accelerating voltage $V = (c/10.5)^2/(2e/m)$, or 1,800 volts. (It would be more exact to use the relativistic formula given in Appendix 1, Chapter 1, but the above approximate calculation is sufficient at the moment). The signal wave travelling along the helix produces a corresponding travelling wave of electric field strength along the axis. The electrons which at A happen to find themselves in an accelerating field move with this field and experience velocity modulation

lasting over many periods of the alternating voltage. Similarly, the electrons which at A happen to be in a decelerating field are decelerated over many periods. This velocity modulation produces bunching as in the klystron, the centre of each bunch being at a point where the field is zero and is changing from decelerating to accelerating. The bunches themselves induce a corresponding series of travelling charges in the helix and these travelling charges, or currents, since they are in the proper phase, constitute an amplified signal. This amplified signal still further bunches the electrons which induce still larger variations of charge in the helix. The action of the electron bunches on the electrons in the wire of the helix robs the electrons in the bunches of some of their energy and so the conditions for amplification are satisfied. Direct-current energy has been transformed into signal-frequency energy.

Attenuation in the Travelling-Wave Amplifier

Referring back to Fig. 11.29, it will be noted that there is nothing to prevent the amplified travelling wave from moving back towards A and so giving rise to positive feedback in the amplifier. In fact such positive feedback occurs, transforming the amplifier into a self-oscillator, unless the attenuation between A and B is sufficient to prevent it. So, where the beam current is small and the energy fed into the valve is therefore less than that required to maintain oscillation, true amplification may be obtained. But, if the valve is used as a power amplifier (and this is its principal use at the moment), attenuation must be deliberately introduced into the helix to maintain the figure of merit of the amplifier. This attenuation may take the form of resistance in the wire or in the glass envelope surrounding the wire. It is usually confined to a section of the helix near the middle.

Bandwidth of the Travelling-Wave Amplifier

It will be noted that the valve itself cannot be a resonant system. Apart from the attenuation inherent in the wire of the helix, the fact that the valve is of the order of 100 wavelengths long electrically makes its action almost aperiodic. This is borne out by experiment. A travelling-wave amplifier made for a mean frequency of 3,000 Mc/s may have a bandwidth of 1,000 Mc/s and this is a most valuable property for certain applications. For comparison it should be noted that a two-resonator klystron amplifier, requiring skilled tuning, has a bandwidth less than 10 Mc/s at 3,000 Mc/s.

Noise in the Travelling-Wave Amplifier

The early promise of the valve has not been fulfilled in respect of noise. With rather critical adjustments of the h.t. voltage and beam current certain valves have shown a slight improvement in noise factor over the silicon crystal used as a mixer. But except in very special cases, the use of a complex valve, requiring a h.t. voltage of about 1,500 and an associated solenoidal magnetic field would not be worth while from an engineering standpoint.

At the moment small numbers of valves are being made by several manufacturers. At least one American firm has gone into production with a valve suitable for power amplification in the 4,000–5,000 Mc/s region. A power output of 10 watts is obtained with a gain of 100. Such a valve can readily be made a self-oscillator by the addition of positive feedback through a frequency-selective circuit.

CHAPTER 12

Noise

The Theoretical Limit to Measurement

IMAGINE THAT the attempt is made to detect and measure an extremely small electric current by means of a sensitive moving-coil galvanometer. By taking various precautions the galvanometer may be made so sensitive that an appreciable movement of the suspended coil occurs when a current of the order of 10^{-12} A flows in it. Such an instrument can detect such a current, but it is found impossible to measure the current accurately because the zero of the instrument continually varies. This variation may be shown to be inherent in the galvanometer itself and not to be due to any accidental movement of the supports. The random variations of the zero are in fact due to the thermal movements of the ions and electrons constituting the wire of the suspended coil.

No advantage is gained by making a more sensitive instrument. Greater sensitivity increases the effects due to the thermal movements of the electrons in the same ratio as the effect due to the current to be measured. The theoretical limit to the measurement of current by this type of galvanometer has been reached. Since the movement of the coil due to the thermal movements of the electrons is an unwanted and random effect, masking the effect of the signal to be measured, it is designated 'noise'. Noise constitutes the theoretical limitation to all detection and measurement, and must therefore be studied in some detail. The most fundamental form of noise, common to all forms of detecting and measuring apparatus is what is called 'thermal noise', and an example of this has just been considered in the case of the sensitive galvanometer.

Thermal Noise

The bar of copper AB shown pictorially in Fig. 12.1, comprises positive ions and electrons in equal numbers per unit volume. The ions form a crystalline lattice and can only vibrate about their positions of static equilibrium. The electrons are almost completely free to move between their 'collisions' with the ions. Putting the matter more accurately, the electrons move from the field of force of one ion to that of another, the average force on each electron being small, but sufficiently high as the electron nears an ion to cause it to change its direction of motion abruptly. This is indicated pictorially in Fig. 12.1 by representing the almost stationary ions by small circles and the 'free' electrons by arrows pointing at random in all directions.

At any given instant the distribution of the electrons in the bar gives rise to a small potential difference between its ends and, since the potential difference is due to electronic movement, it is in fact a thermal e.m.f. At a later instant the e.m.f. is equally likely to be in the opposite direction. The e.m.f., like the electronic motion, is random.

The thermal energy in the metal is proportional to the absolute temperature just as in the case of a gas, and the electrons share this energy equally with the positive ions. If another bar CD is joined at its ends to AB and is imagined to be 'noiseless', then AB will feed electric power to CD by reason of the thermal e.m.f. In fact AB behaves like an electric generator having its own internal resistance R , and generating a definite amount of power in a definite frequency interval B . The power generated (as one might expect) is proportional to T , the absolute temperature, and to B , the frequency interval. The

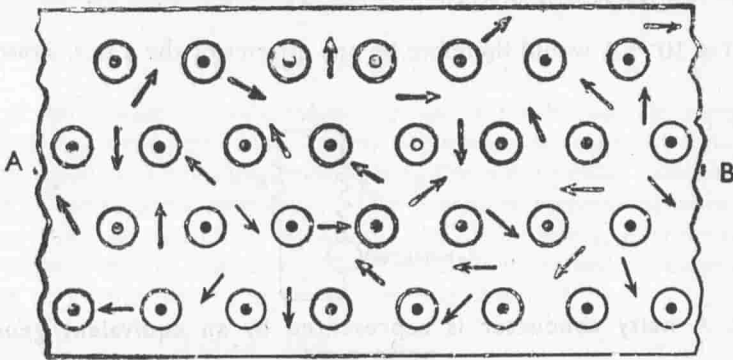


Fig. 12.1. The electrons and ions in a copper bar AB are here represented by arrows and circles respectively. The ions are almost stationary and the electrons move in a random manner.

constant of proportionality may be shown by quantum mechanics to be four times Boltzmann's Constant, k , equal to 1.37×10^{-23} joules/degree. If a more exact investigation of the noise in the bar is made, it is found that the power generated in a frequency interval B is only $4kTB$, as has just been stated, if the mean frequency of the interval and the interval itself are small compared with the natural frequencies of the atoms of the bar. This proviso does not affect the validity of the $4kTB$ formula at ordinary radio frequencies. It might have some importance at frequencies above 20,000 Mc/s.

Noise Power Available from a Resistance R

The total noise power generated by a piece of metal is $4kTB$. Regarding the metal as a generator of e.m.f. and having an internal resistance R , the maximum transfer of power to a 'noiseless' resistance connected across it will be obtained when the noiseless resistance is R also. For in that case the e.m.f. of the equivalent noise generator on short-circuit is given by E_N , where

$$E_N^2/R = 4kTB \quad \text{or} \quad E_N = \sqrt{(4kTBR)} \quad . \quad . \quad . \quad (12.1)$$

But if the noiseless resistance is R also, then the current I_N is given by $I_N = E_N/2R$, and the power available in the noiseless resistance is $I_N^2 R = E_N^2/4R = kTB$.

It is therefore usual to represent a noisy conductor by the equivalent generator of Fig. 12.2, indicating that the maximum noise power available from it is kTB when a noiseless resistance equal to the resistance of the conductor is connected across it. It should be noted that the maximum noise power available from a resistor is independent of its resistance.

Example of Noise Power

Returning to the galvanometer, discussed in a previous paragraph, if it is so made that it responds to frequencies less than or equal to 4 per second, then $B = 4$ c/s, $T = \text{room temperature} = 290^\circ \text{K}$ and $kTB = 1.37 \times 10^{-23} \times 290 \times 4$ watt which is approximately 1.6×10^{-20} watt. Let the resistance of the galvanometer and external circuit be 1,000 ohms. Then the r.m.s. current due to the noise power is I_N , where

$$I_N^2 R = 1.6 \times 10^{-20} \quad \text{or} \quad I_N = 4 \times 10^{-12} \text{ A.}$$

A current of 10^{-12} A would therefore be one quarter of the r.m.s. noise current.

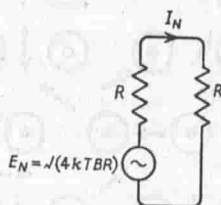


Fig. 12.2. A noisy conductor is represented by an equivalent generator of noise e.m.f.

Shot Effect in Valves

The current in a valve consists of a large number of moving electrons which have been emitted by the cathode in a random manner. This description of the action implies that an exact determination of the current arriving at the anode would disclose random variations corresponding to the randomness of the cathode emission. This is exactly true of a diode valve in which the potential difference between cathode and anode is large enough to produce temperature limitation of the current. In other words it is exactly true of a diode in which there is a negligible space-charge between the electrodes.

The random variations in the anode current of a temperature-limited diode constitute a noise current. If such a valve is used in a radio receiver, this noise current may contribute an unwanted background to the signal. For the effect to be important, (as will be seen later) the valve must be dealing with a very small signal, and so shot noise, as it is called, is of importance only in the early stages of a receiver.

It may be shown by statistics that the maximum r.m.s. noise current I_N obtainable from a diode in a frequency band B is related to the total cathode current I_K by the equation

$$I_N = \sqrt{(2eI_K B)} \quad . \quad . \quad . \quad . \quad . \quad (12.2)$$

where e is the electronic charge. This maximum noise current is only developed in a temperature-limited valve. If the valve is space-charge limited, as is usually the case, the noise current is reduced by a factor F , which depends upon the amount of space-charge. Then

$$I_N = F \sqrt{(2eI_K B)} \quad . \quad . \quad . \quad . \quad . \quad (12.3)$$

and F may have any value between 1.0 and 0.1.

A brief explanation of this space-charge *smoothing*, as it is called, is that the cloud of electrons near the cathode surface acts as a reservoir of charge from which the accelerating electric field can draw more uniformly than from the cathode. The electrons in the 'virtual' cathode formed by the space-charge are at rest. They are therefore drawn off to the anode with a much smaller distribution of velocity and in more constant numbers per unit time.

Equation (12.3) can be given an alternative form for a fully space-charge limited diode. By considering the distribution of velocities among the ejected electrons, it may be shown that the noise power of the equivalent short-circuited generator is

$$I_N^2 r_A = 4k(0.64 T)B \quad (12.4)$$

where r_A is the dynamic anode resistance and T is the cathode temperature. The factor F has here been absorbed in the numerical constant 0.64. If the diode is matched to a noiseless resistance r_A the noise-power available is then $k(0.64T)B$ in the noiseless resistance. This is a more convenient form for most calculations than that given in equation (12.3). However, experiment shows that the values given by equation (12.4) are about two-thirds of those obtained in practice, so that an even simpler equation is obtained

$$I_N^2 r_A = kTB \quad (12.5)$$

showing that a space-charge limited diode behaves like a resistance, equal to its dynamic resistance, at the temperature of the cathode.

Shot Noise in a Triode

In a triode the potential chiefly controlling the current through the valve is that of the grid, and the formula corresponding to (12.4) is

$$I_N^2/g = 4k(0.64T)B \quad (12.6)$$

where the inverse of the mutual conductance ($1/g$) takes the place of r_A for the diode. This noise power appears in the anode circuit, as if it were due to a fluctuating anode current $\sqrt{4k(0.64T)gB}$. Regarding this current as the result of a fluctuating voltage V_N at the grid, since $I_N = gV_N$, we have

$$V_N = \sqrt{4k(0.64T)B/g}$$

This is the e.m.f. developed by a noisy resistance of value R_E , where

$$\sqrt{\frac{4k(0.64T)B}{g}} = \sqrt{4k T_0 B R_E}$$

T being the cathode temperature and T_0 being room temperature. Solving this equation for R_E , we have

$$R_E = (0.64T)/(gT_0),$$

or, taking $T = 1,100^\circ\text{K}$ and $T_0 = 290^\circ\text{K}$,

$$R_E = 2.5/g \quad (12.8)$$

This simple formula is well confirmed by experiment and the result is important. It means that a (noisy) triode amplifier valve may be replaced by an imaginary noiseless valve having a noisy resistance of value $2.5/g$ in series with

the signal in the grid circuit. R_E is known as the equivalent noise resistance of the triode, and it will be noted that a triode having a low mutual conductance has a large equivalent noise resistance. The equivalent noise resistance of an input triode valve is usually between 300 and 3,000 ohms.

Noise in a Multi-Electrode Amplifier Valve

Formula (12.8) does not give the total equivalent noise resistance of a valve in which the current may divide between two or more positively-charged electrodes. When the cathode current divides between, let us say, the anode and screen-grid of a pentode, the current flowing to the screen-grid constitutes a separate random phenomenon in the valve, totally uncorrelated with the variations in cathode emission. This adds a new source of noise to the valve called *partition* noise because it is due to the partition of the current between the two electrodes. It is shown in the next paragraph that the equivalent noise resistance of a pentode in which the cathode current is I_K , the anode current is I_A , the screen-grid current is I_{G2} and the mutual conductance is g

$$\text{is } R_E, \text{ where } R_E = \frac{2.5}{g} \left(\frac{I_A}{I_K} \right)^2 + \frac{20}{g^2} \frac{I_A I_{G2}}{I_K} \quad (12.9)$$

The mutual conductance g is in mA/V and R_E is in kilohms. The factor 2.5 is dimensionless, but 20 has the dimensions of volts⁻¹: This formula is well substantiated by experiment: it leads to values for high-slope pentodes varying from 750–5,000 ohms. Variable- μ pentodes have substantially higher equivalent resistances owing to their lower effective mutual conductances.

Derivation of the Formula for the Equivalent Resistance of a Pentode

► If the electron current flowing to the anode is I_A and that flowing to the screen is I_{G2} , and if $I_K = I_A + I_{G2}$, the mean-square noise current appearing in the screen-grid lead is $2eI_{G2}B$. This is merely a statement that completely random variations are to be expected; there is no space-charge smoothing. Turning now to the corresponding partition noise in the anode lead, it might be expected at first sight that the mean-square current would be the same, $2eI_{G2}B$. But this is only the case if I_{G2} is very small, for it is clear that the noise in the anode lead must be zero not only when I_{G2} is zero but also when I_A is zero. Hence, the mean-square partition noise current at the anode is actually $2e(I_A I_{G2}/I_K)B$, giving zero noise if I_A or I_{G2} tends to zero, and $2eI_{G2}B$ only when I_A tends to I_K ; i.e., when I_{G2} is small. In the anode lead of a pentode, therefore, there are two independent mean-square noise currents which must be added. First, there is the mean-square current due to the variations at the cathode and given by $I_{NK}^2 (I_A/I_K)^2$ where I_{NK} is the r.m.s. noise current in a triode carrying a current I_K . Secondly, there is the mean-square partition noise current, given by $2eI_{G2} I_A B/I_K$. But, $I_{NK}^2 = 4k(0.64T)gB$, [equation (12.6)], whence the total mean-square noise current I_{NA}^2 is

$$I_{NA}^2 = 4k(0.64T)gBI_A^2/I_K^2 + 2eI_{G2} I_A B/I_K.$$

If, as in the paragraph on noise in a triode, we attribute this noise current to an equivalent resistance R_E in the grid circuit of the pentode, then $4kT_e B R_E = I_{NA}^2/g^2$, and, writing $T = 1,100^\circ\text{K}$, currents in mA, g in mA/V,

$$R_E = \frac{2.5}{g} \left(\frac{I_A}{I_K} \right)^2 + \frac{20}{g^2} \frac{I_A I_{G2}}{I_K} \quad (12.9)$$

This is the general equation giving the equivalent noise resistance of an amplifier pentode or screen-grid valve.

Other Sources of Noise in Valves

There are other sources of noise in valves which are of importance only in special circumstances. The presence of gas in a valve may cause an appreciable increase in the noise power. This may be regarded as a rather extraordinary kind of partition noise. The collisions between electrons and gas atoms or molecules may give rise to new electrons, and these, arriving at the anode, contribute a noise component proportional to the equal positive ion current in the valve. Needless to say, this source of noise should be negligible in well-constructed conventional valves and should be suppressed by a proper choice of potentials in gas-filled devices working with small signals.

Another special noise is that encountered at very-high frequencies, when, as has been shown in Chapter 10, transit-time effects produce a finite input conductance proportional to $f^2 g^2 t^2$, where f is the frequency, g is the mutual conductance and t is the cathode-grid transit time. This equivalent resistance in the grid-cathode circuit is not noiseless. Indeed, the fluctuations in the grid potential induced by the fluctuating cathode current produce so much noise in the anode circuit that the equivalent input resistance must be imagined to be at a temperature of the order of $5T_0$, where T_0 is room temperature. This effect may constitute the predominant valve noise at extremely-high frequencies.

In addition to this transit-time noise, partition noise may limit the action of a valve at very-high frequencies by reason of the lead inductance in the cathode circuit. When the impedance of the lead inductance is not negligible, it produces negative feedback with respect to both signal and cathode shot noise, but it leaves the partition noise unaffected. Hence in the limit, the signal is reduced below the partition-noise level. This is the explanation of the failure of multi-electrode valves at ultra-high frequencies.

Noise Factor of an Electronic Device

Regarding any electronic device with its associated circuits as a four-terminal network, as in Fig. 12.3, the input signal voltage V_{IS} and the input noise voltage



Fig. 12.3. A noisy device with an input signal V_{IS} has an equivalent noise input signal V_{IN} .

V_{IN} may be imagined to be determined. Similarly, the output signal voltage V_{OS} and the output noise voltage V_{ON} may be imagined determined. Then the input-signal power/input-noise power when divided by the output-signal power/output-noise power gives what is called the noise factor N of the device or, putting it in the form of an equation,

$$\frac{V_{IS}^2}{V_{IN}^2} = N \frac{V_{OS}^2}{V_{ON}^2} \quad (12.10)$$

This definition of noise factor is of universal application and may be used for an amplifier stage or a mixer stage or a complete amplifier of many stages or for a receiver itself. The definition itself is easy to understand, but the actual determination of the signal to noise ratios at input and output may be a matter of extreme difficulty.

It should be noted that a valve by itself does not possess a noise factor, since the signal to noise ratios depend upon the circuits associated with it and, in particular, on the frequency of the signal. However, if the frequency is specified and if, therefore, the input impedance of the valve as an amplifier may be calculated, an optimum noise factor may be estimated for an amplifier in which the circuits are loss-less.

The noise factor of a device is usually expressed in decibels. For example, an amplifier incorporating the EF50 pentode valve and operating at a frequency of 45 Mc/s should have a noise factor of about 4 db. Similarly, an amplifier incorporating the CV354 and operating at 1,000 Mc/s cannot have a noise factor less than 9 db. These figures have been calculated by methods beyond the scope of this book.

- It is perhaps worth while to examine the noise factor of a signal amplifier in more detail. Fig. 12.4 represents the input circuit of a valve, showing signal and noise e.m.f.s. Only r.m.s. values are used. The signal applied to the valve input circuit may be regarded as due to a generator of open-circuit e.m.f. E_{IS} and internal resistance R_S . This is applied across the valve input circuit at AB, assumed to be a parallel-tuned circuit operating at resonance, so that it may be represented by the resistance R_D .

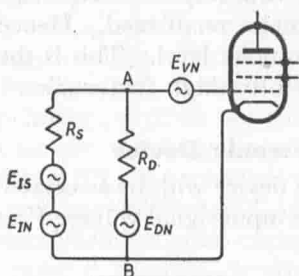


Fig. 12.4. The equivalent input circuit of a valve showing the generators which represent the signal and noise voltages.

There are three noise generators present in the input circuit. First, the signal generator, whether it be an aerial or another tuned circuit, gives rise to a noise e.m.f. E_{IN} . Secondly, the dynamic resistance of the tuned circuit in parallel with the internal valve input resistance forms R_D , and this gives rise to a noise e.m.f., E_{DN} . Thirdly, the valve itself gives rise to an equivalent noise e.m.f., E_{VN} in series with the signal voltage. These three noise generators are shown in Fig. 12.4.

These separate noise generators combine to produce a single noise voltage at the grid V_{GN} and, similarly, the signal e.m.f. E_{IS} gives rise to a signal voltage at the grid V_{GS} . Since the valve may be considered noiseless, the valve noise

having been transferred to the grid, $V_{GS}/V_{GN} = V_{OS}/V_{ON}$, the signal to noise ratio at the output, and, consequently,

$$\frac{E_{IS}^2}{E_{IN}^2} = N \frac{V_{OS}^2}{V_{ON}^2},$$

where N is the noise factor of the amplifier. It will be noted that, as mentioned earlier, the value of N depends both upon R_s and upon R_D , so that the valve must be considered along with its circuits.

In this analysis any noise arising from the output circuit of the valve has been neglected. This is usually permissible, since the stage gain is considerable. But it does not affect the definition of noise factor in any case, and only makes it necessary, if it need be considered, to increase the value of R_E to suit. ◀

The Noise Factor of a Multi-stage Amplifier

A signal amplifier may consist of several stages each comprising a valve and its associated circuits. By so connecting stages in cascade a total power amplification of 10^8 (80 db), or more, may be obtained. In such an amplifier the first stage contributes by far the greater part of the noise for, since noise and signal are both amplified together, if the power amplification per stage is A , then the noise N_1 appearing at the grid of the first valve becomes AN_1 at the grid of the second, and this is likely to be much larger than any additional noise produced by the second stage. If, however, A is small or, as in some cases, is less than unity, the noise due to the second stage may be important. This is the case in many radar receivers, where the noise contributions of the mixer stage and the first stage of the i.f. amplifier may be of the same order of magnitude. It may easily be shown that, if the noise factors of the various stages are N_1, N_2, N_3 , etc. and their power amplifications are $A_1, A_2, A_3, \dots$, the over-all noise factor of the amplifier is N , where

$$N = N_1 + \frac{N_2 - 1}{A_1} + \frac{N_3 - 1}{A_1 A_2} + \frac{N_4 - 1}{A_1 A_2 A_3} + \dots \quad (12.11)$$

from which it may be seen that if the power gain is 100 per stage, and $N_1 = N_2 = N_3$, then $N = N_1$ to within 2%.

► To prove equation (12.11), let the signal at the input be S and the thermal noise due to the input resistor be P_N . Then the equivalent noise at the grid of the first valve is $N_1 P_N$, imagining the stage to be noiseless. Of this noise, $N_1 P_N$, the part due to the stage itself is $(N_1 - 1)P_N$. Similarly, the noise equivalent of the second stage at the grid of the second valve is $(N_2 - 1)P_N$ and so on. The signal at the output of the first stage is $A_1 S$, and at the output of the n th stage is $A_1 A_2 \dots A_n S$.

The noise at the output of the first stage is $A_1 N_1 P_N$.

The noise at the output of the second stage is $A_1 A_2 N_1 P_N + A_2 (N_2 - 1)P_N$

The noise at the output of the n th stage is $A_1 A_2 \dots A_n N_1 P_N + A_2 A_3 \dots A_n (N_2 - 1)P_N + A_3 A_4 \dots A_n (N_3 - 1)P_N + \dots + A_n (N_n - 1)P_N$.

Therefore, if the original signal to noise ratio is S/P_N , the final signal to noise ratio at the output of the n th stage is $A_1 A_2 A_3 \dots A_n S / \{A_1 A_2 \dots A_n N_1 P_N + A_2 A_3 \dots A_n (N_2 - 1) P_N + \dots + A_n (N_n - 1) P_N\}$, whence the noise-factor N of the amplifier is

$$N = \frac{1}{(A_1 A_2 A_3 \dots A_n) / \{A_1 A_2 \dots A_n N_1 + A_2 A_3 \dots A_n (N_2 - 1) + \dots + A_n (N_n - 1)\}}$$

or $N = N_1 + \frac{N_2 - 1}{A_1} + \frac{N_3 - 1}{A_1 A_2} + \dots + \frac{N_n - 1}{A_1 A_2 \dots (A_n - 1)} \dots$ (12.11) ◀

Noise in Mixer Stages

The simplest case is the triode mixer, where partition noise is not present. The treatment of the noise factor of the mixer is then similar to that of the amplifier and leads to a formula for the equivalent noise resistance similar to equation (12.8). First, it is again assumed that the mean-square noise current in the anode circuit is I_N^2 , where

$$I_N^2 = 4k(0.64T)gB, \text{ and } T \text{ is the cathode temperature.}$$

But in this case g is the average value of the conductance taken over one cycle of the local oscillator. Secondly, the equivalent noise resistance is defined, as before, to be R_E , where the r.m.s. noise voltage at the grid is V_N , and

$$V_N = \sqrt{(4kT_o R_E B)}$$

Finally, using the conversion transconductance g_c as defined in Chapter 6, we have for zero anode load $I_N = g_c V_N$. Combining these three equations,

$$R_E = \frac{4k(0.64T) g B}{4kT_o B g_c^2} = \frac{0.64T}{T_o} \frac{g}{g_c^2}$$

Also, using $T = 1,100^\circ\text{K}$, and writing the conductances in mA/V,

$$R_E = \frac{2.5 g}{g_c^2} \text{ kilohms}$$

or, if $g_c = 0.28g_m$, where g_m is the maximum value of the mutual conductance, and if $g = 0.5g_m$

$$R_E = \frac{16}{g_m} \dots \dots \dots (12.12)$$

This is the commonly used formula for the equivalent-noise resistance of a triode mixer. In pentode, hexode and other mixers, where partition noise occurs, greater equivalent noise resistances are obtained. In a hexode, for example, R_E is about sixteen times the triode value. From the point of view of sensitivity, therefore, a triode is the best form of mixer, but considerations other than ultimate sensitivity govern the design of many receivers. Moreover, as explained in the preceding paragraph, where the mixer stage is preceded by a useful amount of signal amplification, the effective mixer noise is much reduced.

Local-Oscillator Noise

In a superheterodyne receiver it is possible for an appreciable fraction of the noise to be generated in the local oscillator and the reason for this is as follows:—

If the e.m.f. supplied by the local oscillator is not exactly sinusoidal, but contains noise fluctuations, it may be regarded as being made up of a strong signal at the true local-oscillator frequency f_o and a series of much smaller signals at frequencies both higher and lower than f_o . Indeed, the local-oscillator e.m.f. may be regarded as consisting of a continuous spectrum of noise frequencies superimposed upon a strong signal at f_o . The noise components do not directly contribute to the noise factor of the receiver, since they are not at either the signal frequency or at the intermediate frequency. But when the local-oscillator e.m.f. is transferred to the mixer, the non-linear action of the latter brings into being signals at frequencies equal to the differences between f_o and the noise components. Some of these difference frequencies may fall within the bandwidth of the i.f. amplifier and will there contribute noise to the receiver.

Local-oscillator noise is particularly troublesome at ultra-high frequencies. This is for two reasons. First, u.h.f. valves are particularly prone to produce noise, and secondly, the local oscillator and signal are generally fed into the same mixer circuit which does not attenuate the noise sidebands to anything like the same extent as do the selective circuits used at lower frequencies. This is particularly true where the intermediate frequency is a small fraction of the signal frequency. It may be remembered that it was stated in Chapter 11 that the magnetron oscillator is 'noisy', and is therefore unsuitable for use as a local oscillator. Certain other valves exhibit similar effects, particularly those reflex klystrons in which electrons may oscillate between the cathode and the reflector several times before disappearing into the anode.

Local-oscillator noise can be greatly diminished by the use of appropriate circuits. An obvious precaution is to reduce the coupling between the oscillator and the mixer, simultaneously increasing the available oscillator power. This diminishes the loss of signal power to the local-oscillator circuit. Other more elaborate devices can be adopted which will cancel out almost all the local-oscillator noise and leave the wanted signal unaffected.

Measurement of Noise: The Noise Diode

In an earlier paragraph it was stated that the r.m.s. shot noise current I_N available in the anode circuit of a temperature limited diode is given by

$$I_N = \sqrt{(2eI_KB)} \quad \dots \quad (12.2)$$

where e is the electronic charge, I_K the cathode current, and B the bandwidth. Imagine such a diode connected as in Figure 12.5.

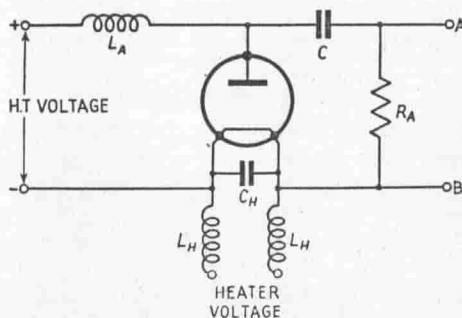


Fig. 12.5. Basic circuit of a noise measuring instrument.

The three chokes, L_A and L_H , are so chosen that they effectively prevent the passage through them of components of current at frequencies between f_1 and f_2 , where $f_2 - f_1 = B =$ the bandwidth under consideration. Then the noise current produced by the diode passes back to the cathode via the capacitor C and the resistor R_A . The mean-square noise voltage developed across R_A is V_N^2 , where

$$V_N^2 = e2I_K BR_A^2 \quad . \quad . \quad . \quad . \quad (12.13)$$

If the resistance R_A is now used as a noise source, and a receiver or other noisy device is connected across AB , a variable known noise power may be fed to the receiver merely by varying I_K . This may be done by varying the heater current in the diode valve, provided that the latter has a tungsten or thoriated-tungsten cathode. As mentioned in Chapter 2, oxide-coated cathodes cannot be used consistently when the valve is temperature limited.

The resistance R_A is usually chosen to be suited to the input impedance of the receiver or other device connected across it. Then if X is the ratio of the noise power output from the receiver with the diode switched on to the noise power output with the diode switched off, the signal to noise ratio at the output is $X - 1$. The corresponding signal to noise ratio at the input is $V_N^2/4kT_o BR_A$. Hence the noise factor of the receiver is N , where

$$N = \frac{V_N^2}{(X - 1)4kT_o BR_A} = \frac{eI_K R_A}{2kT_o(X - 1)} \quad . \quad . \quad . \quad (12.14)$$

This method of measuring the noise factor of a device is gradually gaining in popularity. It eliminates the necessity for an expensive and accurately-calibrated signal generator. The method can only be easily applied where the noise factor is less than about 30 db. If a receiver is really bad, the noise diode will not deliver enough noise power to give a useful value for X .

CHAPTER 13

Gas-Filled Valves

THROUGHOUT the preceding chapters only valves from which gas has been purposely excluded have been considered. It has all along been assumed that the gas pressure in the valve is so small that any effects due to collisions between electrons and gas molecules or atoms can be neglected. There is a large and growing class of valve from which gas has not been excluded and in which the action depends upon the effects of electron-atom collisions. Such valves generally contain a specified gas (not air) at a comparatively low pressure. Indeed, the majority of gas-filled valves operate at a pressure less than 10 mm of mercury, or, using the m.k.s. system, at a pressure less than 1,334 newtons/sq. metre. Normal atmospheric pressure is about 10^5 newtons/sq. metre.

Gas-filled valves are used for many purposes. At the one end of the scale, hot-cathode mercury-vapour diodes are employed for the rectification of a.c. at power frequencies. At the other end of the scale, cold-cathode gaseous-discharge valves are employed as switches in centimetric radar equipment. Some forms of marine electronic equipment contain only gas-filled valves, the latter performing all the functions usually belonging to high-vacuum devices.

Ionization in Gases: Fundamental Processes

An electron which has been accelerated from rest by a potential difference, V , between two electrodes has gained kinetic energy, $mv^2/2$. This has been derived from the loss of potential energy, Ve , whence $mv^2/2 = Ve$. If such an electron collides with a gas atom, it has a finite chance of detaching an electron from the atom (or of ionizing it) if its kinetic energy is greater than a certain critical value. For kinetic energies less than the critical value no measurable ionization can occur. The critical energy which an electron must have varies from one gas to another and is always expressed in electron-volts. The critical energy is called the *ionization potential* of the atom or molecule concerned. Table 13.1 gives the ionization potentials (V_I) of the gases commonly used in valves.

TABLE 13.1

Atom or Molecule	..	He	Ne	A	Kr	Xe	H	Hg	H ₂	N ₂
V_I (electron-volts)	..	24.5	21.5	15.7	14.0	12.1	13.5	10.4	15.8	16.3

It will be noted that an ionization potential is given for atomic hydrogen (13.5) and a different ionization potential for molecular hydrogen (15.8). This means that if the hydrogen molecule is already dissociated, a free electron

can detach the bound electron from the atom when the energy of the free electron is 13.5 electron-volts. If, however, the free electron detaches a bound electron from the hydrogen molecule, the free electron must possess an energy of at least 15.8 electron-volts.

There is no clear evidence that moderately fast-moving positive ions, when they collide with a gas atom or molecule, can detach an electron. Apparently such ionization by collision is a property of the electron alone. Positive ions may, however, when they collide with a metal surface (the cathode) produce secondary electrons. They may also, on being neutralized, produce radiation which is itself capable of ionizing the gas. Therefore, directly or indirectly, both electrons and positive ions can produce new ions when their energies at collision are high enough.

Consider now a diode, containing enough gas to ensure that most of the electrons leaving the hot cathode will collide with a gas atom or molecule before reaching the anode. If the potential difference between anode and cathode is less than the ionization potential of the gas, the valve behaves in much the same way as a normal high-vacuum diode. More heat is required to raise the cathode to its normal emitting temperature because of the heat conductivity of the gas and the effect of the gas on the emitting surface, but when emission has been achieved, the anode current is roughly what one would expect in a corresponding evacuated valve. When the anode to cathode potential difference is greater than the ionization potential of the gas, the electrons have a finite chance of making an ionizing collision before reaching the anode. Each ionizing collision gives rise to a new electron and a new positive ion. Hence, the total current in the valve is increased, and may be multiplied many times for a comparatively small increase in the cathode-anode voltage.

Ionization in a Triode

The actual ionization potential of a gas may be found very easily using a gas-filled triode. Imagine such a valve containing the rare gas argon at a pressure of about 20 newtons/sq. metre. Under these conditions an electron on the average travels a distance of about a quarter of a centimetre between collisions.

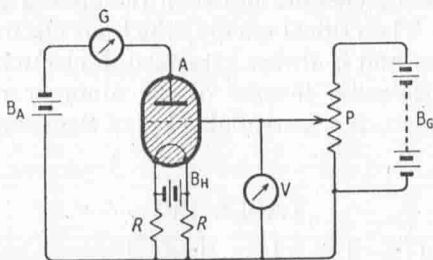


Fig. 13.1. Circuit for determining the ionization potential of a gas triode.

Let the triode be connected to supply batteries as shown in Fig. 13.1. The valve is represented by the usual symbol for a triode with cross-hatching to show that it contains gas. The valve filament, which is preferably of tungsten, is heated by the battery B_H , across which two equal resistances, R , are placed in series, their centre point being earthed. The negative of a 30-volt battery,

B_G , is connected to earth and a potentiometer, P , is placed across this second battery. The grid is connected to the potentiometer tapping and a voltmeter, V , indicates the grid-cathode potential difference, V_G , measured to the cathode centre-tap. The anode is connected via a galvanometer, G , to the negative end of a 2-volt battery, B_A , the positive of which is earthed.

If the voltage, V_G , is less than the ionization potential, V_I , of the gas argon, no current flows in the galvanometer, G , for only positive ions can find their way to the anode and there are no positive ions produced. But when V_G reaches V_I , current begins to flow. Positive ions reach the anode and are there neutralized by the absorption of a free electron. Hence the cut-off value in the I_A - V_G characteristic measures V_I directly. There are small errors due to the potential variation along the filament and to the variable speeds with which the electrons leave the cathode.

Cold-Cathode Gas Discharge

If a diode valve does not have a hot cathode, but only two similar plates placed opposite one another in an envelope containing a gas at a low pressure, the valve is known as a cold-cathode gas-filled discharge valve, and is represented by the symbol shown at D in Fig. 13.2.

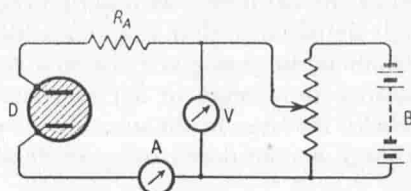


Fig. 13.2. Cold-cathode diode in a circuit for demonstrating its characteristics.

Imagine such a valve connected as shown, where A is a milliammeter, V is a voltmeter and the battery B has an e.m.f. of, say, 400 volts. R_A is a resistance of the order of 100 kilohms. If the potential difference across the valve, V_A , is gradually increased from zero, no apparent current flows in the meter, A , until V_A is greater than a certain value, say 350 volts, but as soon as the current shows on A , the voltmeter reading falls to about 300 volts and a glow appears in the valve. If the gas in the valve is neon, the glow is pink; if the gas is argon, the glow is blue.

When the voltmeter reading is 300 and the current indicated by A is, let us say, 1 mA, the potential drop in R_A must be $10^5 \times 10^{-3} \text{ V} = 100 \text{ V}$, and V_A must be 200. We have here the curious fact that the gas remains practically non-conducting until $V_A = 350$, but maintains a current of 1 mA at a potential difference of 200 volts. The gas must have broken down in some way at the so-called *striking* voltage of 350, but carries a comparatively large current at $V_A = 200$.

There is no physical action occurring in the valve which has not been mentioned in the previous paragraph. As V_A is increased, there comes a point at which the potential difference is sufficient to give an electron which happens to be in the valve sufficient kinetic energy between one collision and the next to ionize a gas atom or molecule. Only average values are of importance for,

when V_A is just greater than V_I , the very few electrons which by chance make almost the whole journey from cathode to anode without a collision produce ionization: but this ionization is negligible. On the average, for the particular valve under consideration, V_A has to rise to nearly 350 volts before the ionization produced ceases to be negligible. Two questions remain to be answered. (a) "Where do the initial electrons come from?", and (b) "Is this ionization by electronic collision sufficient to explain the large current of 1 mA which flows after the valve has begun to conduct?" The answer to (a) is that there are always a few free electrons in any volume of gas. These may be produced by cosmic radiation or by radiation from active material in the neighbourhood or by several other processes. So it is quite reasonable always to assume that at least one electron is present between the two plates of the valve. The answer to (b) is that ionization by electronic collision is *not* sufficient by itself to cause the large current which flows when the valve conducts.

Ionization by electronic collision can only produce a current proportional to the number of electrons originally present, and a current, moreover, which will cease to flow if the initial ionization is prevented. The discharge current in D bears no relation to the random variations in the natural ion current. For a glow to appear and for a comparatively large current to flow it is necessary for the second action to come into play: the positive ions must produce new electrons by their action at the cathode. This action requires a slightly higher voltage across D to give it significance than does ionization by electrons, and so it is only after the electronic action has given rise to a considerable ion current (1 μ A, say) that the positive ions begin to act as generators, producing what is called a *self-sustained electric discharge* in the gas. In Appendix 13.1 the theory of initiation of the discharge is considered in more detail.

Electric-Discharge Characteristic

Referring back to Fig. 13.2, if the resistance R_A can be varied between wide limits and if the battery B is capable of giving large currents, it is possible to plot the complete current-voltage characteristic of the gaseous discharge. This is extremely complicated and a slightly idealized version will be described to bring out the points of importance for gas-filled devices. Fig. 13.3 is a complete (simplified) discharge characteristic in which the current scale is logarithmic to allow the whole curve to be drawn in a reasonable space. Various parts of the characteristic have been given special names and the description will be given under these titles.

1. *Corona Characteristic.* This is the name given to the part of the characteristic between A and B. The current density is small near A and increases rapidly towards B, causing the electric field between the electrodes to deviate considerably from the electrostatic field. At B the space-charge is so considerable that the greater part of the potential rise from cathode to anode occurs in a very short distance near the cathode. Beyond this in the *anode glow* the electric field is small. The rapid fall of potential near the cathode is known as the *cathode fall*. Where the cathode fall occurs the discharge is non-luminous.

2. *Normal Glow Characteristic.* This is the characteristic from B to C. Its most remarkable feature is the fact that in this region the current increases considerably with hardly any increase in potential difference. In this region

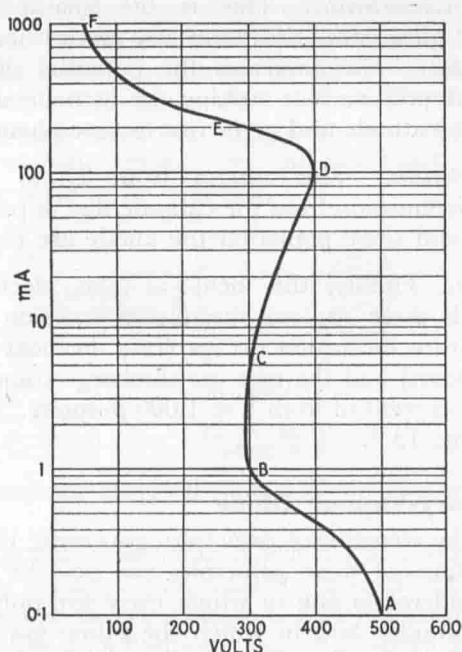


Fig. 13.3. Current-voltage characteristic of cold-cathode gas-diode.

the current density remains constant, but the glow covers a greater and greater cross-section of the valve and extends over a greater and greater area of the cathode surface. The glow very near the cathode, so close to it that it looks like a skin, is known as the *cathode glow*. Between the cathode glow and the *negative glow* (Fig. 13.4) is the cathode fall of potential. From the negative glow the luminosity falls off at first towards the anode and then increases in the *positive column*. The dark space between the negative glow and the positive column is known as the *Faraday dark space*. In the normal glow discharge the cathode fall of potential remains constant. It depends only upon the nature of the gas in the valve and on the nature of the cathode and is usually between 100 and 400 volts. Fig. 13.4 shows the distribution of regions in a typical normal glow. The anode glow may be striated.

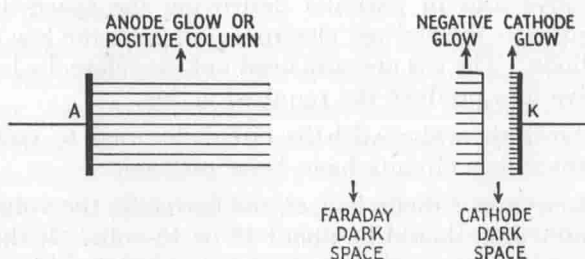


Fig. 13.4. The glow regions of a cold-cathode valve under conditions of normal glow.

3. *Abnormal Glow Characteristic.* This is the characteristic from C to D, where the cathode glow cannot increase any more in area and the current density has to increase. This increases the potential difference across the valve and leads to the positive ions striking the cathode with greater energy. This in turn heats the cathode and gives rise to new phenomena.

4. *Glow-Arc Characteristic.* This ranges from D to E and represents the onset of thermionic emission from the cathode due to positive-ion bombardment. The cathode and some points on the anode are now extremely hot.

5. *Arc Characteristic.* Finally, the heat at the electrodes becomes so great that not only is there copious thermionic emission but the electrodes evaporate. Temperature ionization occurs (i.e., the heat energy is sufficient to ionize the metal atoms) and the true arc discharge comes into being. This discharge can carry a current of from 1 to 1,000 amperes. It is represented by EF on the curve of Fig. 13.3.

Hot-Cathode Mercury-Vapour Diode

Having discussed the elementary principles governing the operation of gas-filled valves, applications of these principles can now be examined, and the first valve to be considered is one in which there are only two electrodes—a hot cathode and an anode—and in which the filling gas is mercury vapour. Such diodes are much used as rectifiers, having the important advantage that the voltage across them when they are conducting is small—about fifteen volts.

Imagine such a diode containing mercury vapour at a pressure of the order of 1 newton/sq. metre. The cathode, which is usually a directly-heated tape of tungsten coated with the usual oxides, emits electrons copiously. If the potential across such a diode is gradually increased from zero, the current at first obeys Child's Law but, when the voltage becomes slightly greater than the ionization potential of mercury, ionization by collision begins. The positive ions which are produced, having much greater masses than the electrons, travel much more slowly towards the cathode where they encounter and partially neutralize the electronic space-charge. This neutralization of the space-charge in the vicinity of the cathode releases a much greater number of electrons, so that the current can no longer be regarded as space-charge limited. Consequently, a very small increase in the voltage across the valve after the ionization potential has been reached causes a large increase in the valve current. This current is still almost entirely due to the thermionic electrons: the direct increase due to ionization is small. It is the indirect effect of the slow-moving positive ions in partially destroying the space-charge near the cathode which liberates the emitted electrons and gives the low dynamic resistance to the gas diode. The gas pressure need not, therefore, be large, for a small number of positive ions produce the required action.

Figure 13.5 shows the valve with its circuit as used to rectify alternating current. The smoothing circuits have been omitted.

When a mercury-vapour diode is operated correctly, the voltage drop across it, when it is conducting, should be about 15 or 16 volts. If the valve is over-run (i.e., if the anode current is increased beyond that which the cathode is capable of supplying) the voltage drop across the valve, when it is conducting, increases. This increase in voltage is necessary to give the increase in the

ionization current to supply part of the current. The cathode which has already been stripped of its protective space-charge is then bombarded by positive ions, and this gradually destroys the oxide coating and so shortens the life of the valve. A mercury-vapour diode should therefore never be asked to produce a current greater than that suggested by the manufacturer, who has provided adequate cathode emission for the published performance, but not enough to allow the valve to be over-run.

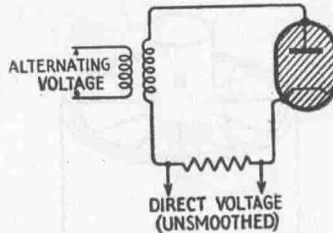


Fig. 13.5. Hot-cathode mercury-vapour diode connected as a rectifier.

When the diode is non-conducting it will act as a very perfect insulator to a very large inverse voltage. For example, a small mercury-vapour rectifier will usually withstand an inverse voltage of 6,000 at normal power frequencies. The inverse voltage at which flash-over occurs is determined by the nature and pressure of the filling gas and by the distance apart of the electrodes. In the non-conducting half-cycle the emitted electrons are, of course, turned back to the cathode which is temporarily the anode, and the valve behaves as if both electrodes were cold. As stated in the paragraph on the cold-cathode gas discharge, a very high electric field is required to produce a self-sustained glow. So long as the peak inverse voltage is less than the cold-cathode *striking potential*, the valve remains non-conducting.

A final point about mercury-vapour valves should be noted. The vapour in the bulb is in equilibrium with a small quantity of liquid mercury at the operating temperature. The vapour pressure of mercury varies from 10^{-4} newton/sq. metre at -40°C to about 2 newtons/sq. metre at 50°C . Mercury-vapour valves do not, therefore, start at their correct operating point if the ambient temperature is low: it is only when the effective temperature of the mercury is about 40°C that the vapour pressure is adequate. Fortunately, the hot cathode generally brings the bulb to the required temperature in a very short time.

Heat-Shielded Cathodes

In a conventional high-vacuum valve, the cathode emitting surface must face the other electrodes of the valve if the space-charge near it is not to be impossibly large, and so in conventional valves, if the cathode is the inner cylinder of a system of concentric cylindrical electrodes, it must be coated on its outside surface. But in a hot-cathode gas-filled valve it is not necessary to have the maximum electric field (in the absence of space charge) near the cathode, for the positive ions neutralize the space charge in any case. Therefore, it is worth while to make the emitting surface of the hot cathode in a gas-filled diode the inner surface of a cylinder and even to shield the hot electrode by another cylinder of high reflecting power.

Such structures are called heat-shielded cathodes and Fig. 13.6 shows a possible arrangement. H is the actual heater, but the inner cylinder placed round it is coated on its inner surface and on the attached fins. Surrounding this again is another cylinder of high reflecting power which tends to throw back the heat to the actual cathode. In this case the anode is placed at one end of the cylindrical cathode and takes the form of a cap, placed near and over it.



Fig. 13.6. Form of heat-shielded cathode.

Mercury-Arc Rectifier Valve

Another mercury-vapour rectifier which should be mentioned in passing, since it is sometimes used in radio work, is the mercury-arc rectifier. In its simplest form this consists of a large bulb containing a pool of mercury which is the cathode and is connected by means of an electrode to the outside of the valve. At some distance from the mercury pool are the anodes, one, two or three, for different purposes. But before the valve can operate the arc must be 'struck'. This requires an additional electrode in close proximity to a second small pool of mercury and, to ensure the maintenance of the arc as the load fluctuates, there must be two additional anodes near the mercury which are known as 'keep alive' electrodes. Fig. 13.7 is a diagrammatic representation.

It was mentioned in an earlier paragraph that the volt-ampere characteristic of the self-sustained discharge consists at its low-voltage end of an 'arc' regime in which the heating effect at the electrodes vapourises them. In the mercury-arc rectifier the evaporation occurs in the mercury pool which becomes an extremely hot and efficient cathode, emitting electrons copiously. The refractory anodes do not give rise to electrons to anything like the same extent. The mercury-arc rectifier is therefore, in principle, an extremely efficient hot-cathode device in which independent cathode heating is not necessary. The striking electrode is operated from an independent winding on the supply transformer, and produces the initial arc which sets the main arc into operation.

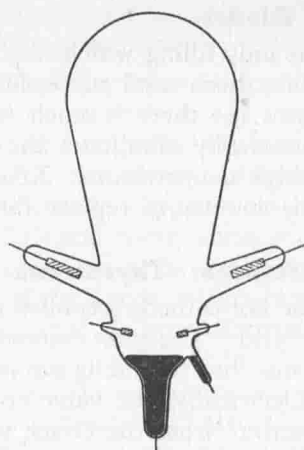


Fig. 13.7. Mercury-arc rectifier valve.

The 'keep-alive' electrodes are also connected through resistances to the supply, so that the current they take is small compared with that flowing in the main gap. Using such valves direct current of the order of 1,000 amperes can be obtained. By means of adequate cooling the inverse voltage can be maintained without inverse current.

Ignitron

A modification of the mercury-arc rectifier which is replacing it is a valve containing an igniter for the arc which carries a comparatively small power but which can be used to control the vastly greater power produced by the main arc. A diagram of such a valve (in principle) is shown in Fig. 13.8.

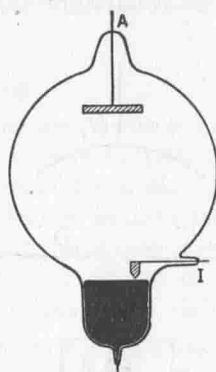


Fig. 13.8. General form of Ignitron rectifier valve.

The current flowing between the mercury pool and the igniter electrode, I, starts the arc which then passes to the main anode, A. The igniter can be used at any appropriate phase of the voltage by means of external circuits and the necessity for 'keep alive' electrodes is removed. This improves the inverse voltage which can be withstood by the valve.

Other Hot Cathode Gas Diodes

Mercury vapour is not the only filling which may be used in a hot-cathode diode. Other rare gases have been used successfully, particularly the gases argon and xenon. For service use there is much to be said for the use of a permanent gas, for it automatically eliminates the danger of failure due to extremely low or extremely high temperatures. Xenon-filled diodes have been successfully developed in this country to replace the mercury diode.

Hot-Cathode Gas-Filled Triodes: Thyratrons

The useful properties of the hot-cathode gas-filled diode immediately suggest the introduction of a control grid. The most common valve of this type is one which contains mercury vapour, but the filling gas may be helium or argon or a mixture of inert gases. Electrically, the valve acts as a switch and so has been given the name 'thyatron' from the Greek word meaning a door. It opens the way to the electric current.

The thyatron grid controls the flow of electrons from the cathode to the anode, but only in the sense of an ON - OFF switch. When the grid is sufficiently negative, no electrons can pass through it to the anode, and so the current is cut off. When the grid is made less negative, there comes a stage at which a certain number of electrons can pass through it. If the anode is then sufficiently positive with respect to the cathode, these electrons produce positive ions by collision with the gas or vapour molecules. As in the case of the gas diode, these positive ions partially destroy the space-charge near the cathode, and current flows through the valve freely. When the current has started, however, the grid can no longer control it, for making the grid more negative only increases the potential difference between it and the anode and draws to it a greater number of positive ions. These positive ions form a space-charge which masks the potential of the grid wires and allows electrons to pass freely through from the cathode. Since, therefore, the grid loses control of the current which it has initiated, to render the thyatron non-conducting again

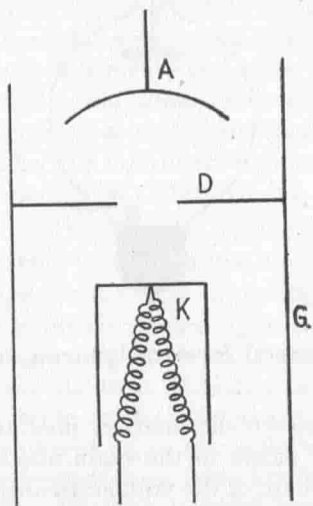


Fig. 13.9. Typical electrode arrangement of a thyatron.

it is necessary to cut off the positive potential applied to the anode. When the current ceases, the grid regains control.

A thyatron is not constructed in the same manner as the conventional high-vacuum valve. The usual structure is shown diagrammatically by the axial section drawn in Fig. 13.9, where the cathode is an indirectly-heated cylinder, K, coated with oxides. The grid, G, is a cylinder surrounding the cathode and containing an annular disc, D. Opposite the aperture in the disc the cup-shaped anode is placed. This odd construction is used for two reasons. First, as in the gas diode, it is unnecessary for the emitting surface of the cathode and the collecting surface of the anode to be placed opposite each other, and secondly, since an extremely small primary electron current is sufficient to start the valve conducting, the grid control must be made as strong as possible. In fact *no* electrons should be able to reach the anode when the grid is at a reasonably high negative voltage. The grid cylinder, G, therefore almost completely surrounds the cathode, leaving only a narrow aperture near the anode. The critical grid voltage, V_{gc} , at which the valve begins to conduct (or 'strikes', as it is called) varies with the anode voltage, V_a . Typical values of V_{gc} for various values of V_a are shown in Table 13.1.

TABLE 13.1

$V_a(\text{V})$	40	100	1,000
$V_{gc}(\text{V})$	0	-2.25	-6.5

However, with a valve containing mercury vapour, the critical grid voltage also varies with the temperature, for the mercury-vapour pressure determines the onset of ionization to a certain extent. For example, if $V_{gc} = -6.5$ at $V_a = 1,000$ and at an effective temperature of 40°C , V_{gc} might be -10 at a temperature of 70°C and -3 at 35°C . This variation of the striking voltage with temperature can be a grave disadvantage where large changes in temperature are experienced.

Thyatron with Shield Grid

In large thyatrons, where the anode current is many amperes, it is extremely difficult to prevent a few electrons from getting through the control grid even when the latter is very negative, and it is also impossible to prevent the control grid from carrying a considerable positive-ion current after the valve has 'struck'. Both difficulties have been overcome by replacing the control grid of Fig. 13.9 by a shield normally connected to the cathode and by introducing a small control-grid cylinder at exactly the right place in the valve.

The diagram of such a thyatron with a shield grid is shown in Fig. 13.10. The cathode cylinder, K, is surrounded by the shield cylinder, G_2 , which contains two annular discs between which the control-grid cylinder, G_1 , is situated. The anode, as before, is opposite the two apertures. This arrangement produces a much more efficient repulsive field at G_1 , so that the number of electrons arriving at the anode for a given negative grid voltage, V_{g1} , is much reduced. Also, when the valve conducts, since G_1 is now small compared with G_2 , and G_2 is at cathode potential, most of the positive ions find their way to the shield. The control-grid current is correspondingly reduced.

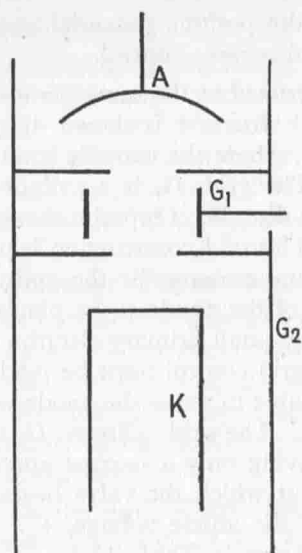


Fig. 13.10. Form of shield-grid thyatron.

Ionization and De-ionization Times

The positive ions produced in a gas- or vapour-filled valve are (as has already been stated) extremely sluggish compared with the electrons. For this reason there is a measurable time lag between the application of a positive voltage to the control grid of a mercury-vapour thyatron and the flow of current to the anode. This time, called the *ionization time*, is of the order of 10 microseconds, and is not usually serious. But when the anode voltage is cut off to bring the valve back to the non-conducting state, the positive ions present in the valve are not only sluggish, but they are in a force-free space, so that their movements are not only slow but random. Under these circumstances the ions persist for a considerable time, a time depending upon the mass of the ions—the heavier the more sluggish—and more indirectly upon the chemical nature of the gas. In a mercury thyatron the *de-ionization time*, as it is called, or the time between cutting off the anode voltage and the effective disappearance of the ions, may be about 1,000 microseconds. This is a severe limitation upon the usefulness of the valve as a switch. It means in fact that the valve cannot be used to produce current pulses at a frequency greater than 1,000 per second, owing to the persistence of the ionization.

Inert Gas Thyatrons

The mass of the argon atom is only one-fifth of the mass of the mercury atom, and therefore argon positive ions might be expected to be $\sqrt{5}$ times more active than mercury ions. Thyatrons containing argon do, indeed, show much decreased ionization and de-ionization times, and for that reason are used wherever rapid operation is important. Mixtures of rare gases can also be used, and modern gas-filled thyatrons generally possess de-ionization times in the neighbourhood of 10 microseconds. There is, however, a corresponding disadvantage which may or may not be serious. The anode-cathode drop in a mercury-vapour thyatron when it is conducting is about 16 volts. For

an argon thyatron with an ionization potential of 15.7 as compared with 10.4 for mercury, the anode-cathode drop is not less than 24 volts. In valves containing helium (ionization potential 24.5 volts) the drop during conduction may be as great as 33 volts. For many applications, however, the valve drop is comparatively unimportant. For example, gas-filled thyatrons can be used in time-base circuits at frequencies up to 50 kc/s.

Hydrogen Thyatrons

The use of hydrogen as a filling gas is becoming very common. The small inertia of the gas molecule is desirable, and the ionization potential is comparable with that of argon. The hydrogen tends to disappear during operation, and so some valves are fitted with automatic 'replenishers' to maintain a constant gas pressure during life. Hydrogen thyatrons may replace all others for small and medium power use when the problem of gas 'clean up' has been satisfactorily solved.

Thyatron Operation

As in the case of hot-cathode diodes, thyatrons must not be operated so that the valve drop is greatly in excess of the marker's figures, and in particular should never be allowed to rise to $4/3$ times the published value. The reason for this is that the latter is about one and a half times the ionization potential. Therefore if the drop is four-thirds of the published value, it is twice the ionization potential. This corresponds to the onset of double ionization in heavy atoms, and these doubly-ionized atoms are very destructive when they strike the cathode. So, to preserve the life of a thyatron, the rated current and voltage should not be exceeded. It should be noted in passing that any thyatron can be used as a gas-filled rectifier valve by connecting the grid to the cathode.

Gas-Filled Voltage-Stabilizing Valves

Referring back to the theory of the cold-cathode gas-discharge given in earlier paragraphs, it will be remembered that once a discharge has been struck there is a régime known as the 'normal glow' in which the voltage between the electrodes hardly varies at all as the current is increased by a factor of four or five to one. This régime is represented by the part BC of the volt-ampere characteristic in Figure 13.3.

If a diode valve is made with a very large cathode area, the normal glow persists as the current is increased, until all the cathode area is covered by the negative glow. If the cathode is in the form of a large cylinder, surrounding a smaller cylindrical anode, coaxial with it, all parts of the cathode are equidistant from the anode, and the voltage remains very constant indeed, as the current is varied. A valve of this type,—a gas-filled cold-cathode diode—operates as a voltage stabilizer when the current through it is of the correct order of magnitude.

Let the valve be represented by V in the network shown in Fig. 13.11. Then a change in the voltage across AB produces a change in current in R and V but no change in voltage across CD nor in the current in R_L , the load resistance. Similarly, a change in the load resistance, R_L , induces a change in current in R and V but no change in the voltage across CD. Therefore, to a first approximation, the voltage across CD is stabilized by the valve, V .

Such voltage-stabilizing valves are often filled with neon. The gas pressure is fairly high compared with the thyatron, and the (constant) voltage produced is about 100 volts. Valves are made in different sizes, covering different current ranges from 1 mA to 600 mA. They may be used in parallel with certain precautions. More recently valves have been produced giving extremely constant voltage output. These are used as voltage-reference valves, and replace standard cells in some electronic equipment. Their temperature coefficient is very small and aging only changes the voltage by about 1%. In most cases the current through the valve is small and must be kept fairly constant.

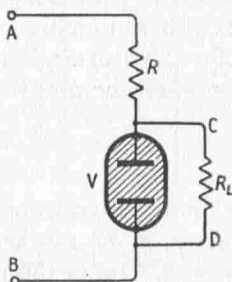


Fig. 13.11. Circuit of voltage-stabilizing valve.

Cold-Cathode Gas-Filled Triodes

Various devices have been developed in which a third electrode is introduced into a cold-cathode diode in such a way that the action of the control grid of a high-vacuum triode is more or less simulated. Clearly it is possible to make a cold-cathode thyatron by operating the valve at an anode-cathode voltage just less than the striking potential, but such a valve has the same disadvantage as the thyatron: the anode voltage must be drastically reduced to return the valve to the non-conducting state.

A more advanced design makes use of the 'corona' characteristic (see Fig. 13.3, section AB), where the effect of space-charge is much reduced and the grid has a better control when the valve is conducting. Where valves have to operate as switches only, it is possible to arrange such cold-cathode triodes in appropriate circuits to replace conventional high-vacuum thermionic valves. But the field of application is still very limited.

Gas Switches: TR and ATR Valves

In a radar system it is almost always necessary to disconnect the receiver when the pulse of energy is going out from the transmitter and in many cases it is also advisable to disconnect the transmitter when the receiver is in operation. To perform these functions special gas switches have been developed. The accepted names for these switches are by no means logical. The switch used to disconnect the receiver during transmission is known as a transmit-receive or TR valve, while that used to disconnect the transmitter during reception is known as the anti-transmit-receive or ATR valve. TR and ATR valves are very similar in construction and in certain cases the same valve can be used for either operation. The general principles of such gas-filled valves will now be considered.

TR Gas Switch

Suppose that a transmitter and receiver are connected to a common aerial as shown in Fig. 13.12. The connections may be by means of open line, coaxial

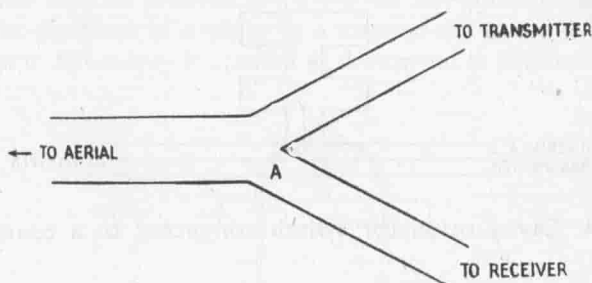


Fig. 13.12. Transmitter and receiver connected to a common aerial; the TR switch is connected at A.

cable or waveguide. The peak power of the transmitter is always large, and the voltage generated in the connecting line is also large. If a gas-filled diode valve is connected across the receiver line at A, the gas pressure may be arranged so that a discharge takes place when the transmitted pulse occurs. If the voltage is not high enough, the valve may be incorporated in a circuit resonant at the frequency of the transmitter, so that the voltage is stepped up to a suitably high value. The discharge, starting as a corona, very rapidly changes to a glow and then to an arc, practically short-circuiting the line to the receiver. This effectively disconnects the latter from the aerial and the transmitter. When the pulse is over, the discharge ceases and the line is ready to receive a signal from the aerial.

Fundamentally, a TR valve is thus a gas-filled cold-cathode diode which ignites when the transmitter is ON and extinguishes itself when the transmitter is OFF. As might be expected, however, there are a good many other conditions which the valve must fulfil if it is to operate satisfactorily.

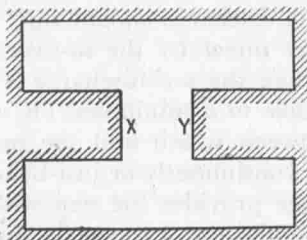


Fig. 13.13. Cavity resonator form of TR switch.

As used at centimetric wavelengths, the TR valve usually consists of a resonator similar to those discussed in Chapter 11. Indeed, the first British TR valve consisted of the same resonator as was used in the first British reflection klystron. Fig. 13.13 shows an axial section of such a resonator; the gap, XY, between the 'posts', which is used in a klystron to accelerate or decelerate the electron beam, is the part of the valve at which the discharge occurs when the

resonator is part of a TR valve. This resonator is coupled in some way to the receiver line. For example, if the latter is a coaxial cable, the resonator may be coupled to it as shown in Fig. 13.14. When the transmitter pulse occurs, a

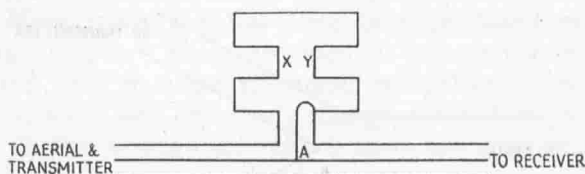


Fig. 13.14. Cavity-resonator switch connected to a coaxial line.

high voltage builds up across the gap XY which is arranged to be electrically half a wavelength from A. The build up of voltage is greatly magnified by the resonance of the cavity, and a discharge occurs. This is equivalent to a short-circuit at XY and at A, half a wavelength away. There are, however, various other factors involved.

(1) It is necessary that the discharge once started should stop as soon as possible after the cessation of the transmitter pulse, for, so long as the discharge continues, the receiver is unable to detect an incoming signal. Hence a de-ionization time of the order of 1 microsecond is desirable in a TR valve, and to obtain such a recovery time special precautions must be taken with the filling gas. Present-day practice is to introduce a little water-vapour, or some other gas having a high electron affinity, to cause the electrons to disappear rapidly. It is not necessary for the electrons to recombine to render the TR valve non-conducting. It is sufficient for them to be captured by neutral gas atoms or molecules to form negative ions. These are so sluggish in their movements that they do not impart any appreciable conductivity to the gas at very high frequencies. Commonly, a mixture of equal volumes of hydrogen or argon and water-vapour is used for TR valves, at a total pressure of about 3,000 newtons/sq. metre. This gives a recovery time of the order of 5 microseconds in typical cases.

(2) It is essential that the discharge should start promptly. Great damage can be done to the receiver mixer by the so-called 'spike' of power which by-passes the TR valve before the arc discharge is properly established. To reduce the power in the spike to a minimum TR valves are provided with a 'keep-alive' electrode, between which and the main resonator gap, XY, a d.c. discharge occurs either continuously or just before the advent of the transmission. This d.c. discharge provides the gap with an adequate number of electrons which hasten the build-up of the main discharge. Fig. 13.15 shows an axial section of a typical centimetric TR cell. The electrode K is the 'keep-alive' electrode and the position of the glass envelope is indicated by the shaded part of the drawing.

(3) It is essential that the transmitter power should not be dissipated in the TR valve. When the radio-frequency discharge occurs, it very rapidly becomes an arc, and the voltage across the arc is 1 to 3 volts, depending upon the gas filling. This almost complete short-circuit is such a bad match to the transmitter that little power is available in the receiver arm.

(4) Most important of all, the power transmitted through the TR valve must be *extremely* small if the receiver crystal valve is not to be damaged. It is the chief purpose of the TR valve to avoid such damage. The energy necessary to injure a crystal may be as small as 0.1 erg, or 10^{-8} joule, when this is applied to the crystal in the form of a pulse in a time of the order of 10^{-8} sec. In a radar transmission the energy is applied to the crystal in this form, for it is in the

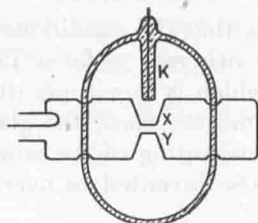


Fig. 13.15. Centimetre-wave TR cell with 'keep-alive' electrode.

initial stages of the pulse, when the TR valve is not properly ignited, that the energy gets through to the crystal. An energy of 10^{-8} joule excited in 10^{-8} sec. corresponds to a power of 1 watt. If the radar transmitter is using a magnetron emitting a peak power of 1 megawatt, the fraction getting to the crystal should always be less than 10^{-6} or, putting it in the more usual form, the attenuation introduced by the TR valve should be 10^6 or 60 db. It is a rather difficult matter to ensure this attenuation during the early stages of the discharge.

(5) Lastly, the TR valve itself should produce a negligibly small attenuation of the signal when the discharge is not present. The cavity with its glass walls should be as free from loss as possible. Great care is taken of this point in the development of these devices.

ATR Gas Switch

Usually, a TR valve can be used as an ATR valve with suitable modifications to its associated circuits. Consider once more Fig. 13.12. It is required that there should be no connection to the transmitter when the receiver is in operation, but that the transmitter to aerial circuit should be completed when the transmitter is ON. There are many ways of producing this effect. If, for example, a quarter wavelength of line, short-circuited at its termination is inserted as a series element into the transmitter line (see Fig. 13.16), the transmitter is effectively isolated from the aerial and receiver. If an ATR (or TR) valve is inserted at B, the firing of the valve short-circuits the series stub line and puts the transmitter through to the aerial. When the valve is non-conducting there is an infinite (or very high) impedance at B.

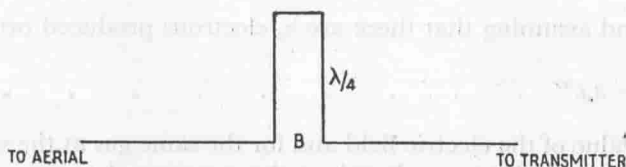


Fig. 13.16. An ATR valve is connected at B. When it strikes it removes the open circuit introduced by the quarter-wave line.

As in the case of the TR valve, the ATR valve should have a short recovery time: but it does not require to start so rapidly and so a 'keep-alive' discharge is unnecessary. The valve should not dissipate much power when ignited: but the small fraction of the power absorbed when it is extinguished is of little account, and so its 'cold' losses are unimportant. It does not much matter how much power the ATR valve passes when it is ignited, so long as there is a change of power level of the order of 20 db between transmit and receive.

It will be noticed, therefore, that the conditions for the operation of an ATR valve do not appear to be so stringent as for a TR valve. There is, however, one additional requirement which is sometimes difficult to meet at the highest frequencies. Where waveguides are used, the glass coupling windows present problems. Cracking and over-heating of the windows are common faults and new techniques have had to be invented to overcome them.

APPENDIX 13.1

Theory of Initiation of Self-Sustained Discharge

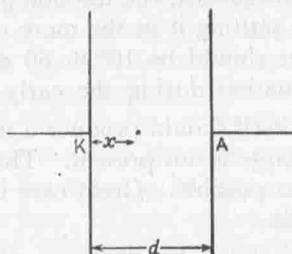


Fig. 13.17. Cathode and anode separated by the distance d .

Let the distance between the plane parallel electrodes of a gas-discharge valve be d , and the potential difference between them be V . Then we suppose the electric field in the gap when no current or a negligible current is flowing to be V/d . Corresponding to this value of the electric field strength, and for a given gas in the valve at a given pressure, let the number of pairs of ions produced per electron per metre be α . Then if n is the number of electrons per cubic metre at distance x from the cathode, the increase, δn , in this number between a plane at x and a plane at $x + \delta x$ is

$$\delta n = \alpha n \delta x$$

Integrating, and assuming that there are n_0 electrons produced originally at the cathode,

$$n = n_0 e^{\alpha x} \quad (13.1)$$

For the same value of the electric field and for the same gas at the same pressure, let β = the number of ions produced on the average when one positive ion is neutralized at the cathode. Then the total number of pairs of ions produced in the gas can be calculated in steps as follows:—

n_0 electrons at the cathode produce $n_0 \epsilon^{\alpha d}$ pairs of ions per sq. metre at the anode
 $n_0 \epsilon^{\alpha d}$ positive ions therefore arrive back at the cathode.

$n_0 \epsilon^{\alpha d}$ positive ions produce $\beta n_0 \epsilon^{\alpha d}$ pairs of ions on neutralization

$\beta n_0 \epsilon^{\alpha d}$ electrons (from these new ions) produce $\beta n_0 \epsilon^{2\alpha d}$ pairs per sq. metre at the anode.

$\beta n_0 \epsilon^{2\alpha d}$ positive ions (which arrive at the cathode) produce $\beta^2 n_0 \epsilon^{2\alpha d}$ pairs.

$\beta^2 n_0 \epsilon^{2\alpha d}$ electrons produce $\beta^3 n_0 \epsilon^{3\alpha d}$ pairs at the anode.

$\beta^3 n_0 \epsilon^{3\alpha d}$ positives produce $\beta^4 n_0 \epsilon^{3\alpha d}$ pairs, etc., etc., etc.

The total number of ion pairs produced is found by adding. It is N , where

$$N = n_0 \epsilon^{\alpha d} (1 + \beta \epsilon^{\alpha d} + \beta^2 \epsilon^{2\alpha d} + \beta^3 \epsilon^{3\alpha d} + \dots) \\
= \frac{n_0 \epsilon^{\alpha d}}{1 - \beta \epsilon^{\alpha d}} \quad (13.2)$$

This number is finite and proportional to n_0 so long as $\beta \epsilon^{\alpha d}$ is less than 1. As $\beta \rightarrow 1/\epsilon^{\alpha d}$, the number tends to infinity, representing the initiation of the self-sustained discharge. This régime in the discharge, where the effects of space-charge can be neglected, and the current is built up, is known after its chief investigator as the Townsend régime. Townsend, however, believed that positive ions produced others by collision, a theory which is not held to-day. The mathematical analysis is similar on the two theories, but the significance of β is different.

CHAPTER 14

Semi-Conductors

Elementary Discussion

THE THEORY OF semi-conductors is much more difficult than that of the ordinary valve if it is to explain precisely how they operate and why they possess certain characteristics. However, the devices employing them are no harder to understand because their characteristics can be represented by means of families of curves, which, in many cases, bear a close resemblance to those of thermionic valves. Such curves are used in exactly the same way as ordinary valve curves and from the point of view of the designer, let alone the user, of radio equipment it is rarely necessary to know much about what goes on inside the semi-conductor.

In this chapter an attempt is made to explain the theory of semi-conductors, but it is recognised that it is difficult and that many readers must content themselves with the summary of the characteristics of semi-conducting devices which now follows and which precedes the theory.

A major practical use of semi-conductors is for rectification as an alternative to the thermionic diode. The main types are the selenium and the copper-oxide rectifiers. At power frequencies the selenium type is now usual because it can operate at a higher temperature than the copper-oxide. The latter has more stable characteristics, however, and is used where this is important as, for example, in alternating-current measuring instruments.

Both types have characteristics of the same general form as the thermionic rectifier and are more or less direct alternatives to it. The differences are mainly that the current rating is limited by the temperature rise, so that they are especially useful under conditions of high peak current; that the back resistance is not infinite; and that there is appreciable capacitance across the rectifier.

The finite back resistance means that the metal rectifier passes some reverse current. The resistance is not constant but decreases with increasing voltage and it sets a limit to the voltage which may safely be applied to a rectifier. A basic selenium rectifier unit can usually be operated up to 25 volts peak-inverse at which the peak density is of the order of 1 A/sq. in. A normal r.m.s. forward current density is 0.3 A/sq. in.

Higher voltage requirements are met by stacking basic units in series, while the current needs are catered for by choosing the area of the units appropriately. Cooling fins are often fitted to the layer units.

Practical rectifiers range from assemblies of a few very large discs for low-voltage high-current operation to assemblies of very many very small discs for high-voltage low-current operation. As examples, a rectifier for battery charging may have 16 discs measuring 5 to 6 inches over the cooling fins and

be rated for an output of 5 amperes at 6 volts. At the other extreme, there are assemblies no more than 7/16 inch in diameter and under 1.2 inches long which will withstand 19 kV peak inverse and provide a rectified output of 8 kV at 0.5 mA.

As the operating frequency is increased the capacitance of the rectifier becomes more important. In many cases, however, the current-handling capacity that is required also becomes small. The contact area of the rectifier can then be reduced and also the capacitance. Suitably designed types are widely used for signal rectification and modulation in carrier telephony, as limiters in a.f. circuits and sometimes as demodulators up to frequencies around 1 Mc/s.

Usually, however, the germanium point contact rectifier is more suitable for frequencies over a few hundred kilocycles and can be employed up to about 100 Mc/s. Commercial types have a back resistance of 50,000 Ω or more, (1 M Ω or more in selected types) with a forward resistance of a few hundred ohms. They will withstand a peak inverse voltage of about 100 volts and pass a forward current of up to 50 mA. Their capacitance is a few picafarads only and they can replace the thermionic diode in most receiver signal applications where a high back resistance is not essential.

The silicon point contact rectifier is used up to the highest frequencies, where it is superior to the thermionic diode. It finds wide applications as a mixer in centimetre-wave equipment. It is a relatively fragile device, for it is easily destroyed by overloading, and the rectifier current must rarely be allowed to exceed 1 mA.

All these metal rectifiers—the selenium and copper-oxide contact rectifiers and the germanium and silicon crystals—find their main uses as alternatives to the thermionic diode. Usually, they perform very similarly and they always have the advantage of requiring no cathode-heating power. Their performance is not identical but in many applications the difference is not significant. In other cases where the differences are appreciable the metal rectifiers are not necessarily inferior; they are sometimes superior.

Semi-conductors are used also in protection devices, where large surge voltages are liable to occur. They can have a resistance which varies inversely as a high power of the applied voltage. Such a device will have a high resistance at low voltage but a low resistance at high voltage. If it is connected in shunt with a piece of equipment which normally operates at low voltage but which may be subject to high-voltage surges, it will not appreciably affect the normal operation but will tend to short-circuit the equipment, and so protect it, when a surge of voltage comes along. Such surge limiters are made of silicon-carbide (Metrosil) and are widely used in switchgear.

Although just beyond the development stage, the germanium triode is a particularly interesting device which in time may well replace the thermionic amplifying valve in some of its applications. In one form it is a germanium crystal with three contacts and exhibits characteristics similar to those of the triode valve. At the time of writing, its electrical characteristics are considerably inferior to those of the thermionic valve, but it needs no cathode-heating power and it is smaller than most valves. So far it is unsuitable for frequencies above 10 Mc/s and the problems of quantity production have not been solved. It is, therefore, definitely a device of the future.

The way in which all these semi-conducting devices operate is by no means easy to understand for it demands a knowledge of the basic relationships between matter and energy. An attempt is made in the following pages to reduce the theory of these devices to its simplest form. While this theory is essential to anyone concerned with the development of semi-conductor devices themselves, it is not so important to those who are only concerned with using them. Those who are content to treat them merely as devices having certain specified characteristics without enquiring into why they have these characteristics, need read none of the marked paragraphs in this chapter.

Introduction to Theory

- In Chapter 1 an elementary discussion of the cause of metallic conduction was given in which the valency electrons, i.e., the electrons in the outermost shell of the atom, were regarded as being relatively 'free' to move under the action of an electric field. It is now necessary to consider more closely the 'freedom' of the valency electrons and to apply the results obtained for metals to non-metals and particularly to crystals. It will be found that three headings are required to classify all substances—conductors, semi-conductors and non-conductors or insulators—and that the intermediate class of semi-conductors is rather complex. The growing importance of semi-conductors in electronics warrants a careful study of their properties, for semi-conducting crystals in association with conductors have now found applications as diodes and triodes, as photo-electric or secondary-emission devices, and as cathode-ray screens. Some of the applications are new and some are old; but the explanation of the effects encountered is in all cases of fairly recent origin. For example, copper-oxide rectifiers have been in use for many years but the exact physical explanation of their action is not yet complete.

Electric Conductivity

Referring back to the Periodic Table (p. 5), it will be recalled that the electrons in atoms are arranged in shells. Quantum theory shows that these shells are themselves arranged in sub-shells, and the exact shell and sub-shell structure of most atoms is now known both from theory and experiment. Atoms of the alkali group (lithium, sodium, potassium, rubidium, caesium and francium) contain electrons arranged in complete shells plus one additional electron. This valency electron exists in a sub-shell which, when it is complete contains two electrons, so that the alkaline earth atoms (beryllium, magnesium, calcium, strontium, barium and radium) all contain two valency electrons, forming a complete sub-shell.

It was stated in Chapter 1 that the high electrical conductivity of the alkali metals is due to the 'free' valency electrons and that the 'freedom' of the electrons is an indication that they are loosely bound to their parent ions. Further investigation has shown that the freedom of the valency electrons is due rather to the fact that the sub-shell containing them is incomplete, so that an electron can easily pass from one atom to another, temporarily filling up the sub-shell of the latter, and may as easily be passed on to a third. This picture of conduction (which does not differ markedly from that given earlier) suggests that the partially vacant sub-shell forms a kind of conduction channel along which electrons may be passed from one atom to another. The picture

is applicable to all solid and liquid substances, elements, compounds and mixtures, whether they are good or bad conductors. Quantum physics leads us to speak of the 'free' electrons being in a conduction band, meaning that there are a large number of possible states of the electron, differing very little in energy from each other, and all in a vacant or partially vacant sub-shell. A minute increase in energy is therefore sufficient to transfer an electron from one state in one atom to another state in the same or in another atom. This energy can easily be derived from an applied electric field and will be given back to the field when the electron moves on to a third atom in its original state. This passing of the electron from one atom to another in the electric field is electrical conduction.

In general terms a substance is a conductor if the valency electrons exist in a conduction band; and a conductor is then alternatively defined as a substance in which the number of 'free' electrons per unit volume is of the same order of magnitude as the number of atoms per unit volume. The closer the atoms are to each other, the more easily the electrons can pass from one atom to another, for the effect of packing the atoms is shown by wave-mechanics to be to increase the possible states of energy of the valency electrons.

On first consideration it might be thought that the alkaline earths would be non-conductors, since their two valency electrons completely fill the sub-shell. This is quite true of the alkaline earths when the atoms are moderately far away from each other; but when the atoms are crowded together, as in the crystalline solid, the sub-shells tend to over-lap, making it possible for an electron to be passed from one *sub-shell* to the *next* in the same or a neighbouring atom without first acquiring any appreciable additional energy. To give a specific example, the four electrons of beryllium in the stable state are arranged in two complete structures—two in the so-called K-shell and two in the first L sub-shell. To move an electron from the first L sub-shell to the next L sub-shell requires several electron-volts, and the atom would then be termed an 'excited atom'. But if a large number of beryllium atoms are brought together in a crystal, each sub-shell divides up into a wide band of possible energy levels and these bands over-lap.

The action may be illustrated most easily by a diagram. In Fig. 14.1, the filled shell of the valency electrons in the beryllium atom is shown on the right-hand side of the figure by an energy level at s . The next higher level to which an electron can move is at p , also on the right-hand side. But, if the distance between the atoms of beryllium is decreased, the single level at s breaks up into

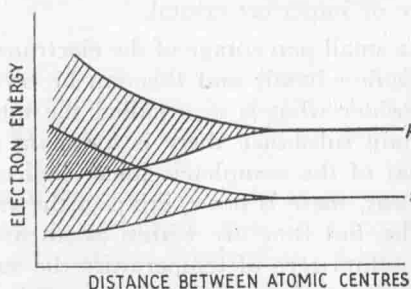


Fig. 14.1. Energy levels in the beryllium atom.

a band of levels as shown in the middle of the figure and the same breaking up effect occurs with the p levels. When the distance between the atoms is as small as it is in a crystal of beryllium, the s and p bands overlap as shown on the left-hand side. In the figure the abscissa is the average distance between atomic centres and the ordinate is electron energy, and it should be noted that, since we cannot distinguish between atoms, the electron energy is no longer referred to one particular atom but to the crystal lattice as a whole.

When the sub-shell bands overlap, it is possible for an electron to pass from a complete sub-shell (a 'filled band') to an incomplete sub-shell (a 'vacant band') by absorbing a small amount of energy from an electric field, and so the alkaline earths in the crystalline form are conductors. This criterion for conductivity can be applied to all substances and is found to be in accord with experiment.

It should be noted in passing that according to our theory of electrical conduction, metallic conductivity should increase with decrease of temperature, for any additional thermal energy given to the electrons can only increase the work which must be done (on the average) to pass an electron from one atom to another. This is in agreement with experimental fact. Metallic conductors in the pure state have a positive coefficient of increase of resistance with temperature.

Semi-Conductors

A semi-conductor is a substance in which the number of 'free' electrons is small compared with the number of atoms per unit volume. Instead of there being about 10^{28} free electrons/cu. metre, as there are in copper, there are only about 10^{21} per cu. metre in a typical semi-conductor, such as silicon or germanium. In a semi-conductor, the majority of the electrons are unable to move from their stable states in a filled sub-shell because the energy required to move them to a vacant sub-shell, forming a conduction band, is too large. In a non-conductor or insulator this is the case with all the electrons.

A good example of a typical non-conductor is a pure crystal of sodium chloride. Here we have an ionic crystal in which the sodium atom has already been bereft of its valency electron which has formed a complete shell with the chlorine electrons. Therefore in each case the electrons are in closed shells and moreover, in spite of the close packing of the atoms in the crystal, the next higher energy band due to the next higher sub-shell does not overlap the filled shells. Consequently, the energy necessary to remove an electron to a vacant level is high and conductivity is zero. It should be noted that this does not apply to an impure or imperfect crystal.

In a semi-conductor a small percentage of the electrons can rise into a vacant band (forming a conduction band) and this can be brought about in several ways. First, let us consider what is now called the intrinsic semi-conductor. Suppose that in a certain substance there is a vacant sub-shell at a slightly higher energy than that of the completely-filled shell containing the valency electrons. In other words, there is not overlap of the vacant and filled bands as in the alkaline earths, but they are within about an electron-volt of each other. Then at the absolute zero of temperature the substance has zero conductivity, since the electrons only possess their atomic energy which is exactly sufficient to maintain them in their stable band. But if the temperature is

raised, there is then a finite chance that an electron will cross the barrier to the vacant band, and the chance will increase as the temperature increases. Therefore the substance will be an extremely poor conductor with a negative coefficient of increase of resistance with temperature.

The mechanism by which the conductivity is achieved is identical with that already described in the case of thermionic emission, and such a substance is known as an intrinsic semi-conductor. Silicon, germanium, tellurium, lead sulphide and cuprous oxide all fall into this category. The difference in energy between the stable filled band and the vacant band varies from 0.38 electron-volt for tellurium to 1.4 electron-volts for cuprous oxide.

Summing up, the explanation of intrinsic semi-conductivity is very similar to the explanation of thermionic emission, and the effect might with justification be called thermionic semi-conductivity. In both cases the supply of electrons is due to their escape from one state to another of higher potential energy by reason of their excess kinetic energy. It should, however, be noted that when electrons have escaped from the filled to the vacant band, both bands then behave as a single conduction band, since electrons can be passed from one atom to another in both of them. The movement of *one* electron to a higher vacant band produces *two* free 'electrons'. 'Electrons' is here placed in inverted commas, because it is really the absence of an electron from the filled band which gives half the conductivity.

All semi-conductors are not of the *intrinsic* type; in fact at one time it was thought that a pure crystal could not possess any conductivity other than metallic. The second type is known as an impurity semi-conductor for the simple reason that the conductivity is due to the presence of an impurity. Imagine that a crystal contains a small percentage of the right kind of impurity; that is, of an impurity in which the stable filled band is about 1 electron-volt below a vacant crystal band. Fig. 14.2(a) shows the relative energy levels in such an impurity semi-conductor.

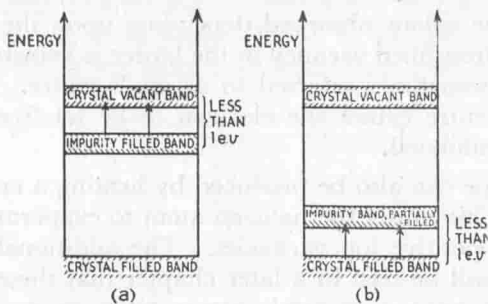


Fig. 14.2. Energy levels in an impurity-type semi-conductor; (a) is for a stable filled band 1 eV below a vacant crystal band, and (b) is for an impurity which itself has a partially-filled band.

The thermal energies of the electrons of the impurity allow a certain fraction to reach the vacant band, so that conduction throughout the crystal is possible, although the number of electrons involved is, of course, dependent on the percentage of impurity present. Substances which exhibit this effect are known as N-type impurity semi-conductors. Alternatively, if, as in Fig. 14.2(b), the

impurity itself has an only partially-filled band which is not much above the stable band of the crystal, electrons under the influence of heat can rise from the stable crystal band to the impurity band. This leaves 'electron holes' in the stable crystal band, positive charges due to the positive ions. Electrons can therefore be passed on from one of these holes to another, giving a certain degree of conductivity. Substances exhibiting this effect are known as P-type semi-conductors.

Crystal imperfections may also produce conductivity. Few natural crystals are perfect: they may possess additional atoms in the crystal lattice or may have vacancies in the lattice. For example, in a natural crystal of NaCl there are almost equal numbers of Na^+ and Cl^- ions missing. Crystal imperfections sometimes lead to the appearance of new energy bands which may act as conduction bands. Semi-conductors produced by such imperfections may be N or P-type.

Finally, it should be noted that a crystal may exhibit both intrinsic and impurity semi-conductivity simultaneously. Indeed most crystals if taken to a high enough temperature while still remaining solid must exhibit intrinsic conductivity.

Colour Centres

A paragraph may be inserted at this point describing a crystal phenomenon related to semi-conductivity in its mechanism, but very different in its application. This is the formation of what are known as colour centres in some crystals, notably the alkali halides, such as KCl.

It has already been mentioned that a natural crystal of an alkali halide contains an almost equal number of negative and positive ion vacancies. If such a crystal is bombarded by electrons some are captured by the crystal and take the places of Cl^- ions. This restores locally the electric stability of the lattice, but, of course, alters many of its other properties. The most interesting change is that the crystal at each of these electron-filled vacancies absorbs visible radiation, the colour absorbed depending upon the crystal used. For that reason the electron-filled vacancy in the lattice is known as a colour centre, although it is also sometimes referred to as an F-centre. The absorption of light by a colour centre causes the electron to be set free again. Thus the colour centre is annihilated.

Centres of this type can also be produced by heating a crystal in the vapour of its own alkali. This causes the halogen atom to evaporate, while alkali ions fill up some of the positive ion vacancies. The additional electrons produce colour centres. It will be seen in a later chapter that these colour centres are used to produce an opaque or semi-opaque trace on a screen of potassium chloride crystals when the latter is bombarded by electrons. The device is known as the *skiatron*.

Devices Employing Semi-Conductors

The explanation of the action of devices containing semi-conductors is not easy, for the physical mechanism involves not only properties of the conduction bands of semi-conductors and insulators, but also the properties of free electrons in moving from one energy level to another. A knowledge of thermionic emission and contact difference of potential is therefore required. Before

proceeding to the explanation of the action of specific electronic devices, a brief recapitulation of the relevant physical facts described in this and earlier chapters will therefore be given. These may be catalogued as follows:—

1. A conduction band is a possible set of electronic levels which are not completely filled with electrons. In a conducting metal *all* the valency electrons are in the conduction band. Conductivity at absolute zero is theoretically infinite. (Chapter 14).
2. Semi-conductors possess *some* electrons in the conduction band, but the number of such free electrons per unit volume is only a small fraction of the number of atoms per unit volume (Chapter 14).
3. Insulators have no electrons in the conduction band, but electrons may be injected into the conduction band of an insulator from outside. (Chapter 14).
4. Electrons in a conduction band—free electrons—have thermal energies which allow them to leave their parent substance. They may be thermionically or photo-electrically emitted, or may be given the additional energy by other impacting electrons thus producing secondary emission (Chapter 1).
5. The energy which must be given to an electron in the conduction band to free it from the conductor or semi-conductor may be measured in electron-volts. This is the work function of the substance, usually denoted by ϕ (Chapter 1).
6. To a rough approximation the contact difference of potential between two metals, two semi-conductors or a metal and a semi-conductor is the difference between their work functions. (Chapter 1).

Copper-Oxide and Selenium Rectifiers

The so-called copper-oxide rectifier, and its more modern counterpart, the selenium rectifier, extensively used in radio work in place of valve rectifiers, have been shown to consist of several layers as exhibited in Fig. 14.3. The methods by which the layers have been identified need not concern us.

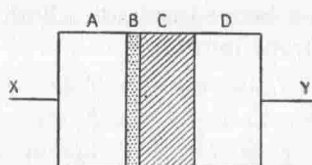


Fig. 14.3. The layers of a copper-oxide rectifier: A, copper electrode; B, layer of cupric oxide; C, layer of cuprous oxide; D, copper or lead electrode.

In the case of copper-oxide rectifier, A is a copper electrode in intimate contact with an extremely thin layer B of cupric oxide which merges into the main layer C of cuprous oxide. The latter is in intimate contact with the other copper or lead electrode, D. The cupric oxide layer is about 10^{-4} cm thick and is an insulator. Cuprous oxide is a P-type semi-conductor due to the existence of many Cu^{++} ions in the crystal lattice. These are positive holes which can pass electrons on. The junction between the cuprous oxide and

the second copper electrode is as perfect as it can be made. Such an arrangement acts as a rectifier, passing current more easily when X is negative than when it is positive. To understand the action a potential energy diagram must be drawn, showing the levels of the conduction bands in the various layers.

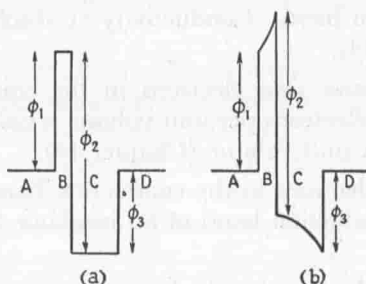


Fig. 14.4. Potential energy diagrams of copper-oxide rectifier; (a) initial conditions, (b) stable conditions.

First imagine that X and Y are connected by a copper wire, so that they are at the same potential. Then when the contact is just made the energy levels are as in Fig. 14.4(a). Taking the copper electrode A as being at zero potential, the conduction band, B, (empty) of the insulator is at a much higher potential. In fact B is only an insulator because the electronic energy corresponding to its conduction band is several volts higher than that corresponding to its filled band. The conduction band of the semi-conductor, C, (its valency band) is at a lower potential energy. D is at the same potential as A. However, some electrons can pass from A to B (due to their thermal energies), a very few from C to B, while a great many can flow from D to C directly, causing C to become negative. But B is an insulator and C a semi-conductor, so most of the potential change due to space-charge occurs in B. Hence the equilibrium state of affairs is represented by Fig. 14.4(b), where ϕ_1 , ϕ_2 , ϕ_3 remain as before, but there is a potential difference across the insulator B and a smaller potential difference across the semi-conductor C. The potential difference across B is greater simply because it is a better insulator. Both potential differences tend to prevent the flow of electrons into C.

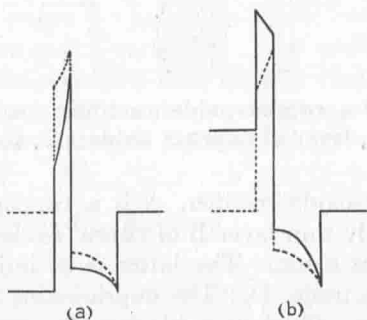


Fig. 14.5. Diagrams similar to those of Fig. 14.4; (a) when X of Fig. 14.3 is positive to Y, and (b) when X is negative to Y.

Now imagine that X is made positive with respect to Y (Fig. 14.3). The potential energy diagram is then as shown in Fig. 14.5(a), where the dotted line is a reproduction of Fig. 14.4(b) and the full-line corresponds to X positive. The flow of electrons from right to left, required by the external potential difference between X and Y is impeded by the magnitude of ϕ_2 compared with ϕ_1 and ϕ_3 . It must be borne in mind that $\phi_2 = \phi_1 + \phi_3$, but that the number of electrons able to jump a barrier of height ϕ is proportional to $\exp(-\phi/kT)$ (Chapter 1). This number decreases much more rapidly than ϕ increases, so the number of electrons capable of jumping barrier ϕ_1 is much greater than the number capable of jumping ϕ_2 .

Now imagine that X is made negative with respect to Y. The potential energy diagram is as shown in Fig. 14.5(b). The flow of electrons from left to right is impeded by the magnitude of ϕ_1 and, as we have seen, this is a smaller obstacle than in the previous case. Moreover, by making X sufficiently negative the potential gradient in the insulator has been made favourable to a transfer from left to right.

From these considerations it is clear that the arrangement called the copper-oxide rectifier owes its rectifying action (a) to the various contact differences of potential between the layers and (b) to the existence of the insulating layer of cupric oxide. The rectifying action of the junction between the copper and the cuprous oxide opposes the action of the insulating layer, but it is of minor importance. It can be reduced still further by the proper choice of a metal for the anode to replace copper.

The rectifier is known as a barrier-layer rectifier and we shall meet several other examples of this type in succeeding paragraphs. There are two barrier layers in the copper-oxide device. First, there is the insulator itself—a thick barrier layer which electrons can only surmount by becoming free electrons in the conduction band of the insulator and by diffusing through the layer under the action of the external field. Secondly, there is the barrier layer between the metal and the semi-conductor. This has an effective thickness of the order of 10^{-7} cm and electrons can be propelled across the barrier by auto-electronic emission alone. The two barriers are very different in their properties and although the metal to semi-conductor junction can be neglected in the case of the copper-oxide rectifier, it is of greater importance in other applications.

The potential difference which may be impressed across a single copper-oxide rectifier is small, for the resistance of the oxide layer causes a rise in temperature when current flows. This rise in temperature tends to destroy the rectifying action at the cupric-cuprous oxide junction. Selenium rectifiers are much superior in this respect. They are able to work at higher temperatures and they can withstand a higher peak inverse voltage. Fins are usually attached to rectifiers handling power to keep the temperature low. To increase the voltage which may be applied to a rectifier, single units are placed in series; to increase the current-carrying capacity the area of the plates in contact is increased. However, as Fig. 14.3 suggests, the insulating layer between the conductor and the semi-conductor acts as a capacitance, and increasing the plate area to increase the current carrying capacity also increases the capacitance. Copper-oxide rectifiers are not, therefore, suitable for power rectification at high frequencies: they may be used as demodulators, however, where the

current is small, by reducing the plate area until the capacitance becomes tolerable.

Barrier layer rectifiers do not necessarily involve a P-type semi-conductor. If an N-type semi-conductor is used, the potential energy diagram corresponding to Fig. 14.4(a) becomes that of Fig. 14.6, where $\phi_1 = \phi_2 + \phi_3$.

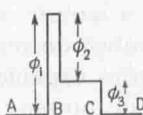


Fig. 14.6. Potential energy diagram of N-type semi-conductor.

In this case, as the reader may easily verify, the current passes more readily when A is positive with respect to D. The polarity is reversed in comparison with the P-type semi-conductor. The explanation of the action is as before, and depends upon the fact that ϕ_1 is greater than ϕ_2 or ϕ_3 , and therefore electrons flow more readily from right to left than from left to right.

Silicon and Germanium Crystal Valves

One of the most important fields in modern electronic development is concerned with 'crystal valves'. The number of types of crystal valve is continually increasing, but we shall confine our attention to the following:—

- Silicon metal-rectifier valves, used in mixer units of superheterodyne receivers or for small-signal demodulation.
- Germanium metal-rectifier valves, used in place of vacuum diodes for dealing with larger signals. For example, such valves may be used in demodulator units of superheterodyne receivers.
- Germanium metal triodes, used in amplifiers and oscillators where the noise factor of the unit is unimportant. These valves can only handle small powers, being limited as oscillators, for example, to the generation of powers of the order of 10 milliwatts. Such crystal valves are known as point-contact transistors.
- Germanium metal triodes which may be used in the future as small signal amplifiers or as medium power oscillators. This is a more modern type of crystal valve, known at present as the junction transistor.

In more general terms the applications of semi-conductors are only beginning to be realised. Their original use as non-linear circuit elements covers (a) and (b) above. In association with two small metal contacts placed very close together and a third larger contact some distance away, they give rise to (c) above. Finally, and now opening the way to many new applications, a single crystal of germanium may be treated so that different regions in it exhibit N-type and P-type conductivity. Such devices simulate the action of an ideal triode valve very closely. This covers (d) above.

Silicon Crystal Mixer Valve

Silicon containing a small percentage of aluminium, boron or beryllium as an impurity is a P-type semi-conductor. Such crystals are used in the mixer valves which are about to be described. The percentage of impurity determines

among other things the conductivity of the finished valve: a typical impurity percentage is 0.25% by weight of aluminium and 0.2% by weight of beryllium.

Crystal valves have been made in various forms, but by far the most common construction is shown in the diagrams of Fig. 14.7.

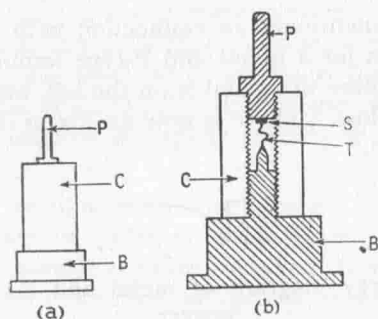


Fig. 14.7. External form (a) and internal construction (b) of silicon crystal valve.

The external appearance of a crystal mixer valve is shown in Fig. 14.7(a). P is a metal pin, one of the two electrodes. C is a cylindrical ceramic insulator, and B is the metal base, forming the second electrode. The total length is about three-quarters of an inch. The internal arrangement is shown in Fig. 14.7(b). The top pin is soldered or otherwise connected to the silicon crystal, S, which is in electrical contact with a 'whisker' of tungsten wire, T. The latter is soldered to the metal base. After the valve has been assembled and adjusted, wax is forced into the cavity in the ceramic cylinder; this holds everything in place and provides protection against mechanical shock and damp. The crystal valve so formed behaves as a diode rectifier, passing current more easily when the pin is positive than when it is negative.

A typical current-voltage characteristic for a crystal mixer valve is shown in Fig. 14.8. From this it will be seen that there is a very rapid variation in the slope of the characteristic in the vicinity of zero voltage, but that there is an appreciable current for any finite applied voltage whether it be positive or negative.

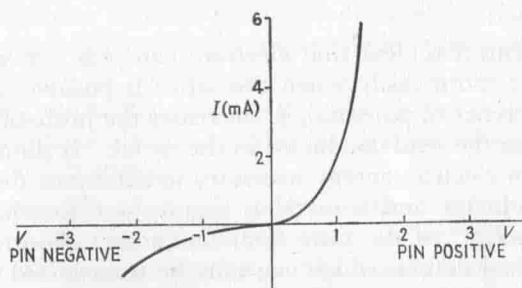


Fig. 14.8. Characteristics of silicon crystal valve.

► A complete theory of the silicon-tungsten crystal unit is not yet possible, but the rectifying action can be explained in sufficient detail for our purpose in terms of what has already been said about semi-conductors and contact

difference of potential. The rectifying action of the silicon-tungsten crystal valve is due in general terms to the asymmetry of the arrangement. One metal semi-conductor contact is over a very small area, the point of the tungsten whisker. The other metal semi-conductor contact is over a comparatively large area. It is this asymmetry of contact which produces the rectification.

As has already been mentioned in connection with the metal rectifier, the potential energy diagram for a metal and P-type semi-conductor in contact is as shown in Fig. 14.9, where the metal is on the left, and equilibrium has been established by electron flow. There is now an excess of electrons in the semi-



Fig. 14.9. Potential-energy diagram of metal and P-type semi-conductor in contact.

conductor and, since these can only exist in positive holes which are relatively infrequent, the space-charge is distributed over a considerable volume. In the metal, on the other hand, where there are as many free electrons as there are atoms in any volume, the electron deficiency is confined to a very small space.

As in our consideration of the copper-oxide rectifier, if we now imagine a potential difference established between the semi-conductor and the metal, when the latter is positive the diagram of Fig. 14.10(a) is obtained. When the metal is negative, the diagram of Fig. 14.10(b) is obtained. These have been derived from Fig. 14.9 simply by raising and lowering the potentials of the metal and semi-conductor.

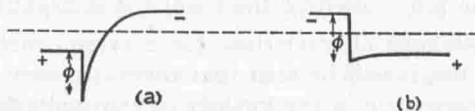


Fig. 14.10. Potential-energy diagram of metal and P-type semi-conductor; (a) metal positive, (b) metal negative.

From the diagrams it is clear that electrons can pass between the metal and the semi-conductor more easily when the latter is positive, and the direction of the contact difference of potential, ϕ , decreases the probability of an electron finding its way from the semi-conductor to the metal. It should be emphasized that the increase in electron energy necessary to surmount the barrier between a P-type semi-conductor and a metal is mainly due to what may be called the 'ionization energy' of the pure semi-conductor. Positive holes are convenient mathematical fictions which can only be transmitted across the barrier by (a) 'ionizing' an atom in the semi-conductor and (b) moving the electron so formed from the semi-conductor to the metal. The energy required for (b) may be positive or negative, depending upon the external voltage. That required for ionization is always positive, and for silicon is just over 8 electron-volts. In general the total energy required will be positive and will therefore have the direction shown in the diagrams above.

If there were only one metal-to-semi-conductor contact in the crystal valve, one would expect, therefore, that rectification would occur and that the current-voltage characteristic would be of the shape shown in Fig. 14.8. But there are two such contacts, for the crystal must back on to a metal holder. In this case, however, the contact is spread over a much larger area and the space-charge in the semi-conductor for a given current is much smaller. The resistance *for both directions of flow* is therefore much smaller for the large area contact, and so the resistance of the small area contact is the factor determining the action of the valve. This holds for all values of the current of any importance and explains why it is possible to neglect entirely the rectifying action of the larger area contact.

The silicon-tungsten rectifier is another example of a barrier-layer rectifier, but in this case the effective width of the layer is intermediate between the values for the two copper-oxide rectifier contacts. It is of the order of 10^{-6} cm. This layer width leads to certain complications in the theory which need not concern us at the moment.

Operation of the Crystal Mixer

The equivalent circuit diagram for a crystal valve is shown in Fig. 14.11, where r is the variable resistance of the metal-to-semi-conductor contact, R is the (approximately) constant resistance of the semi-conductor itself and C is the effective capacitance at the contact due to the barrier layer. This is physically an over-simplification, but electrically it gives a useful picture of the effects which are to be expected. The capacitance, C , depends upon the area

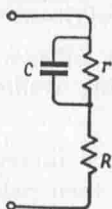


Fig. 14.11. Equivalent circuit of silicon crystal valve.

of contact which in turn depends upon the pressure exerted by the tungsten whisker on the crystal surface. Adjustment of the crystal valve during manufacture is chiefly aimed at producing a suitable pressure. C also varies with the voltage applied across the valve, as may be deduced from a study of the electron-energy diagrams already given. But for potentials of the order of 1 volt the variation in C is small.

Two crystal parameters which are easily measured are of great assistance to the user in assessing whether a valve is 'good' or not. These are the *forward* and *back resistances* of the unit. The forward resistance is the resistance measured when the applied potential difference is in the direction in which the current flows more easily; it is usually of the order of 250 to 500 ohms for a potential difference of 1 volt. The back resistance is the resistance measured with the p.d. reversed and should be at least ten times this. In fact a rough guide as to whether a crystal valve is good or bad is to measure the ratio of 'back' to 'front' resistance. If this is less than 10/1 the valve is suspect.

When crystal valves are used in mixer units for frequency conversion, it is usually at signal frequencies of the order of 3,000 to 10,000 Mc/s. The advantages obtained from their use all stem from their small dimensions and simplicity of construction. Since the only essential feature of such a mixer valve is the metal-to-semi-conductor contact, lead inductances can be reduced to a minimum. Moreover, since the barrier layer is only of the order of 10^{-6} cm thick, and since this is where the principal time lag between the application of a voltage and the flow of the corresponding current occurs, transit-time effects are much reduced. These advantages have been found in practice to outweigh the disadvantages of using what is essentially a diode rectifier as a frequency changer. The crystal valve has a poor conversion efficiency. If the ratio of the i.f. output power to the signal input power is measured, the crystal valve shows a loss of from 5–9 db at 3,000 Mc/s under the most favourable conditions. This is only acceptable because the noise factor of the device is correspondingly small.

In Chapter 6 it was shown that the conversion efficiency of a diode or triode mixer valve varies with the local-oscillator power as shown in Fig. 14.12. To

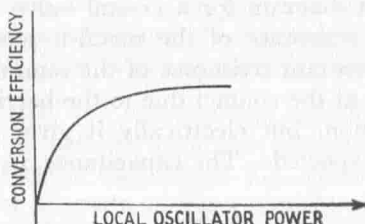


Fig. 14.12. Variation of conversion efficiency with local-oscillator power for a mixer using a silicon crystal.

begin with, the conversion efficiency increases rapidly as the local-oscillator power is increased from zero, and then, when the linear part of the current-voltage characteristic has been reached, further increase in local-oscillator voltage produces no improvement. It may also be shown experimentally and theoretically that the effective noise temperature of a crystal valve, regarded as a resistor, varies with the local-oscillator power as shown in Fig. 14.13. The effective noise temperature is defined as the ratio of the noise power available from the valve to the noise power for a resistor at room temperature (290°K); i.e., kTB .

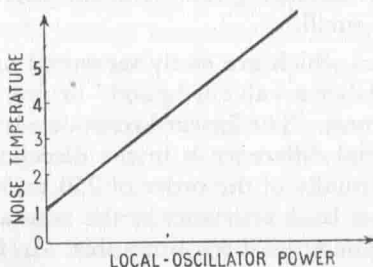


Fig. 14.13. Variation of noise temperature with local-oscillator power.

As a direct consequence, the over-all effective noise factor for a small crystal mixer varies with the local-oscillator power as shown in Fig. 14.14. When the local-oscillator power is small, both conversion efficiency and the noise temperature are small, but the former is negligible while the latter cannot be less than unity. So the noise factor of the device is infinite. As the local-oscillator

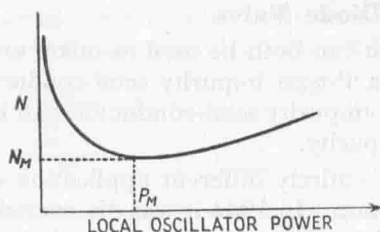


Fig. 14.14. Variation of noise factor with local-oscillator power.

power increases, the conversion efficiency increases more rapidly than the noise temperature, causing the noise factor to decrease. As the rate of increase of the conversion efficiency falls off (Fig. 14.12), the noise temperature again asserts itself, producing a minimum, N_M , in the noise factor at some optimum value, P_M , of the local-oscillator power. Fortunately, the minimum is very flat, giving a fairly wide tolerance in local-oscillator adjustment. In general terms, the optimum local-oscillator power for a bandwidth of a few megacycles in the average i.f. amplifier is about half a milliwatt. With the usual mixer adjustments this leads to a d.c. component in the crystal rectifier of about 400 μ A.

For a crystal valve in a frequency converter operating at a frequency of 3,000 Mc/s into an i.f. amplifier having a noise factor of 5 db, the overall noise factor is between 10 and 16 db. For more exact calculation a convenient formula is

$$N = L + 10 \log_{10} (N_{IF} + t - 1)$$

where N = over-all noise factor (db)

L = conversion loss of valve and circuit (db)

t = noise temperature of crystal valve (numerical ratio)

N_{IF} = noise factor of i.f. amplifier (numerical ratio)

The value of t is usually between 2 and 4 for optimum local-oscillator input.

Burn-Out of Crystal Mixer Valves

It has already been stated in Chapter 13 that the purpose of TR and ATR gas switches is to reduce the power flowing into the crystal mixer of the radar receiver during a transmission, and that the reduction in power is absolutely necessary, because the direct action of the transmitter on the crystal mixer is to diminish its efficiency and to increase its overall noise factor. Such changes in the rectifying action of a crystal valve are named *burn-out*.

Burn-out is due to current flowing in the forward direction. This has been proved by experiment. It is nowadays generally due to the 'spike' of energy transmitted through the TR cell before the latter becomes properly conducting. It has also been shown experimentally that the change in the crystal is probably due to a localized thermal effect. The high temperature generated at the

metal-to-semi-conductor contact during the absorption of the spike energy causes gradual changes in the nature of the semi-conductor at or near the metal 'point' contact. The resulting gradual loss of efficiency can be checked by observations made from time to time on the ratio of the back to forward resistances of the crystal valve.

Germanium Crystal Diode Valve

Silicon and germanium can both be used as mixer crystals. When silicon is so used, it is always as a P-type impurity semi-conductor; when germanium is used, it is as an N-type impurity semi-conductor. In the latter case antimony has been used as the impurity.

There is, however, an entirely different application of germanium which is of at least equal importance. In 1944 it was discovered by S. Benzer, working in U.S.A., that germanium containing certain other impurities, such as tin, produces a metal-to-semi-conductor contact having the interesting current-

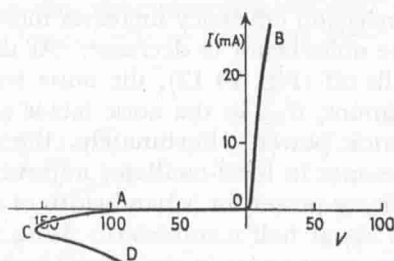


Fig. 14.15. Characteristics of germanium crystal valve.

voltage characteristic shown in Fig. 14.15. This characteristic is noted for two features. First, between A and B it represents an almost perfect rectifier, and secondly the negative resistance part, CD, does not occur until a very high back voltage is applied to the crystal. The part OB also exhibits a negative resistance. But this is of little importance for many rectifier applications.

Valves with characteristics of this type are now used as 'high level' rectifiers in many circuits. By 'high level' is meant input voltages of the order of 1-10 volts. The back voltage necessary to reach the negative-resistance part of the characteristic can be made almost anything up to 100 or even 200 volts. Commonly, it is specified that OC shall be greater than 50 volts. But the manufacture of such germanium crystal diodes is as yet more of an art than a science and the explanation of many of their properties is not yet fully known. The advantages of such crystal rectifiers are obvious. First, no heater voltage is required. Secondly, the front resistance is small (of the order of 150 ohms at 1 volt) and the back resistance is almost as high as for a valve (about 100,000 ohms at 1 volt). Thirdly, the current-voltage characteristic passes through the origin. Fourthly, the crystal valve need not take up more space than a quarter-watt resistor. Fifthly, the inter-electrode capacitance can be very small, much smaller than that of a corresponding thermionic diode. Lastly, it can be soldered directly into the circuit. Such valves can be used at frequencies up to about 100 Mc/s. They cannot be used at low signal levels, since their noise factors are high, and they cannot be used at temperatures in excess of 70°C.

Point-Contact Transistors

Certain crystals, notably germanium, silicon and lead sulphide (galena), which, either as impurity or intrinsic semi-conductors, give rise to pronounced rectification when in contact with a fine metal wire, can be used in conjunction with three metallic contacts as crystal triode valves, called *transistors*. The contacts may be arranged as shown in Fig. 14.16, where A and B are two point contacts very close together (less than 0.1 mm apart in germanium) and C is a large area contact at some distance from A and B. Fig. 14.16 is only one

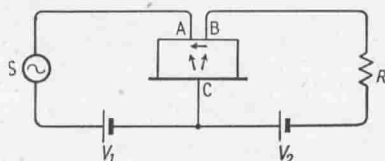


Fig. 14.16. Transistor in its basic circuit.

example of how such a triode valve may be made. So long as A and B are close together and C is a more distant large area contact, the actual crystal shape is of little importance.

If A is made positive with respect to C by means of the direct voltage, V_1 , and B is negative with respect to C by reason of V_2 , it is to be expected that an N-type semi-conductor, such as germanium, will allow a larger current to pass in the AC circuit than in the BC circuit, for the current will be forward across AC and backward across CB. But the proximity of A to B, associated with the space-charge in the semi-conductor, causes an appreciable current to flow between A and B. This can only return to earth via the resistance, R , so the potential of B depends to a certain extent upon the potential of A. There is, therefore, a mutual transconductance in the device similar to that found in the thermionic triode valve. The potential of B depends both on the potential of A and on that of C.

► In order that this triode effect should occur the carriers of current in the semi-conductor must be able to pass from A to B without being lost. Now it has been shown that for germanium containing certain impurities the current carriers between A and B are positive holes. In this case it is necessary that A and B should be so close together that a positive hole can be passed from A to B without recombining with an electron. This condition is fairly well satisfied in existing commercial transistors if AB is less than 0.1 mm. ◀

A wholly acceptable theory of the metal-crystal transistor has not yet been given. It is clear from experiment that the effect depends little if at all on the nature of the metals forming the contacts, and also that the germanium near the surface must simulate the behaviour of a P-type semi-conductor, as mentioned in the last paragraph. At greater depth the germanium retains the characteristics of an N-type semi-conductor. Whether this surface effect can be ascribed to oxygen absorbed at the surface is yet to be determined: but for our purpose the description already given must suffice.

Returning to Fig. 14.16, A and B are now known as the emitter and collector respectively, and the voltage amplification obtained from a transistor can best be demonstrated by means of a graph showing the collector voltage, V_C ,

plotted against the emitter voltage, V_E , for a given value of the supply voltage, V_2 . The curve depends, of course, on the value of R , as it does in the case of the thermionic valve. As drawn in Fig. 14.16, $V_E = V_1$ and the signal voltage supplied by the generator S is supposed to be zero. Varying V_E gives a curve of the shape shown in Fig. 14.17 and varying V_2 produces the other curves of the family.

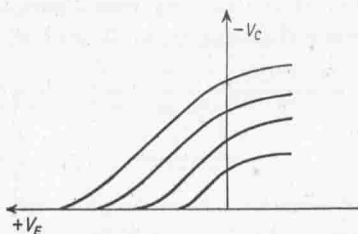


Fig. 14.17. Transistor characteristics. Note that the scales are the reverse of the usual ones.

To bring the characteristics into the usual positions on the paper the directions of the axes of V_C and V_E have been reversed from the usual mathematical convention. If the characteristics are drawn for a P-type semi-conductor such as lead sulphide, they do not require the direction of the axes to be reversed.

It will be noted that a change in V_E inside a certain range produces a change in V_C , and that the curves are very similar to those obtained using a thermionic triode valve. The voltage amplification depends in the usual manner on the value of R and compares favourably with the values obtained from thermionic devices. The transistor valve has small inter-electrode capacitances, but the voltage gain depends in the usual way on the frequency. At frequencies above a megacycle the gain begins to diminish, so that this type of valve so far can only be regarded as a low-frequency device.

The possible uses of the transistor valve are innumerable and they are now being employed experimentally in low-frequency and carrier-frequency communication equipment. They suffer from two disabilities both concerned with their manufacture. First, the germanium used is an impurity semi-conductor with all its attendant difficulties. Secondly, the proximity of the two 'point' contacts presents a certain difficulty so far as reproducibility is concerned. For certain switching applications point-contact transistors can already be used instead of thermionic valves with profit. Their temperature limitation is, however, serious.

The Junction Transistor

Recently, members of the Bell Telephone Laboratories in U.S.A. have reported a new type of transistor, which, if it can be made in quantity, will have many important uses. The Bell Telephone Laboratories have led the world in transistor development since one of their members discovered the 'triode' effect in 1945.

The new transistor comprises a crystal of germanium consisting of two regions, each exhibiting N-type semi-conductivity. Three electrodes are attached to the crystal, none being 'point' contacts, but rather 'large' contact

areas of approximately equal ohmic resistance. One is attached to each N-type region and the third to the P-type region. Fig. 14.18 shows diagrammatically the arrangement, connected as an amplifier.

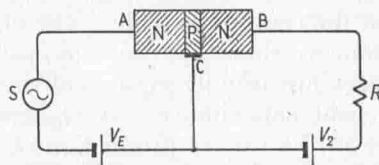


Fig. 14.18. N-P-N transistor.

The electrons from A (the emitter) diffuse through the P-type layer to B (the collector). For a given value of V_E , the current to the collector increases at first very rapidly as V_2 is increased and then saturates. When the current flowing in the collector circuit is saturated, the potential of B, V_C , varies in a linear manner with V_E over a very wide range of the latter. This gives a family of curves strongly reminiscent of the pentode, as can be seen from Fig. 14.19.

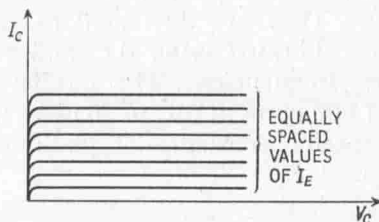


Fig. 14.19. Characteristics of N-P-N transistor.

The figures claimed for the performance of these transistors are remarkable. First, let us consider a valve made to handle power at low frequencies. In one case the P-layer was made 0.7 mm thick. Such a valve is said to give about 2 watts of undistorted power, operating as a class A amplifier, at an efficiency very close to the theoretical value. The valve, excluding leads, is smaller than a 1/4-watt resistor. In such a transistor the limiting factor in the frequency response is the capacitance of, and transit-time in, the P-layer. The cut-off occurs at about 10,000 c/s. Secondly, let us consider a valve made to operate as a signal amplifier with a P-layer thickness of 1.5×10^{-2} mm. This gave a power gain of 50 db at low signal levels and had a noise factor of 10 to 15 db, measured at 10 kc/s. Oscillators have been made, using these transistors, which have operated with one microwatt d.c. input. The same transistor gave more than 100 mW output as an amplifier.

If such N-P-N or P-N-P transistors can be made economically in quantity, they bid fair to replace the thermionic valve in very many of its applications. Their high-frequency behaviour appears, however, to be inherently less satisfactory than that of a thermionic device, so the replacement will probably occur first at low and medium frequencies. The temperature limitation applies equally to junction transistors where these are made of germanium.

Thermistors

As was indicated earlier in the chapter, a metallic conductor may be expected to decrease in conductivity with rise of temperature due to increased disorder

in the free electrons, but a semi-conductor must always increase in conductivity since an increase in the thermal energy represents an increase in the *number* of electrons or positive holes in the lattice. The two effects are of a different order of magnitude, for a semi-conductor may double its conductivity for a 20° rise of temperature on the centigrade scale. The effect is used to construct thermally-sensitive resistances, shortened to thermistors, which are much used in radio. Thermistors are usually very small beads or discs of a semi-conductor having small heat capacitance. A change of current in such a device produces a change in the rate of production of heat (due to the resistance) which changes the temperature and therefore the resistance itself. They can therefore be used in one arm of a resistance bridge to give a very sensitive indication of temperature or of the current flowing.

The great difficulty in manufacture is, as might be expected, the control of impurity to give reproducibility. Titanium oxide in one of its forms is an N-type semi-conductor which has been used successfully: in other cases mixtures of oxides have been sintered to form a bead.

One of their most important uses in radio work is in the measurement of power at ultra-high frequencies. They are then used as variable resistances in a bolometer which can be calibrated using d.c. and which will maintain its calibration at the highest frequencies. The specific resistance of a typical thermistor changes from 1,000 ohms at 0°C to 50 ohms at 100°C. For the same temperature range the specific resistance of platinum changes from about 10^{-5} ohm to 2×10^{-5} ohm.

CHAPTER 15

Electron Optics

IN EARLIER chapters there have been occasional references to 'electron guns', 'electron optics' and 'electron lenses'. It is the purpose of this chapter to explain what is meant by these terms and to describe how the direction and cross-section of a beam of electrons may be accurately controlled. This discussion will act as an introduction to the study in Chapter 16 of cathode-ray tubes and similar devices.

The reader is familiar with the action of a gun on a projectile. It shoots the projectile from the muzzle with a particular speed in a particular direction. An electronic device which shoots electrons at a particular speed in a particular direction is known as an electron gun. But whereas an ordinary gun shoots out projectiles one at a time, an electron gun shoots out anything from 10^{12} to 10^{20} electrons per second. An electron gun is an essential part of every cathode-ray tube and klystron; it is finding new uses every day in television camera tubes, travelling-wave valves and storage tubes.

The reader is also familiar with the action of a convex or concave lens on a beam of light. A convex lens tends to bring the light passing through it to a focus, while a concave lens tends to spread out the beam. Electron beams can be influenced in an exactly similar manner by certain electrostatic and magnetostatic fields, produced by electric potentials applied to shaped electrodes or by constant electric currents circulating in shaped coils. Consequently such arrangements of electrodes or coils are known as electron lenses, their functions being analogous to those of lenses used for light. Moreover, it has been shown that the laws governing the movements of electrons in electric and magnetic fields are formally identical with the laws governing the behaviour of light in moving from one medium (such as air) to another medium (such as glass) or in passing through an aperture. The name electron optics has therefore been given to the study of electron motion in such fields. The correspondence between light optics and electron optics is not fortuitous. There is a fundamental correspondence between the phenomena associated with a train of electromagnetic waves and those associated with a stream of fast-moving particles. Indeed, it has been demonstrated that a stream of electrons, each of mass m and moving with speed v can be regarded as a train of waves of wavelength h/mv , where h is Planck's constant. All the properties of such a train of waves are exhibited by the stream of electrons.

However, despite the fundamental connection mentioned above, and despite the use of the term "Electron Optics" as the title for this chapter, it is not necessary for our purpose to stress the wave-like character of electron streams. All the simpler phenomena of electron lenses can be understood quite easily without reference to wave-motion. In the subsequent paragraphs the ex-

planations of these important qualities of electron beams will be given in purely classical electro-mechanical terms, just as one can describe the simple actions of lenses and mirrors in light without using wave theory at all.

Typical Electrostatic Lenses

Suppose (Fig. 15.1) that electrons are emitted by a disc-shaped hot cathode, K, and that a certain number pass through the aperture, P, in an anode, A, which is at a potential V_1 relative to K. The beam of electrons passing through P diverges in the same way as light diverges in similar circumstances. Also, the electrons passing through P do so at a speed $v_1 = \sqrt{2V_1 e/m}$. Let the beam of electrons now pass axially along the metal cylinder C_1 , connected to A, and therefore at the same potential V_1 .

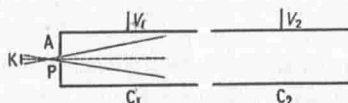


Fig. 15.1. Simple electron gun and electrostatic focusing system.

The first cylinder, C_1 , is placed opposite and coaxial with a second cylinder, C_2 , of the same diameter but maintained at a different potential, V_2 , relative to K. Then it will be shown that the space between the two cylinders acts as a converging electron lens, tending to bring the diverging electron beam to a focus. This is one of the simplest electron lenses and its action is easy to understand.

Drawing the lens on a larger scale, as in Fig. 15.2, the lines of electric force between the opposing ends of the two cylinders are as shown by the broken

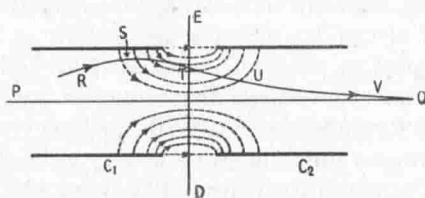


Fig. 15.2. Electron lens showing the electric field between the cylinders when C_2 is at a higher potential than C_1 .

lines. The common axis, PQ, of the two cylinders is also shown. Then an electron, moving away from the axis along the path RS, comes gradually into the electric field between the cylinders.

Consider first the case where V_2 is greater than V_1 , so that the electrons are accelerated in moving from C_1 to C_2 . Then the arrows on the lines of force indicate the directions of the forces acting on the electrons. The electron at S is acted on by a force tending to move it *towards* the axis, and this converging force continues until the electron has crossed the plane ED, mid-way between the two cylinders. Consequently, the electron path or 'trajectory' is bent towards the axis, the electron moving along RST. On the right-hand side of the plane ED, the force acting on the electron tends to move it *away* from the

axis, as is shown by the arrows. The *diverging* force continues until the electron is out of the field altogether and produces a bending away from the axis, giving the path TUV. The electron beam therefore tends to converge on the left-hand side of ED and tends to diverge on the right-hand side, and at first sight it might appear that these two tendencies would cancel each other, leaving the beam unaffected. This is not so. It must be noted that the electrons acquire energy in moving from C_1 to C_2 , since V_2 is greater than V_1 , and so the speed of each electron gradually increases as it moves from P to Q. But the deflection actually produced by the radial force between the two cylinders depends upon the time spent by the electron in the field of this force. The electron spends less time in the field on the right-hand side than it does in the field on the left-hand side, since its speed is continually increasing, and so the actual deflection produced between R and ED is greater than that produced between ED and V. Hence the over-all effect of the electric field is to cause the beam to converge. The lens is a converging lens analogous to a convex lens in light.

Consider now the case where V_2 is less than V_1 . The diagram corresponding to Fig. 15.2 is shown in Fig. 15.3. In this case the electron beam first tends to diverge and then tends to converge. But the speed of the electron is reduced

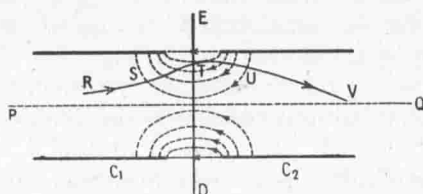


Fig. 15.3. Electron lens similar to that of Fig. 15.2 but with C_2 at a lower potential than C_1 .

as it moves from C_1 to C_2 , and therefore the actual deflection produced on the right-hand side of ED is greater than that produced on the left-hand side. So, again, the lens is a converging lens, despite the reversal of potentials.

The case just examined is that of the cylindrical electrostatic lens, which is always convergent. Such an arrangement has considerable practical significance and is used extensively in cathode-ray tubes. Modifications are quite numerous, and each may be examined in the same way. But care must always be taken to allow for *all* the electric fields in the vicinity of the electrodes, if a complete result is to be obtained.

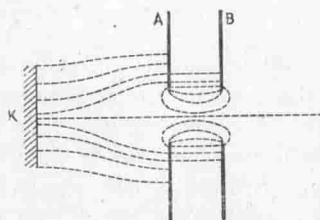


Fig. 15.4. Electron lens using plane electrodes instead of cylinders.

As an example of this of considerable practical importance, consider what is generally known as a 'single-aperture' electrostatic lens. Suppose the anode A in Fig. 15.1, instead of being followed by a cylinder, is followed, as in Fig. 15.4, by another plane electrode B, having a coaxial aperture. In general, the geometrical arrangement of all three electrodes and their several potentials must be taken into account in estimating the lens action, but if the field due to K can be neglected near A and B in comparison with the field due to the relative potentials of A and B, then the lens action is easily studied and can be shown to be convergent. The lines of force for this case are shown in Fig. 15.4, the symmetry of the field implying that the effect of K can be neglected.

The Electron Gun

The practical case of greatest importance is the arrangement of electrodes in an electron gun. A simple example of this is shown in Fig. 15.5, where all electrodes are supposed to possess cylindrical symmetry about the line PQ. K is the disc cathode, surrounded by the coaxial cylinder, C, known as the control electrode. This has an aperture opposite the cathode and about the same size. Facing the end of this cylinder is the plane anode, A, having a smaller coaxial aperture. The control electrode is maintained at a potential equal to or negative with respect to K, while the anode is maintained at a high positive potential. Assuming, for example, that C is negative with respect to K, the lines of force are shown by the dotted lines in Fig. 15.5, the arrows showing the direction of the forces acting on an electron. An electron leaving the cathode finds itself in a field accelerating it parallel to the axis and radially towards the axis. But as it nears the anode, the radial field diminishes and finally reverses in direction, so that the electron tends to move away from the axis. As in the previous cases, the slowly-moving electron near K is deflected more *towards* the axis than the fast-moving electron near A is deflected *away* from it: hence the over-all action of the gun is to produce a converging beam. Indeed, the effect of the gun is to bring all the electrons leaving the cathode to an approximate focus at Q on the axis.

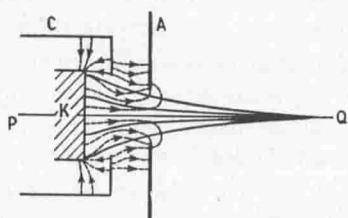


Fig. 15.5. Simple electron gun.

Two important points should be noted. First, an important physical action has been neglected in the discussion: the space-charge in the vicinity of the cathode modifies considerably the electrostatic field so that the lines of force cannot be exactly drawn by normal geometric means. This modifies to some extent the action of the gun. Secondly, the negative potential on the control electrode may have such a considerable effect that the focus F may actually occur between the two apertures, or almost exactly at A. This condition is shown in Fig. 15.6: it is the normal state of affairs in a cathode-ray tube gun.

As a result, the electrons are actually diverging as they move away from A, and another electron lens is required to bring them to a focus.

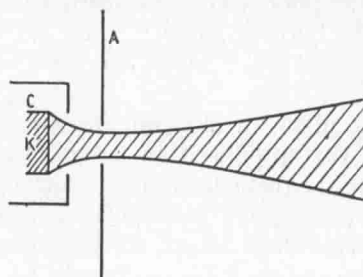


Fig. 15.6. A negative-control electrode C can bring the focal point of the beam to the plane of the anode A.

The Rubber Model and Electrolytic Tank

In certain simple cases, such as that shown in Fig. 15-2, it is possible to calculate the exact degree of convergence or divergence introduced into an electron beam for given electrode potentials. What is done in this case, as in the case of lenses for light, is to calculate the focal length of the arrangement, showing where a parallel beam would be brought to its smallest cross-sectional area. Where the focal length cannot be calculated, two equivalent experimental techniques are available for measuring it. One of these methods, the rubber model, is easy to manipulate and gives answers very quickly. The other, the electrolytic tank, is capable of much greater accuracy, but requires much painstaking work to yield any answer at all. Only the rubber model will be described.

Imagine a sheet of extremely thin rubber held in a horizontal plane in a slightly stretched condition. This may be done by pegging the rubber sheet to a wooden frame. It may then be shown that the electrostatic field acting on an electron between electrodes is almost exactly imitated by the gravitational field acting on a small spherical ball between simulating electrodes. The sheet is pushed up or down by solid strips made to represent sections of the electrodes. For example, considering Fig. 15-5, the sheet is held at one particular level at one place (cathode potential) by two strip clamps above and below the rubber, each equal in length to the line K in the diagram. Two right-angled strips are made to represent C and these are pushed upwards against the rubber at the appropriate place. The height to which the rubber is forced upwards by the strips is made proportional to the negative potential applied to C. Two straight strips are made to represent A and are pushed downwards against the rubber at the appropriate place, simulating a proportional positive potential. So long as the deformations of the rubber sheet are small, the electrostatic field is represented three-dimensionally by the model, the vertical axis being electrostatic potential.

Therefore, if a small spherical ball is started from the cathode strip, its subsequent motion corresponds to that of an electron in the field and by starting balls from different points on the cathode strip, the focus of the corresponding electron gun can quickly be found. Also, the electrodes can readily be adjusted to give any required focus. Such a rubber model is very useful for teaching

purposes and for rough design. It should, however, be noted that in common with the electrolytic tank, the model suffers from the disability that it takes no account of space-charge phenomena. In many cases this is of little importance but it must always be borne in mind.

Magnetic Focusing

A magnetic field, acting upon an electron, can produce a deflection of the latter corresponding to that produced by an electric field. Consequently, devices can be made, using magnetic fields only, which focus a beam of electrons: such devices are known as magnetic electron lenses. In some ways magnetic lenses are superior to electrostatic lenses, for they may be (and usually are) entirely outside the envelope of the valve or cathode-ray tube and may therefore be readily moved about to adjust the position of the electron beam in the vacuum. Where exact alignment of an electron beam is required, a magnetic lens is technically superior: but the fact that it requires magnetic material such as soft iron and either a permanent magnet or an independent electric current makes it unsuitable for some simple uses.

Effect of a Solenoid

Imagine a beam of electrons, A, diverging from a small aperture on the axis, AC, of a long solenoid which is carrying an electric current (Fig. 15.7). Considering one electron only, let its path be AK as it enters the solenoid, and let



Fig. 15.7. Long magnetic lens. The divergent beam of electrons from A traverses the long helical coil towards C.

the axial magnetic field, due to the current in the solenoid, be B . Then B acts parallel to the axis and is independent of the distance of K from the axis (see Volume 1). In Fig. 15.8 the velocity of the electron has been resolved into u , parallel to the axis, and v at right-angles to the axis. Since $\angle KAC = \theta$

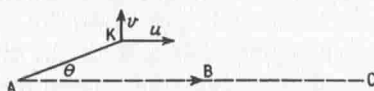


Fig. 15.8. An electron has velocity u parallel to the axis and v at right angles to the axis. The magnetic field B has no effect on the component u .

is usually small, v is small compared with u . The magnetic field, B , has no effect on the component u , but it produces a mechanical force Bev on the electron, acting at right-angles to the plane of the paper, and due to the electron current ev at right angles to B (see Volume 1). This force causes the electron to spiral about the axis AC, introducing a component of speed, w , at right-angles to both u and v .

As soon as the electron has a component speed at right-angles to u and v , the magnetic field produces a mechanical force acting towards the axis. By suitably choosing the value of B , this force may be made to cancel out v , the

diverging speed, and to drive the electron towards the axis. Fig. 15.9 shows the net effect on an initially diverging electron of an axial magnetic field. Beyond D the process repeats itself, so that the electron proceeds in a helix, the axis of which is parallel to AC. The pitch of the helix is AD, for it should be noted that at X the electron is not on the axis, but some distance behind the paper.

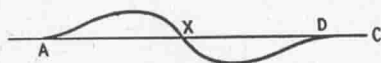


Fig. 15.9. Plan of the helical path taken by an initially diverging electron in an axial magnetic field.

In practice, it is possible to maintain a narrow beam of electrons over a very considerable distance by applying a uniform axial magnetic field which can be done by using either a solenoid or a permanent magnet. It may be shown also by exact calculation that the distance AD (Fig. 15.9) is nearly independent of θ , the diverging angle, when the latter is small. It is thus possible by means of a solenoid to bring the beam to an approximate focus at a certain distance from the aperture. The focal spot, however, is no smaller in area than the aperture itself.

A solenoid is used to maintain an electron beam of small cross-section in all travelling-wave tubes and similar devices. The reader should make himself familiar with the principle involved, for a similar action occurs in all magnetic lenses. A more exact treatment is given in Appendix 1.

Ideal Magnetic Lens

If we now imagine a beam of electrons diverging from the axis at A, as shown in Fig. 15.10, and coming suddenly into a constant magnetic field parallel to the axis at X, the divergent beam tends to converge. If the magnetic field

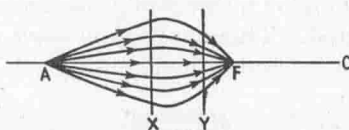


Fig. 15.10. Effect on a divergent electron beam of a magnetic field between X and Y.

ceases abruptly at Y, the electrons proceed thereafter undeflected to an approximate focus at F, and the region of the magnetic field acts as an electron lens. It is possible by increasing the field and by decreasing XY to make a device which corresponds closely to a convex optical lens, and the focal length of the lens can be varied simply by varying the strength of the magnetic field.

Real Magnetic Lens: Simple Coil

It is not possible to obtain a magnetic field which suddenly increases from zero to a finite value at X and as suddenly disappears at Y, but a simple coil with its axis along the axis of the electron beam functions in much the same way. A coil of wire, indicated in section by the small circles of Fig. 15.11, when a direct current is passed through it, gives rise to the magnetic field shown qualitatively by the lines drawn round the coil.

If a diverging beam of electrons passes through the coil from A to C, the effect is qualitatively the same as for the ideal field shown in Fig. 15-10. This is most simply shown by considering separately the two components of the magnetic field, one axial and the other radial. So far as the axial field is concerned, it gradually increases from zero to a maximum value at Z and then falls

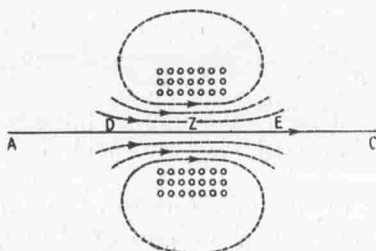


Fig. 15.11. Form of magnetic field around a simple focusing coil.

off again towards C. The integrated effect of the field is to cause the diverging electrons to converge, for the force is always towards the axis. The radial field first deflects an electron to one side at D and then to the other at E. Since the electron speed parallel to the axis is unchanged by the passage through the lens, the effects of the radial field cancel out exactly. The radial field has no effect at all.

Here we have an electron lens which can be mounted outside the evacuated container and the focal length of which can be varied simply by varying the current in the coil. The reader should consider the various possibilities which present themselves, using such coils to control an electron beam.

Shielded Coil Lenses

To obtain the maximum magnetic focusing effect from a given direct current in a coil it is necessary to employ a shield of a magnetic material to concentrate the field where it is required. There are many ways of doing this, one of the most popular being shown in section in Fig. 15-12.

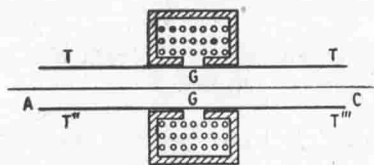


Fig. 15.12. Common form of magnetic focus coil. The winding has an iron surround with a gap G through which the flux emerges.

AC is the axis of the electron beam enclosed in a tube the walls of which are shown at TT' T'' T'''. The section of the coil is indicated by the small circles. It is surrounded by a shield of magnetic material with a small circular gap, G. It is at this gap that the magnetic field is used. By this means an extremely powerful magnetic field can be made to act over a very limited distance, giving a strong electron lens. Double lenses can be formed by two such coils and shields mounted together, and by shaping the gap suitable corrections can be introduced to the electron optical system.

Defects in Electron Lenses

Returning to our analogy of the optical lens, when a clear picture is required using a lens system, care has to be taken to avoid distortion. We are all aware of how such distortion, due to defects in optical lenses, can affect work with cameras and projectors. Electron lenses are subjects to similar defects, and where an exact reproduction of a picture is required (as, for example, in television) the distortions produced by these defects may be a limiting factor.

Chromatic aberration, very troublesome in optical systems, is not important in electron lenses, for the wavelength of the light, giving rise to coloured images, corresponds to the electron velocity in the electronic system. The velocities of the electrons in the beam are usually identical except for their emission velocities from the cathode. This is a negligible effect in most cases.

To obtain clear pictures, using electron beams, two conditions must always be satisfied. First, the magnification of the image must not vary with the distance of the image from the axis, and secondly, images at different distances from the axis must come to focus on the same plane. Varying magnification over the picture is due to radial variations in the electric or magnetic fields in the lenses. For example, the electric or magnetic field at B in Fig. 15.13 should bear a certain relationship to the electric or magnetic field at A. Should this not be the case, two effects follow. The magnification and the focal length vary according to the part of the lens through which the electrons pass, and distortions result.

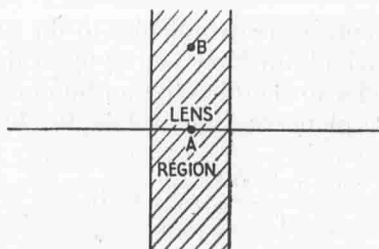


Fig. 15.13. The magnetic field strength at a point B of the axis must be properly related to the field on the axis A if distortion is to be avoided.

An additional effect is experienced using magnetic lenses. When passing through the latter, individual electrons are deflected in azimuth; i.e., they suffer an angular displacement about the axis. Should this displacement depend upon the radial distance of the image, a 'spiral distortion' will result. It may readily be corrected, however, by the use of two magnetic lenses (or a double lens) causing the electrons to spiral in opposite directions. The methods used to test the fidelity of an electron-lens system are described in the following chapter. These methods involve the use of devices which have not yet been discussed.

APPENDIX 1.

Magnetic Focusing

Suppose that a beam of electrons, originating at a cathode, K, emerges from a small aperture in an anode, A, the anode potential being V (Fig. 15.14). Then the electron speed at the anode is $v = \sqrt{2Ve/m}$, where e and m are the electronic charge and mass respectively. Let a constant magnetic field of strength B act parallel to the axis of the beam, CZ. The electron speed v has two components $v \cos \theta$, parallel to CZ, and $v \sin \theta$, perpendicular to CZ, where θ is the angle which the electron trajectory at A makes with the axis. In the beam there are electrons with all values of θ up to a certain critical value determined by the geometry of the cathode and aperture.

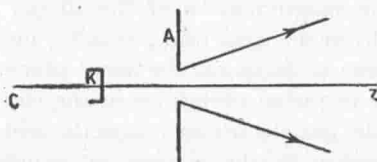


Fig. 15.14. A beam of electrons originating at K emerges as shown from a small aperture in the anode A.

The element of electron current perpendicular to the direction of the magnetic field is $ev \sin \theta$, and the resultant force acting upon the electron is $evB \sin \theta$. This force is at right-angles to the direction of motion and therefore gives rise to a circular path in a plane at right-angles to the axis. By elementary mechanics

$$\frac{m(v \sin \theta)^2}{r} = evB \sin \theta,$$

where r is the radius of the circular path. Hence

$$r = \frac{v \sin \theta}{Be/m}$$

The electron performs one complete cycle in time T , where

$$T = \frac{2\pi r}{v \sin \theta} = \frac{2\pi}{Be/m}$$

During this cycle the electron moves a distance $Tv \cos \theta$ along the axis. After one cycle, therefore, the distance of the electron from its starting point is

$$D = \frac{2\pi v \cos \theta}{Be/m}$$

If θ is small, $\cos \theta = 1 - \frac{\theta^2}{2} + \dots$ is nearly unity, and D is approximately $2\pi v/(Be/m)$. Therefore all the electrons in the beam come to a focus at approximately the same point.

CHAPTER 16

Cathode-Ray Tubes

THE PURPOSE of the cathode-ray tube is to show visually how certain electrical quantities, such as a voltage or a current, or both, vary with time. Since it is nowadays possible, by some means or another, to transform any physical parameter into a voltage or a current, the cathode-ray tube is extensively used to exhibit the variation with time of any physical quantity. The applications of the c.r. tube cover the whole field of science, and in television it has also become an adjunct to visual art.

All applications of the c.r. tube require the same basic elements, and it is the purpose of this chapter to describe these elements, to show how they are put together to form the completed device and to examine how they must be specified to meet the requirements of radio engineering. The elements of a c.r. tube may be listed as follows:—

- (a) a means of producing a narrow beam of electrons
- (b) a means of varying the intensity of (or electric current in) the beam or, using other words, a means of modulating the beam.
- (c) a means of making visible the small area where the beam impinges upon the end of the tube. The active area which can become luminous where the beam strikes it is known as the screen.
- (d) a means of deflecting the beam by applying to the tube a voltage or current; i.e., deflecting plates or coils.

The beam of electrons in the form of a narrow pencil, P (Fig. 16.1) is brought to a focus, F, on the screen, S, by means of an electron-lens system. The spot,

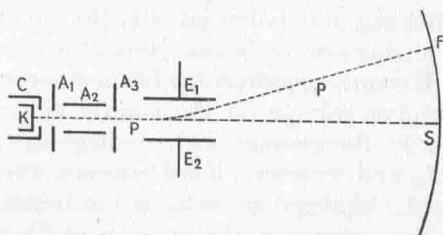


Fig. 16.1. Diagram showing the elements of a cathode-ray tube.

F, is then caused to move across the screen by the application of deflecting forces. These may be due to the electrostatic field between the plates, E_1 and E_2 , across which there is a potential difference, V . On the other hand, the deflection may be caused by the magnetic field due to a pair of coils carrying a current, and not shown in the diagram. The spot where the beam strikes the screen is made visible by means of a thin layer of luminescent material

deposited on the inside surface of the glass end of the tube. By using two deflecting systems the spot may be made to traverse the screen in two dimensions, its path depending upon the voltage or current applied to the deflecting plates or coils.

The Electron Beam

On the left of Fig. 16.1 is shown one method of producing an electron beam. The cathode, K, is coaxial with a control cylinder, C, and opposite the circular aperture in the first anode, A_1 . Beyond this anode is a coaxial cylinder, forming the second anode, A_2 , and beyond that again is another annular disc, A_3 , forming the third and final anode. As described in Chapter 15, such a system of electrodes may produce a convergent beam from the electrons emitted by the cathode. The electron-optics are shown more clearly in Fig. 16.2.

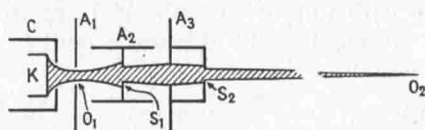


Fig. 16.2. In this section through a c.r.-tube the electron beam is shown shaded.

Considering the cathode to be at earth potential, such a beam-forming system might operate as follows:—

$C = -20$ volts, $A_1 = +1,000$ volts, $A_2 = +250$ volts, $A_3 = +1,000$ volts these figures indicating only the order of magnitude of the voltages concerned. When the potential of C is properly adjusted relative to K and A_1 , the beam is brought to a focus about O_1 , diverging again towards the right. The last lens, between A_2 and A_3 , again produces convergence, bringing the beam to an approximate focus at O_2 on the screen of the c.r. tube. However, such an electron-optical system is unsatisfactory, unless the area of the beam is defined by limiting apertures or 'stops'. These are shown at S_1 and S_2 . They cut off the peripheral zone of the beam and allow a much sharper spot to be obtained at O_2 . As will be seen immediately, they also make it possible to modulate the beam (to vary the beam current) without changing its focus. Increasing the negative potential of C (in the absence of the limiting apertures), causes two changes in the beam. First, the beam current is reduced. Indeed, if C is made sufficiently negative, the beam current is effectively reduced to zero, when the negative voltage on the control electrode is known as the cut-off voltage. Secondly, the change in the voltage of C slightly changes the position of O_1 and of O_2 , and therefore, if a sharp spot is required on the screen, the voltage of A_2 must be changed to re-focus the beam. Using the limiting apertures, S_1 and S_2 , the change in the position of O_2 due to changing the potential of C is much less marked, and for practical purposes, the effects of the potential of C and of the potential of A_2 are quite independent. This gives a beam the intensity and focus of which can be independently controlled. This is what is required for most c.r. tubes.

Fig. 16.2 is a diagram of only one example of a c.r. tube gun system. It is not necessarily the best system which can be devised, but it illustrates all the principles. In the manufacture of a c.r. tube it is obviously necessary to exercise considerable care in 'lining up' the various electrodes. If one aperture or

cylinder is badly out of line with another, the electron beam will be assymetrical and distortion effects will appear on the screen. A common example of this, not usually serious, is that the spot on the screen is not a uniform circle. It may have a slight tail. This may, however, be due to other causes.

Modulation of the Beam

For many applications of the c.r. tube it is necessary to vary the intensity of the spot on the screen. The intensity may have to vary with time in a regular manner: for example, in a common case the spot has to be suppressed for a certain period of time at regular intervals. Alternatively, the variation in the spot intensity may be required to produce a picture on the tube screen: the outstanding example of this is a television picture tube. In the latter case the intensity of the spot must be accurately controlled, and this is done by applying a variable voltage to the cylinder, C (Fig. 16.2), surrounding the cathode. A smooth and (if possible) linear modulation characteristic is required for this, implying that the intensity of the spot on the screen should vary directly as the voltage applied to C. In other applications, where the spot, for example, has to be suppressed completely at certain times, it is necessary to arrange that a finite voltage on C, negative with respect to the cathode, should cut off the beam current completely. There are, therefore, two conditions which must be satisfied by a good c.r. tube electron gun. First, linear voltage control of beam intensity and secondly, a finite and reasonably small cut-off voltage suppressing the beam altogether.

Essentially, the control cylinder C behaves like the grid of a triode valve, and, as in the case of the triode, the closer C is to the cathode, the stronger is the control it exercises. Therefore, if fine variations in the spot intensity are required, the control of the cylinder is weakened by increasing its distance from the cathode. For many radar display tubes, where the beam must be cut off sharply, the control is strengthened by keeping the cylinder as close to the cathode as possible. By analogy with the triode valve such electron guns are known as low- μ and high- μ respectively. Also by analogy with the triode valve, the control cylinder is usually called 'the grid'.

The Screen

As has already been mentioned, the area of impact of the cathode-ray beam with the end of the tube is made luminous by coating the inner surface of the tube end with a certain powder. Any substance which becomes luminous when bombarded by electrons is known as a phosphor. Cathode-ray tube screens are made from a variety of phosphors, chosen on account of special properties appropriate to the particular application of the tube.

The phosphors used in the screens of c.r. tubes act in two distinct ways. When light is emitted during the bombardment of the material, the phenomenon is known as fluorescence, and all c.r. tube screens show this effect. When light is emitted after bombardment has ceased, the phenomenon is known as phosphorescence. All c.r. tube screens exhibit this effect also, but to a greater or lesser degree. Where the luminous spot produced by the bombardment is required to vanish quickly after the bombardment ceases, the persistence of the phosphorescence must be reduced to a minimum and the phosphor chosen accordingly. This is done where very rapid variations in time are to be dis-

played and the luminous trace produced by the moving spot is to be photographed. Where it is desirable that the luminescence should linger after bombardment has ceased, a longer period phosphorescent material is employed. This is done where the time variations are slower and particularly where the pattern produced by the moving spot has to be observed by the eye.

A screen phosphor for a c.r. tube comprises a semi-conductor, but two components are used in its manufacture, (a) a metallic oxide or combination of oxides and (b) an activator consisting of a pure metal, which is very often manganese. For example, the commonly used willemite screen is zinc orthosilicate, $\text{ZnO}(\text{SiO}_2)$ with a trace of manganese. The luminescent characteristics of the screen depend upon both constituents and upon the percentage of the activator present, so that a very wide range of phosphors is available. By properly choosing the screen, the luminescence may be made of almost any colour, and in the case of television tubes it is white. For laboratory use in c.r. oscilloscopes green is the preferred colour, since the eye is more sensitive to green than to any other colour. By making a second choice out of the selection of screens giving the same desired colour, the luminescence may be made to persist for a long or a short period. Radar display screens generally require a trace which persists as long as possible. In this case it is usual to employ a double screen. The first layer of this, furthest away from the glass end, is energized directly by the electron beam and emits strongly during bombardment and immediately after it. The second layer, beneath the first and next to the glass end, is very little energized by the electrons themselves, which cannot penetrate to it, but is mainly energized by the light emitted by the first layer. The second layer is phosphorescent with a long persistence; it may be said to store up the energy given to it quickly by the first screen and to re-emit this over a longer period.

Using such a double screen, the emitted light may be made to persist so that its intensity falls to 1% of its maximum in a time of the order of 500 milliseconds. A short-persistence screen falls off by the same fraction in a time of the order of 10 microseconds. A medium- or long-persistence screen is generally used for oscilloscopic work, where visual examination is required. A short-persistence screen is sometimes required where a photograph of the trace is made or in colour television.

It is now necessary to consider the electrical characteristics of the screen. If a cathode-ray tube is made with the electrodes arranged as shown in Fig. 16.1, the electrons can only return to the cathode via the screen, and the positive potential applied to A_3 would also require to be applied to the end of the tube. This can be done in certain modern metallic screens, where the latter is bounded by an extremely thin aluminium film, for then there is a good conducting path for the bombarding electrons. But where the screen is not metallized, the screen material is a poor conductor, and some additional collector must be arranged. In the majority of tubes made for Service use the screen is not metallized, but the inside of the glass envelope is coated with colloidal graphite from A_3 (Fig. 16.1) all the way to S. This coating forms the collecting electrode for the electrons, and the mechanism by which electrons are transferred from the screen to the collector is secondary emission.

It may be remembered from Chapter 1 that the number of secondary electrons emitted by an electrode is a function of the speed of the primary bombarding

electrons. Expressing this in another way, the graph of Fig. 16.3 is obtained, where the electrode is at potential $+V$ relative to the cathode, and the ratio of the secondary electron current to the primary electron current (I_s/I_p), is given as a function of V .

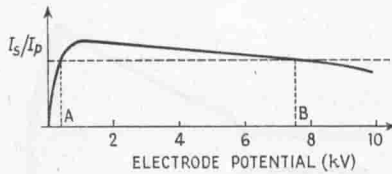


Fig. 16.3. Variation of secondary to primary electron-current ratio with voltage.

When V is small, so that the electron bombarding speed, $\sqrt{2Ve/m}$, is also small, very few secondaries are produced. As V increases, the number of secondaries per primary increases rapidly, reaching a one to one ratio in the case of the average c.r. tube screen at about two or three hundred volts. The ratio, I_s/I_p , increases still further to a maximum of about 1.3 at a thousand volts or so, and then gradually decreases. A one to one relationship is again established at a voltage in the neighbourhood of 7,000.

Now suppose that a screen is insulated from, but close to, a collector electrode in the form of a graphite coating over the tube surface. If the screen potential is initially between A and B (Fig. 16.3), when the beam strikes the screen, the potential of the latter will tend to increase, since it loses more electrons than it gains. But if its potential rises above that of the collector, it will tend to collect its own secondaries, thus reducing its potential again. This is the usual balancing action which was studied in the case of the tetrode valve, and it causes the screen to remain effectively at the same potential as the collector. If, however, the collector potential is increased above B (Fig. 16.3), the secondary to primary ratio falls below unity, and the screen tends to become less positive. It therefore 'sticks' at this potential no matter how much the collector voltage is increased. This is a fundamental limitation on the final-anode voltage of a simple c.r. tube.

It should be noted in passing that there is a potential below which the screen will not collect electrons at all. If the potential of the screen is less than A, the 'blocking potential', bombardment causes it to become more negative and so to turn the electrons back before they reach it.

As mentioned above, modern tubes are sometimes metallized. After the screen has been deposited in the usual fashion, a mirror-like thin film of aluminium is laid down over the whole surface. Thus the end of the tube, moving from the inside to the outside, consists of (a) an aluminium layer, (b) one or more layers of phosphor, (c) the glass envelope. The aluminium layer is so thin that it does not seriously absorb the electron beam, but it is sufficiently smooth and mirror-like to reflect the light produced by the beam. This increases considerably the light output of the tube. It also allows the screen to conduct away the electronic current without employing secondary-emission effects. Consequently, the high-voltage limitation produced by 'sticking' is removed and a tube may be made to work at a voltage of 30 kV, if desired. There are other secondary advantages of using an aluminized screen which will be mentioned in succeeding paragraphs.

It should also be mentioned at this point that modern large television-viewing tubes are often made with a steel body in place of a glass one, the glass screen being sealed to the steel at the end. A section of such a tube is shown in Fig. 16.4, indicating where the glass and metal parts of the envelope lie. The

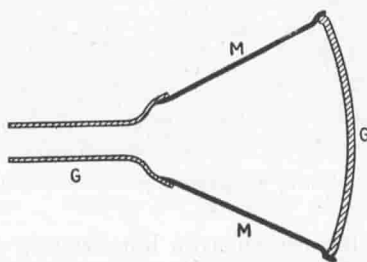


Fig. 16.4. Section of a 'metal' tube. Only the part M is of metal and is sealed to glass at its ends.

advantages, apart from mechanical strength and ease of manufacture, lie mainly in the magnetic screening obtained by the use of a steel body surrounding the beam. Tubes with screens of 16 inches diameter are made in this way and the optical characteristics are better than can be achieved using the all-glass envelope. It is of interest to note that the air pressure gives rise to a force of about $1\frac{1}{2}$ tons acting on the face of a 16-inch tube.

Before leaving the subject of screens for c.r. tubes a most important point must be discussed, which is of considerable interest to Service users. Although a c.r. tube is evacuated to a pressure less than 10^{-9} of a metre of mercury before it is sealed, there are always a certain number of gas ions present, particularly near the cathode. The negative ions generated there are accelerated like the electrons and, if nothing is done about it, arrive at the screen with the electronic energy. Such massive particles can do considerable damage to the screen material in course of time, producing an ion spot which ultimately emits no light under bombardment. This is a real problem in high-voltage tubes and two different methods have been adopted to solve it.

First, returning to the aluminized screen, the aluminium layer has a protective effect so far as ion bombardment is concerned. Certain aluminium film thicknesses which pass the electron beam stop gas ions almost completely. So a secondary advantage of the aluminized screen is the reduction of the danger of ion spots. The second method of eliminating ion spots is much more complicated and takes the form of a modification to the electron-gun structure, usually called an ion trap.

When a charged particle in a c.r. tube is deflected by means of an electric field, the deflection is independent of the mass of the particle. This may be easily shown as follows:—

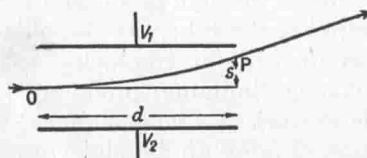


Fig. 16.5. Illustrating electric deflection of the c.r. beam.

Let a stream of particles of mass m and charge e enter the electric field, E , between two parallel electrodes at potentials V_1 and V_2 (Fig. 16.5). Then the force acting on the particle at right angles to its direction of motion is eE , and this is its mass $\times$ acceleration. Hence, where α is the acceleration,

$$m\alpha = eE \text{ or } \alpha = \frac{e}{m} E$$

If the force remains constant until the particle leaves the electric field at time t , the particle will move a distance $s = \frac{1}{2}\alpha t^2$ at right angles to its original direction.

$$\text{Therefore } s = \text{deflection} = \frac{1}{2} \frac{e}{m} Et^2$$

But the particle of mass m derived its original speed from the accelerating potential V between the cathode and the final anode. Therefore its speed along the axis is $v = \sqrt{2Ve/m}$. Now the deflection is small, and therefore, if the length of the deflecting plate is d , the particle will traverse d at speed v , whence $t = \text{time between the plates} = d/v$ and $t^2 = d^2/v^2 = d^2/(2Ve/m)$

$$\begin{aligned} \text{Therefore the deflection } s &= \frac{1}{2} \frac{e}{m} Et^2 = \frac{1}{2} \frac{e}{m} E \frac{d^2}{2Ve/m} \\ &= \frac{Ed^2}{4V} \end{aligned}$$

This expression for s does not contain m showing that the deflection will be the same for all particles originating at the cathode and carrying charge e , no matter what their masses may be.

It is quite different when magnetic deflection is considered, for there the deflection due to a field B at right angles to the beam results in a deflecting force evB and gives rise to movement in a circle of radius r , where

$$m \frac{v^2}{r} = evB \text{ or } r = \frac{v}{Be/m}$$

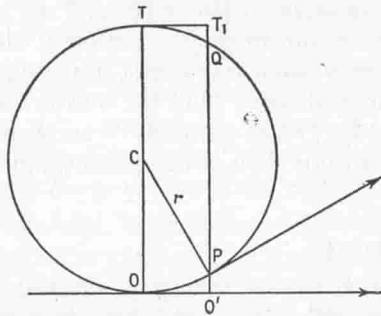


Fig. 16.6. Diagram to show the action of a magnetic field in deflecting the c.r. beam.

In Fig. 16.6 the beam is moving along OO' and comes into the magnetic field B at right angles to the paper at O . The field continues to act until the electron arrives at P , where $CO = CP = r$ in the above equation. Then $O'P =$

deflection = s , and $OO' =$ distance over which field acts = d . By producing OC to the point T where it cuts the circle, and producing $O'P$ to Q where it cuts the circle and to T' where it cuts the tangent at T , it may be seen that

$$\begin{aligned}(OO')^2 &= O'P(O'Q) \\ &= O'P(2r - O'P) \\ \text{or } d^2 &= s(2r - s).\end{aligned}$$

But s is usually small compared with r , and thus to a first approximation, $d^2 = 2rs$

$$\text{Substituting for } r = \frac{v}{Be/m}, \quad d^2 = \frac{2vs}{Be/m}$$

Also, as before $v = \sqrt{2Ve/m}$,

$$\text{whence } d^2 = \frac{2s\sqrt{2Ve/m}}{Be/m} = \frac{2s}{B} \sqrt{\left(\frac{2V}{e/m}\right)}$$

In this case s is *not* independent of m , so particles of different masses are deflected by different amounts.

The ion trap makes use of the fact that a magnetic deflecting field differentiates between ions and electrons to deflect the electron beam away from the negative ion beam. A method of common occurrence will now be described. The idea is a fairly obvious one. The electron gun is bent as shown in Fig. 16.7. The

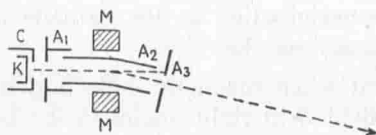


Fig. 16.7. The elements of a simple ion trap.

electrodes of the gun are as before, there being three anodes. But the second-anode cylinder is bent at an angle of about 10° and the electron beam is correspondingly bent by means of the magnetic-deflecting coils shown at M . The ions, deflected through a very small fraction of this angle owing to their greater mass, strike A_2 and are neutralized. Thus the ions in the beam produced near the cathode are eliminated. Other methods of accomplishing the same result involve the use of an electrostatic lens cut at a slant and an associated magnetic field.

Electric Deflecting Plates

An elementary discussion of the theory of electric deflection has already been given in connection with ion traps. Fig. 16.5 shows how an electron beam may be deflected and the accompanying theory shows that the total deflection, measured in units of length at right angles to the undeflected beam direction, is $s = Ed^2/4V$, where E is the electric-field strength between the plates, d is their length parallel to the beam and V is the final anode potential relative to the cathode, determining the energy of the electrons as they enter the deflecting field.

Redrawing Fig. 16.5 as Fig. 16.8 to show where the beam impinges on the screen, T, the actual screen deflection is $O'Q$, and the deflection $s = Ed^2/4V$ is PS. Producing the straight line QP to cut the axis OSO' at R, the deflection at the screen, $QO' = PS \times RO'/RS$.

Let $RO' = D$, then $QO' = Ed^2 D / \{4V(d/2)\} = EdD/2V$.

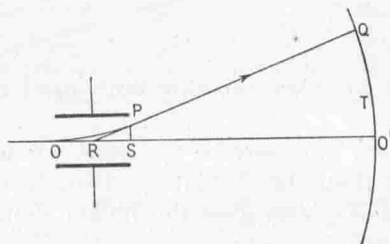


Fig. 16.8. Fig. 16.5 is here re-drawn on a smaller scale to include the screen of the tube.

Hence the deflection is

- (i) Proportional to E , the electric-field strength
- (ii) Proportional to d , the length of the plates
- (iii) Proportional to D , the distance between the plates and the screen.
- (iv) Inversely proportional to V .

These results are almost intuitively obvious, but they have important repercussions on the design of a c.r. tube.

It is often the object of the designer to make a tube which has a high deflection sensitivity, so that a small voltage applied across the deflecting plates will give a considerable deflection. To do this he must increase d or D , or decrease V . To increase d or D beyond a certain point makes the tube clumsy and in the case of increase of d may introduce an undesirably high capacitance, since the plates form a parallel-plate capacitor. To decrease V beyond a certain point reduces to an undesirable degree the light emission from the screen, for the light over a considerable range varies directly with the electronic energy at impact. Taking these two factors together, a compromise must always be reached, and in practice this leads to a figure of the order of one-tenth of a millimetre deflection at the screen for a change of one volt in the potential difference between the deflecting plates. This figure is characteristic of most standard cathode-ray oscilloscopes, used for measurement purposes.

Since $QO' = EdD/2V$, and assuming that $E = V_x/a$ where V_x = voltage between the plates and a = their distance apart, then $QO' = V_x dD/2Va$. The deflection sensitivity of the c.r. tube is defined as $dD/2Va$. As mentioned above, this is rarely greater than 0.1 mm/volt for a standard tube.

If a c.r. tube is to be operated entirely by deflecting voltages, two pairs of deflecting plates are required, giving deflection in two directions at right angles. These two pairs of plates cannot be placed at the same distance from the screen, since their electric fields would then interact in an undesirable manner. The plates are therefore placed as shown in Fig. 16.9, one pair being nearer the cathode than the other. Often the plates are referred to as the X-plates and

Y-plates, since they give deflection parallel to x and y rectangular co-ordinates on the screen.

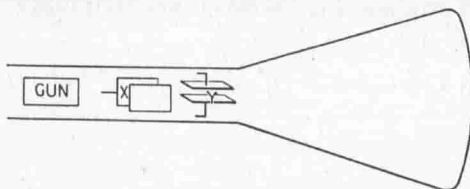


Fig. 16.9. Sketch of c.r. tube showing both pairs of deflector plates.

The deflection sensitivity can only be the same for the two pairs of plates if the Y-plates are longer than the X-plates. Usually no attempt is made to obtain equality and the X-plates give the higher deflection sensitivity.

Nowadays, deflection plates are shaped to allow a greater deflection to be used with plates which are fairly close together. The plates, instead of being plane and parallel, are shaped as shown in Fig. 16.10, so that electrons deflected



Fig. 16.10. In modern c.r. tubes the deflector plates are shaped in the fashion shown here so that the beam does not strike them when the deflection is large.

through a large angle do not strike the plate. Although the electric field is weakened owing to the increase in the distance between the plates, this can be counteracted by increasing their length while maintaining a reasonable value for their capacitance.

Double-Beam C.R. Tube

A most important application of electron-optical principles to electric deflection is to be found in some double-beam oscilloscopes. A single-cathode gun structure is used in the tube, but the beam is so split that two almost entirely independent indicator beams are obtained. The principle of operation may be described as follows:—

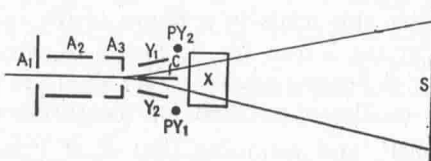


Fig. 16.11. Elements of a double-beam c.r. tube. The beam is split into two by the plate C.

The beam emerges from a conventional gun at A₃ (Fig. 16.11), converging towards the screen. Immediately in the path of the undeflected beam is a plate C which splits the beam into two. One half is deflected by the voltage between C and Y₁, the other by the voltage between C and Y₂. Both beams are then equally affected by the X-plates placed as shown. Since there is a residual electric field acting on both beams between the Y-plates and the

X-plates, two pins, marked PY_1 and PY_2 are placed outside the beams and are connected to the opposite deflecting plates. That is, PY_1 is connected to Y_1 and PY_2 to Y_2 . The additional fields produced by these pins almost exactly compensate for the electrostatic leakage from Y_1 to Y_2 in the space between the Y-plates and the X-plates.

By this elegant means the operator is given a c.r. tube in which he has two beams available and can make direct visual or photographic comparisons between electrical phenomena. Other methods of doing the same thing exist, employing two completely separate cathode-gun-deflector-plate systems. These are more costly and more complicated to operate.

A word should be said about some less common electric systems. It is often convenient to deflect the spot on the screen radially from the centre rather than laterally. In short, it is sometimes useful to employ polar co-ordinates (r, θ) instead of Cartesian co-ordinates (x, y) . To do this, radial-deflecting electrodes have been invented. These take several forms, the simplest of which is a small cylindrical electrode, R, running from the centre of the screen axially down the tube (Fig. 16.12).

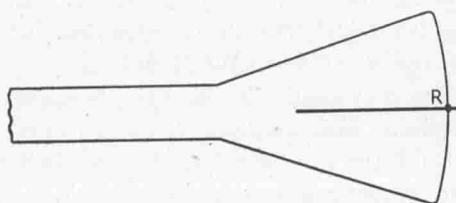


Fig. 16.12. Central electrode R for radial deflection.

A potential negative with respect to the final anode applied to R produces a radial deflection of the beam. This has useful applications in certain radar-display systems.

Magnetic-Deflecting Coils

Theoretically, anything which can be done by electric-deflecting plates can be done by the proper arrangement of coils carrying the appropriate currents, and since such coils are external to the tube and can therefore be moved, magnetic deflection offers certain apparent advantages over electric deflection. The elementary theory has already been given with regard to the action of ion traps and Fig. 16.6 shows that in a uniform magnetic field the beam is deflected by an amount

$$\begin{aligned} PO' &= s = d^2 B / 2 \sqrt{\left(\frac{2V}{e/m} \right)} \\ &= \{ d^2 B \sqrt{(e/m)} \} / \{ 2 \sqrt{(2V)} \} \\ &= \{ d^2 B \sqrt{(e/m)} \} / \sqrt{(8V)} \end{aligned}$$

Extending Fig. 16.6 to the screen, Fig. 16.13 is obtained, where QS is the deflection at the screen, and $RS = D$ is the approximate distance of the deflection coil from the screen. Then $QS = PO' \times RS/RO' = PO' \times D/(d/2)$

$$= \frac{dDB \sqrt{(e/m)}}{\sqrt{(2V)}}$$

Hence the deflection is

- (i) Proportional to B , the magnetic-flux density
- (ii) Proportional to the distance over which the magnetic field acts.
- (iii) Proportional to D , the distance of the coil from the screen.
- (iv) Inversely proportional to $\sqrt{V}$.

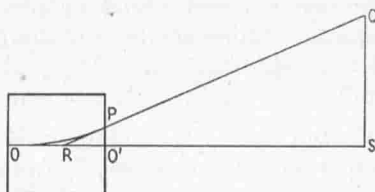


Fig. 16.13. Illustrating deflection with a magnetic field.

As already observed, the deflection varies with the mass of the particle, m , so that, if magnetic deflection is employed, there will be as many spots on the screen as there are carriers of the negative charge, e . In practice this means that, unless an ion trap is used, there is an ion spot near the centre of the screen. This rapidly burns away the screen at that point.

To increase the deflection sensitivity without seriously affecting the light emission the only possibility is to increase D or d . The difficulties associated with each of these possibilities are design problems which cannot be discussed here. It should be noted, however, that the correct position for a magnetic coil which gives large deflections is determined by the flare on the c.r. tube. As indicated by Fig. 16.14, if the coil is at X' , the electrons will tend to strike the wall of the tube. If it is at X , the flare allows the electrons to proceed to the screen.

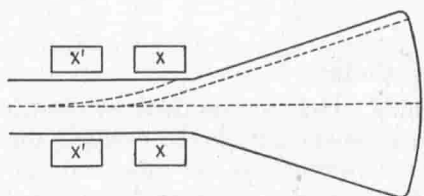


Fig. 16.14. The correct position for a deflector coil is at X ; if it is placed at X' the beam will hit the side of the tube as shown when the deflection angle is large.

In contradistinction to the electric-deflection system, it is possible to mount coils at the same distance from the cathode to produce deflections in two dimensions. If a core of magnetic material is not employed, it is usual to form the coils, each to a saddle-shape, fitting the cylindrical surface of the tube, but capable of being moved. The structure of a typical magnetic deflection system is illustrated by Fig. 16.15. Each turn of each coil has the shape shown in diagram (a), where the arrows indicate the direction of flow of the current, giving a saddle-shaped coil. Two of these, placed opposite each other on the neck of the c.r.t. appear in section as shown in (b) and give rise to a magnetic field in the tube as shown by the arrows. If two-dimensional deflection is required, the coils are arranged on an iron core as shown in (c), where X

refers to one pair of coils and Y to the other. By suitably adjusting the area and number of turns for each pair of coils, the x and y direction sensitivities can be adjusted as required. The complete set of coils with their mounting is known as the *yoke* which can be adjusted on the neck of the tube to the most favourable position.

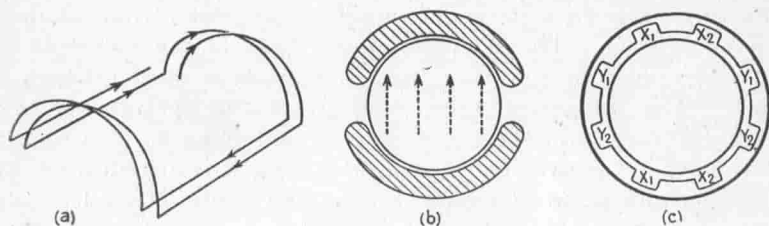


Fig. 16.15. Form of deflector coils. Each coil has its turns of the shape shown in (a) and an end section of a pair for horizontal deflection is shown at (b). Two pairs of coils are used for the two directions of deflection and usually have an outer surround of iron. Sometimes the iron ring is slotted to contain the windings, as in (c).

Many forms are taken by magnetic deflection yokes. In certain cases, where high-frequency response is important, dust-cores or cores made of a ferrite (notably Ferroxcube) have been employed. The core may take the form of a square (Fig. 16.16) round the neck, each side of which carries a coil. It should be noted that in this case the magnetic field due to coils X, X' is as shown by

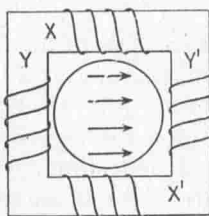


Fig. 16.16. A deflector-coil assembly in which the coils are wound around the limbs of a rectangular iron core is sometimes used.

the arrows across the neck of the tube. When the arrangement shown in Fig. 16.16 is employed, care must be taken with the windings to ensure that the currents induced in X and X' due to currents in Y and Y' are in opposition, so that there is no effective interaction between the coils. This is only a question of the direction taken by the currents. Coils, X, X' are in series as are coils Y, Y' . The arrangement is not much used owing to its low efficiency. A most important feature of the magnetic deflecting coil is that it may be rotated mechanically and may therefore reproduce on the c.r. tube the angular position of some external component (such as an aerial) which is directly coupled to the deflection coils. This is the principle employed in the simple Plan Position Indicator used in radar. The direction in which the spot on the screen is deflected from the centre represents the bearing of the rotating aerial, while the distance of the spot indicates the range of the target.

Particular Uses of Cathode-Ray Tubes

As will be seen from the foregoing, the possible modifications of the c.r. tube to suit different applications are very numerous. Indeed, if a table were drawn up, showing the possible combinations of electron guns, deflection systems and screens, it would contain hundreds of different tubes, each of which would have some advantage in a particular application. But certain tendencies are now well established. The television tube tends to use magnetic focusing, magnetic deflection, an ion trap, and a composite screen which is often aluminized. The measuring tube, used in oscilloscopes, tends to employ electrostatic focusing and deflection with a screen giving a green trace and the maximum deflection sensitivity. Quite often, too, it is a multiple-beam tube, so that comparisons between screen pictures can easily be made.

With certain modifications the cathode-ray tube can be transformed into a storage tube, a device into which an electrical signal may be fed and stored and out of which the signal can be obtained at some subsequent time. The only example of this which will be discussed here is one which goes under the trade name of the Graphecon. A simplified diagram of a modern form of this is shown in Fig. 16.17.

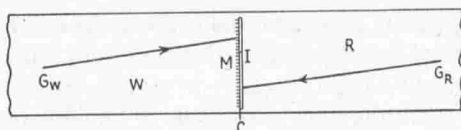


Fig. 16.17. Simplified form of the Graphecon storage tube.

First, let us consider the left-hand side of the figure. This comprises a device whereby a cathode-ray beam, originating at a 'writing gun', G_W , produces a distribution of electric potential on a target, MI, corresponding to the intensity of the beam when it impinges upon the target. The beam is made to 'scan' the target by a two-dimensional deflection system. Usually, this consists of magnetic coils. But the intensity of the beam at the target is varied according to the signal which is being investigated. For example, if the writing gun is building up a television picture, the intensity of the beam is modulated to give the appropriate picture intensity at the appropriate part of the scan. But instead of obtaining a picture from the target, the latter has imprinted on it a corresponding variation of the potential. How this is done will be considered subsequently. Now turning to the right-hand side of the figure, this comprises a device whereby the potential distribution on the target may be 'read off', the reading taking the form of electrical signals appearing at the collecting electrode, C. Again the beam (this time the reading beam from the reading gun, G_R) scans the target synchronized with, but not in phase with, the writing scan. When the beam impinges on the target, MI, it produces a signal at C, the magnitude of which is proportional to the potential at that part of the screen. The important feature of the device is that the reading can be done many seconds after the writing, and indeed the pattern may be 'read off' many times.

The explanation of the action at the target is in terms of the secondary-emission properties of an insulator. The target comprises a very fine metal

mesh, M (about 500 threads to the inch), which is virtually transparent so far as electrons are concerned, and which is the base upon which is built a thin layer (about 10^{-5} cm thick) of a good insulator, I. When the writing beam falls on the target, it causes the insulator to become negative with respect to the mesh, for the writing beam employs electrons having energies above 6,000 electron-volts, and for this energy the secondary-emission ratio of the insulator is less than unity. The reader should compare this statement with the explanation given earlier in the chapter of the secondary-emission action at an ordinary c.r. tube screen. As a result, when the writing beam has passed on, the potential of the insulator at each point is negative and the negative potential is very roughly proportional to the beam intensity up to a certain saturation value. The 'reading' beam employs electrons of much lower energy—usually about 1,000 electron volts; and so when it impinges upon the screen, it tends to bring the insulator back to equilibrium—to the same potential as the mesh. The transient current produced by this change in the charge on the insulator appears at the collector electrode which is connected to the mesh, and forms the output signal. By using a weak reading beam and a high insulator-to-mesh capacitance, the signal may be read off many times before it is ultimately erased. Also, since the insulation is moderately good, the signal may be stored for minutes before being read.

Applications of such a storage tube are too numerous to mention. It may, for example, be used to store and to distribute radar information: it may be used in computers: it may be used in certain forms of radio communication. Its storage time is adequate for almost any purpose.

Certain other special forms of c.r. tube which involve photo-electric action will be discussed in the next chapter. Since the most important of these also involves a built-in secondary-emission amplifier, this also will not be discussed at the moment.

A picture tube which was mentioned in Chapter 14 and which is used in certain radar equipment is the skiatron. The principle involved in the skiatron is the production of a persistent and intensified picture on a screen due to the action of a cathode-ray beam.

As stated in Chapter 14, the action of a beam of electrons on a screen consisting of potassium chloride crystals is to produce certain changes in the lattice which result in a change in optical properties. This change is observed as an absorption of certain frequencies in visible radiation transmitted through the screen. If, therefore, a tube is arranged as shown in Fig. 16.18, so that the screen, S, may be observed by transmitted light and a picture of it is thrown on to a projection screen, P, by means of lenses, a controllable contrast and persistence may be obtained. The contrast depends upon the beam intensity

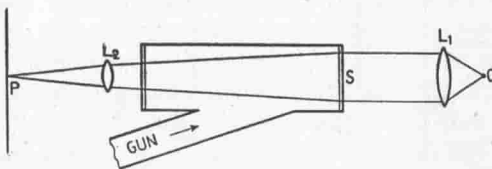


Fig. 16.18. General form of the skiatron in which the light transmission of the screen S is varied by the electron beam.

and the screen thickness, while the persistence is largely a function of the temperature. The tube is used to produce large radar pictures and may be regarded also as a simple short-term storage device.

Distortion

There are many types of distortion and defocusing effects associated with c.r. tubes which may or may not be of importance depending upon the application. Certain simple distortions are associated with the screen itself. First, there is halation. The screen is on the inside surface of the glass. The outside surface of the glass reflects part of the light back into the tube. If total internal reflection occurs, secondary fluorescence is produced in the screen, giving rise to a halo round the central spot. Fortunately, total internal reflection only occurs when the angle of incidence at the glass surface is large, and so the halo is usually weak, but it can have an appreciable effect on the contrast in certain cases.

Secondly, if the screen is not uniform in material or uniformly thick, both the intensity and size of the spot will tend to vary from place to place. This is not usually serious. Thirdly, if the screen is flat, a beam which is focused at the centre of the tube will be slightly out of focus near the periphery. For a fine focus all over the tube face, the latter should be a part of a sphere having its centre somewhere in the vicinity of the gun structure. This is rarely the case in tubes with large diameter screens, where the gun is designed to give an appreciable depth of focus.

The electron gun itself may introduce distortion. If the elements of the gun are not properly aligned, the spot will not have circular symmetry. The size of the spot is, of course, directly controlled by the potentials of the gun electrodes. But in addition to this, the exact position of a magnetic focusing coil can affect the spot size appreciably. It is as well for the operator to find the best position, remembering that every time the coil is moved, the current required to produce the best focus is changed. Deflection defocusing is a cause of several distortions. If the magnetic field due to a deflection coil is not uniform over the cross-section of the beam, the spot is non-circular. Elliptical spots are often caused by such non-uniformity. 'Pin-cushion' and 'barrel' distortion are names given to certain patterns obtained with faulty electron-optical systems or with flat screens. If a deflection system is arranged to produce movements of the spot along two directions at right angles to each other, it is possible to imagine voltages applied to the deflecting coils or plates to produce a pattern on the screen corresponding to (a) of Fig. 16.19. This

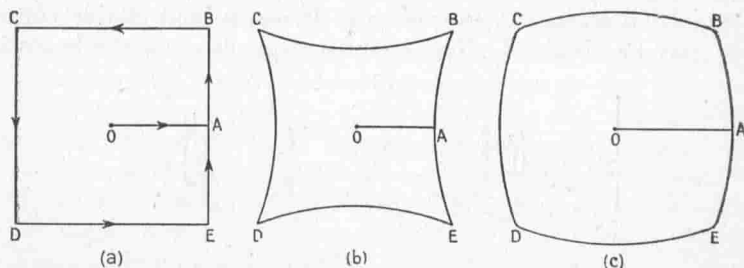


Fig. 16.19. Ideal rectangular trace on the screen of a tube (a). It takes the forms (b) or (c) if pincushion or barrel distortion is present.

is a square which could be produced by a gradually increasing x -deflection followed by a constant x and increasing y , followed by a constant y and decreasing x , etc. In certain cases when this procedure is followed, a pattern like (b) of Fig. 16·19 is obtained. This is a 'pin-cushion' distortion and indicates that (if the remainder of the system is perfect) the fields through which the beam has passed are non-uniform across the section of the tube, but that the non-uniformity is radial. 'Barrel' distortion is shown in (c) of Fig. 16·19. This is again due to radial non-uniformity of the opposite type. All types of distortion can be shown to have optical analogues and so we have spherical aberration and astigmatism. Chromatic aberration has no electronic analogy in c.r. tubes for the wavelength of light is analogous to $1/\sqrt{V}$ in the case of the electron beam, where V is the accelerating potential. In the c.r. tube V is well-defined and the wavelengths of the electrons are contained within very narrow limits.

The Magic Eye or Tuning Indicator

To conclude this chapter some mention must be made of a device which is not usually considered a cathode-ray tube, although it actually is a special form of one. This device is the so-called Magic Eye and it comprises a simple combination of a cathode-ray tube and a triode. It is used as a simple visual indication of the strength of a signal.

The simplest form of the tube is shown in Fig. 16·20, which is an axial section, the axis being shown as OO' . The cathode, K , is cylindrical with a non-

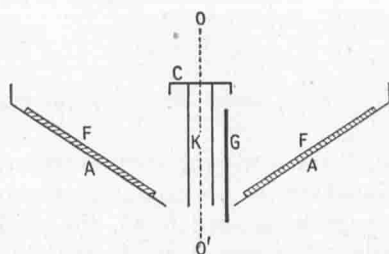


Fig. 16.20. Section through the electrode system of a magic eye.

emitting cap C, at the viewing end. The anode, A, is conical in shape and its inner surface is coated with a fluorescent material. The rudimentary grid is simply a single rod, G, placed near the cathode. The electrode system is therefore cylindrically symmetrical except for the grid. Imagine that the cathode is emitting and a voltage is applied to the anode. Clearly there is a certain voltage which, when it is applied to the grid, does not disturb the electrical field, so that a symmetrical distribution of light is obtained looking from O. This particular grid voltage is positive with respect to the cathode. For more negative voltages the grid repels electrons and casts a shadow on the anode. This in consequence produces an asymmetrical sector of light when the anode is viewed from O. By careful design it is possible for a swing of grid potential of about 50 volts to vary the sector of light by 100° . Hence, if the tube is biased in the absence of a signal to give 100° of shadow on the anode, a voltage derived from the signal will cause the shadow to disappear when its value is about 60 volts. This enables the observer to tune the device to maximum signal.

Modifications of this indicator may contain two or more grid wires. These can be used either to intensify the shadow or to give a more complicated pattern on the anode. Fig. 16.21 shows some possibilities. Diagram (a) shows the use

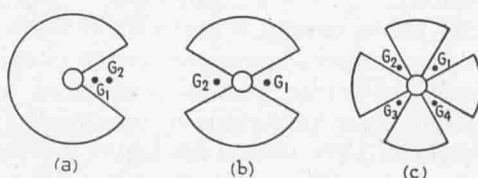


Fig. 16.21. Forms of shadow obtained with different arrangements of grid rods.

of two grids, G_1 and G_2 , to intensify the shadow. Diagram (b) shows the use of two grids opposite each other to obtain a double pattern. Diagram (c) shows the use of four grids to obtain a four-spoked pattern.

The magic eye just described is usually combined with a triode amplifier valve in the one envelope. The reason for this is that a potential swing of the order of 60 volts is not readily available in a radio set. It is much more convenient to operate the indicator from a control of a few volts.

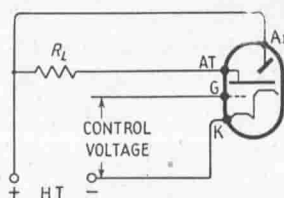


Fig. 16.22. Circuit used with typical magic eye.

As shown in Figure 16.22, the grid of the actual tuning indicator is connected to the anode of the associated triode. K comprises two cathodes with a common heater. The control voltage is applied to G, the grid of the triode section, and the anode of the triode, AP, is connected internally to the control electrode of the indicator. The anode of the indicator, AI, is brought out to a separate connection. The circuit arrangements are shown in the figure. The load resistance, R_L , produces the voltage changes at AT.

CHAPTER 17

Photo-Electric Devices

PHOTO-ELECTRIC devices transform light energy into electrical energy and are used in several ways in radio equipment. For example, photo-cells, as they are called, can be used in protection gear to detect a change in light intensity and thereupon to actuate relays cutting off the power supplies. They can be used in conjunction with lamps of special construction to measure the high-frequency power in a radio transmitter and, most important of all, the photo-electric cell is an essential part of a television camera, transforming light signals into electrical impulses.

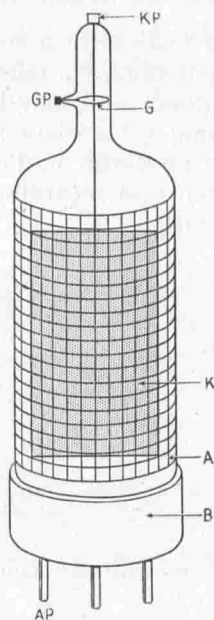


Fig. 17.1. General form of photo-emission cell.

Photo-cells are of two distinct types, differing entirely in their mode of operation. In the photo-emission cell light falling upon a metal surface produces the emission of electrons: this action has already been described in general terms in Chapter 1. In the photo-conductive cell light falling upon a thin layer of a metal, or a metallic compound, changes its resistance. All photo-conductive materials are semi-conductors and the action of the light on the semi-conductor has been described in general terms in Chapter 14. In both cases the light energy produces a change in the electric current flowing through the cell, but the two types are used in very different applications.

Photo-cells of the emissive variety are often combined with electron multipliers in the same glass envelope. An electron multiplier is a device in which secondary emission from a metallic surface is used to multiply the electron current. Secondary emission has been described in Chapters 1 and 16. The secondary-emission multiplier is becoming a most important electronic device, particularly where amplification of a small signal produced by photo-electric action is required.

Vacuum Photo-Emission Cells

The first type of photo-cell to be considered takes the form of a diode in an evacuated envelope. The simplified structure of a typical vacuum photo-cell is shown in Fig. 17.1. Essentially, the cell comprises an anode, A, consisting of a cylindrical mesh of fine wire, surrounding a rectangular cathode, K, the surface of which is light sensitive. The anode may be brought out to a pin, AP, in a valve base, B, while the cathode is generally brought out to a lead KP at the opposite end of the envelope. To avoid electrical leakage across the glass, two guard rings, G, in external and internal contact with the glass are placed near the cathode. The guard rings are earthed.

When such a cell is exposed to light of a suitable wavelength, the anode being maintained at a positive potential, V_A , relative to the cathode and earth, electrons are emitted by the cathode and may be drawn towards the anode. The characteristic curve, showing the variation of the cathode current, I_K , with the anode voltage, V_A , is as shown in Fig. 17.2, the scales being left arbitrary, since nothing has been said regarding the wavelength of the light or the material of the cathode surface.

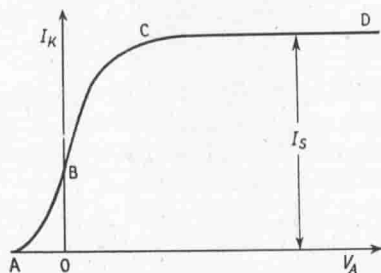


Fig. 17.2. Characteristic of photo-emission cell for constant light intensity.

The explanation of this curve may be given in terms of the general description of the photo-electric effect to be found in Chapter 1. Suppose, first, that there is no potential difference between the cathode and anode. The light, falling on the cathode, causes the emission of electrons, if the quantum of energy hf is greater than the energy required to release electrons from the metal surface; here h is Planck's constant in watt. sec. and f is the frequency in cycles/sec of the light radiation. Emission will occur if hf is greater than $e\phi$; where ϕ is the work-function of the metal in electron-volts and e is the electronic charge in coulombs. In Chapter 1 it is shown that for the metal caesium, where $\phi = 1.75$ electron volts, $e\phi = hf$ when the wavelength of the radiation is about $7,000 \text{ \AA}$. Therefore, for a caesium photo-cell one would expect photo-emission with any colour of visible radiation.

If the potential V_A is made negative, the electrons emitted by the light tend to be turned back towards the cathode and, as V_A is made more negative, fewer electrons attain the anode. Assuming that the kinetic energies of the electrons in the metal cathode are small compared with hf , which is true of the majority (see Chapter 1), and neglecting contact difference of potential, it will be impossible for electrons to travel to the anode against the opposing electric field when $V_A = -hf/e\phi$. Therefore $AO = hf/e - \phi$.

For a given light intensity the current increases rapidly as V_A changes from AO to zero and then becomes positive, and at a small positive value of V_A the current becomes stationary and does not change for further increase in the anode potential. This simply means that the total number of electrons emitted from the cathode under the action of the light is being collected at the anode, and this number is a measure of the light intensity. Let I_s = the saturation current produced by a light intensity L . Then I_s is proportional to L over a wide range of the latter, making the vacuum photo-cell an excellent photometer. Indeed, modern photometry relies almost exclusively on vacuum photo-cells.

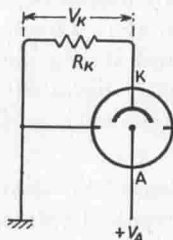
It has been said in a previous paragraph that the response of the cell to light of a given wavelength depends upon the surface of the cathode. Caesium is the pure metal having the smallest work function and it is, therefore, the metal which gives the best response at the red end of the spectrum. Each alkali metal gives its maximum response as shown in Table 17-1.

TABLE 17-1.

Metal	Cs	Rb	K	Na	Li
Max. Sensitivity in angstroms	5,400	4,800	4,400	4,200	4,100

It is, therefore, unsuitable to attempt to use a lithium cell in the near infra-red, although it would give a small response. Caesium responds reasonably well up to 6,800 angstroms which is not far from the theoretical value quoted in an earlier paragraph.

To obtain the maximum photo-electric current from a cell it is necessary to prepare the surface carefully. The maximum current in the case of a caesium cell is obtained from a surface of caesium on an extremely thin layer of caesium oxide on a silver base. The techniques employed in forming photo-sensitive surfaces are delicate and sometimes complicated. They will not be discussed further as they are a matter for the designer.

Fig. 17.3. Circuit of photo-cell used with a load resistance R_K .

To use a vacuum photo-cell for measurement it is only necessary to connect a resistance in series with the cathode. The circuit is shown in Fig. 17.3. The guard-ring, G , is connected to earth and to one end of the load resistance, R_K . The cathode is connected to the other end of the load resistance. The anode is connected to the positive potential, V_A . The potential V_K developed across the resistance, R_K , is a measure of the light intensity. A load line can be drawn across the characteristic curves as shown in Fig. 17.4, where the different curves correspond to different light intensities. Since I_K is only of the order of $1\text{--}10\ \mu\text{A}$, R_K can be very large ($10\text{--}50$ megohms). The voltage appearing across R_K can then be amplified in the conventional manner.

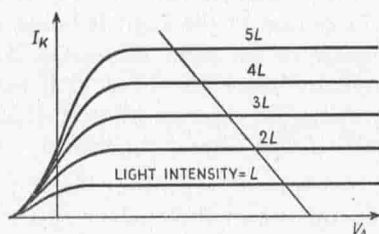


Fig. 17.4. Characteristic of photo-cell with a load line superimposed.

Two points should be mentioned about the vacuum cell. First, its bandwidth is extremely large. It will reproduce accurately variations in the light intensity at frequencies up to 100 Mc/s . Secondly, its sensitivity is limited by the noise due to the resistance, R_K . Since this is of the order of $10\text{--}50$ megohms, the noise voltage appearing across it, $V_n = \sqrt{(kTBR_K)}$, is also large. It will be remembered that k is Boltzmann's constant, T is the temperature and B is the bandwidth.

Gas-Filled Photo-Emission Cells

The maximum photo-electric current obtained from an evacuated photo-cell is quite small, being of the order of $50\ \mu\text{A/lumen}$. Since a lumen is the light power falling on one square inch of surface at a distance of 1 inch from a standard candle, a very high light intensity is required to obtain a current greater than $100\ \mu\text{A}$ from a vacuum photo-cell. For many purposes this small current is inconvenient, particularly where a light sensitive detector is to be used to actuate a switch.

To increase the sensitivity, cells are made which contain a small quantity of gas. The voltage between the cathode and the anode is then adjusted so that the photo-electrons, emitted by the action of the light, form pairs of ions while they are travelling towards the anode. This ionization by collision has already been discussed in Chapter 13, and it has been shown that the total current arriving at the anode may be many times the primary photo-electric current. The cell then comprises a photo-emissive surface and an ionization amplifier or multiplier.

A gas photo-cell does not possess the simple characteristic of its vacuum counterpart. Fig. 17.5 shows a typical volt-ampere curve. The variation of current, I , with voltage is similar to that observed in a vacuum cell when the potential difference between the anode and cathode, V , is small. But when V

becomes several times the ionization potential of the filling gas, the current increases, the value varying rapidly with V . An amplification factor of 10 can sometimes be obtained by using the cell at a high enough voltage. There is a limit to the amplification which may be obtained, for, if the voltage is increased sufficiently, a self-sustained discharge occurs and the current is then

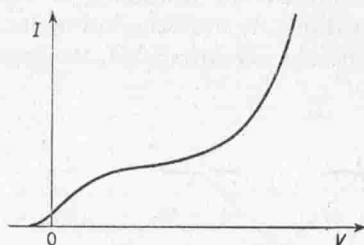


Fig. 17.5. Typical form of characteristic obtained with a gas-filled photo-emission cell.

independent of the photo-electric current. Indeed, when a self-sustained discharge has been initiated, it persists after the light has been cut off from the cell. Gas-filled photo-cells can therefore only be used over a limited range of voltages, and the greater the amplification at which a cell is operated, the more voltage sensitive does it become. As already mentioned, it is not usual to try to use a cell at an amplification greater than 10. Moreover, since the incidence of a high light intensity might initiate a self-sustained discharge unexpectedly, a gas-filled cell must never be operated without a high resistance in series with it. If the discharge is allowed to degenerate into a glow-arc or an arc, the cell is ruined.

Only certain gases can be used in filled photo-cells. Almost all gases are unsuitable because of their chemical action at the cathode surface. Indeed, only a rare gas can be used and of these argon is the most popular. Argon combines a relatively low ionization potential with a moderate atomic weight. The former makes for a high amplification factor, while the atomic weight should be small to avoid injury to the cathode surface due to bombardment with positive ions.

Gas photo-cells cannot be used at high frequencies, for the inertia of the gas ions makes their response slow compared with that of a purely electronic device. At frequencies above 10,000 c/s, gas cells begin to show a diminished response. The inertia of the ions shows itself not only in a lower response at higher frequencies, but in a failure to reproduce a wave form containing high-frequency components. This is what might be expected.

Summing up, the gas-filled cell is to be preferred so far as sensitivity is concerned and particularly at low frequencies. It is an excellent light-operated switch and can be made to trigger a relay directly. If, however, measurements are to be made or high-frequency effects are to be examined, the vacuum cell must be used.

Photo-Multiplier Cells

An alternative method of directly amplifying the photo-electric emission from a cathode is to cause the primary photo-electrons to produce secondaries

by the bombardment of a target. By suitably choosing the target material and the voltage, a secondary to primary ratio of 3 or 4 to 1 can be obtained. This idea has been in existence for a large number of years, but it is only fairly recently that photo-multiplier cells have become practical propositions. The lag between conception of the idea and its practical application has been largely due to the difficulties inherent in obtaining a stable secondary electron-emissive surface. A lesser difficulty has been in the development of a satisfactory electron-optical system to take advantage of the secondary-electron emission ratio.

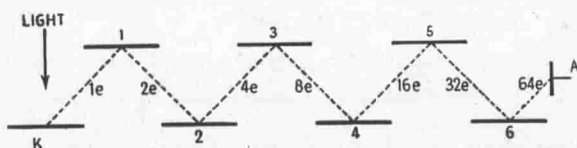


Fig. 17.6. Diagram illustrating the principle of photo-multiplication. Practical forms differ considerably.

A simple photo-multiplier cell to exhibit the principle, but not the practice, is shown in Fig. 17.6. Forgetting, temporarily, the electron-optics of the device, suppose that K is a photo-electric cathode on which light can fall as shown. If electrode 1 is at a potential of, let us say, 100 volts positive relative to K, then the photo-electrons emitted by the light travel to 1 and there produce secondary electrons. If 2 is 100 volts positive relative to 1, and there produce new secondaries. If the secondary emission ratio is r and there are n electrodes (6 in the diagram), the final current reaching the collecting anode, A, is the original photo-electric current I_0 multiplied by r^n . If r is 4, which is not impossible, and n is 6 as in the diagram, the amplification factor is $4^6 = 4,096$ approximately. This is an excellent amplification to be obtained from a single cell, and in practice it can be comfortably increased to 10,000 by increasing the number of electrodes.

As might be expected, there are many difficulties about the manufacture and use of such photo-multipliers, and the results obtained by means of them are not so striking as the figures in the last paragraph would lead one to believe. First, there is considerable difficulty in ensuring that the surfaces of the secondary emitting anodes remain stable and constant in output. The 'processing' of such multiplier cells calls for refined and elaborate vacuum techniques. Secondly, the shape and disposition of the electrodes is a problem. Clearly, the surfaces indicated in Fig. 17.6 would be useless, since electrons would tend to move from 1 to 3 and thence to 5. Several practical electrode arrangements have been designed, one of the most popular being accommodated in a cylindrical envelope of the conventional valve pattern. A third difficulty lies in the fact that the sensitivity of the device to unwanted effects is considerable. A small thermionic emission from electrode 1 appears as a considerable current at A. Similarly, positive ions falling on the cathode give rise to secondary electrons there which produce a multiplied current. Altogether, the 'dark' current due to unwanted effects is large. For a cell giving half an ampere per lumen, the dark current may be a third of an ampere, and this must be balanced out before true readings can be taken.

Despite this large 'dark' current, the photo-multiplier cell possesses a remarkable advantage over the conventional vacuum cell and amplifier. This lies in the low noise level inherent in the multiplier. A vacuum cell operates into a high load resistance (R_k of Fig. 17.3). This contributes noise to the circuit before any amplification has occurred. Using a photo-multiplier, the only noise to be expected is 'shot' noise, and so long as the 'dark' current is not too large, this 'shot' noise may well be much smaller than that due to a high resistance. In practice it is possible to produce a 40-db improvement in the signal-to-noise ratio, when the signal is small, by using a multiplier cell in place of a conventional amplifier. There is reason to believe that this figure will show a certain improvement in the future.

It should be noted in passing that the photo-multiplier is almost an aperiodic amplifier. Since the cell contains no gas, it is very frequency insensitive, at least up to very high frequencies (v.h.f.). An amplifier with a total gain of 40 db and a completely flat frequency response is a device of considerable potential value.

Television Camera Tubes

The Iconoscope. Combining the information on cathode-ray tubes given in Chapter 16 and that given on photo-electric devices in the present chapter, it is now possible to consider the operation of the camera tubes used in television. These rather complicated devices make it possible to transform a visual picture into a corresponding static pattern of electric charge and then into a series of electrical signals corresponding to the charge pattern. Three actions are involved in any camera tube: first, there is the transformation of light energy into electric current; secondly, there is storage of the electric charge in the form of a charge mosaic; and thirdly, there is the scanning of this mosaic by an electron beam to liberate a signal corresponding to each part of the pattern.

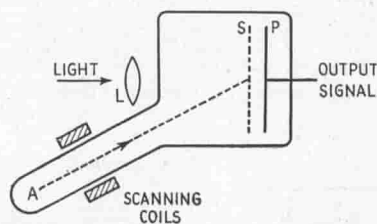


Fig. 17.7. General forms of iconoscope television camera.

The iconoscope is the simplest camera tube to explain, but in its original form it suffered from many defects. Fig. 17.7 shows the essential features of the elementary form. An image of the visual picture is formed by the lens system, L , on the screen, S . This screen consists of a fine mosaic of small photo-electrically active surfaces, each insulated from its neighbours. Each element of the mosaic emits electrons, and its final electric charge depends upon the intensity of the light falling upon it. Thus a replica of the visual picture is obtained in terms of electric charge density. Immediately behind the photo-electric screen is a metal plate, P , insulated from the elements of the screen, but forming with the latter a multiplicity of capacitors. The quantity of electricity on each element, due to illumination of the screen

depends not only on the intensity of the light, but on the capacitance due to the proximity of the plate, P. For, only if each elemental capacitance were infinite, would the voltage remain zero for all light intensities, and only under these circumstances would the charge be exactly proportional to the light intensity. With a finite capacitance, the voltage change produces a decrease in the photo-electric current which becomes more marked the greater the voltage. It is, therefore, desirable that the elemental capacitances should be as large as possible.

A conventional cathode-ray tube is included in the iconoscope at A, arranged so that the beam may be caused to scan the screen, S. As the beam traverses each element, it partially discharges it, and an electric impulse is transmitted to the plate, P. This impulse constitutes the outgoing electric signal from the iconoscope.

The most critical part of the iconoscope is the photo-electric screen and the method of manufacture is one requiring careful control. The mosaic consists of small globules of silver with a caesium surface on a mica sheet. The electrode P is a continuous conducting layer of silver on the opposite side of the mica sheet. The globules are very small compared with the cross-sectional area of the cathode-ray beam which scans the screen and consequently many globules are irradiated simultaneously by the electrons. Neglecting many complicating factors, the light falling on the mosaic gives rise to photo-electric emission so that each globule tends to become positive; it loses electrons. The scanning beam restores the original electron density, simultaneously giving rise to a pulse of current in the circuit connected to the plate, P. The complications are due to space charge and to secondary emission. Taking these into account, the iconoscope is not as sensitive as might be expected and is very insensitive at high light intensities. The variation in light intensity over the screen is not correctly reproduced by the electric signal.

The Image Orthicon.—One of the most interesting devices, combining photo-electric action, secondary-electron multiplication and electron-beam scanning, is the image orthicon which is a very sensitive television camera tube. Fig. 17.8 shows diagrammatically how the device operates.

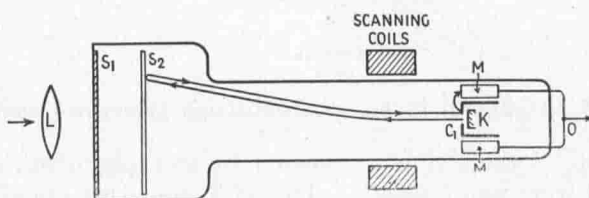


Fig. 17.8. Image orthicon television camera.

An image of the picture is formed by the lens system, L, on the continuous photo-electric surface, S_1 , on the end of the tube. Immediately opposite this photo-electric surface and not far from it is a double mosaic, S_2 , usually called the target. The photo-electrons emitted by S_1 move directly under the influence of a strong electric field to S_2 , where they produce secondary electrons on impact. This secondary-electron emission charges the mosaic and pro-

duces an electric charge picture corresponding to the visual picture. At the other end of the image orthicon is an electron gun from which electrons are directed towards the target. By means of magnetic coils the beam is made to scan the second mosaic of the target in the usual fashion. This scanning beam is comprised of low-velocity electrons: indeed the velocity is adjusted so that the elements of the mosaic reflect the electrons where no light has fallen on the corresponding part of S_1 , but collect the electrons where the light has been intense. There is therefore an electron current from the target towards the electron gun, and the focusing and scanning coils bring this return beam to the first collecting electrode, C_1 . This is the first electrode of a secondary-emission multiplier, M , the final electrode of which is the output terminal, O_1 so the reflected beam gives rise to a very much amplified current at the output.

The image orthicon is extremely sensitive and is easy to set up and to use. By substituting a screen sensitive to infra-red radiation for one sensitive to visual radiation, an infra-red camera tube is immediately available. This has certain Service applications.

There are other camera tubes in existence, used for special purposes. The principles are covered by the two discussed above, the iconoscope and the image orthicon.

Photo-Conductive Cells

When light falls on certain semi-conductors, the resistance of the surface layer is altered. This effect can be used to detect and measure light, by embodying the semi-conductor in a resistance bridge. It can also be used to produce a new kind of electric battery, activated by light.

It has already been explained in Chapter 14 that the absorption of light by a semi-conductor or insulator may move electrons in the crystal to a higher energy level which is not 'filled' and which may therefore behave as a 'conduction band'. There are many ways in which this may occur and the detail is not important. The classical example is the photo-conductivity of selenium, and selenium cells have been used for many years to detect slow variations in light intensity. But the resistance of selenium is high and is much affected by impurities: as in the case of other semi-conductors, notably germanium, it is only recently that the factors affecting its action have been fully realized. Recently other substances have been used as photo-conductive devices, particularly in the infra-red part of the spectrum. Very sensitive infra-red detectors can now be made and these are sensitive to radiation of very long wavelength. The possibilities of photo-conductive devices are only beginning to be explored, and it may well be that in the near future they will find many Service uses, particularly as d.c. generators.

Index

A

- Aluminized screen, 233.
- Amplification factor of triode, 41.
- Amplifier, class A, 46.
- class B, 46.
- class C, 46, 84.
- earthed grid, 135.
- more exact theory of, 71.
- push-pull, 74.
- push-pull, characteristic of, 76.
- push-pull, theory of, 75.
- Amplitude, demodulation, 108.
- modulation, 100.
- Angstrom unit, 10.
- Anode, a.c. resistance, 43.
- glow (discharge), 183.
- modulation, 103.
- transmitting valve, 120.
- Applegate diagram, 143, 152.
- Atomic, number, 1.
- shells, 2.
- structure, 1.
- A.T.R. switch, 192.

B

- Beam tetrode, 56.
- double, 117.

C

- Camera tube, 253.
- Capacitances, inter-electrode of triode, 47, 50.
- inter-electrode of tetrode, 53.
- Cathode, dark space (discharge), 183.
- glow (discharge), 183.
- heat shielded, 185.
- heaters, 22, 24.
- oxide coated, 19, 22.
- thoriated tungsten, 19, 21.
- transmitting valve, 119.
- tungsten, 21.
- types of thermionic, 20.
- virtual, 34, 57.
- Cathode-Ray tube, 229.
- aluminized screen, 233.
- beam formation, 230.
- "blocking" potential, 233.
- deflecting coils, 239.
- deflecting plates, 236.
- deflector coil assemblies, 241.
- double beam, 238.
- grid, 231.
- ion trap, 234.
- metal body, 234.
- modulation, 231.
- persistence of trace, 232.
- phosphor, 232.
- pincushion and barrel distortion, 244.
- screen, 231.
- "sticking" potential, 233.

C

- Characteristics, pentode, 60.
- screen-grid, 54.
- transmitting triode, 121.
- triode, 42.
- Circuits for use at u.h.f., 133, 140.
- Colour centres, 204.
- Conductance, conversion, 95.
- mutual, 42.
- Conduction band in a crystal, 200.
- Copper-oxide rectifier, 198, 205.
- Cross-modulation, 111, 113.
- Crystal-mixer, 95.

D

- Deflection in c.r. tubes, electric, 234, 236.
- magnetic, 235, 239.
- De-ionization time, thyratrons, 190.
- Demodulation, 108.
- Demodulator, amplitude, 108.
- frequency and phase, 111.
- linear, 110.
- square law, 109.

- Diode, 32.
- crystal, 208.
- current-voltage characteristics, 33.
- double, 115.
- dynamic resistance, 36.
- hot-cathode mercury vapour, 184.
- mixer, 95.
- noise, 177.

- Diode rectifier, action of, 36.
- space-charge limited, 36.
- static resistance, 36.

- Discharge, cold cathode, gas, 181.
- anode glow, 183.
- cathode dark space, 183.
- cathode glow, 183.
- characteristic, 182.
- Faraday dark space, 183.
- negative glow, 183.

- Disc-seal valve, 127.

- Distortion, calculation of, 70-73.
- due to curvature of characteristics, 67, 80.
- due to grid current, 77.
- due to use of two control grids, 80.
- in cathode-ray tubes, 244.
- of grid modulated wave, 112.
- of sine-wave signal, 69.

- Distribution of velocities, 7.
- in a gas, 7, 14.
- in a metal, 8, 14.

- Double, beam tetrode, 117.
- diode, 115.
- diode-pentode, 118.
- diode-triode, 118.
- triode, 116.

- Dynamic characteristic of triode, 72.

E

- Earthed-grid amplifier, 135.
- Electric deflection in c.r. tubes, 234, 236.
- Electrolytic tank, 223.

E

- Electron, 1.
 - beam formation in c.r.t., 230.
 - emission, 9.
 - energy levels, 203.
 - focusing by magnetic field, 224, 228.
 - gas, 7.
 - gun, 219, 222.
 - lenses, 219.
 - lens, defects in, 227.
 - lens, electrostatic, 220.
 - lens, magnetic, 225.
 - lens, shielded coil, 226.
 - optics, 219.
 - shells, 2.
 - sub-shell, 201.
 - volt, 6.
- Electronic switch, 107.
- Elements, table of, 3.
- Emission, 9.
 - auto-electronic, 9.
 - field, 11.
 - photo-electric, 10.
 - secondary, 10.
 - thermionic, 10, 16.
- Emitted electrons, velocities of, 12.
- Emitters, 18.
 - alkaline earth oxide, 18.
 - atomic film, 18.
 - pure metal, 18.
- Energy distribution in a gas, 14.
- Energy of a mass, 13.

F

- Faraday dark space (discharge), 183.
- Fermi-Dirac distribution, 8.
- Flying-lead valves, 28.
- Frequency changers, 88.
 - heptode, 92.
 - octode, 92.
 - triode-hexode, 93.
- Frequency modulation, 103, 107.
- Frequency demodulator, 111.

G

- Gas, discharge with cold cathode, 181.
 - inert, 2.
 - switches (TR and ATR), 192.
- Gas-filled, cold-cathode triodes, 192.
 - photo-emission cell, 250.
 - voltage-stabilizer, 191.
- Germanium, point contact rectifier, 199, 208, 214.
 - triode, 199, 208.
- Graphecon storage tube, 242.
- Grid, control, 38.
 - emission, 122.
 - modulation, 101.
 - modulation distortion, 112.
- Grid-bias, automatic, 78.
- Grid current, distortion due to, 77.
- Grounded-grid (see Earthed grid).

H

- Heaters for cathodes, 22-24.
- Heptode, frequency changer, 92.
 - mixer, 91.
- Hexode mixer, 90.

I

- Iconoscope, 263.
- Ignitron, 187.
- Image orthicon, 254.
- Impedance, input at high frequency, 125, 137.
- Intermediate frequency, 88.
- Ion, 6.
 - trap, 234.
- Ionization, in gases, 179.
 - in a triode, 180.
 - time in thyatron, 190.
- Isotopes, 2.
 - table of, 3.

K

- Keep-alive electrode, 187, 194.
- Klystron, amplifier theory, 145.
 - amplifier practice, 146.
 - cavity resonators, 146.
 - oscillator, 149.
 - oscillator modes, 150.
 - reflex, 152.
 - reflex modes, 153.
 - reflex, power v. frequency for, 154.
 - reflex, tuning of, 155.

L

- Langmuir-Child law, 35.
 - derivation of, 48.
- Leads, importance of at high frequencies, 124.
- Lens, electron, see Electron.
- Load line (pentode), 64.
- Load line (triode), 62.
- Local oscillation, 88.

M

- Magic eye, 245.
- Magnetic deflection in c.r. tubes, 235, 239.
- Magnetic focusing of electron beam, 224, 228.
- Magnetic lens, 225.
- Magnetron, effect, 156.
 - frequency pulling in, 165.
 - modes of oscillation of, 161.
 - multi-resonator, 159.
 - output coupling, 163.
 - rising-sun, 162.
 - split-anode, 158.
 - strapping, 161.
 - transit-time effect, 157.
 - typical operating figures, 165.
- Maxwell distribution, 7.
- Mercury-arc rectifier, 186.
- Mercury-vapour diode (hot cathode), 184.
- Metal, surface forces, 8.
- Metallic elements, 6.
- Metrosil, 199.
- Mixers, 88.
 - diode, 95.
 - heptode, 91.
 - hexode, 90.
 - pentode, 93.
 - silicon crystal, 95, 199, 208.
 - silicon crystal, back to front resistance of, 211.
 - silicon crystal, burn-out of, 213.
 - silicon crystal, characteristic of, 209.

M

- Mixers, silicon crystal, conversion efficiency of, 212.
- silicon crystal, equivalent circuit of, 211.
- silicon crystal, noise temperature of, 212.
- single electrode, 93.
- triode, 93, 176.
- two electrode, 88.
- Modulation, 100.
 - amplitude, 100.
 - anode, 103.
 - cathode-ray tube, 231.
 - control grid, 101.
 - cross, 111, 113.
 - frequency and phase, 103, 107, 114.
 - pulse, 107.
 - suppressed carrier, 106.
 - suppressor grid, 102.
- Modulator, balanced, 105.
- Multiplier, photo, 251.
- Mutual conductance, 42.

N

- Negative glow (discharge), 183.
- Neutrons, 1.
- Noise, 29, 168.
 - diode, 177.
 - resistance of pentode, equivalent, 172.
 - resistance of triode, equivalent, 171.
 - factor, 173.
 - factor of amplifier, 174.
 - in mixer stages, 176.
 - in a multi-electrode valve, 172.
 - local oscillator, 176.
 - measurement of, 177.
 - power available from resistance, 169.
 - shot, 29, 170.
 - shot, in a triode, 171.
 - thermal, 29, 168.

O

- Octode frequency changer, 92.
- Oscillators, 87.
- Oxide emitter, 19.

P

- Pentode, 59.
 - amplifier, power gain of, 61.
 - characteristics of, 60.
 - equivalent noise resistance of, 172.
 - load line, 64.
 - mixer, 93.
 - operation at reduced screen voltage, 60.
 - suppressor-grid distortion, 80.
 - transmitting, 83.
 - variable-mu, 65.
- Periodic Table of elements, 2, 4, 5.
- Phase, modulation, 103, 107.
 - demodulator, 111.
- Phosphors, 232.
- Photo-conductive cell, 255.
- Photo-electric emission, 10.

P

- Photo-emission cells, 247.
 - gas-filled, 250.
 - gas-filled, useful amplification of, 251.
 - spectral response of, 249.
 - vacuum, 248.
- Photo-multiplier cell, 251.
- Planck's constant, 10.
- Plate (of valve), see Anode.
- Plates, deflecting, 236.
- Protons, 1.
- Pulse modulation, 107, 114.
- Push-pull amplifier, 74.

R

- Reactor valve, 104.
- Rectifier, copper-oxide, 198, 205.
 - germanium point contact, 199, 214.
 - mercury arc, 186.
 - selenium, 198, 205.
- Reflection (reflex) klystron, 152.
- Relativity correction to mass, 13.
- Reliable valves, 27.
- Resistance, diode dynamic, 36.
- diode static, 36.
- triode dynamic, 43.
- Richardson-Dushman equation, 30.
- Rising-sun magnetron, 162.
- Rubber model, 223.

S

- Screen-grid, 52.
- Secondary emission, 10.
 - effects in tetrodes, 54.
- Selenium rectifier, 198, 205.
- Self-sustained discharge, 182, 196.
- Semi-conductors, 198.
 - elementary discussion of, 198.
 - N-type, 203.
 - P-type, 204.
 - thermistors, 217.
- Silicon-carbide (Metrosil) limiter, 199.
- Silicon crystal mixer, see Mixers.
- Skiatron, 204, 243.
- Space-charge, limitation of current, 33.
 - limited diode, 35.
- Split-anode magnetron, 158.
- Storage tube (Graphecon), 242.
- Suppressor grid, 59.
- Surface forces on a metal, 8.
- Switch, electronic, 107.

T

- Television camera tube, 253.
- Temperature limitation of current, 32.
- Tetrode, 52.
 - beam, 56.
 - beam, characteristics of, 58.
 - beam, construction of, 57.
 - beam, virtual cathode in, 57.
 - inter-electrode capacitances of, 53.
 - screen grid, 52.
 - screen grid, characteristics of, 54.
 - secondary emission effects, 54.

T

- Thermionic, emission, 10, 16.
 current, space-charge limitation of, 33.
 current, temperature limitation of, 32.
 emitters, 18.
 work function, 17.
- Thermistors, 217.
- Thoriated tungsten emitter, 19.
- "Three-halves" law, 35.
- Thyratron, 188.
 de-ionization time, 190.
 hydrogen, 191.
 inert gas, 190.
 ionization time, 190.
 with shield grid, 189.
- TR, switch, 192.
 "keep-alive" electrode, 194.
 recovery time, 194.
 "spike" energy, 194, 213.
- Transconductance (see Mutual conductance).
- Transistors, characteristics of junction, 217.
 characteristics of point contact, 216.
 junction, 216.
 point contact, 215.
- Transit time, of electron, 127.
 effects, 129.
 effect on input impedance, 131.
 effect on output impedance, 132.
 effects, reduction of, 132.
 in space-charge limited diode, 138.
 in temperature-limited diode, 128.
- Transmitter, anodes, 120.
 cathodes, 119.
 grids, 121.
- Travelling-wave, valve, 165.
 amplifier, 166.
 attenuation in, 167.
 bandwidth of, 167.
 noise in, 167.
- Triode, 37.
 amplification factor of, 41.
 amplifier, more exact theory of, 71.
 amplifier, power gain of, 61.
 anode resistance, 43.

T

- Triode, characteristics, 38, 42.
 cold-cathode gas-filled, 192.
 difficulties of, 50.
 disc-seal, 127.
 double, 116.
 equivalent circuit of, 45.
 equivalent noise resistance of, 171.
 germanium, 199.
 grid, action of, 39.
 inter-electrode capacitances of, 46, 50.
 large-signal operation of, 45.
 load line, 62.
 mutual conductance of, 42.
- Triode-hexode frequency changer, 93.
- Tungsten thermionic emitters, 18.
- Tuning indicator, 245.
- Twin valves, 115.

U

- U.H.F., circuits for use at, 133, 140.

V

- Valve, bases, 25.
 flying-leads, 28.
 general construction of, 24.
 manufacture, 26.
 reliability, 27.
- Variable reactor, valve as, 104.
- Velocity, modulation, 141.
 bunching, 142.
 bunching of reflected electrons, 151.
 drift space, 144.
- Velocities of electrons emitted by cathode, 12, 15.
- Virtual cathode, 34, 57.
- Voltage stabilizer (gas-filled), 191.

W

- Work function, 7.
 net, 8.
 of metals, 9.
 thermionic, 17.

Cathode ray tubes
 Germanium diodes
 Silicon diodes
 Photoelectric cells
 Geiger-müller tubes
 Barretters
 Voltage stabilisers
 Neon indicators
 Cold cathode triodes

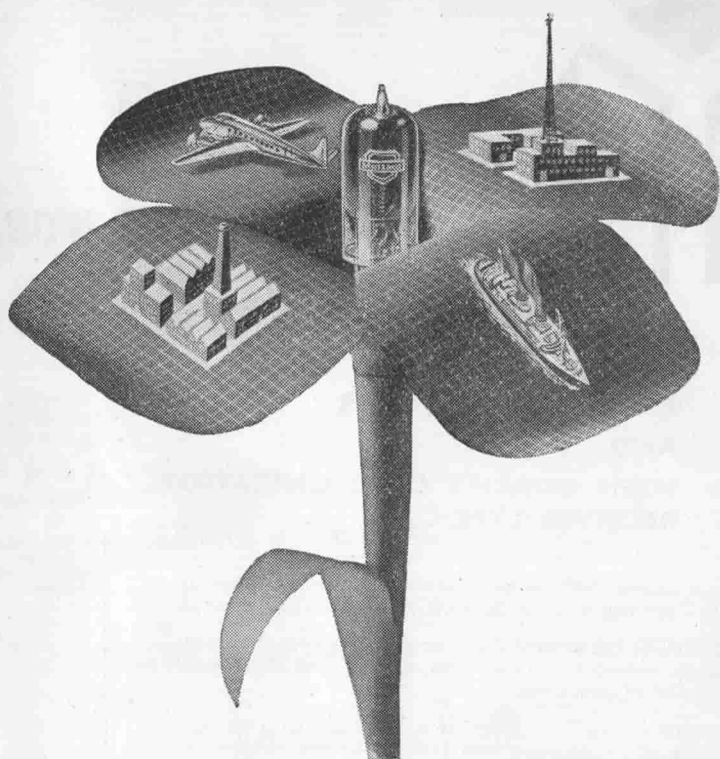
G.E.C.
ELECTRONICS

The range includes many types of special interest to the laboratory, including: Noise diodes, electrometer triodes, high speed cathode ray tubes including p.d.a. and double gun-types.

Information on these and on standard receiving valves can be obtained from: *The Osram Valve and Electronics Dept.*

Osram
VALVES

Receiving
Transmitting
Industrial types
for all purposes



..at the Centre of things...

At the centre of the unfolding science of electronics is the electronic valve. In the air, on land, on and under the sea, electronic valves and tubes are at the heart of equipment which is having an increasingly profound effect on twentieth century ways of life.

Mullard are making a major contribution to this electronic progress with the manufacture of high quality valves and tubes for the many branches of communications, industry and research. Whatever the application which interests you, an authoritative Advisory Service at the address below will welcome your enquiry.

ELECTRONIC VALVES AND TUBES by

Mullard



MULLARD LTD., COMMUNICATIONS & INDUSTRIAL VALVE DEPT.,
CENTURY HOUSE, SHAFTESBURY AVENUE, W.C.2
MVT 166A

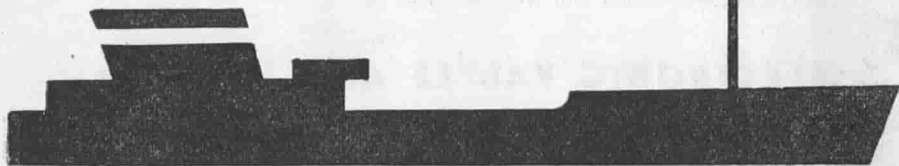


SUPPLIES THE NAVIES OF THE WORLD

ADMIRALTY TYPE 619 H.F./M.F. TRANSMITTER AND HIGH QUALITY COMMUNICATIONS RECEIVER TYPE C.A.T.

Designed as a general purpose wireless set to a British Admiralty specification, Type 619 is surely the most versatile set in the world.

- ★ Receiver C.A.T. can receive A1, 2 or 3 signals over the whole band 30 mc/s—160 kc/s (10—1900 metres), and A1 signals down to 60 kc/s (5000 metres).
- ★ It is specially designed to suppress re-radiation with its consequent risk of interception and D/F by enemy submarines in war.
- ★ Transmitter 619H can radiate A1, 2 or 3 emissions from 1.5 mc/s to 16 mc/s (200—19 metres) at a mean power of 40 watts.
- ★ Transmitter 619M can radiate A1 or 2 emissions from 330—550 kc/s (910—545 metres) at a mean power of 15 watts.
- ★ Type 619 can be matched into any aerial system. The transmitters are so designed that they cannot overload if the aerial is shot away or short circuited in war or by accident, and remain ready for instant service when an alternative aerial can be connected.
- ★ A specially designed vacuum relay provides listening through on a single aerial.
- ★ Type 619 is robust and small. Mounted vertically it occupies the space of a tall man. In cramped wireless offices it may be mounted as four separate units.



Designed and manufactured for the Royal Navy by

PYE MARINE LIMITED

One of the Pye group of Companies

ENQUIRIES TO: PYE LIMITED · CAMBRIDGE · ENGLAND

Pioneers' progress

In 1902 Cossor pioneered the first British cathode ray tube. Today, the name of Cossor is foremost and first for

AIRFIELD CONTROL RADAR

MARINE RADAR

Cossor made the first radar receiver in the world.

ELECTRONIC INSTRUMENTS

Cossor oscillographs are used in a wide variety of industries the world over.

RADIO AND TELEVISION

COSSOR

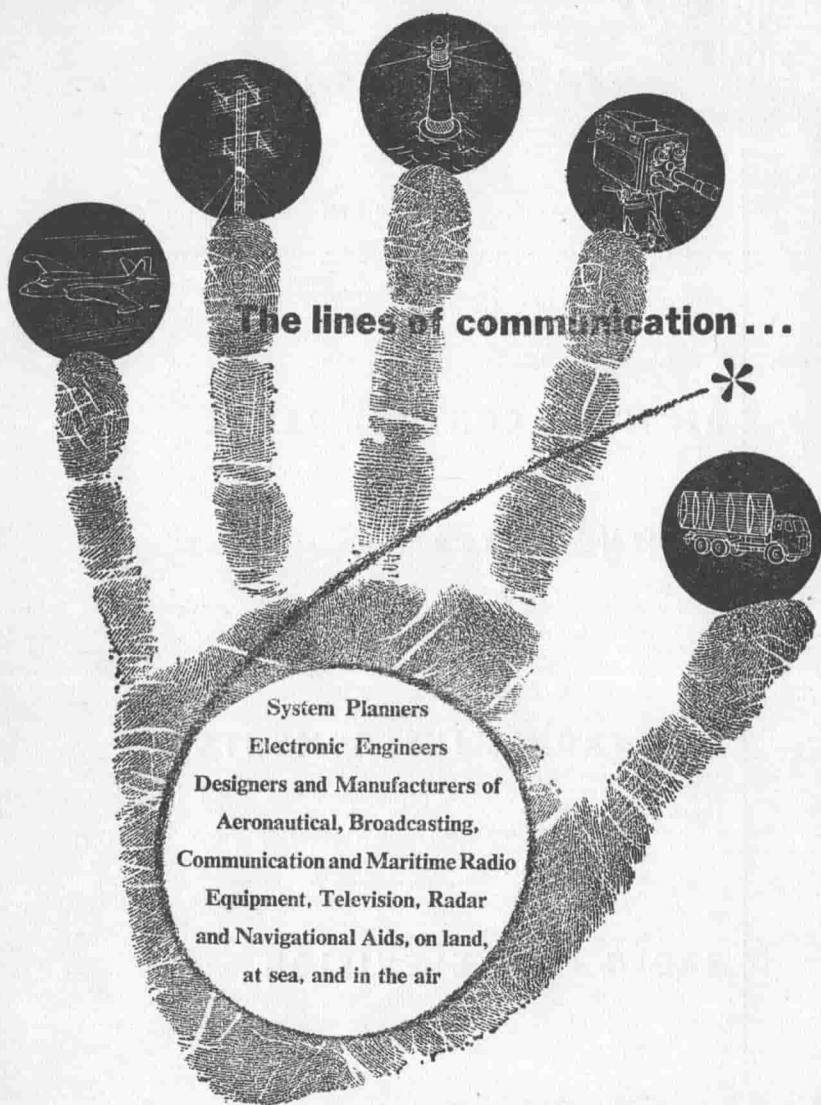
THE COSSOR GROUP OF COMPANIES

HIGHBURY GROVE • LONDON • N.5

A. C. COSSOR LIMITED - COSSOR RADAR LIMITED

STERLING CABLE CO. LIMITED - BEST PRODUCTS LIMITED

COSSOR (CANADA) LIMITED - BEAM INSTRUMENTS CORPORATION (U.S.A.)



The lines of communication ...

System Planners
Electronic Engineers
Designers and Manufacturers of
Aeronautical, Broadcasting,
Communication and Maritime Radio
Equipment, Television, Radar
and Navigational Aids, on land,
at sea, and in the air

...are in capable hands

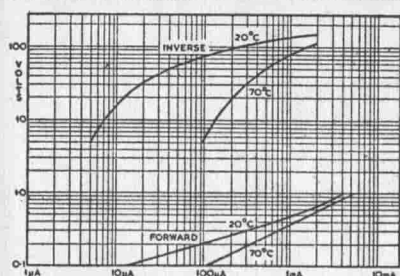
MARCONI

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED
CHELMSFORD · ESSEX · ENGLAND



GERMANIUM RECTIFIERS

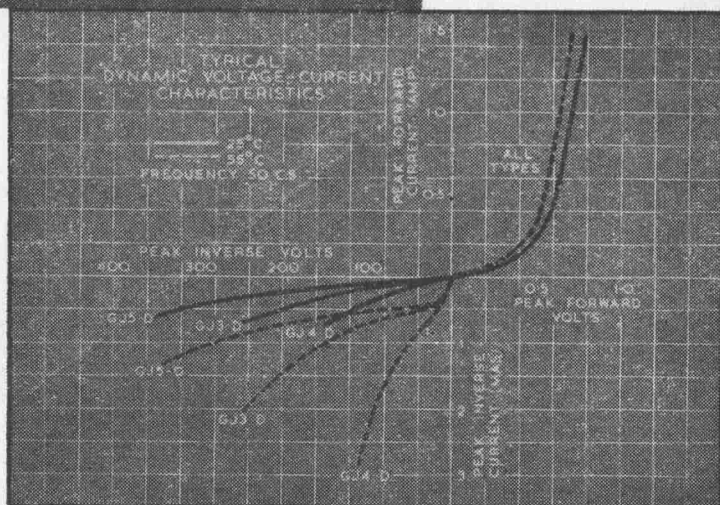
Compact • Reliable • Inexpensive • Efficient



TYPICAL CHARACTERISTIC CURVES

POINT CONTACT RECTIFIERS

A wide range is available, covering applications for television and VHF radio receivers, electronic computers, measuring instruments, etc.



JUNCTION RECTIFIERS

These new rectifiers combine very low forward resistance with high back resistance, to give almost 100% rectification efficiency. Their small size makes them eminently suitable for telecommunication circuits, blocking relays, instrumentation applications, and miniature power packs.

BRITISH THOMSON-HOUSTON

THE BRITISH THOMSON-HOUSTON COMPANY LIMITED • RUGBY • ENGLAND

Member of the AEI group of companies

A4850

PUBLISHER'S NOTE
THE SERVICES' TEXTBOOK OF RADIO

As each volume in this series deals with a specific branch of Radio theory and technique, it has been decided to issue the Volumes as ready rather than to defer publication until the series is complete. Accordingly, Volume III, Electronics, is the first to be published. It is expected that Volume I, Electrical Fundamentals, will be the next to be published.

H.M. STATIONERY OFFICE
November, 1955.

E5646 Wt. 3620-9702 5,250 11/55 Gp. 8 F. & C. Ltd.

ELECTRONICS

HMSO