PHILIPS TECHNICAL REVIEW

13. NOV. 1952
Bibl. Nat. Lab.

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

VOLUME 13

1951/1952

PHILIPS RESEARCH LABORATORIES - EINDHOVEN

CONTENTS

VOLUME 13, 1951/1952

LIGHT

Page

Flicker in television pictures, by J. Haantjes and F. W. de Vrijer 55

A new high-pressure mercury lamp with fluorescent bulb, by J. L. Ouweltjes, W. Elenbaas and K. R. Labberté 109

New phosphors for fluorescent lamps, by J. L. Ouweltjes 346

High-pressure mercury vapour lamps for photo-copying processes, by W. Elenbaas and K. R. Labberté 323

TELECOMMUNICATION

New electronic tubes employed as switches in communication engineering, by J. L. H. Jonker and Z. van Gelder,
I. Contact tubes 49
II. Switch tubes 82

Flicker in television pictures, by J. Haantjes and F. W. de Vrijer 55

The image iconoscope, a camera tube for television, by P. Schagen, H. Bruining and J. C. Francken 119

The wire capacitor and other composite drawn products, by J. L. H. Jonker and P. W. Haaijman 145

A simple ionosphere sounder, by R. Aschen and P. Gaillard 152

Velocity-modulation valves for 100 to 1000 watts continuous output, by B. B. van Iperen 209

A through supergroup filter for carrier telephone systems on coaxial cable, by J. F. Klinkhamer 223

Delta modulation, a new modulation system for telecommunication, by J. F. Schouten, F. de Jager and J. A. Greefkes 237

Some applications of Ferroxcube, by W. Six 301

Flywheel synchronization of saw-tooth generators in television receivers, by P. A. Neeteson 312

Thermionic emitters under pulsed operation, by R. Loosjes, H. J. Vink and C. G. J. Jansen 337

X-RAYS

Special X-rays tubes, by B. Combée and P. J. M. Botden 71

Collimating X-rays in beams of very small divergence and high intensity, by J. A. Lely and T. W. van Rijssel 96

Fluorography with the aid of a mirror system, by W. Hondius Boldingh 269

SOUND

New developments in the gramophone world, by L. Alons 134

ELECTROTECHNICS

The wire capacitor and other composite drawn products, by J. L. H. Jonker and P. W. Haaijman 145

Automatic welding with Contact electrodes, by W. P. van den Blink, H. Bienfait and J. A. van Bergen 247

Electric fencing of grazing land, by F. H. de Jong and L. W. Roosendaal 328

MEASURING EQUIPMENT

Measuring the dielectric constant and the loss angle of solids at 3000 Mc/s, by M. Gevers . 61

Measuring rapidly fluctuating gas temperatures, by B. H. Schultz 104

Apparatus for testing transistors, by P. J. W. Jochems and F. H. Stieltjes 254

Geiger-Müller counters, by N. Warmoltz . 282

An indentation meter for paint, by J. Hoekstra and J. A. W. van Laar 352

MATERIALS AND THEIR RESEARCH

	Page
New views on oxidic semi-conductors and zinc-sulphide phosphors, by E. J. W. Verwey and F. A. Kröger	90
Ageing phenomena in iron and steel after rapid cooling, by J. D. Fast	165
Internal friction in iron and steel, by J. D. Fast and L. J. Dijkstra	172
The magnetic and electrical properties of Ferroxcube materials, by J. J. Went and E. W. Gorter	181
Ferroxdure, a class of new permanent magnet materials, by J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout	194
The structure of glass, by J. M. Stevels	293
Some applications of Ferroxcube, by W. Six	301
Thermionic emitters under pulsed operation, by R. Loosjes, H. J. Vink and C. G. J. Jansen	337
New phosphors for fluorescent lamps, by J. L. Ouweltjes	346
An indentation meter for paint, by J. Hoekstra and J. A. W. van Laar	352
Dielectric losses in glass, by J. M. Stevels	360

MISCELLANEOUS

	Page
Philips diamond jubilee	1
Scientific research of Philips' Industries from 1891 to 1951, by W. de Groot	3
Books written by scientists of the Philips' Research Laboratory	48
In memoriam Dr. h.c. A. F. Philips	81
A simple ionosphere sounder, by R. Aschen and P. Gaillard	152
Automatic welding with Contact electrodes, by W. P. van den Blink, H. Bienfait and J. A. van Bergen	247
Electric fencing of grazing land, by F. H. de Jong and L. W. Roosendaal	328

BOOK REVIEW

B. G. Dammers, J. Haantjes, J. Otte and H. van Suchtelen, Application of the electronic valve in radio receivers and amplifiers. A. F. amplification, output amplification and power supply . . . 180

VOL. 13 No. 1-2, pp. 1-48 JULY-AUGUST 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

PHILIPS' DIAMOND JUBILEE

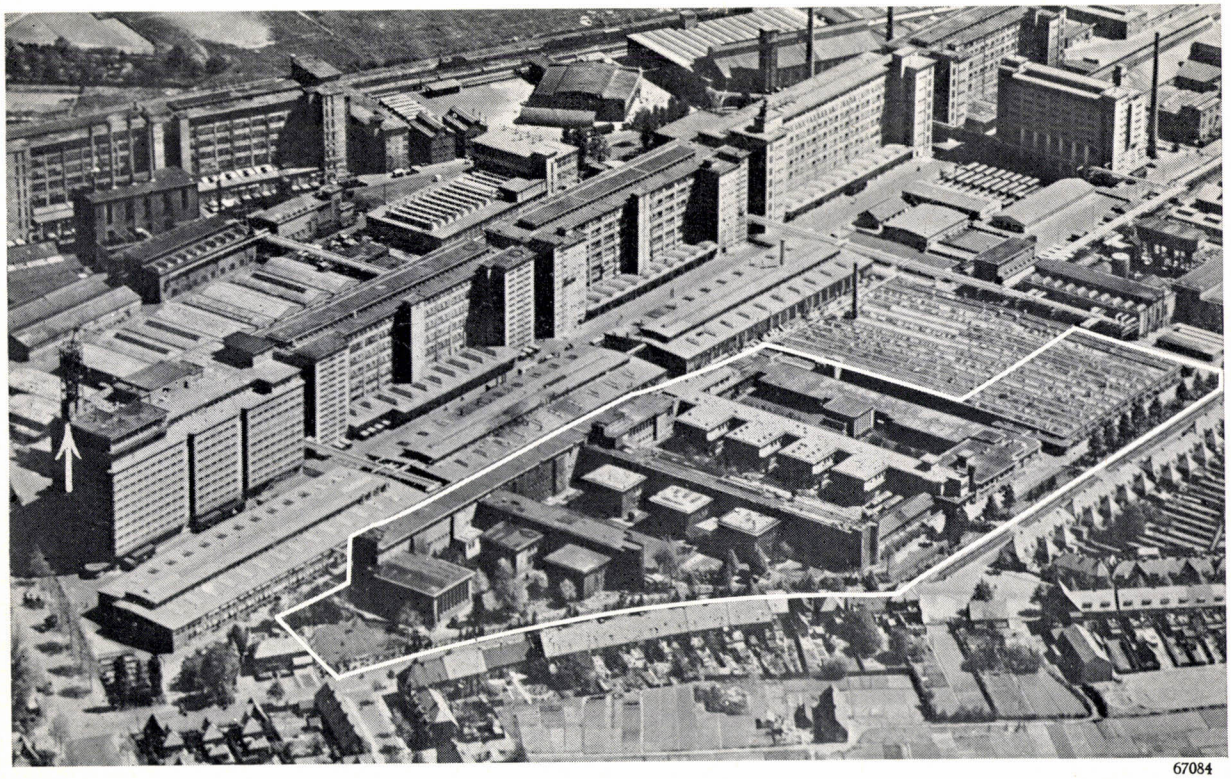
It was sixty years ago, in the month of May, that the Philips' Lamp Works were established at Eindhoven. The Editors of Philips' Technical Review wish to take part in celebrating this memorable event by issuing this Jubilee Edition. In the 1941 volume of this Journal one might seek in vain for any mention of the commemmoration of the 50th year of our Concern's existence. Under the conditions of enemy occupation it had been decided merely to celebrate that event unostentatiously at a meeting of the Management and the executives. Nevertheless that day in May 1941 turned out to be one of boisterous merry-making. Quite unexpectedly the tens of thousands of Philips workers spontaneously downed tools and set out in procession to give vent to their feelings of joy — and to their sense of national pride. Within a few hours almost the entire population of Eindhoven had enthusiastically joined in that demonstration. Although these festivities were abruptly brought to an end by the threat of armed intervention and prohibitions, there are many who regard that May day of 1941 as one of the most memorable days of their lives. Now that, on the occasion of this Diamond Jubilee, liberty and fraternity again reign supreme in our country, there is every inducement for celebration when looking back upon the past, as we shall do in this Jubilee issue. No attempt will be made here to set forth the history of the Philips' concern as a whole; this is to be done in some other way. It has been deemed

Spontaneous jubilation on the 50th anniversary of the foundation of the Philips' Works.

best to confine the scope of this special issue to some aspects of the research work which, from about 1910 onwards, has been carried out mainly in the Physical Research Laboratory, and to throw light upon the importance of that work for the enterprises of our Concern. Dr. W. de Groot, who has been connected with the laboratory ever since 1923 and thus has been personally associated with the development of this research work for the greater part, having himself furnished valuable contributions towards it, was found prepared to write a review on the lines indicated above. It will not lay any claim to being a complete summing up of all research work carried out — such might well prove to be too dry reading — but rather it is to be regarded as an illustrated story about the laboratory, showing in particular the multifarious nature of the problems dealt with and the often unexpected new possibilities emanating therefrom.

In conclusion, on behalf of the whole of the editorial staff, we extend our sincere congratulations, on this Diamond Jubilee, to Dr. A. F. Philips, one of the two founders of the Concern, and further to the Board of Management and to all who have worked so hard to contribute towards the growth and prosperity of Philips Industries.

THE EDITORS



Aerial photograph of part of the Strijp quarter of the Philips' Works at Eindhoven.
The Physical Research Laboratory is outlined. On the extreme left, above the arrow, the tower with television aerial.

SCIENTIFIC RESEARCH OF PHILIPS' INDUSTRIES FROM 1891 TO 1951

INTRODUCTION

With the invention, at the beginning of the nineteenth century, of the voltaic pile and the galvanic battery, providing a source of current with an almost unlimited voltage and a relatively small internal resistance, in principle also the possibility of electric light was created. In 1808 *Davy* succeeded in passing current through air between two carbon rods of a few millimetres diameter. The tips of the carbons and the gas discharge radiated a bright light. Upon the carbon tips being spaced a couple of centimetres apart, owing to convection the discharge assumed the shape of an arc (an inverted U), which name came to be used to describe all gas discharges with a high current density and with a thermionic cathode.

In a certain sense the carbon arc is the precursor of modern discharge lamps and, in so far as the light originates from the glowing carbon tips, also of the incandescent lamp. A second forerunner of the incandescent lamp is the platinum wire brought to incandescence by an electric current, an experiment which was also carried out by *Davy* (1802). Carbon-arc lamps were known about 1850 and, as was subsequently disclosed in the famous law-suit between *Edison* and *Goebel* (1893), also electric incandescent lamps had already been made at that time. These sources of light were fed from galvanic batteries. It was not, however, until the invention of the dynamo (*Siemens* 1866, *Gramme* 1869) that electric light came to be introduced on any large scale. In 1879, on the one hand the differential-arc lamp of *Von Hefner-Alteneck* made its appearance and, on the other hand, *Edison's* carbon-filament lamp, which was demonstrated at the Paris exhibition of 1881, together with the machines for generating the current and the means of distributing it. That exhibition aroused interest everywhere, and it was shortly after this that *E. Rathenau* founded the A.E.G. in Germany.

All this greatly interested the young Dutchman Gerard Leonard Frederik Philips, born 9th October 1858 at Zaltbommel, who was studying at the Polytechnical School (now the Technical University) at Delft, where he graduated as an engineer in 1883. Shortly after that he was given the opportunity to install electric lighting in some ships at Glasgow, and there he made the acquaintance of *William Thomson (Lord Kelvin)*, who was likewise greatly interested in electric light. Later on he conducted negotations for *Rathenau* in Berlin between the A.E.G. and the municipality of Amsterdam about an electric-lighting project, but these negotiations broke down on account of the high price per kWh (fl. 0.60) asked for by the A.E.G.

In 1890, when 6800 arc lamps and 118,000 incandescent lamps were already burning in Paris, four electric power stations in Berlin were supplying current to some 3000 arc lamps and 70,000 incandescent lamps, and in London a generating station was being built for feeding 600,000 lamps of 10 candle power, whilst in the U.S.A. 23,500 arc lamps and 2,800,000 incandescent lamps had already been installed, Ir. Philips conceived the idea of starting a new incandescent-lamp factory in the Netherlands, where others had already been established, among which were the *Pope* and the *De Kothinski* works. This plan materialised in the month of May 1891 with the opening of the factory at Eindhoven. In the course of time G.L.F. Philips and his younger brother Anton Frederik Philips, who joined the firm in 1895, turned this small factory, with a starting capital of 75,000 guilders, into a world-wide concern, the parallel of which is but scarcely found.

It is not the intention to enter here into the history of this concern, for this will be found in a book which is to be published shortly. Instead of that, a review will be given of the development of the scientific research, both in the pure and in the applied sciences, of the Philips Industries and in particular that which has been carried out in the Physical Research Laboratory founded in 1914.

The first research work connected with the Philips business was carried out by Gerard Philips himself. For some time before the factory was started he was studying the preparation of the carbon filament and various other processes in a primitive workshop at home. In November 1890, when writing to one of the many people with whom he was negotiating for the establishing of the new factory, Ir. Philips wrote: "I am able to produce perfectly homogeneous cellulose filaments on a business scale". Considering, however, what the management of a rapidly growing young industry involved, and bearing in mind that up to 1895 both the commer-

Ir. G. L. F. Philips engaged with his first experiments for the manufacture of carbon filaments from cellulose (1890).

cial side of the business and the management of the works were in the hands of one man, obviously this research work was at first limited to what was absolutely essential. Technical and scientific help soon became necessary, especially when from 1903 onwards new materials, such as osmium, tantalum and tungsten, the last of which was destined to be the filament material of the future, came to be used in the manufacture of incandescent lamps. About 1907, at the time that incandescent lamps began to be made with squirted tungsten wire, P. N. L. Staal, A. de Broekert and H. Gooskens*) joined the firm as Gerard Philips's assistants, these being followed, respectively in 1908 and 1909, by the chemical engineers J.C. Lokker and A. de Graaff, and shortly afterwards by the mechanical engineers H. de Jong, H. Reufel and W. Koning.

In December 1911 the brittle squirted wire was successfully replaced by the more rigid drawn wire.

Thus a technical staff of some size was formed, comprising people trained in mechanics and in chemistry, but it was not long before also the need of physicists began to be felt when, in 1913, a new development was announced in the lamp-making world, in the form of the gas-filled lamp.

In the General Electric Company's laboratory at Schenectady (U.S.A.), where *Edison*'s work was carried on and physicists were available, in 1909 *Coolidge* had found an entirely new method for drawing strong, thin wires of tungsten. There soon followed, in 1913, in the same laboratory, *I. Langmuir's* invention whereby a coiled filament could be brought to incandescence in an inert-gas atmosphere, thereby counteracting vaporization of the tungsten and thus making it possible to heat the filament to a higher temperature, so that the luminous output of the lamps could be greatly increased. Within a very short space of time a number of other inventions followed from the same laboratory, such as X-ray tubes with heated cathode and gas-filled rectifying tubes.

In November 1913 also Philips brought gas-filled tungsten lamps on the market, under the name of "half-watt lamps". In the manufacture of these lamps and also, for instance, in their photometry so many problems arose which lay more in the domain of physics than in that of chemistry or mechanics that Ir. Philips decided to engage a physicist. Thus on 2nd January 1914 G. Holst joined the firm in that capacity, and this was the beginning of the

Prof. Dr. G. Holst, after a painting by *S. Schröder.*

*) In this review the names of people connected with the Philips Concern are spaced out, while those of others are printed in italics.

foundation of a physical laboratory. Some months later E. Oosterhuis joined the staff of this laboratory as second physicist.

It will be shown how the work of the laboratory established in 1914, starting with the problems of the manufacture of incandescent lamps, has expanded into such a multifarious programme, from which such a wide variety of other products have been born. In this growth three main periods are to be distinguished:

1) 1914-1923, the period which saw the first world war and during which the laboratory was located in a part of the lamp factory;

2) 1923-1940, beginning with the opening of a new laboratory specially equipped for the purpose (in 1929 it was considerably enlarged) and ending with the occupation of the Netherlands by enemy forces in the second world war;

3) 1940 up to the present day, covering the years of occupation and the revival of activity after the country's liberation.

Dr. E. Oosterhuis, after a painting by *S. Schröder.*

THE LABORATORY OF 1914 - 1923

After studying at the Technical University at Zürich — where he took his doctor's degree in 1914 — Dr. Holst worked for a number of years in the laboratory of *H. Kamerlingh Onnes* at Leyden, where, inter alia, he took an active part in the discovery of superconductivity. Thus, in addition to his bent for purely scientific work, Dr. Holst was also technically interested in taking up his new appointment with the Philips' laboratory. Although it was considered as his main task to study the physical questions arising in the manufacture of the tungsten lamp, it was felt at the same time that it should not be left at that, but that it was necessary to penetrate to the very roots of the phenomena to be studied. From the tackling of subjects on such a wider basis physical science was primarily benefited, but then this in turn bore fruit for technical science, often in a surprising manner, in the form of improvements of existing products and the opening up of new fields of activity.

The study of the tungsten lamp was therefore the starting point: first the behaviour of tungsten wire in the processing, and further its behaviour upon being heated by an electric current to a high temperature in a glass bulb either in vacuo or in the atmosphere of an inert gas; the current had to be passed into the bulb via vacuum-tight and fused-in wires.

Owing to the extensive programme of work the staff soon had to be enlarged. Dr. Holst received successively the assistance of P. G. Cath, S. Weber, C. Bol, H. C. Burger, G. Hertz, Balth. van der Pol, A. Bouwers, A. E. van Arkel, W. Geiss, J. H. de Boer, P. Clausing, R. Vermeulen, W. de Groot, C. Zwikker and others, whilst contributions towards this research were likewise furnished by L. Hamburger, D. Lely and J. A. M. van Liempt, who belonged to the chemical staff of the works.

As regards the physical problems directly related to the incandescent lamp, first of all there was the photometry to be studied, particularly of the gas-filled lamps, since their luminous flux is distributed in space in a less surveyable manner than that of vacuum lamps with a linear filament. In a series of "Communications of the Philips' Laboratory" (1918-1919), which in a sense may be regarded as the forerunners of this journal,

conceptions in the field of illumination were explained and the basic elements of utility lighting, fittings, projectors, car lamps, sunlight lamps, etc. were dealt with.

Of a more fundamental nature was the investigation of radiation, such as the spectral energy distribution of incandescent lamps. At that time *Planck's* formula was still regarded with some scepticism and there was no thought of an "international relative luminosity curve" or an "international temperature scale". Much doubt existed as to the constant c_2 in *Planck*'s formula, which it was important to know in connection with optical pyrometry.

Owing to the gradual decline in the luminous intensity of the incandescent lamps, and particularly of the vacuum lamps, due to precipitation of tungsten on the inside of the bulb, a matter which will be reverted to in the next section of this review, the study of such thin layers of metal and of the chemical processes taking place in residual gases was taken in hand (Hamburger).

In connection with the filament problems, the electrical conductivity and the thermal conductivity of graphite and tungsten were investigated, as also, in connection with the gas filling, the thermal conductivity of gases and the vaporization of tungsten in a gaseous atmosphere (Weber). Further, following upon the investigations which Holst had himself carried out and had seen conducted by others in the cryogenic laboratory at Leyden, extensive tests were made with the argon-nitrogen system, the mixture of gas used for filling the half-watt lamps.

As far as the metallurgy of tungsten is concerned, with the aid of the then new methods of X-ray spectroscopy and X-ray diffraction a start was made with the study of the structure of tungsten and other metals (Burger, Van Arkel), of the physical (e.g. elastic) properties of tungsten (Geiss) and of the chemistry of tungsten compounds (Van Liempt and others).

Of importance, for its future consequences, was the investigation of electrical conductivity in gases and their emission of light when an electric current is passed through them. This came about as a result of the method applied in the factory for testing the degree of evacuation in vacuum lamps and the purity of the gas in gas-filled lamps with the aid of a *Tesla* discharge (spark test). Another, no less important, reason for this investigation being started was the occurrence of electrical breakdown between the extremities of the filament in gas-filled lamps, both when the lamps were burning and when the filament was broken.

To combat this breakdown it was necessary to make a thorough study of the passage of electricity through gases. The investigation of gas-discharge phenomena, already studied by *Faraday* and more intensively by *Plücker, Hittorf, Crookes, H. Hertz, Lenard, J. J. Thomson, Townsend* and others, reached a new stage when it appeared that these phenomena could be related to *Bohr*'s new theory of the atom (1913). This theory, just as was the case with *Planck*'s quantum theory, was at first received with a certain reserve. Moreover, the outbreak of the 1914-1918 war held back more or less the propagation of *Bohr*'s ideas. Of importance for the laboratory in this connection were the lectures given in Eindhoven by *P. Ehrenfest,* since 1912 professor in theoretical physics at the Leyden University and a keen advocate for the ideas held by *Bohr,* whilst in 1920 the laboratory staff was supplemented by G. Hertz, who in 1914, together with *J. Franck,* had established the relationship between *Bohr*'s theory and the phenomena of electric conductance in gases, for which, in 1926, they were both awarded the Nobel prize. Hertz's investigations will be discussed in the next section.

The most important result of the investigation of gas discharges in the period 1914-1923 was the deeper insight thereby gained into the phenomena of electrical breakdown, in particular at low pressure, in rare gases between plane, parallel, cold metallic electrodes, and the deeper knowledge gained of the transition from the non-self-sustained to the self-sustained discharge (Holst, Oosterhuis). One of the facts thereby established was that the production of electrons in the glow is due to the action of the positive ions upon the cathode and not, as *Townsend* imagined it to be, as a result of direct ionization of the gas by the positive ions.

The first practical outcome of the investigation into gas discharges was the appearance in this period of the neon glow lamp (1917) and the tungsten arc lamp with neon-filling (1920).

The investigation of rare gases was facilitated by the fact that since 1916 the Philips' works had been making (under the guidance of H. Filippo) their own liquid oxygen and nitrogen, from which the rare gases could be distilled.

Since during the war the importation of X-ray tubes was stopped medical practitioners in the Netherlands sent their tubes to the Philips' works for repair. Thus interest also came to be taken in these tubes, not only in respect to the gas-filled

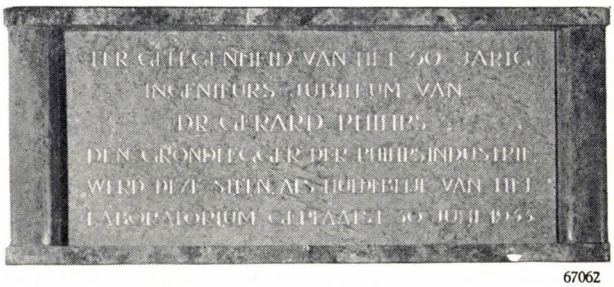
tubes that were then being used but also in the new Coolidge vacuum X-ray tube with a heated tungsten filament as electron source.

The study of the manner in which the metal leads could be passed through the glass of the lamp bulb led to the production of ferrochromium alloys which have about the same coefficient of thermal expansion as that of glass and which can be fused onto glass. Prior to this, the leads were of platinum wire, later of iron wire coated with copper; for other glass-metal joints pure copper was used, which differs in expansion from glass but owing to its softness changes shape while cooling. With the new alloys (Bol, Bouwers, B. Jonas) there was no need to make allowance for any deformation during cooling, so that the leads and connections could be made much stronger and fusing-in was no longer confined to thin-walled tubes. This led to the construction of metal transmitting tubes and metal X-ray tubes. In connection therewith it was of importance that in 1916 Philips had started a glass works of their own (P. J. Schoonenberg).

In 1883 *Edison* had discovered that an electric current flowed through the vacuum between the poles of the filament in his incandescent lamps, a fact which we now know to be caused by electron emission. Following upon *Richardson*'s investigation of electron emission, in 1904 *Fleming* had invented the diode and in 1907 *Lee de Forest* had made the first triodes by adding a grid. Simple receiving triodes were produced by Philips in 1917, these being followed by transmitting triodes and diodes for rectifying alternating currents. Soon a systematic study of the phenomena in the radio valve was started. In 1922 Balth. van der Pol (a former pupil of *H. A. Lorentz*) was charged with radio research*).

In connection with the manufacture of radio valves, whereby use was made of the electron-emitting properties of tungsten, further investigations were carried out with a view to finding materials which could replace tungsten and give a greater emission for a smaller filament power. In 1921, as a result of new ideas about the theory of atoms, *Coster* and *Hevesy* discovered the element hafnium, related to zirconium. Great expectations were held about the thermionic emission of this element, and the preparation and study of this new material was energetically taken in hand in the Philips' laboratory. Though this did not culminate in the important results that were expected of it, the chemical and metallurgical experience thereby gained was useful in many respects.

The ever-increasing amount of research work to be carried out demanded more space, so that in 1922 it was decided to build a new laboratory, which was completed and taken into use in 1923.



67062

Memorial tablet in the hall of the Research Laboratory presented to Ir. G. L. F. Philips on the occasion of his 50th anniversary as Engineer in 1933.

On 1st April 1922 Dr. Ir. G. Philips, who during the first world war had had the honorary degree of doctor in the technical sciences conferred upon him at Delft, resigned his co-directorship at the age of 63. He died at The Hague on 26th January 1942.

THE NEW LABORATORY 1923 - 1940

The new laboratory, situated in a part of Eindhoven which in 1923 was still on the outskirts of the town, was taken into occupation in November that year by a staff of 15 graduated physicists, chemists and engineers with about 20 assistants, instrument-makers, glass-blowers, etc. By 1939 there were 106 scientists and 360 assistants. The number of publications issued by the laboratory in the period 1914-1923 was about 150 and in the period 1923 - 1940 about 1500. A review of research done in this latter period must therefore be limited to the main lines and some outstanding points.

The work begun in the old laboratory was continued in the new under more favourable circumstances. With the larger space available and the increase of staff there was more opportunity for deeper and more far-reaching investigations. It should be mentioned that Dr. Holst always left his scientific staff a high degree of freedom.

*) Since 1949 Prof. Dr. Balth. van der Pol has been Director of the Comité Consultatif International des Radiocommunications at Geneva.

It is impossible, in this short survey, to give a proper account of the various ways in which Dr. Holst stimulated his many co-workers. The few places where Holst is explicitly named here, certainly give an imperfect impression in this respect. The same applies to Dr. Oosterhuis, who (until 1946) held the position of Vice-Director and supervised a team of workers engaged in practical radio research.

Of course research work was still bound to a certain extent to the factories and their production. The laboratory was expected not only to evolve new ideas leading to new products or to the improvement of existing ones — ideas which were suitable for patenting — but these ideas had also to be brought to fruition for manufacture. Further, a certain amount of service was called for, such as the calibrating of measuring instruments, examination and testing of materials, etc. Thus, side by side with the purely scientific work, more and more work of a different nature had to be done. This is the reason why the number of publications issued year by year did not grow in proportion to the growth of the staff.

For a better comprehension of the vastness of the laboratory work this has been divided into five main groups, viz:

I. Light and the production of light, including gas discharges.

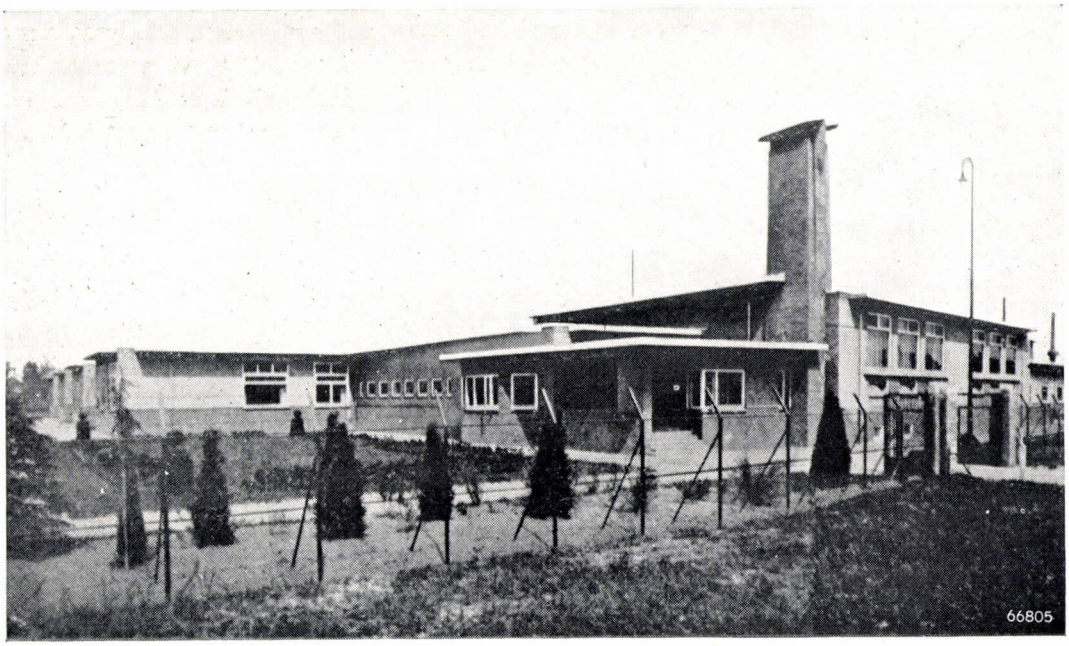
II. Electrotechnics and radio, including acoustics.

III. Chemistry, including metallurgy.

IV. X-ray investigations.

V. Mathematics and mathematical physics.

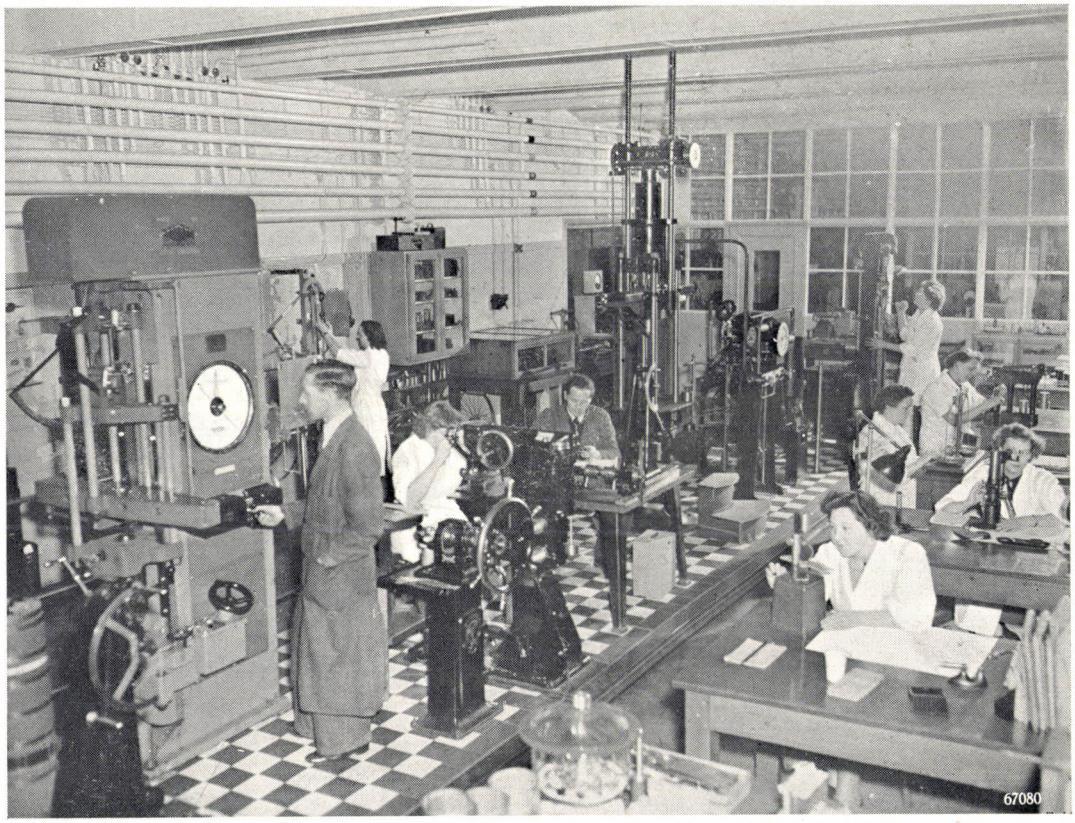
These main divisions have to be taken broadly, and furthermore there are important connections between them. The X-ray examination of crystals, for instance, forms a link between the groups III and IV, the examination of magnetic materials links up groups II and III, whilst the problems coming under the heading V mostly arise from other groups, especially from group II.



The Physical Research Laboratory in 1923.

I. Light and gas discharges

Investigations in connection with incandescent lamps, as far as the new laboratory is concerned, were more or less terminated about 1925 by Zwikker's thesis on the physical properties of tungsten as functions of temperature, which work can be placed side by side with similar work in the U.S.A. In the lamp factory, where a physico-chemical laboratory continued to be maintained, for many years much important work was done by Van Liempt, Geiss and others in the field of the chemistry of tungsten compounds and the metallurgy of tungsten.



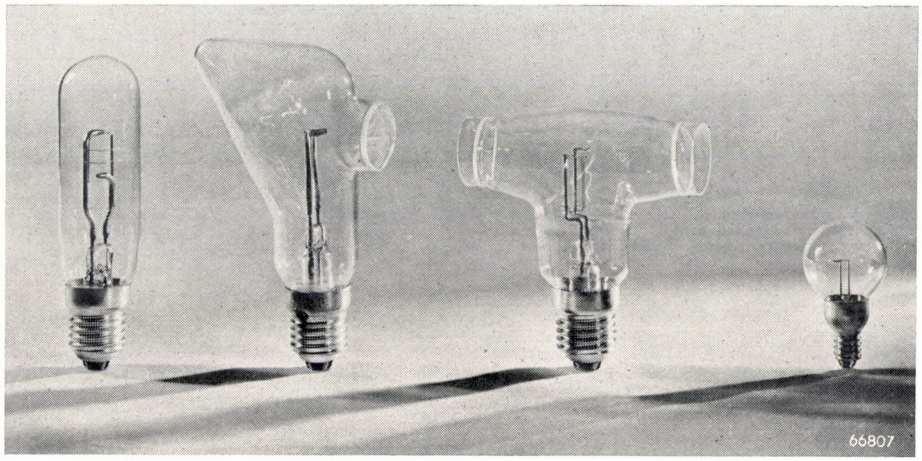
Department for testing materials used in the Philips' Works.

The investigation of gas discharges, which had been begun by Holst, Oosterhuis and Hertz in the old laboratory, was continued in the new one on a wider scale. Hertz had worked out methods for accurately measuring excitation and ionization potentials, mainly of rare gases, and established a relationship between these quantities and the spectrum (term diagram). The resonance lines of all rare gases were photographed with the vacuum spectograph and their wavelengths accurately determined. Furthermore, following up the investigations of Holst and Oosterhuis, the low-tension arc was investigated, when it appeared that an arc discharge can take place in a rare gas when there is a potential difference between anode and cathode lower than the lowest excitation potential. Other investigations concerned the negative glow discharge, especially the heating of the anode when electrons enter it, and the thermal effects arising at the electrodes of the tungsten arc lamp. Important work was by done by F. M. Penning in regard to electrical oscillations in a D.C. discharge in mercury vapour of low pressure; in deviation from *Langmuir's* findings (1925) it was proved that abnormal velocities of electrons in such a discharge are due to the said oscillations.

For Hertz's experiments an equipotential cathode was used which had been coated with barium oxide, obtained by oxidation of barium applied to the cathode in the form of an azide and then thermally decomposed. The experience gained with these oxide-coated cathodes was of direct use for the manufacture of radio valves. Tungsten as electron-emitting substance was very soon replaced by materials (dull-emitters) which already give adequate thermionic emission at a lower temperatue.

Oxide-coated cathodes soon came to be used also for rectifying tubes filled with argon and mercury and for gas-discharge lamps (mercury lamps, neon tubes).

The study of the spectra of rare gases had revealed that the atoms of these gases, namely of argon and neon, may be in an excited state which cannot be transformed into the ground state by radiation, such a state being called a metastable state. After the departure of Hertz in 1925, H. B. Dorgelo, who in collaboration with *Burger* and *Ornstein* had been studying in Utrecht the intensity rules for multiplets in spectra, which study was continued in Eindhoven, took up the study of metastable atoms.



Tungsten-ribbon lamps made in the laboratory to serve as light sources for various optical methods of measuring. The second lamp from the left has been so shaped that the light reflected by the bulb does not pass through the (plane-parallel ground) window and thus does not interfere with the measurement. The third lamp from the left has two flat windows placed somewhat obliquely, this being useful for optical pyrometry.

These atoms are capable of absorbing and re-emitting certain spectral lines which the normal gas allows to pass through unhindered. With the aid of this absorption the lifetime of metastable atoms was investigated with respect to the influence of temperature and gas pressure.

Owing to their energy of excitation, in a gas mixture metastable atoms of one atomic kind (say neon) may ionize other atoms (say of argon), such being possible under the condition that the ionization potential of the second gas is lower than the potential corresponding to the metastable levels of the first gas. In such gas mixtures various factors, such as the breakdown voltage, are greatly influenced by small amounts of the readily ionizable component.

This was further investigated by Penning. The fact that the lowering of the breakdown voltage is due to the action of metastable atoms was proved by demonstrating that the breakdown voltage of neon, lowered by the addition of argon, rose again when the neon was irradiated with light which is absorbed by the mestastable atoms, as a consequence of which these atoms, while emitting resonance radiation, show a transition to a non-metastable state and then to the ground state. It was also found possible by this means to raise the voltage drop of a positive column in a mixture of neon and argon by irradiation with neon light, and even to quench the discharge if the voltage applied is low enough.

The effect of the ionization through metastable atoms is also apparent when investigating the breakdown voltage V_s between plane parallel plates as a function of the product p_0d (p_0 = pressure reduced to 0 °C, d = distance between electrodes), thus determining the *Paschen* curve. Instead of just one minimum (optimum ratio of ionization and excitation) this curve then shows two minima, owing to an increase of metastable atoms and thus the formation of "secondary" ions accompanying increased excitation.

Another anomaly of the *Paschen* curve occurs at values of $p_0d < (p_0d)_{min}$. Normally V_s increases monotonically with decreasing p_0d, or, what amounts to the same thing, p_0d decreases monotonically with increasing V_s. In the case of helium it is remarkable that p_0d as a function of V_s shows a minimum and a maximum, so that in a certain range of p_0d values three critical values of the potential are found, V_1, V_2 and V_3. Breakdown only takes place if the applied voltage V_a answers to $V_1 < V_a < V_2$ or $V_a < V_3$. The same anomaly is found when in the case of stronger currents the voltage is investigated as a function of the current.

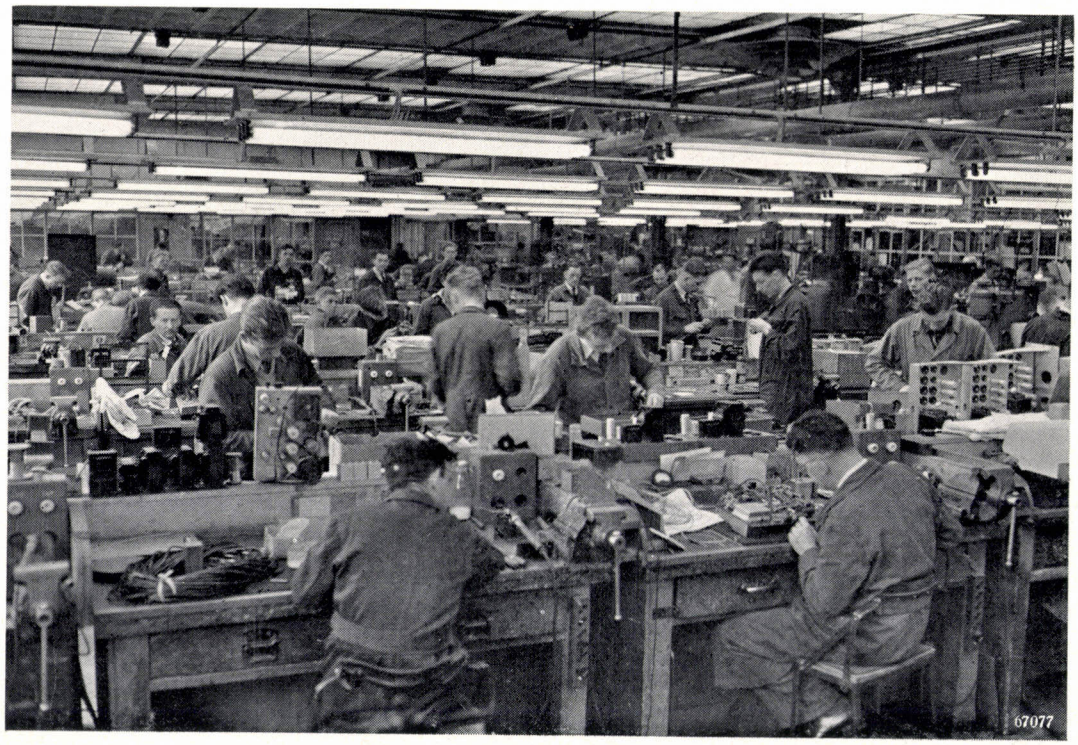
The study of the relation between the current i and the voltage V in discharges, as described above, led also to interesting data being collected in regard to the stability of gas discharges, a subject which

was further investigated later in the Laboratory for Technical Physics at Delft University by Prof. *H. B. Dorgelo*, *Chr. van Geel* and *C. Verhagen*.

Finally there were interesting investigations into the influence of magnetic fields upon discharges. Apart from the fact that the discharge as such forms a "current conductor" which in the magnetic field is subject to a force perpendicular to the current and to the magnetic lines of force, there is the influence of the magnetic field upon the paths of the individual electrons. The latter is particularly manifested at low pressures, where the electrons have a long free path. Following upon Penning's investigations into this field a vacuum meter was constructed in which the current in a gas placed in a magnetic field serves as a measure for the pressure. With this meter direct readings can be taken of pressures between 10^{-5} and 10^{-3} mm Hg and, for example, the improvement of the vacuum in pumping installations (such as in the case of the electron microscope and the cyclotron) can be followed from one minute to the next.

We must now consider the development of gas-discharge lamps. In 1923, following *Claude's* example (1910), the manufacture of neon tubes for advertising purposes was begun. These tubes had (cold) iron electrodes. Instead of neon, which gives a typical red light, other rare gases, such as helium, were used and also mixtures of a rare gas and mercury vapour. The colours could be given more variation by using coloured glass for the tubes. Later on, also fluorescent glass was used and fluorescent powders were applied to the inside of the tubes so as to produce new colour effects. With the introduction of the oxide-coated cathode, already applied for rectifying valves, it was possible to make neon tubes for stronger currents and a lower voltage, which came to be used as beacon lights for airfields and as light sources for special purposes (irradiation of plants).

Scientific research kept pace with this development. As is known, in the case of a discharge in an elongated tube *Faraday* had already distinguished a cathodic part (glow discharge) and an anodic part (positive column), the two being separated by „*Faraday*'s dark space". It is the positive column that produces the light in neon tubes, the glow discharge at the cathode being of no importance in this respect. The positive column was thoroughly investigated, theoretically as well as empirically, by M. J. Druyvesteyn, after *Schottky* had yielded an important contribution to the theory of this discharge. An important concept is that of the electron temperature, which characterizes the velocity distribution of the electrons in the column and can be measured with the well-known probe method of *Langmuir*. Druyvesteyn was able to deduce that in the absence of cumulative processes the electron temperature is, to a first approximation, proportional to the ionization potential of the gas.



Central workshop of the Physical Research Laboratory.

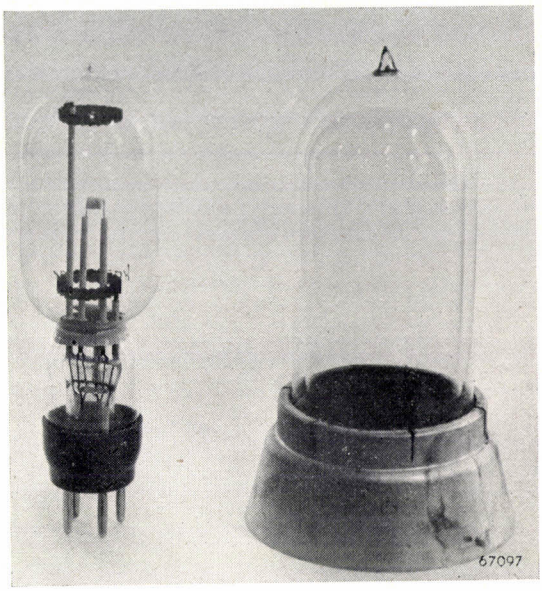
Among the discharge lamps with a filling of rare gas and metallic vapour the sodium lamps occupy a special place. It is well worth while outlining the development of the sodium lamp in the Philips' laboratory. Discharges in sodium vapour had already been investigated and used. In 1919, for instance, such a discharge was described by *R. J. Strutt (Lord Rayleigh*'s son). In 1923 *A. H. Compton* and *Van Voorhis* (Westinghouse) investigated a large number of gases and vapours for their ability of producing light and thereby found also the high light-yielding capacity of sodium vapour. They even patented a special kind of glass resistant to sodium, but this did not lead to any useful lamp being developed.

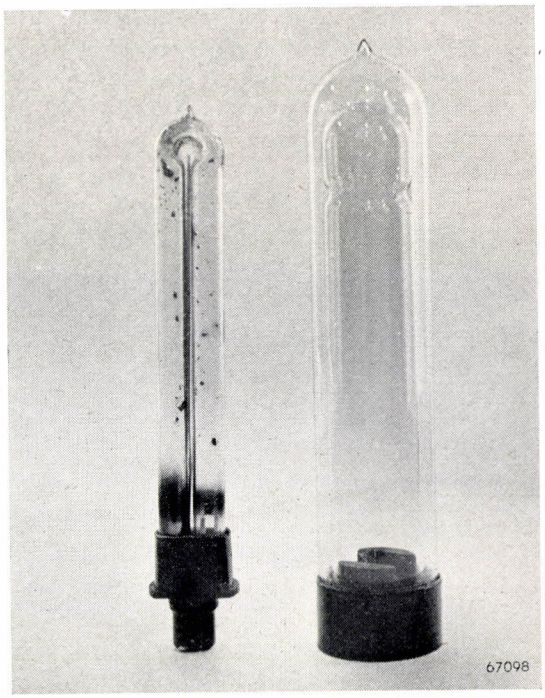
In 1925 *G. Hertz* demonstrated in the laboratory a low-tension arc in sodium vapour in a bulb of about 7 cm diameter. The bulb was fitted with an oxide-coated cathode and an anode, and the sodium was introduced into the bulb by electrolysis through the glass wall, with the cathode acting as negative pole and a bath of molten $NaNO_3$ as positive pole (*Warburg*'s method, 1890). This lamp was placed in a furnace with a temperature of about 250 °C in order to maintain a sufficiently high vapour pressure of the sodium.

Attention was again drawn to sodium as a source of light when G. Zecher set out to improve, in the Philips' laboratory, the yield and the colour of the light of neon tubes by introducing a small amount of lithium into the tube. This was an obvious means, since from the works of *Kirchhoff* and *Bunsen* it had become known that lithium gives a flame a bright red colour. However, something quite unexpected happened: the lithium attacked the wall of the tube and thereby released sodium, so that a sodium lamp was formed unintentionally. This lamp and a number of similar tubes, in which sodium was purposely added to a rare gas, have been used as polarimeter lamps in the laboratory for a number of years.

Meanwhile investigations conducted by *Pirani* in the Osram works had shown that under favourable conditions more than 95% of the energy absorbed by the column could be converted into sodium light, and the Osram works also brought a polarimeter lamp on the market.

At that time the investigations carried out in Philips' laboratory had turned in another direction. The use of the mercury lamp as a source of light for medical ray treatment had led to a closer study being made of the connection between vitamin D and rachitis. It appeared that vitamin D could be produced by irradiating ergosterine with light of a wavelength between 2800 and 2900 Å, which could be produced by a mercury lamp or a magnesium spark. This complexity of phenomena was investigated by E. H. Reerink and A. van Wijk. The demand arose for a special source of light for the production of vitamin D and for anti-rachitic ray treatment. Magnesium proved to be the most suitable element for this. This led to the construction of a low-tension arc lamp with a mixture of rare gas and magnesium vapour; the magnesium





An old sodium lamp (for direct current) and a lamp for connection (via a choke) to 220 V alternating current, each with its vacuum envelope. The D.C. lamps, in groups of 30 to 40 connected in series and fed from a rectifier, were used for the first installations for road lighting with sodium lamps (1932).

was contained in the anode and evaporated into the discharge.

This opened the way to the construction of other metallic-vapour lamps by the same means. In connection therewith, for instance, sodium was investigated anew and this led to the production of a low-tension arc lamp fed with direct current and filled with a mixture of neon and sodium vapour, which has a favourable luminous efficiency (50-60 lumens per watt) (De Groot and E. G. Dorgelo). In order to give the sodium sufficient vapour pressure the bulb was thermally insulated by placing it inside another, evacuated, bulb. Experimental lighting systems installed in a corridor in the laboratory and later on in one of the roadways on the factory site soon demonstrated the excellent qualities of sodium light for road-lighting purposes. After the construction of the lamp had been improved and it had been given the form of a single bulb contained in a *Dewar* flask, these experiments were continued on a more extensive scale, inter alia in cooperation with Prof. Dr. *H. C. J. H. Gelissen*, on stretches of road forming part of the Dutch network of highways (1932). Meanwhile Druyvesteyn and W. Uyterhoeven had been further investigating the positive column in mixtures of sodium and rare gas, whereby it was found that this discharge, in a tube placed in a *Dewar* flask, is less susceptible to changes in the ambient temperature than the low-tension arc lamp. It is to the merit of Bol, who had also given the low-tension arc lamp a shape suitable for manufacture, that the A.C. column lamp with vacuum envelope was developed into a practical unit which for the greater part can be manufactured mechanically, also as far as the glass envelope is concerned. The discharge tube proper has the shape of an elongated U. By 1940 more than 100,000 of these lamps had been installed. Abroad too, where similar lamps had likewise been developed, numerous lighting projects were carried out with sodium lamps, some of which by Philips.

Lighting with mercury lamps has also to be discussed. Compared with sodium lighting, at first Philips took little interest in mercury lighting. Since the investigations of *Küch* and *Retschinsky* (1907), which resulted in the appearance of the quartz-mercury lamp with mercury-pool electrodes (artificial sun), it had been found that with mercury vapour under high pressure (1-3 atm) a source of light of high efficiency could be obtained. The spectral composition of this light shows a striking lack of red and as a consequence colour rendering is inadequate. But with sodium there is still less colour rendering and when this drawback came to be generally accepted for the sake of the other, favourable, qualities of sodium light (efficiency, visual acuity, contrast) there was every inducement to try out also the mercury lamps. Great Britain set the example by installing experimental lighting with high-pressure mercury lamps, with oxide-coated cathodes, on some highways. Eindhoven, too, very soon started producing these lamps. But meanwhile further developments were taking place. The experimental and theoretical investigations carried out by W. Elenbaas very soon made it possible to survey the whole field of mercury discharges as functions of the various parameters (dimensions of the tube, mercury pressure, current and voltage). For instance, a principle of similarity could be worked out, whereby the number of essential parameters could be reduced and it could easily be predicted what the behaviour would be of a lamp with certain dimensions and containing a certain amount of mercury.

It was again Bol who succeeded in drawing practical conclusions from this study and arrived at a very compact construction of mercury lamps, which, however, only became possible after H. J. Lemmens had worked out a method of fusing tungsten leads to quartz, using only one intermediate glass. These lamps, only a few centimetres in length and with an internal diameter of 2 mm, were so designed that part of the mercury remained in the liquid state. Owing to the considerable heating of the wall of the tube these lamps were cooled with water. The internal pressure amounted to 100 atm and more. These lamps are being used for cinema projectors, in television studios and in searchlights, in general wherever there is no serious objection against the installation of a water-cooling system.

Another type, with a determined quantity of mercury which entirely evaporated during the operation of the lamp, had no need of any forced cooling. The equilibrium pressure was 5 to 10 atm. This type of lamp found extensive use for road lighting. In order to limit the preheating time the lamp proper, which has an internal diameter of 10 mm and a length of 35 mm, was placed in a normal incandescent lamp bulb filled with an inert gas to prevent oxidation of the leads. Notwithstanding the fact that at this high pressure the mercury spectrum shows, in addition to the much widened mercury lines, a continuous background extending into the red, colour rendering is inadequate, just as is the case with the lamp with a pressure of 1 atm. This drawback can be remedied by admixing incandescent light, but then the efficiency of the

whole is reduced. Later it was found possible to improve the colour rendering by coating the inside of the outer bulb with substances which fluoresce under the influence of the ultra-violet rays of the mercury light.

Even when corrected with incandescent light or by means of a (red) fluorescing substance, however, for various reasons these lamps are not suitable for indoor lighting. In this direction gas-discharge lamps appeared in a different form. Starting from the low-pressure mercury discharge, whereby the emission of the visible mercury lines is only small and mainly the ultra-violet resonance lines of 2537 Å and 1849 Å are produced, these ultra-violet rays were converted into visible light by means of fluorescence. Philips contributed much towards the development of these "TL" tubes. By coating the inner wall of the discharge tube with a suitable mixture of fluorescent substances ($MgWO_4$, $(Zn,Mn)_2SiO_4$, $Cd_2B_2O_5$-Mn, $(Zn,Be,Mn)_2SiO_4$, and later also other substances, such as halophosphates) it was possible to produce a white light with a spectral energy distribution sufficiently approximating that of daylight or of incandescent light.

Such lamps as these, with a gross output of 40 to 50 lm/W, are being used more and more for all forms of utility lighting, such as in the home, workshops, offices, shops, etc., and for road lighting.

J. Voogd and others have been occupying themselves with the photometry of gas-discharge lamps.

The development of gas-discharge lamps with their spectral energy distribution differing so greatly from incandescent light, their extensive applications for road lighting and the problems of colour rendering all called for a profound study of the properties of the human eye and the faculty of seeing under different levels of brightness. In the Philips' laboratory these problems were energetically tackled by P. J. Bouma and A. A. Kruithof. Bouma revived interest in the more or less forgotten work of *A. König* (1891) on the subject of seeing at low brightnesses and he strougly advocated the ideas of *Schrödinger* (1920) in regard to the use of the colour space for colorimetric problems. His division of the visible spectrum into eight sections as a means of judging colour rendering has come to be of almost universal use, at least in Europe. For the practical application of the eight-section analysis a special photometer was constructed by P. M. van Alphen.

II. Electrotechnics, radio and acoustics

From the manufacture of incandescent lamps a number of important electrotechnical products have emerged which are related to the property of the filament to emit electrons, an effect which, as already remarked, was discovered in principle by *Edison* and subsequently investigated quantitatively by *Richardson*.

Library of the Physical Research Laboratory.

As a result of this investigation there appeared, as we have seen, first rectifying valves and radio valves with tungsten filament in vacuo, whilst Philips followed the G.E.C. in producing gas-filled rectifying valves with coiled tungsten filament. An important improvement was the replacement of the helicoidal tungsten filament in these valves by an oxide-coated cathode (J. Bruynes).

With the introduction of the oxide-coated cathode many more uses were found for rectifying valves, because on the one hand larger emission currents were obtained, thereby extending the field of application to heavy currents, whilst on the other hand the lifetime of these tubes was so extended as to be dependent only upon the chance of accidental damage (breakage or breakdowns) and not upon exhaustion of the thermionic properties of the filament (J. G. W. Mulder).

The field of rectifiers is an interesting example of a case, often seen, where a phenomenon first found as a small effect, so small that its existence might be doubted, ultimately finds application on a scale precluding any doubt as to its reality. The currents studied by *Richardson* could at first only be measured with a sensitive galvanometer. Now a rectifying valve with an emission current of 100 A is nothing rare.

Rectifying valves are now being made for high as well as low voltages. Valves for low voltages (some tens of volts) are used, for instance, for charging batteries and feeding cinema arc lamps (D. M. Duinker). In X-ray practice, on the other hand, rectifying valves are employed which can withstand voltages exceeding 100 kV, thanks to their being given a suitable shape. Another important product was the welding rectifier, characterized by a very heavy current (H. A. W. Klinkhamer).

Within the scope of rectifier research mention is to be made also of the work done in developing blocking-layer rectifiers, and in connection therewith the investigations with selenium (W. Ch. van Geel, N. W. H. Addink).

With the advent of the triode as receiving and transmitting valve round about 1914, radio entered upon a new era, a development to which Philips laboratory contributed in no small degree.

It is remarkable how often in the field of radio practice has been far in advance of the theory. The triode proved to answer its purpose well and was already being applied on a large scale ever before the relative problems, such as the calculation of the field between the electrodes and the behaviour of the electrons in that field, had been really mastered; the propagation of the radio waves over long distances had been found practicable long before scientists had got to the bottom of the actual reason for it and before the electromagnetic equations governing the propagation along the earth's surface had been satisfactorily solved. In the long run, however, it is essential to gain the fullest possible insight into the theory of the phenomena, and for that reason Philips have always devoted much attention to their theoretical investigation. Many other problems remained to be

A corner of the glass-blowing shop of the Physical Research Laboratory.

solved in connection with the practical application of radio valves, and so in the new laboratory radio investigations were divided among a number of groups of research workers. B. van der Pol was charged mainly with the conducting of theoretical radio investigations, whilst the more practical investigations were carried out by E. Oosterhuis, P. R. Dijksterhuis, Y. B. F. J. Groeneveld, H. Rinia, B. D. H. Tellegen and many others.

Van der Pol had begun in 1922 with the study of the triode. It had been found that the flow of electrons to the grid and the anode, respectively i_g and i_a, is a function of the voltages V_a and V_g respectively at the anode and at the grid with respect to the cathode. A three-dimensional plaster model was constructed with which this relationship could be visualized. Then attention was paid to the paths followed by the electrons under the influence of the field and to the secondary emission caused by the electrons impinging on the anode, the effect of which can be demonstrated with the plaster model. Further problems were the distribution of the electron stream between grid and anode and the effect that space charge has upon the field; for the study of these problems a foundation had been laid by the theoretical investigations of *Langmuir* and *Epstein*.

The problem of the motion of the electrons, which has always demanded attention in connection with the construction of radio valves, was subsequently investigated by P. H. J. A. Kleynen and J. L. H. Jonker with the aid of a model employing small steel balls made to roll over a sheet of rubber. This device has proved to be of great value in cases where the mathematical approach to the problem is too complex. Another means of circumventing mathematical difficulties was the measuring of fields of complex electrode systems with the aid of the electrolytic tank.

In connection with the phenomena arising in the radio valve mention is also to be made of the extensive investigation of fluctuation phenomena (noise) carried out by C. J. Bakker and M. Ziegler, which later on was extended to higher frequencies by M. J. O. Strutt, A. van der Ziel and others.

Further, there was the investigation made by H. Bruining into the secondary electron emission of solids, not only in connection with the occurrence of this emission in ordinary radio valves but also with a view to the construction of special valves in which secondary emission is brought about purposely in order to produce special effects.

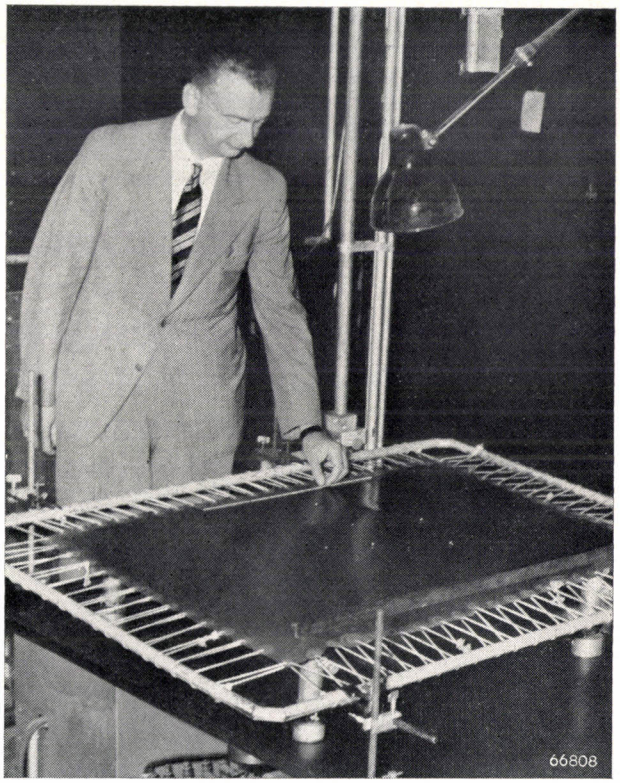
Following upon the study of the radio valve as such, its behaviour was investigated when employed as an amplifier or as an oscillator in a network comprising capacitances, resistances, self-inductances and mutual inductances.

An electric network is a system which as a rule is governed by a number of linear differential equations. With the introduction of radio valves in the network not only are negative resistances introduced, which make it possible for oscillations to be generated, but also non-linear terms then enter into the equations and make the problem more complicated.

Van der Pol succeeded in working out a non-linear differential equation in a simple form (the Van der Pol equation) which incorporates all essential data involved in the case of oscillations and the limiting of oscillations in networks. This equation,

$$\ddot{v} - \varepsilon(1 - v^2)\,\dot{v} + v = 0\,,$$

contains as parameter the factor ε, which to a considerable degree determines the behaviour of the solution. It appeared that within a certain range of values ($\varepsilon \gg 1$) the oscillation bears a character differing greatly from the known behaviour. Van der Pol named these relaxation oscillations. As opposed to "ordinary" oscillations as may occur when the system comprises mainly C's and L's, and which have a sharply defined cycle, while the amplitude depends upon various secondary conditions, relaxation oscillations may arise in systems which are governed mainly by C's and R's. In the latter case the amplitude of the oscillations is determined by the system, while their frequency is highly sensitive to external disturbances. Thus a system showing relaxation oscillations, such as a glow lamp shunted by a capacitor charged by a voltage source via a resistor, can easily be synchronized with a periodical signal, a principle widely used nowadays in television.



Small steel balls made to roll over a sheet of rubber give a picture of the paths followed by electrons in, for instance, electronic valves.

With the aid of relaxation oscillations it is also easy to demultiply frequencies (frequency dividing) and to produce the sub-harmonics, in contrast to the formation of higher harmonics, which can be obtained by connecting non-linear impedances to a normal oscillatory circuit.

As Van der Pol pointed out, the concept of relaxation oscillation is also of biological importance, where the rôle of "resistance" is performed by some diffusion phenomenon governing the biological process. Examples are the reaction of plant foliage to the alternation of day and night and the functioning of the human heart.

Van der Pol and J. van der Mark succeeded in constructing an electrical model of the heart, built up from a number of glow discharge lamps, capacitors and resistors, with which not only the functioning of the normal heart but also certain pathological aberrations could be imitated. To our minds the medical world was too sceptical about the value of that model at the time. Fortunately, however, medical scientists of the present day are showing more and more interest in the results of electro-physical and electrotechnical work, as for instance in neurophysiology. There are many indications that in the near future closer cooperation between the medical practitioner, the physicist and the electrical engineer, with mutual appreciation for the experiences and views of each, will prove to be of great advantage to all concerned, including the patients.

An important part of radio research is that concerning the propagation of electric waves and the interaction between these waves and matter. First of all mention is to be made of the investigations relating to skin effect and to the penetration of electromagnetic alternating fields in conductors. The practical side of this subject was made manifest in the manufacture of radio valves, whereby the valve was placed in a high-frequency magnetic field for degassing various parts mounted in it. It was of importance to be able to predict the degree of heating of the parts (plate, cylinder, grid) as a function of the field strength and of the orientation in the field; this problem was thoroughly investigated in the laboratory by M. J. O. Strutt and others. Of particular interest are the phenomena taking place in bodies of a magnetic material, for there it may happen that above the *Curie* point ($\mu \approx \mu_0$) the product $\mu\delta$ ($\mu = \mu_0\mu_r$ = permeability, δ = depth of penetration) is small with respect to the dimensions and below the *Curie* point ($\mu \gg \mu_0$) large. As demonstrated experimentally by J. L. Snoek, this has remarkable consequences when small objects of magnetic material are subjected to high-frequency heating. Upon the field strength being reduced the temperature remains high, owing to sufficient power being absorbed even when the field is weak. When, however, the temperature drops below the *Curie* point the absorption of power rapidly decreases and the field has to be made much stronger to heat the body to a high temperature again.

The radiation from aerials was investigated both empirically and theoretically. As far as the experimental work is concerned, extensive measurements were taken, for instance, of the field strengths in areas covered by various broadcasting stations all over the Netherlands (R. Veldhuyzen).

Theoretically the problem of wave propagation for a free radiating dipole had already been solved by *H. Hertz*. When the dipole is placed over an infinite, perfectly conducting, "flat earth" the field of the dipole and that of its reflected image can simply be added together. Considering, however, that the conductivity of the earth is finite and that it has a finite relative dielectric constant, the problem is much more complicated. For this case *Sommerfeld* arrived at an exact formula as far back as 1909, but this formula is not suitable for calculations. Therefore at the same time an approximative formula was given for calculating the field strength at the earth's surface. In 1919 the same problem was tackled once more, but in a different way, by *Weyl*, and in 1926 *Sommerfeld* showed that *Weyl's* result agreed with his own. But he then put his approximative formula in a somewhat different form. Much has been written on the question which of the two formulae, that of 1909 or that of 1926, was the "correct" one and what was to be decided by experiment. Important contributions on this subject were given by Van der Pol and his coworker K. F. Niessen (one of *Sommerfeld*'s pupils), who arrived at a strict solution in a new form, and further by the Americans *Norton* and *Burrows*; the exact experimental determinations carried out by the latter proved to be in agreement with the formulae given by *Weyl* and by Van der Pol and Niessen.

Another question of practical importance is what part of the energy radiated by a dipole is dissipated in the earth, since from that the efficiency of a transmitter can be calculated. This mathematically complicated problem was solved by Niessen (1940) by employing *Sommerfeld*'s exact formula.

The problem of the propagation of the waves over a spherical surface has been the subject of an intensive investigation by Van der Pol and H. Bremmer following upon *Watson*'s work. Since in the formulae the dimensions of the spherical surface and the wavelength are not bound to certain values, these apply also in other realms of physics, such as for the interaction between light waves and droplets of water (the rainbow). A remarkable theoretical result, for instance, is that the radiation of the rainbow shows polarisation. This can now easily be verified with the aid of "Polaroid" spectacles; it seems that meteorologists had never noticed it.

In the investigations referred to, the atmosphere was regarded as a homogeneous medium. Later,

Bremmer contributed important information on propagation in an inhomogeneous atmosphere (the ionosphere, "ducts"). In Great Britain the investigation of the ionosphere had been taken up experimentally on a wide basis by *Appleton*, who gave a lecture on this at Eindhoven in 1930. The root of the problem, for which no solution had then been found, lay in the relation between the virtual height (velocity of light in vacuo × half the reflection time) as a function of the frequency and the concentration of free electrons as a function of the height above the earth. From the discussions held at Eindhoven it appeared that, when ignoring the magnetic field, this relationship is given by an integral equation of *Abel*, so that the result can be written explicitly.

At certain frequencies the reflection time is infinite. Van der Pol saw in this a connection with the mysterious "delayed echos" which at that time were puzzling many radio amateurs. It is to be noted, however, that *Störmer* had quite a different explanation for this phenomenon and attributed it to charged particles coming from the sun at a great distance from the earth (10^6 km and more).

A remarkable effect of the ionosphere, discovered at Eindhoven, is the interaction of radio waves of different frequencies. In the empty space, owing to the linearity of the *Maxwell* equations, the principle of superposition applies exactly, so that two wave systems may penetrate each other without any mutual interference. In the ionosphere, however, the interaction is partly of a non-linear nature, with the result that waves propagated through the ionosphere over a powerful transmitter become modulated with the frequencies of that transmitter. Tellegen observed this at Eindhoven in the case of signals from the Beromünster station, which were subject to interference from the powerful Luxemburg station.

In addition to this, for the greater part, purely scientific research a considerable amount of work was directed towards the practical side of radio, in connection with both radio valves and further radio equipment. Above all, as was only natural, radio valves underwent repeated changes. There was a universal rational dimensioning of the various electrodes. In many cases the filament was replaced by an indirectly heated cathode. A second grid was introduced, first as space-charge grid and later as screen grid (*Hull*), so as to render the i_a-V_g characteristic less sensitive to anode voltage fluctuations and, furthermore, with the object of reducing the capacitance between the control grid and the anode, which was found desirable on account of the ever higher frequencies at which transmitting stations were working.

With the tetrode thus formed much trouble was experienced from the secondary emission of the anode. Tellegen therefore introduced between screen grid and anode a third grid (suppressor grid) to hold back the secondary electrons. Thus arose the five-electrode valve or pentode, which at first was used as output valve and subsequently came to be applied also in other circuits. On first sight this may not seem to be a very drastic change, but it has proved to be an exceptionally important improvement, since now, with a very few exceptions, all radio valves are pentodes.

Much work has been done in investigating the behaviour of these and other types of valves in various functions, such as for high-frequency and audio-frequency amplification (A. J. Heins van der Ven, J. van Slooten, H. van Suchtelen, and others).

Considerable attention has also been paid to the construction of transmitting valves, partly in connection with the employment of short waves (15-50 m) for long-distance radiotelephony. An experimental transmitter was built (J. J. Numans), provisionally equipped with a "dummy aerial", in which the energy normally radiated could be dissipated. A milestone was reached in the history of the laboratory in March 1927, when this transmitter was connected to an aerial and communication was established with what at that time was the Netherlands East Indies. This soon gained world fame, especially when, on 1st June 1927, *H. M. Queen Wilhelmina* used the transmitter (station PCJJ) to broadcast an address to the overseas territories.

In connection with the ever higher frequencies that were being used, there was also the development of the magnetron as transmitting tube. Philips Research Laboratory contributed much towards a proper understanding of the working of this tube. The treatises by K. Posthumus on the functioning of the magnetron with split anode still form the basis for all theoretical expositions in this field. Experiments in range finding by means of radio waves of 1 m and smaller generated by a magnetron transmitter were carried out in the laboratory before 1940; in the period 1940-'45 this principle was applied on a large scale elsewhere in the form of "radar".

The designing and manufacture of radio valves calls for great care and special methods. As higher

H.M. Queen Wilhelmina, accompanied by H. R. H. Princess Juliana, addressed Her subjects in overseas territories via the Philips' transmitting station PCJJ on 1st June 1927.

frequencies came to be used the connections had to be shorter and so different methods had to be found for constructing the electrode lead-ins. The solution was found by applying pressed glass or sintered glass (Lemmens) and special metals for the leads. This development came for a large part from the factory laboratories.

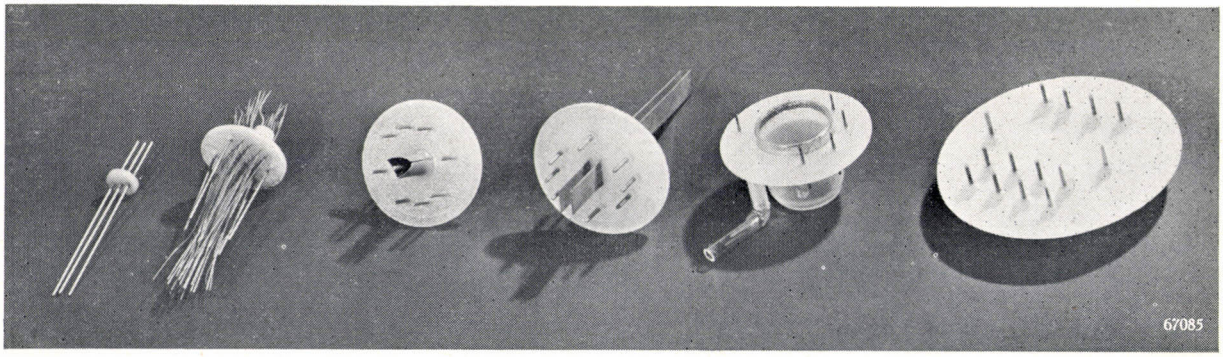
At first radio valves and some radio parts, such as high-tension units, audio-frequency transformers and resistance-capacitance couplings, used by amateurs and set makers, were the only radio products manufactured in the works. However, with the increasing popularity of radio broadcasting there was such a big demand for radio receivers that it became worth while to start manufacturing complete receiving sets. Looking back now, it is hardly imaginable that 25 years ago this was regarded as being something out of the ordinary.

Almost at once the need was felt for an apparatus that could be worked without batteries by feeding it entirely from the mains. This was made possible by employing the indirectly heated cathode which had meanwhile been developed at Eindhoven.

In the construction of radio sets, as already remarked, a great many problems were involved which had to be solved in the laboratory, in the beginning even in all sorts of constructional details (Bol, C. J. van Loon, J. M. Unk). On the one hand there was a need of certain circuits which had to be as efficient as possible while at the same time being easy to make and to repair, and on the other hand accurate methods of measuring were needed for testing the functioning of experimental circuits.

Much attention was devoted to the coils. By making these as loss-free as possible (according to a principle evolved in this laboratory by H. Rinia) greater selectivity was obtained and a "straight set" could be built with four tuned circuits, which for a long time answered the purpose very well. As the "ether" became more and more crowded with the increasing number of stations working within the allotted frequency bands, so that short waves came to be used for broadcasting, receivers had to be built on the superheterodyne principle.

Of great importance was the invention, by



Metal leads and supporting rods can be fused into bases of sintered glass in an almost unlimited number and in any order.

Posthumus, of negative feedback, for which patents were obtained already in 1928. This principle, which reduced distortion due to curvature of the valve characteristics, was very soon introduced in the Philips receivers. It proved, however, to have a much wider scope than this, so that at the present day it is being employed in practically all amplifiers and in all kinds of regulating devices, etc.

For those cases where a highly constant voltage (low internal resistance) is required for feeding radio apparatus a self-regulating high-tension supply unit was developed (Rinia, H. J. Lindenhovius). In connection with smoothing systems for the feeding of radio apparatus mention is to be made of the considerable amount of work done in the field of electrolytic capacitors (Van Geel, A. Claassen).

A difficulty encountered in the manufacture of radio sets lay in the lack of uniformity of electrical networks, especially since there are both direct and alternating current mains. J. W. Alexander constructed vibrator-converters capable of converting direct voltages of 100 to 200 V into an alternating voltage. Such vibrators are now commonly used with car radio sets for converting the battery voltage of 6 or 12 V into an alternating voltage of 220 V.

Another problem lies in the feeding of receivers in places where no mains are available. In the place of accumulators and dry-cell batteries the thermoelectric generation of current was thought of, but this is too uneconomical. This subsequently led to the development of the air engine, which before long will be able to take the place of the petrol engine now used for this purpose.

In the development of radiotelephony there is also the problem of conversion of the radio signal into audible vibrations of the air, thus into speech or music. As a consequence acoustics, and especially electro-acoustics, have become inseparably connected with radio. In 1925, when under the guidance of R. Vermeulen Philips started making loudspeakers, these were built on the electromagnetic principle, whereby the movement of an armature is transmitted to a diaphragm. Then there appeared the moving-coil loudspeaker, in which the cone-shaped diaphragm is connected to a small cylindrical coil through which the varying signal current flows and which is placed in a radial magnetic field. This magnetic field was at first produced by means of a soft-iron circuit excited with direct current. Philips very soon replaced this circuit by a permanent magnet. This development involved intensive research in connection with magnet steel, a subject which will be referred to again elsewhere. The new magnetic materials also made it necessary to give the magnets a different shape, so that attention had to be paid to the designing of magnetic circuits (A. Th. van Urk).

In addition to the work involved in the development of loudspeakers, much work has also been put into the development of power amplifiers, which in turn led to great activity in the field of line telephony (W. Six, H. G. Beljers, J. te Winkel).

In connection therewith attention may be drawn to the various measuring instruments developed in the laboratory, such as a measuring bridge for measuring losses in coils and capacitors, with which phase angles can be measured with an accuracy of 10^{-6} in a frequency range of 10^3 - 10^5 c/s (J. W. Köhler, C. G. Koops).

Further research work led to the development of various types of microphones, gramophone pick-ups and gramophone motors, and finally sound reproduction for the sound film and for broadcasting studios. An important part has been played in this by the Philips-Miller process, whereby a width-modulated track is cut in a celluloid tape coated with a lacquer, after which it is scanned by the known optical means. An advantage of this system was that the recording could be played back at once and thus corrected where necessary.

In this connection mention is to be made of a system, developed by K. de Boer, for stereophonic sound reproduction, both direct and by means of gramophone records or a Miller tape.

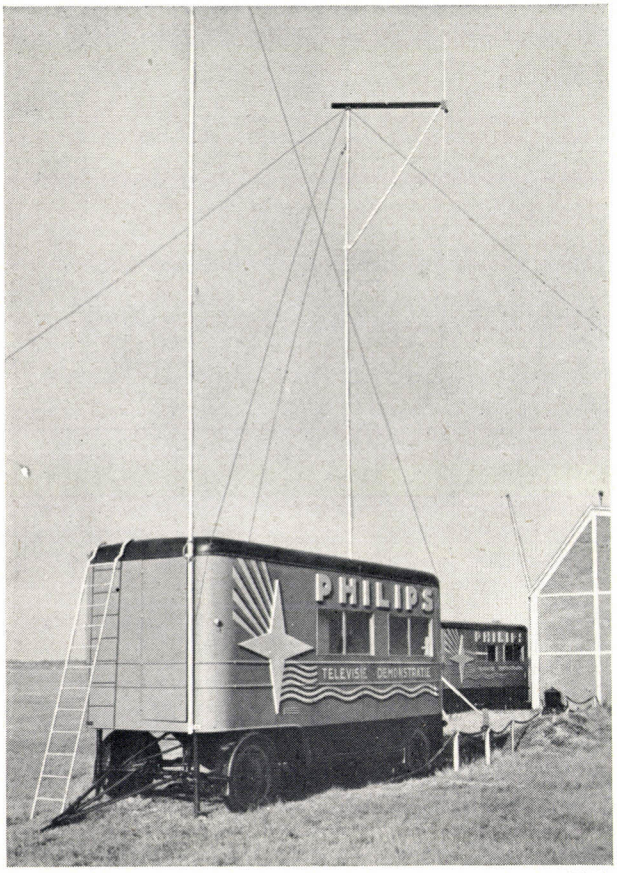
Important work has also been done in the field of the physiology of hearing. Particular mention is to be made of J. F. Schouten's investigations into the validity of the so-called *Ohm's* acoustic law, from which it appeared that in a mixture of frequencies which are a multiple of a certain fundamental frequency the latter can sometimes be heard even if it is not itself present in the mixture (the "residue" theory).

Further, L. Blok designed various signal generators, which have been applied, inter alia, for audiometric investigations.

Other developments in the technique of sound reproduction will be dealt with in the last section of this review.

The development of television began already in the thirties. In its earliest stages a *Nipkow* disc with 48 lines was used both at the transmitting

and at the receiving end. Subsequently a cathode-ray tube was used in the receiver, whilst the *Nipkow* disc, still employed for the transmission, was perfected for televising films (Rinia). Later on the *Zworykin* iconoscope, meanwhile developed in the U.S.A., came to be used for TV transmission, first for the 180-line system and later for larger numbers of lines (Van der Mark, G. Hepp, A. Venis). Experience was gathered in the construction of amplifiers (J. Haantjes and others) and the building of transmitters (W. Albricht). Experiments



67083

In 1937 already a transportable installation for television transmission and reception was completed and taken on tour for demonstrations over a large part of Europe. It could be worked on the system of 405 lines or on that of 567 lines.

were also carried out with projection television by means of lenses, employing a cathode-ray tube with concave screen (M. Wolf). Complete mobile television apparatus was built and installed in two motor-vans, with which demonstrations were given in various places in Western Europe (Utrecht, Brussels, Antwerp, Copenhagen, Stockholm) and also in Eastern Europe (Budapest, Bucarest, Zagreb, Belgrade, Warsaw). Owing to the outbreak of the war these tours came to an abrupt end in 1939.

III. Chemistry

It would be saying too much to maintain that all the chemical problems tackled in the Philips laboratory in the course of time likewise emanated from the incandescent lamp, but for a very large part such is indeed the case, either directly or indirectly.

Disregarding the metallurgy of the tungsten and the making of the glass, in the manufacture of the incandescent lamp chemical problems arise from the action of the gases released from the wall of the bulb and the metal leads upon the glowing tungsten filament, whereby traces of water vapour have a particularly disastrous effect. In the dissociation of the water vapour by the filament oxygen is combined with the tungsten and forms a volatile tungsten oxide which is precipitated on the glass wall of the bulb. The remaining hydrogen in turn reduces this oxide to metallic tungsten. Thus, owing to this "*Langmuir*'s cycle", tungsten is continuously being transported to the wall of the bulb, so that gradually this is blackened. It is therefore essential to remove water vapour and other gases, and this is done by introducing into the bulb a substance, such as phosphorus, which combines with the residual gases into a non-volatile and harmless product. Such a substance is called a "getter". In radio valves, where phosphorus cannot be used on account of its high vapour pressure and residual gases are moreover obnoxious because in the ionized state they affect the oxide-coated cathode and give rise to grid currents, and in X-ray tubes, where residual gases increase the risk of breakdown, metals like barium and zirconium are used as getters.

Even though the bulb of an incandescent lamp may not contain any gases attacking the filament, still evaporation of the tungsten takes place. In course of time this evaporation likewise turns the bulb black and thus the yield of light is reduced. Furthermore, the filament itself is reduced in thickness and as a result eventually collapses. The same applies in the case of the gas-filled lamp. The prevention of the evaporation by the gas-filling as such is offset by the filament being heated to a temperature so much higher that the lamp has about the same lifetime, so that the only advantage gained is the higher efficiency.

The manner in which this evaporation of the tungsten takes place differs according to whether the lamp is filled with gas or evacuated. In a gas-filled lamp the tungsten atoms come into collision with the gas molecules time after time, so that they have an opportunity to combine into aggregates which move about in the gas in the form of sub-

microscopical flocculations and eventually settle upon the wall of the bulb. Owing to convection this shows a preference for that part of the lamp which in the burning position is uppermost, thus not the part through which most of the light passes. With the vacuum lamp the situation is different: the tungsten atoms travel in a straight line from the filament to the inner wall of the bulb. It can be imagined that upon striking the glass wall the tungsten atom is repelled like a minute metal ball, but we know that ultimately it adheres to the glass, so every time the atom strikes against the wall it must lose some of its kinetic energy. *Langmuir*'s conception was that upon collision with the wall the atom temporarily adheres to it and is then, as it were, again evaporated, thus implying a certain "adhesion time". By accurate experimentation this has been confirmed by Clausing. It is true that in the case of cadmium atoms only an upper limit (10^{-6} sec) could be found for this adhesion time, but in the case of argon atoms on glass at temperatures of 80 to 90 °K adhesion times of 10^{-4} to 10^{-5} sec were found.

This matter of the evaporation of tungsten has been dealt with at some length because the investigations carried out in connection therewith formed an introduction to further investigations into the adsorption of tungsten atoms and the resultant absorption of light. So long as the tungsten atoms remain isolated on the wall of the bulb there is no appreciable absorption of light, but the position is different when they form a continuous layer of metal. In order to minimize the absorption of light, therefore, before the mount is fused into the bulb the filament is sprayed not only with the getter but also with a little salt, say CaF_2. As soon as the filament is heated to a high temperature this salt evaporates and is precipitated on the wall of the bulb as an invisible thin layer. The tungsten atoms coming from the filament shoot into this layer of salt, the particles of which keep the atoms separated, so that very much less light is absorbed.

About 1920 little was known with certainty about the effect of such layers of salt. Investigations into their action carried out in Philips laboratory, mainly by J. H. de Boer, extended over a period of more than 15 years. It has thereby been found that the layer of salt is not to be regarded as a homogeneous mass but as an agglomeration of minute crystal lamellae about 10^{-6} cm thick lying criss-cross one on top of the other and thus forming a very large active surface, tens of times greater than the surface of the glass covered by them. The tungsten atoms are adsorbed on the surface of the crystals. The intensive study of the adsorption of atoms and molecules (e.g. caesium and iodine) in such layers of salt (De Boer, C. J. Dippel, C. F. Veenemans) has yielded very important results.

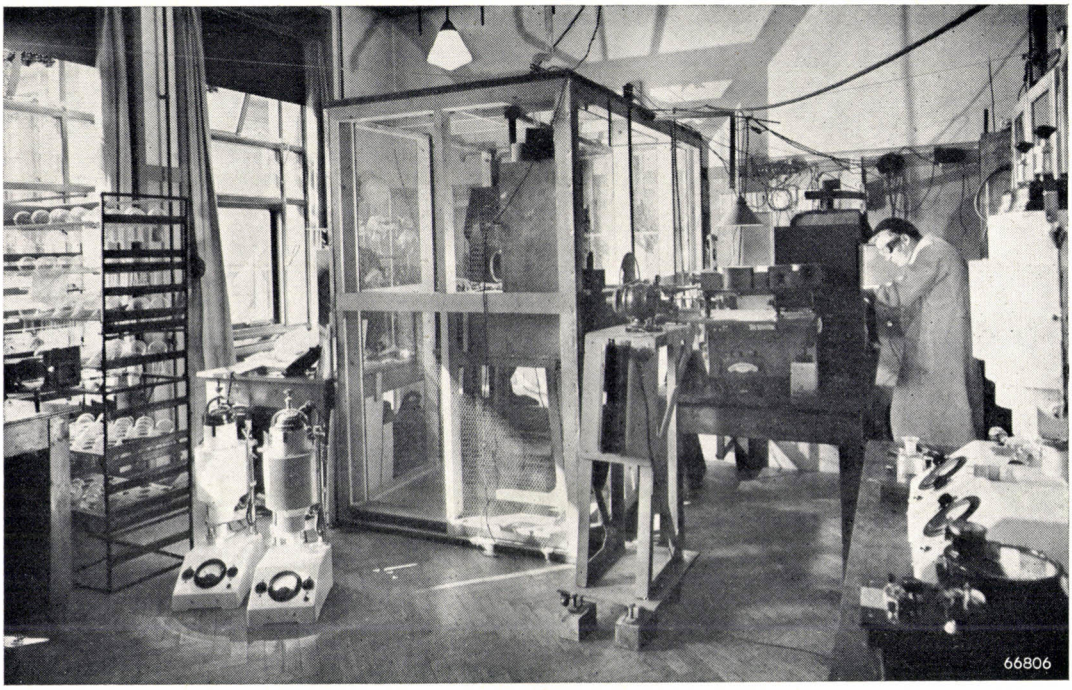
Metals like caesium may be adsorbed also on the surfaces of metals, in which case they have the property of reducing the work function of the metal and thus increasing the electron emission, both the thermionic emission and that brought about by irradiation with light — photo-electric effect. By oxidizing the adsorbed layer and again precipitating caesium onto the layer of oxide it is possible to produce complex layers with an exceptionally strong photo-electric effect (M. C. Teves). The study of this emission was of importance for the construction of photocells, for which there was a need in connection with the sound film. The deeper insight gained into the nature of these layers opened up new possibilities, such as the employment of the "light transformer" for converting infra-red rays into visible light (Holst, De Boer, Teves), the construction of the electron-multiplying valve and, later, the television pick-up tubes.

The investigations into the vaporization of CaF_2 also led to other compounds, such as H_3BO_3, K_2BF_4, being tried out for the same purpose, as a result of which the volatility of various compounds was studied. Why, it may be asked, is NaCl for instance a substance having a high melting point and a negligible vapour pressure at room temperature, whereas WCl_6 is a substance that readily melts and is easily evaporated? From the point of view of the theory of heteropolar chemical compounds the answer is simple. The molecule of NaCl is built up from a positive sodium ion and a negative chlorine ion, which together form an electric dipole, so that two NaCl molecules exercise a very strong attractive force upon each other. In the case of WCl_6, on the other hand, the central metallic ion is surrounded by six chlorine ions screening off the charge of the central ion, so that there is no great attraction between neighbouring molecules. The theory of the heteropolar chemical bond was formulated by *Kossel* in 1920, following upon *Bohr*'s work. It is mainly due to the work of Van Arkel in this direction that a more or less coherent explanation was established for the most important facts in anorganic chemistry, an explanation that is of great value as a basis for chemical thought and for education in chemistry. It is fortunate that this was more or less completed before

the conception of the homeopolar bond came into more prominence as a result of the development of quantum mechanics. Thus the one-sided-heteropolar aspect could gradually be introduced into the new system without losing its value.

In connection with these investigations the energy of formation of a number of molecules and crystal lattices was calculated and the relative stability of various molecule models investigated.

known that WCl_6 and hydrogen on the surface of a heated tungsten filament may react one upon the other and form tungsten, which is deposited on the wire. When a mono-crystalline wire is taken (*Pintsch* wire) also the growing metal becomes a mono-crystal. In the Philips laboratory it was found that dissociation of WCl_6 at elevated temperature takes place also without hydrogen; attempts were made to produce other metals in the pure



Measuring the spectral transmission of various materials in the ultra-violet range. On the right the source of ultra-violet radiation and the monochromator. In the centre, in a screening cage, the measuring apparatus with photoelectric cell.

The above conceptions were also applied to the relation between physical properties of homologous organic compounds, as for instance the relation between the boiling points of CH_4 and of the compounds obtained when replacing in CH_4 one or more hydrogen atoms by atoms of F, Cl, Br or I. Subsequently, after Van Arkel's appointment as professor at Leyden, this investigation was extended by him and his students to a large number of other compounds. Also worthy of mention is the study of the dielectric behaviour of organic dipole molecules in solution, which led, inter alia, to the "Van Arkel and Snoek formula", which was later investigated theoretically by *Onsager* and by *Böttcher*.

The study of the volatility of metal compounds had also important technical consequences. It was

state in this way. Perhaps the most striking result was the preparation of titanium, zirconium, hafnium and thorium from their iodides by precipitating the latter on a thin tungsten wire as core (De Boer and J. D. Fast). In this way titanium and zirconium, known as being greyish brittle substances of a doubtful metallic character, were obtained in the form of fine lustrous metallic products. The rods, consisting of a few large crystals, proved to be highly ductile, so that they could be drawn into wire and rolled into foil. This is particularly of importance in the case of titanium, considering the interest taken in this metal in recent years. Ductile zirconium has found important applications in vacuum technics. Applied to the anode of a transmitting valve it proves to be an excellent getter, on account of the almost unlimited capacity

of this metal to adsorb oxygen and other gases. The oxygen taken up in the metal has a negative charge, as appears from the fact that in an oxygen-charged rod of zirconium through which a direct current is passed the oxygen migrates towards the positive pole.

The knowledge of metals and their interaction with gases proved to be of great value to the laboratory staff when it was decided to manufacture welding rods (J. Sack, P. C. van der Willigen), in addition to the welding rectifiers and rectifying valves already being produced.

The study of tungsten wire and the behaviour of tungsten in processing led to extensive research in the domain of recrystallization. A substance particularly suitable for this is aluminium, which is easily deformed and recrystallized at comparatively low temperatures (< 600 °C), while the process of this recrystallization can be followed by etching in aqua regia after removal of the superficial oxide with caustic soda or hydrofluoric acid. This simple technique, supplemented by crystallographic study (W. G. Burgers, J. F. H. Custers), led to a deeper insight into the essence of the formation of crystal nuclei. This study is still being continued at the present day by Prof. *W. G. Burgers* in the inorganic-chemical laboratory of the Technical University at Delft.

Finally, in connection with applications in the field of electrotechnics and acoustics (transformers, loudspeakers), extensive research has been carried out in regard to the magnetic properties of metals, particularly of iron and iron alloys (G. J. Sizoo, W. F. Brandsma, Elenbaas, Jonas, Snoek, Six, G. W. Rathenau, J. J. Went, H. J. Meerkamp van Emden).

Special products, the fruit of years of study in this domain, were the rolled nickel-iron ("Fernicube") with strong anisotropic properties, for loading coils, and magnet steels with exceptionally high coercivity and high value of the (HB_{max}) product, which are widely used not only in loudspeaker magnets but also in pocket and bicycle dynamos.

Of importance for the investigations both of magnetic and non-magnetic metals was the work done by Snoek in studying the magnetic after-effects of iron, whereby it was found that these effects are related to the presence of traces of carbon and nitrogen, which influence not only the magnetic but also the elastic after-effects. Both these are governed by a sort of diffusion of C and N atoms present in interstitial places to neighbouring interstitial places.

Other materials studied at the time in the laboratory are the ferrites, which since 1934 have been further investigated by Van Arkel, Snoek and E. J. W. Verwey.

Preparation of magnet steel. Emptying the furnace in which the alloy has been melted by induced high-frequency currents.

This led, inter alia, to the view (De Boer and Verwey) that a substance like Fe_3O_4 derives its conductivity from the fact that ions of one and the same element but of different valency are present in crystallographically identical places and thus make it possible for an electron to pass over from one ion to another.

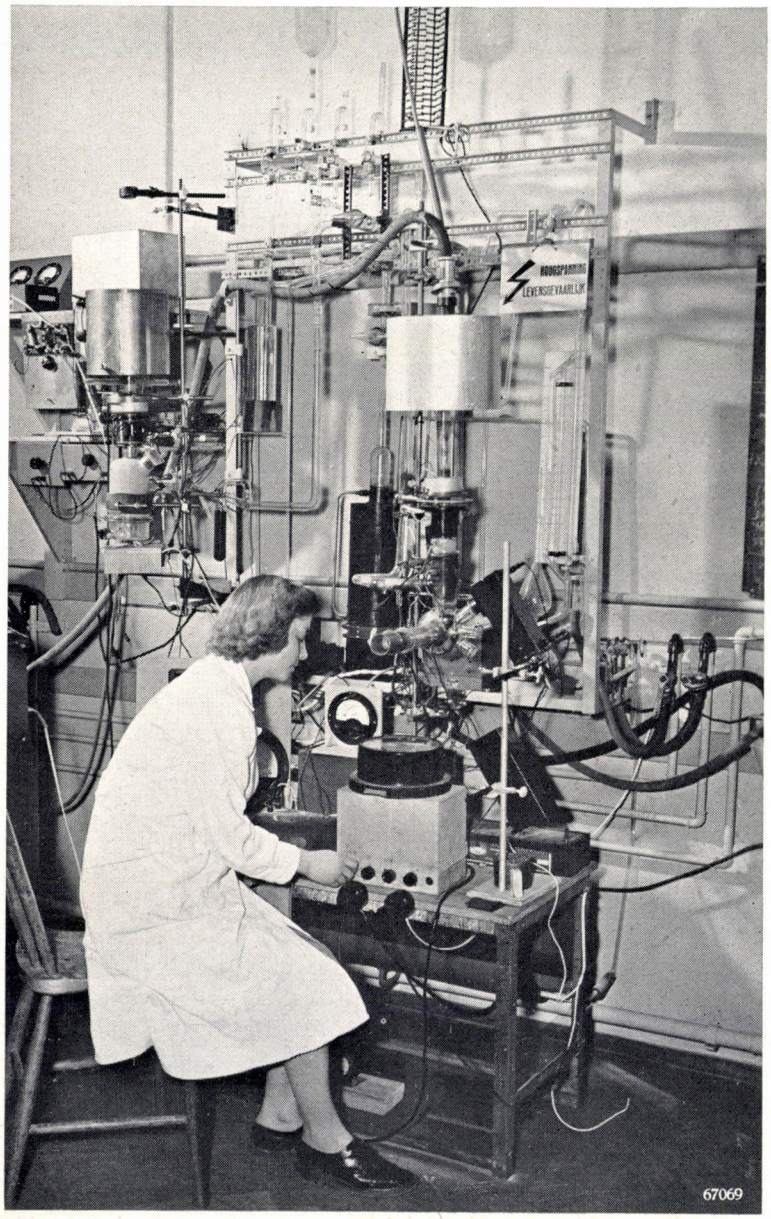
In the course of the resultant investigations our research workers became familiar with various problems of the solid substance, in particular with oxidic systems, of which the spinels form an interesting sub-group. The latter were further studied both in respect to their electrical properties (Verwey and others) and with regard to their magnetic properties (Snoek). The study of the electrical properties led to the development of semi-conducting materials, while that of the magnetic properties yielded the important magnetic material "Ferroxcube", which will be dealt with later.

The previously mentioned work by De Boer and others on the adsorption of atoms on crystals was extended to cases where atoms are built into a crystal lattice, a subject to which much attention was devoted in Germany by *Pohl* and his pupils. What is particularly due to De Boer is the disclosure of the fact that the place of a lacking negative ion in an ion lattice may be occupied by an electron and that, at the cost of only a little energy, this electron may be brought into a state in which it acts as a conduction electron.

It was at that time that the programme of work on solids was given shape, which later on was to occupy the minds of such a large part of the scientific staff of the laboratory and comprised, among others, the investigation of luminescent substances.

The luminescent substances (luminophores or phosphors) used for converting the ultra-violet light of gas-discharge lamps into visible light are prepared according to methods evolved by *Lenard* round about 1890. These investigations in the field of luminescence are to be regarded as classical and have not yet lost any of their value. They revealed in particular the fact that the luminescence of substances like zinc sulphide is due to the presence of extremely small admixtures of certain metals, such as copper and silver (activators).

When it became evident that large quantities of luminescent materials would be needed for lighting purposes Philips began to take up the manufacture of these substances and, at the same time, the study of the phenomena of luminescence. This study covered both the physical aspects — such as the spectral composition and the intensity of the luminescent light as a function of that of the incident rays, the decay of luminescence as a function of time in the case of discontinuous irradiation, the relation between luminescence and temperature, luminescence under the influence of cathode rays and X-rays — as well as the chemical aspects, such as the influence of the composition and of the conditions during preparation upon the luminescence, and the influence of admixtures (quenchers, sensitizers) upon the intensity of fluorescence. Often the chemical and the physical problems are so closely interwoven as to be inseparable, so that close cooperation between physicists and chemists or the combination of physicist and chemist in one person is essential for these investigations.



Phosphors are exposed to an electron beam in a vacuum tube and tested for their fluorescing properties.

Important physical results lay in the deeper insight thereby obtained into the mechanism of fluorescence and phosphorescence, the transfer of excitation energy in phosphors and the relationship between persistence and quenching as a function of temperature (F. A. Kröger, H. A. Klasens).

Chemical results lay in the deeper insight gained into the properties of zinc sulphide and related compounds and of substances such as Zn_2SiO_4-Mn_2SiO_4 and other manganese phosphors (Kröger).

In the foregoing some cases have already been mentioned where organic compounds were studied. Mostly these concerned simple organic molecules which are very closely allied to the inorganic compounds.

But work has also been done in the field of organic chemistry proper, namely in that of compounds with large molecules, such as gelatin, cellulose, proteins and synthetic resins. Gelatin was thoroughly studied because this substance serves as base for the Philips-Miller tape (Dippel, Vermeulen) used for sound reproduction; this tape consists of a celluloid carrier with a transparent coating of modified gelatin, covered by a non-transparent layer of HgS-sol in which the cutter of the recording apparatus traces a "sound track".

Artificial resins were originally studied with a view to the possibility of using these materials for making the bases of radio valves and loudspeaker baffles, and later for making various parts and the cabinets of radio sets from these materials. The phenols hardened with formaldehyde, which had been named bakelite after the Belgian chemist *Baekeland*, were placed on the market by Philips in a large number of varieties under the trade name "Philite".

A third group of organic compounds, the diazo compounds, was intensively investigated in the Philips laboratory when it was contemplated to produce dye-line paper, in addition to the mercury lamps destined for the dye-line process. With the aid of these compounds in various carriers, such as cellophane, paper, cellulose esters, etc., materials for photographic reproduction were produced which, via conversion of the products of dissociation through light into a developable latent metallic image, ultimately yield silver images with a resolving power of 1200 lines per mm, which appear to present interesting possibilities of application (Dippel, R. J. H. Alink, K. J. Keuning).

Attention was also given to the study of colloid-chemical problems. One result of these investigations, which was of importance for the manufacture of radio valves, was a new method of coating cathodes with a layer of oxide, covering them by means of electrophoresis with finely distributed carbonates of barium and strontium, which are afterwards turned into oxides. Also the insulating layer around the filament of indirectly-heated cathodes is applied in this way (De Boer, Verwey, H. C. Hamaker).

This concludes the review of chemical research in the Philips laboratory prior to 1940. A number of investigations carried out with metals and with non-magnetic and magnetic ceramic materials, glass and semi-conductors will be dealt with in the last section.

IV. X-rays

From the radio valve it is but one step to the X-ray tube. Any diode is in principle a source of X-rays, and it does in fact become so when the electrons strike the anode with sufficient energy. With the X-ray tube problems arise which are similar to those encountered with the radio valve, such as the shaping of the electrodes with respect to the nature of the electric field, the paths followed by primary and secondary electrons, and the focusing. To these are added the typical problems connected with high tensions, as for instance the efficient distribution of potential differences. Since only a small fraction of the electron energy is converted into X-rays and the rest is absorbed in the anticathode in the form of heat, it is a great problem how to carry off that heat; by providing for sufficient dissipation of this heat and giving the anticathode a suitable construction it has to be ensured that in the focus the material struck by the electrons does not melt.

Of physical importance is the measuring of the intensity of the X-rays and the strength of the dose. Much attention has therefore been devoted to this problem by A. Bouwers, who took up X-ray research in the Philips laboratory in 1920.

In the construction of X-ray tubes one is confronted with such problems as the safeguarding of the users of X-ray apparatus against scattered rays and high tension, the raising of the specific load, the requirement of easy handling, and the applications in connection with a proper formulation of the optical requirements which the apparatus has to answer; these applications may be of a medical (diagnostics and therapy), a physical (examination of crystals with X-rays) or a technical nature (examination of materials).

An important discovery was the possibility of fusing glass to metal, for which the chrome-iron already mentioned was used. This made it possible for Bouwers to build rugged X-ray tubes, which, since the part where the rays are generated is entirely enveloped in metal, afforded the maximum of safety against undesired radiation and, moreover, allowed of a simple safeguarding of the user against high tension by earthing the metal shield.

Tubes came to be developed for still higher voltages and for larger powers, both for diagnostic purposes and particularly for therapy.

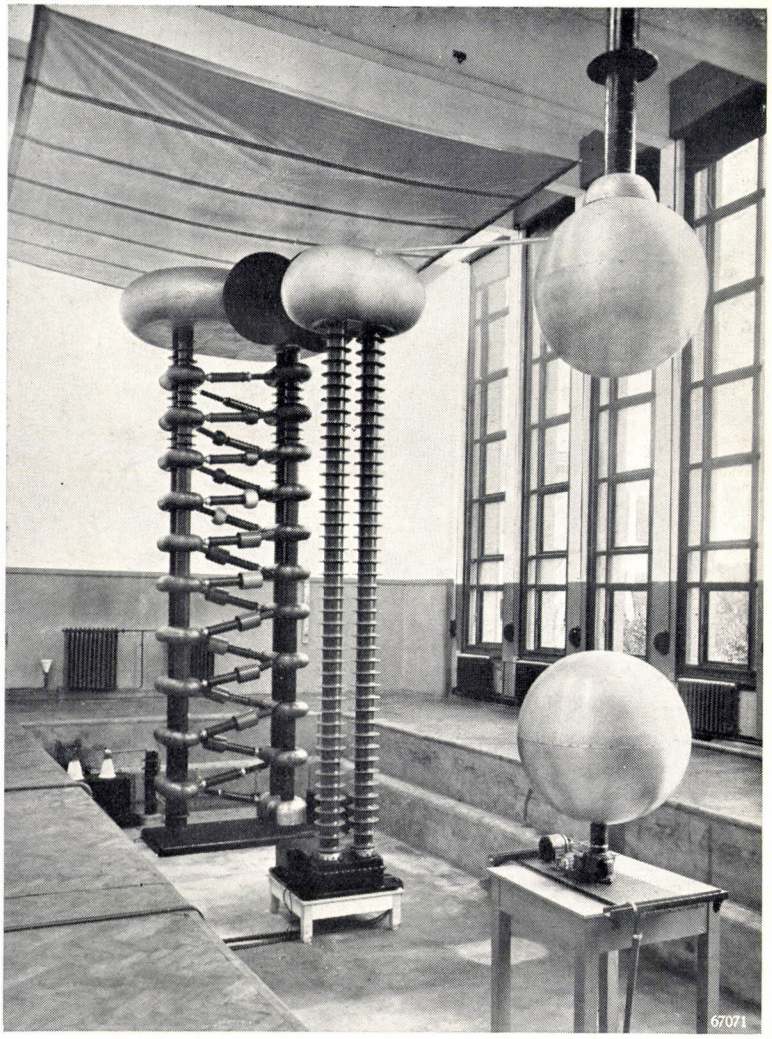
A special form of X-ray tubes which should be

mentioned is the rotating anode tube, which was first developed by Bouwers in cooperation with J. H. van der Tuuk. The problem of the dissipation of heat in these and other anodes was thoroughly investigated by W. J. Oosterkamp.

Further, special tubes were designed for crystallographic research with the aid of X-rays, in which the anticathode was made of special materials, such as molybdenum, copper, iron, etc., because these are required to yield approximately monochromatic rays of a known wavelength.

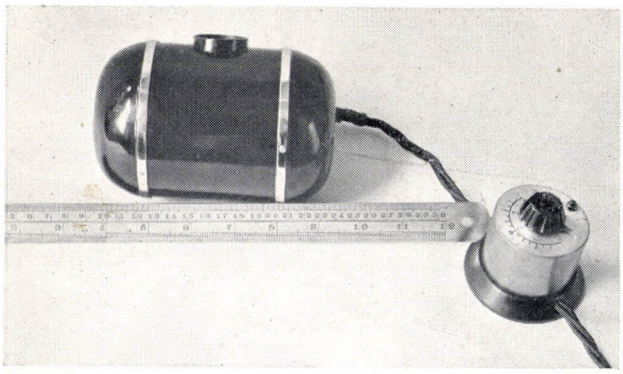
In the beginning all X-ray tubes were made in the laboratory, but later a separate factory was opened. The main features in the manufacture of X-ray tubes are well studied and well applied technology (getters!) and the exercising of extreme cleanliness in handling the materials.

Apart from the development of the X-ray tubes themselves, that of the apparatus for supplying the high tensions required for the working of the tubes, as also that of the control desks, forms a considerable part of Philips' activity in the field of X-rays. Often alternating voltage has to be transformed into direct voltage, for which special rectifying valves are needed, which formerly had a tungsten cathode but now have either an oxide-coated cathode or one of thoriated tungsten.



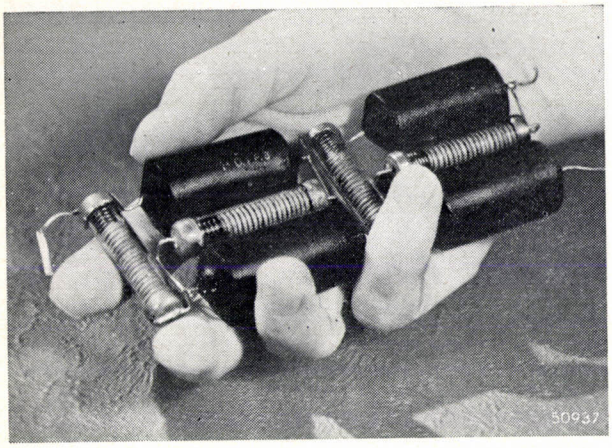
One of the first practical executions (1937) of a cascade generator, for 1.7 MV direct voltage.

Side by side with the aim towards higher tensions and greater powers, provision was also made for cases where a relatively low voltage and a small power suffice. Small X-ray apparatus was therefore developed in which the tube and the transformer are incorporated in one single unit. Such units serve as portable apparatus for diagnostic work and, for example, as X-ray apparatus in dentistry. It has even been possible to produce a complete X-ray apparatus so small that it can be carried in the pocket, and yet it yields a satisfactory beam of X-rays.



One of the first experimental X-ray units of very small dimensions (1933), a forerunner of the "Centralix" and "Oralix" apparatus.

From investigating current sources for high voltages Bouwers arrived at the principle of voltage multiplying by means of a cascade circuit. Subsequently it appeared that this principle had already been recorded by *Greinacher* and that it had also been applied by *Cockcroft* and *Walton* for nuclear-physical research in the Cavendish laboratory. As a special feature of Philips' development in this field is to be mentioned the elegant solution of the heating of the filaments in the valves by high-frequency current (A. Kuntke). A number of high tension generators of this type, for tensions

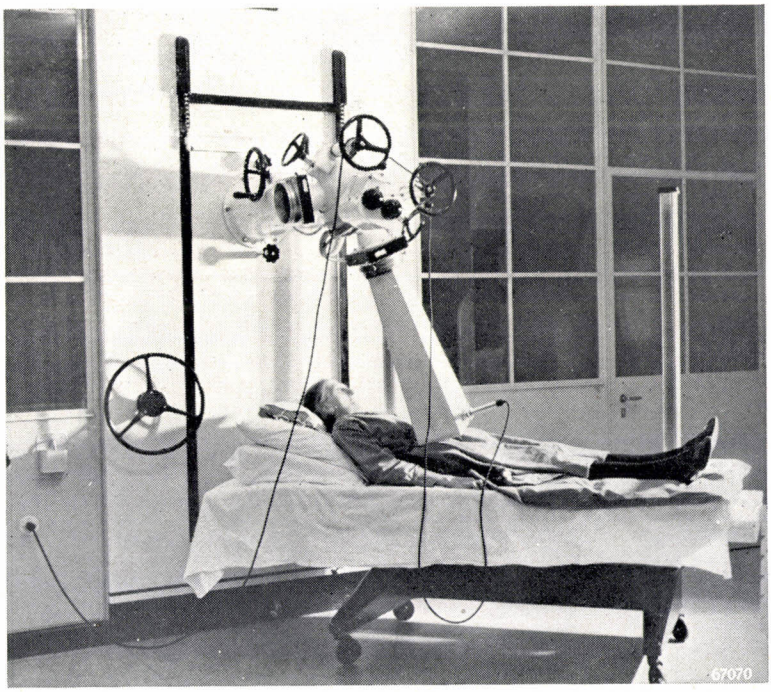


Small rectifier in cascade connection, consisting of selenium valves and capacitors. Connected to 220 V alternating voltage it yields 1200 V direct voltage, with which, for instance, a radiation counter tube can be fed.

up to 2 MV, have been built at Eindhoven for institutions in several countries.

Once it became possible to generate high tensions it was only a logical sequence to take up also research in nuclear physics. Bouwers and F. A. Heyn designed an ion-accelerating tube with which the ions of deuterium, formed in a separate ion source, could be accelerated to an energy of 1.2 MeV and focused upon a target of beryllium or lithium. Thus a powerful neutron source is formed by means of which materials placed in its vicinity can be made radioactive. In order to gain experience in the working of this apparatus, scientific nuclear research work was carried out with it. In the course of that work Heyn discovered the $(n, 2n)$ reaction with the elements Cu and Zn, by which reaction the nucleus is struck by one neutron and yields two neutrons, this being accompanied by the formation of a radioactive isotope of the same atomic number but lighter than the basic isotope. In 1939 A. H. W. Aten, Bakker and Heyn studied the transmutation of uranium and thorium by neutrons.

In this connection mention is also to be made of a neutron tube without a separate ion source constructed by Penning, whereby deuterium ions are accelerated in a gas discharge and the target is placed in the discharge tube itself. In order that sufficiently high voltages can be applied and the free path of the ions made large enough, the tube is filled with deuterium under a very low pressure and the discharge space is placed in a magnetic field so as to constrain the electrons to follow long paths, just as in the case of the manometer described earlier, thus making it possible for the discharge to be maintained under the low pressure.



Installation for X-ray therapy, working with a voltage of 400 kV, supplied in 1941 to the Academic Hospital at Groningen.

V. Mathematics and theoretical physics

With the growth of the laboratory more and more interest was taken in theoretical-physical and mathematical research. Whereas on the one hand, besides development work for practical applications, experimental physical research is

necessary to obtain a deeper insight into the problems, so on the other hand fruitful experimentation is only possible when at the same time the necessary attention is devoted to theoretical physics. This implies that the necessary mathematical apparatus has to be mastered and where possible further developed. This applies more or less to each of the groups already dealt with. Both the study of gas discharges and radio research, as well as the physico-chemical studies and the work on X-rays, had each in turn their own mathematical problems.

In regard to gas discharges there was the mathematical treatment by G. Hertz of the diffusion of electrons in an electric field. The method followed by Hertz has served as an example for many theoretical calculations in this field (Druyvesteyn, Penning, De Groot). This brought us a step nearer to the ideal: to explain the various forms of discharges with the aid of a small number of data concerning elementary processes (the probability of excitation and ionization in the case of collision of electrons and atoms, etc.), in much the same way as the kinetic gas theory relates the measurable quantities, such as pressure and temperature, to the elastic collisions between gas molecules.

Further it has already been seen, for instance, that with the aid of considerations of similarity the phenomena in discharges in mercury vapour of high pressure can be reduced to one single aspect. Once he had become familiar with this method of calculation, and following upon a publication by *Nusselt* (1916), Elenbaas was able to apply similar considerations to the phenomenon of thermal emissivity through natural convection. The need for this arose from the manufacture of blocking-layer rectifiers in connection with the dimensioning of cooling fins, but also other problems of convection can be considered in this light, such as the heat dissipation of horizontal cylinders, which brings us back to the gas-filled incandescent lamp designed by *Langmuir*.

The part of the work programme that lent most stimulation to the practising of mathematics was radio. Mention has already been made of important mathematical problems which arose in connection with the propagation of waves. Van der Pol took up the study of the wave equation and the potential equation, not only in three dimensions but also in the cases of fewer or more than three dimensions.

As is known, the propagation of waves can be treated in two ways, either as a whole, by the solution of a "wave equation", or by an approximative solution of the behaviour of a narrow beam or "ray". The latter method is well known from the theory of light; the image produced by a lens is not usually studied by starting from the representation of a wave but by investigating how the rays of light are refracted by the lens. In the case of radio waves Bremmer has investigated in how far results can be reached with this geometrical-optical approximation.

In addition to radio, particularly the theory of atoms has formed grounds for excursions into the field of mathematics. It is remarkable how closely the mathematics of these problems are associated with the mathematics encountered in the field of radio and acoustics. In 1925 *Schrödinger* showed, for instance, that the motion of an electron under the influence of an electromagnetic field has to be described by a wave equation which shows a certain resemblance to the equation representing the propagation of light or sound in an inhomogeneous medium, and that the classical consideration of an electron as a "particle" describing a "path" is to be compared to that conception as geometrical optics compare to wave optics. The quantity used by *Schrödinger* to play the part of "field strength" or "deformation" of the medium, and which he denotes by Ψ, has the property that $|\Psi^2|$ is a measure for the probability of finding a particle at a certain place. Owing to the similarity between these problems and those of the propagation of waves encountered in radio and acoustics, theorists in the respective fields soon understand each other and find interest in the results of each other's work.

In dealing with the problems connected with oscillations in networks or the propagation of radio waves one mostly has to do with more or less complicated differential equations. *Heaviside* showed that in many cases the differential symbol, $\mathrm{d}/\mathrm{d}t$, can be regarded as an algebraical quantity (usually represented by D), with which ordinary arithmetical operations can be carried out. In this way he was able to derive deep-lying results by simple means.

This operational or symbolic calculus was at first received with much scepticism, but later *Carson* (1926), for instance, found it to be justified. What *Carson*'s formula amounts to is the furnishing of a function $\mathrm{f}(p)$ corresponding to a function $\mathrm{h}(x)$ by the *Laplaec* transformation:

$$\mathrm{f}(p) = p \int_0^\infty \mathrm{e}^{-px}\, \mathrm{h}(x)\mathrm{d}x.$$

The shape of $\mathrm{f}(p)$ depends, of course, upon $\mathrm{h}(x)$. This function $\mathrm{f}(p)$ is said to the be "image" or the

"translation" of the function h(x) and it is denoted by the symbol:

$$f(p) \doteqdot h(x).$$

The functions x^n, e^{-x}, sin x and cos x, for instance, have as their images

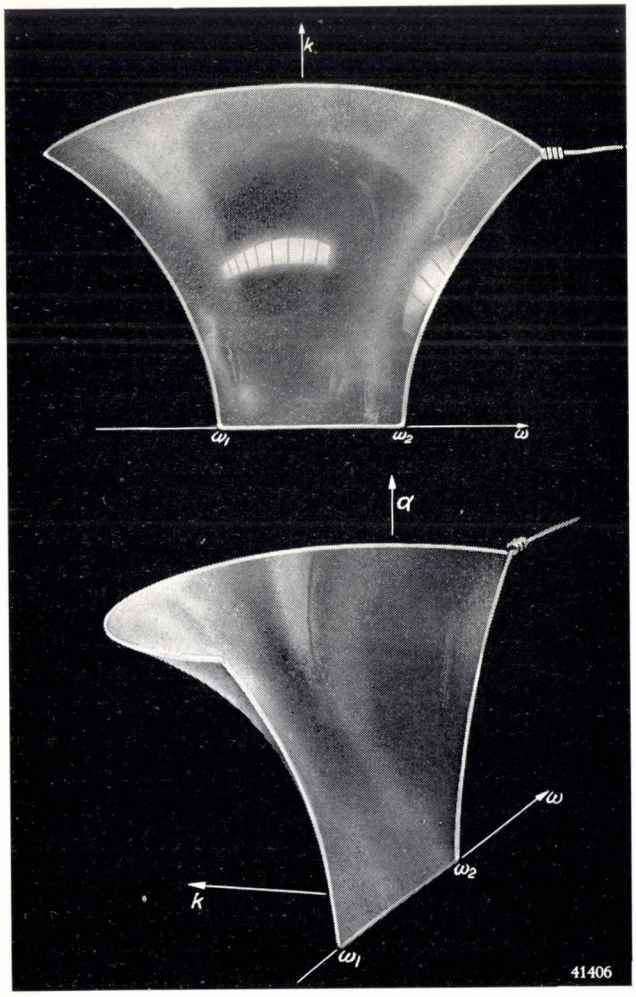
$$\frac{n!}{p^n}, \quad \frac{p}{p+1}, \quad \frac{p}{p^2+1} \text{ and } \frac{p^2}{p^2+1},$$

from which it is to be noted that often the image is a simpler function than the original. Thus relations between different originals can be deduced via simpler relations between their respective images.

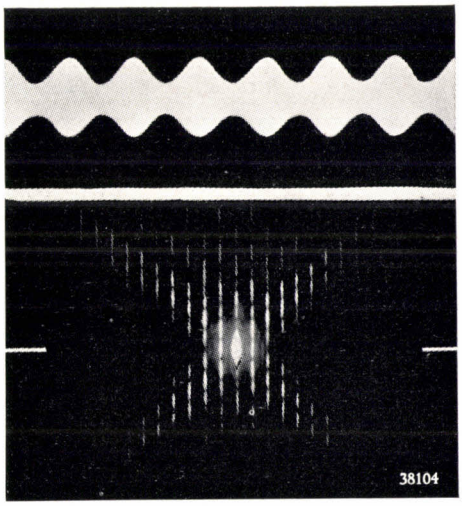
Van der Pol and Niessen have worked out a great many images and from them derived new relationships between functions. Mention has already been made of the problem of the aerial over the flat earth, and for that, too, the symbolic calculus was employed, with the further help of the methods followed in the theory of the functions of a complex variable.

These results were obtained with the aid of the one-sided *Laplace* integral, with the integration extended from 0 to ∞. Later, Van der Pol and Bremmer went deeply into a symbolic calculus based upon the two-sided *Laplace* integral (integration from $-\infty$ to $+\infty$).

The space available does not permit us to enter into details regarding other mathematical work in connection with radio and acoustics. We can only refer to the numerous publications by Strutt on acoustic and antenna problems, Niessen's calculations on aerials and cavity resonators, C. J. Bouwkamp's work on radiation properties of antennae and acoustic and electromagnetic diffraction problems, that of F. H. L. M. Stumpers and Th. J. Weyers on frequency modulation, Tellegen's studies of network synthesis, from which the "gyrator" subsequently appeared as a new network element, and finally Kleynen's investigations of electric fields in radio valves, etc.



In the course of development of carrier telephony it was necessary to investigate how the behaviour of electric filters is affected by losses. This behaviour is governed by the *Laplace* differential equation. A graphical solution of this equation, with given boundary conditions, is obtained by stretching a film of soap between three-dimensional curves corresponding to the boundary conditions. The case illustrated here relates to a bandpass filter (frequency limits ω_1 and ω_2); k is a measure for the losses, α a measure for the damping.



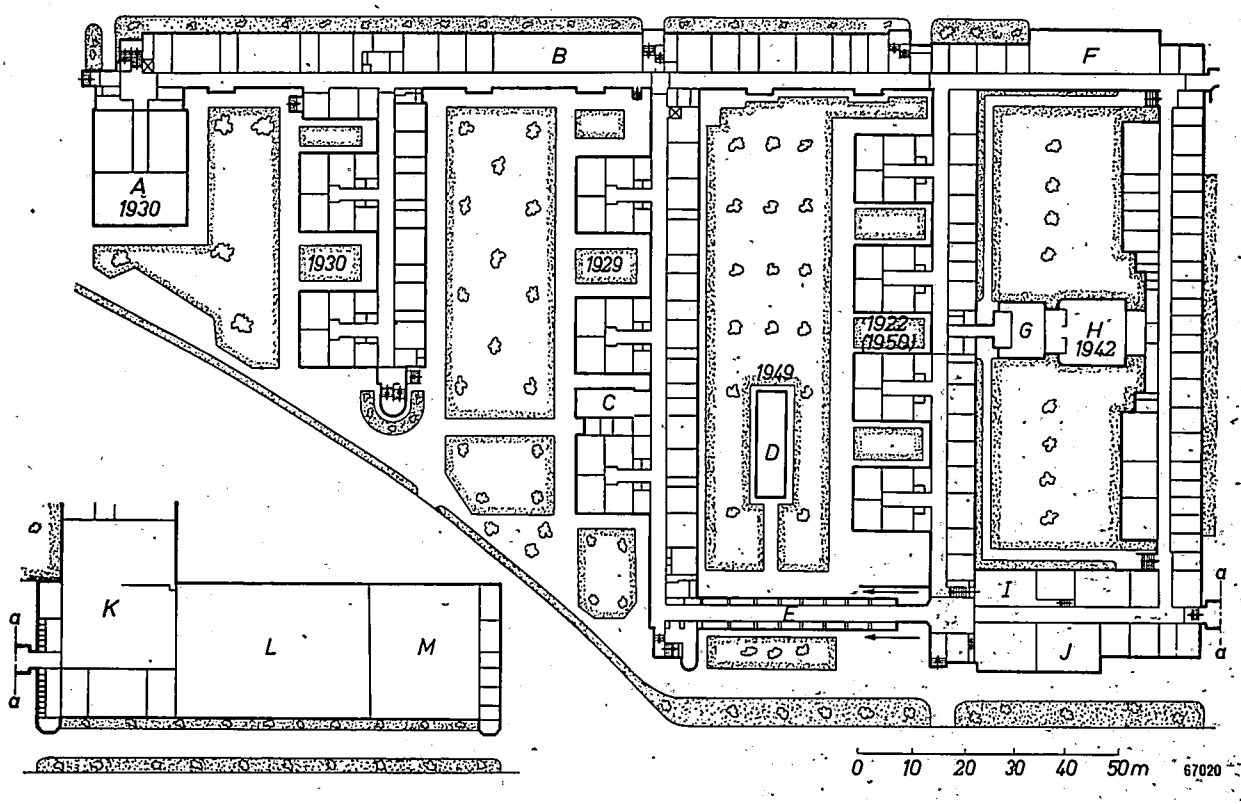
Philips-Miller tape (upper picture) in which a sound track (10 times enlarged) has been cut. When light passes through the tape a diffraction spectum is obtained (lower picture) from which the *Fourier* analysis of the recorded sound can be read. In this case (sinusoidal signal) the spectrum consists only of components of the zero and first orders.

As regards mathematical and theoretical-physical contributions outside the realm of radio, reference is to be made to the investigations of J. Haringx into problems of applied mechanics, the publications by Bouma and G. Heller on the geometry of colour space, and those by Niessen on diamagnetism and by Verwey and J. Th. G. Overbeek on the theory of lyophobic colloids.

THE PERIOD 1940-1951

In the foregoing section an attempt has been made to give a review of the scientific research work undertaken in the period 1923-1940, and incidentally mention has been made of the technical products that emanated from that work.

The development of this research suffered a check when on May 10th 1940 enemy forces invaded the Netherlands and shortly afterwards Eindhoven and the Philips' works came under military occupation.



Ground plan of the Physical Research Laboratory as it has been since 1942. *A* high-tension room, *B* material-testing department, *C* television studio, *D* horticultural glass-house, *E* small greenhouses with artificial climate, *F* one of the chemical departments, *G* one of the battery rooms, *H* library, *I* installation for carrier telephony, *J* engine room, *K* glass-blowing shop, *L* central workshop, *M* room for testing transmitting valves. The wing *K-L-M* adjoins the main diagram at *a-a* on the right. Most of the buildings have either one or two upper storeys.

It is not the place here to enlarge upon the course of affairs during the period of occupation, which, as far as Eindhoven was concerned, lasted until September 1944. As everyone will realize, the state of tension arising from war conditions and the frequent acts of injustice, coupled with more direct causes such as scarcity of foodstuffs, clothing and means of transportation (bicycle tyres), and further a number of air attacks on the works, repeated air-raid warnings, etc., all tended to create an atmosphere adversely affecting work. It would be wrong to suppose, however, that scientific research was thereby brought more or less to a standstill. On the contrary, many investigations not directly concerned with practical applications were widened in scope and in some fields important results were reached, though care was taken to keep them secret from the occupying forces. The long working hours imposed by the enemy administrators were further turned to use for the exchange of experiences in all sorts of domains by organizing lectures, courses of instruction, etc.

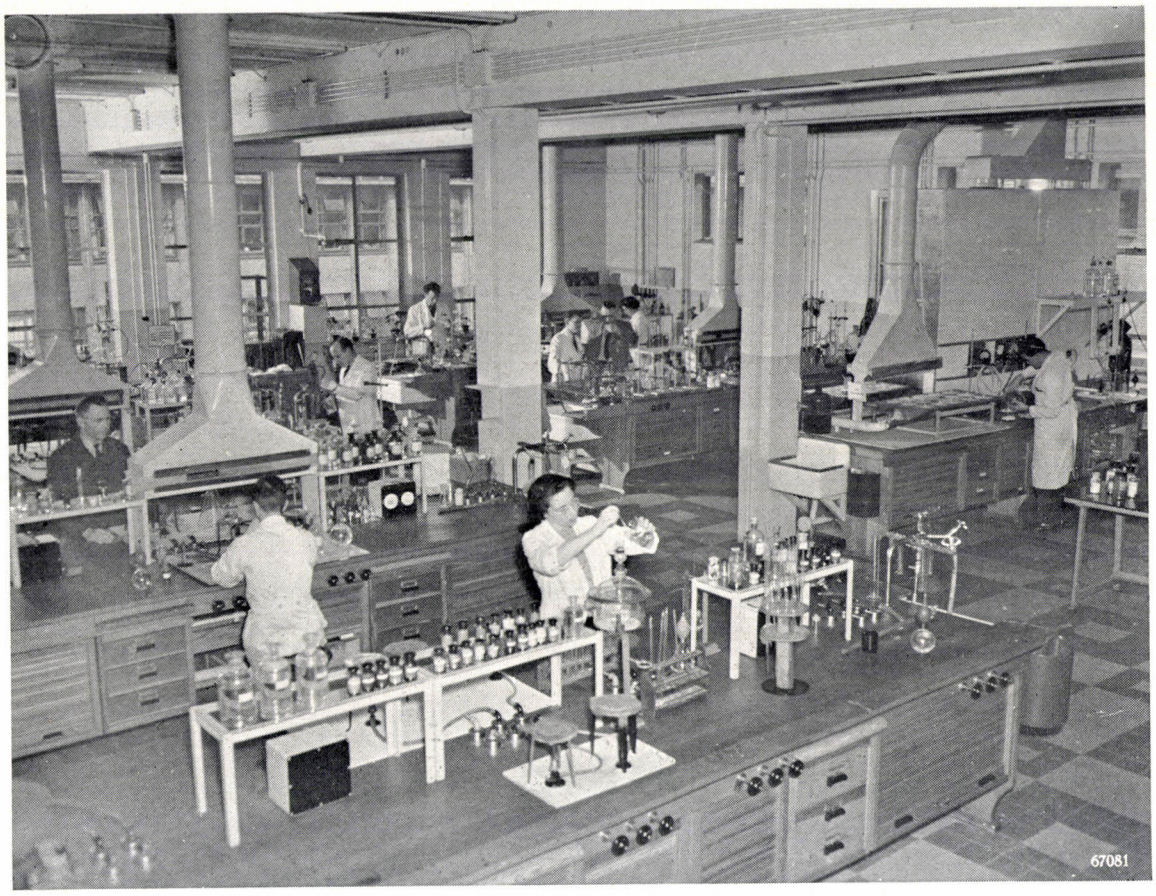
After the liberation of Eindhoven in 1944 it took some time before a return to normal conditions could be established. In the first half of 1945 the northern half of the country was still in enemy occupation and even after its liberation there was no regular contact with the rest of the country owing

to the lack of railway communications. In order to help university students, with permission of the government in February 1945 a temporary academy was set up at Eindhoven where a large number of Philips' scientists assumed the rôle of professor, and this continued up to November 1945.

In this connection it may be well to point out that relations between the Philips laboratory and the higher educational authorities have never been confined to that special occasion. Long before that there had developed in the course of time a more and more intimate relationship between the laboratory and the universities, including the Technical University at Delft, partly on account of the fact that quite a number of scientists have in course of time left our laboratory to take up a professorship.

During the occupation one began to feel more and more the lack of literature from the outside world, so that as far as scientific work was concerned one had the feeling of living as it were in a vacuum. The possibility of publishing was also greatly restricted. The publication of Philips Technical Review, begun in 1936, had to be stopped in 1942, and publications in the Dutch journal "Physica" could only be made in the native language.

It was not until 1946 that journals and books began to come into the country again from abroad. On 1st January 1946 Philips Technical Review appeared again. Meanwhile, in October 1945, the first number had been issued of a new publication under the name of Philips Research Reports, containing scientific articles which bear a decidedly Philips character or which owing to their volume cannot easily be published elsewhere. Furthermore, part of the material suitable for publication which had been collected during the years 1940 to 1945 was published in book form.



Department for analytical chemistry in the new part of the Physical Research Laboratory (1950).

Meanwhile the need of more space was being felt in the laboratory, which in 1929 had already undergone a considerable expansion increasing ten-fold the amount of floor space originally available in 1923. A new wing had already been built in 1942, for the administrative staff and the library, but

until 1945 the latter had been housed elsewhere owing to the risk of fire due to air attacks. In 1950 a second floor was built onto the oldest single-storey part of the laboratory first erected in 1922, and that new floor was destined mainly for chemical work.

In 1946 a change was made in the management of the laboratory, Prof. Holst, whose 25th year of directorship had been celebrated in the laboratory in 1939, retired from that function. As adviser to the concern, he has still at the present day a word to say, be it indirectly, in the course of affairs, whilst at the same time he is devoting himself to the interests of the Technical University at Delft, where he holds of the Presidency of the Board of Governors. The direct management of the Philips laboratory was placed in the hands of a triumvirate formed by the physicist Prof. Dr. H. B. G. Casimir, the electrotechnician Ir. H. Rinia and the chemist Dr. E. J. W. Verwey.

For reasons which will be understood, in this section it will not be possible to do full justice to all the research work which has been undertaken in the Philips laboratory in the course of the last decennium and which, in part, is still in progress. Neither will it be possible to mention the names of all those engaged in the various investigations in the laboratory.

More so than in the preceding sections, here attention will be paid to materials and products and their applications.

Materials

It has been seen that in the period 1923-1940 scientific interest was directed for a large part towards gases, particularly as carriers of electric discharges. In addition, in various ways more and more interest came to be taken in solids. Mention has already been made of the investigation of thin crystal layers adsorbing foreign atoms. Activity in the field of radio demanded a closer study of the dielectric properties of glass and other insulators. For illumination engineering, too, glass had become an important material, as a result of the demand, for instance, for ultraviolet-transmitting glass and for intermediate glasses for combining glass with quartz glass. For the fusing of glass to metal suitable kinds of glass and alloys were required. In the technique of lighting fluorescent substances came to be applied more and more, so that the phenomenon of luminescence had likewise to be studied in the laboratory. Further, attention has also been drawn to magnetic materials, such as magnet steels and ferrites.

Generally speaking it may be said that round about 1940 interest was centred upon solids and the problems relating to the solid state, both from the point of view of purely scientific investigations and in respect of practical applications. In the following a number of materials will be briefly dealt with, but since it will be mainly the results that will be brought forward it is well to point out once more that in many cases these results are due for a large part to purely scientific research. As an example may be mentioned the investigations into "induced valency", which led to improved methods of preparing semi-conductors and luminescent materials, and the insight gained into the structure of "spinels", which resulted in a greater variety of "Ferroxcube" products.

Luminescent substances

The investigation of luminescent substances begun about 1935 led not only to the results already mentioned but also to a wider knowledge of luminescent tungstates and molybdates and of some activators such as Mn, Ti and U in different states of ionization.

Of particular practical importance were the silicates activated with Ti and magnesium arsenate activated with tetravelent manganese. With the latter substance, which shows a strong red fluorescence, it is possible, for instance, to improve considerably the colour of the light from the small high-pressure quartz mercury lamps.

The investigation of sulphides gave a better insight into the part played by "fluxes", such as NaCl, used in the preparation of luminescent substances. It appeared that it is the Cl^--ions that make it possible, for instance, for monovalent Cu^+-ions to be "built into" the lattice. The same can be reached by introducing into the lattice, which consists normally of bivalent ions Zn^{++} and S^{--}, trivalent cations such as Al^{3+}.

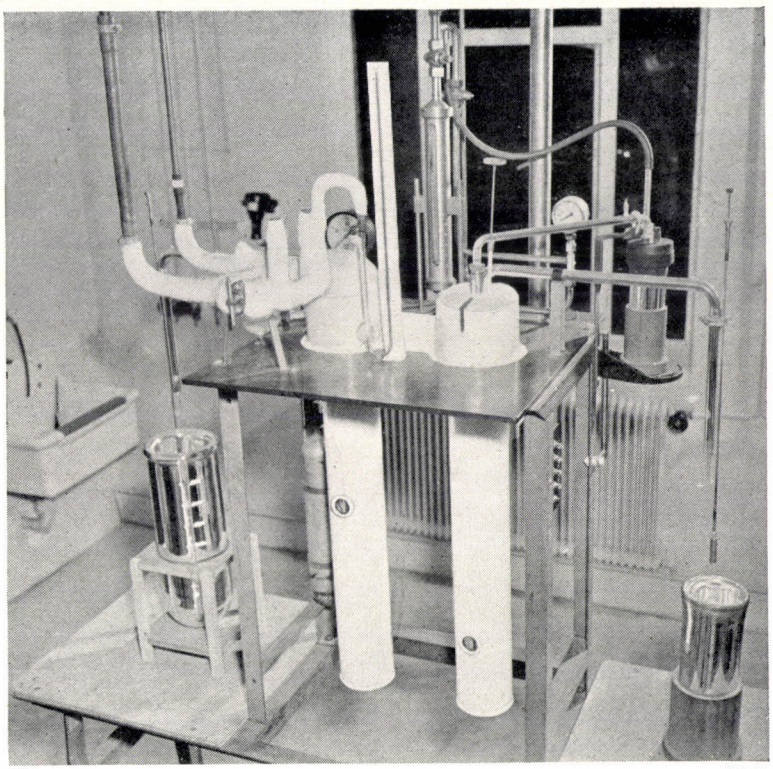
Results were also achieved in respect to luminescence brought about by bombardment with cathode rays; a deeper insight was gained into the phenomenon of saturation with increasing current.

For some purposes, such as radar, it is not the fluorescence but rather the phosphorescence (persistent after-glow) that is of importance. It was found possible to obtain a fluorescent screen with long persistence by making it in two layers, the first of which gives a blue fluorescence under the influence of the cathode rays, this blue fluorescence then being suitable for producing a green phosphorescence in the second layer. For other purposes, such as the televising of films,

substances with a short persistence are desired: a "field" produced on the fluorescent screen by an unmodulated electron beam is imaged onto the film, whilst behind the film is a photocell. This converts the light passed through, which is modulated by the variations in density of the film, into a signal which is passed to the transmitter.

ceramic materials (ferrites), because these have proved to be of great importance for high-frequency technique (radio and telecommunications).

Of theoretical importance was the interpretation of the losses which the magnetic ferrites show at very high frequencies (10^5 to 10^7 c/s): these losses were ascribed to the gyromagnetic effect



67068'

A simple hydrogen liquefactor has been constructed for the examination of solids at low temperature (20 °K). The hydrogen gas is fed in under high pressure and precooled to the temperature of liquid nitrogen, after which, through expansion, by employing heat exchangers according to the *Linde* method, it becomes partly liquid (*Joule-Kelvin* effect). This installation has an output of about 2 litres liquid hydrogen per hour.

Many other important applications of luminescent substances which have been investigated in the laboratory have to be passed over here.

Ceramic materials and glass

Ceramic materials are obtained by sintering together small particles which in themselves have a crystalline structure. We do not refer here to the more common materials like porcelain, steatite, etc., about which little research work has been done, but to special materials, as for instance those with a high dielectric constant, such as rutile (TiO_2), and titanates such as $BaTiO_3$, with which important work has been done in recent years.

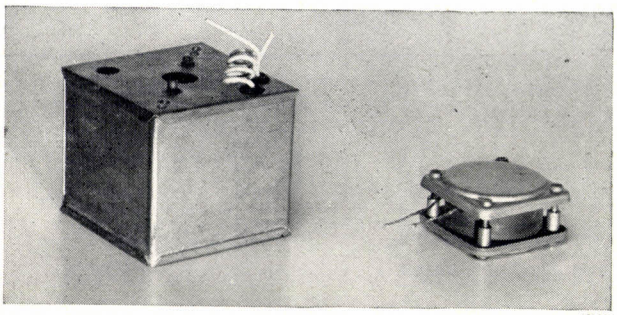
Particular mention is to be made of the magnetic

prophesied by *Landau*, and this conception was confirmed experimentally.

Of practical importance was the resultant knowledge gained of the fact that the less the initial permeability of the substance in the low-frequency range, and thus the greater its crystal anisotropy, the higher is the frequency at which the gyromagnetic losses become perceptible.

The ferritic materials, which combine great hardness and relatively light weight with a very small electric conductivity and favourable magnetic properties, such as low losses in a wide frequency range, are now being manufactured by Philips under the name of "Ferroxcube" in various compositions.

In addition to the ferrites, of equal importance are the ceramic semi-conductors, which will be dealt with below in a separate section devoted to semi-conductors.



48308

Improved quality and saving of space are important advantages of the magnetic material "Ferroxcube" as compared with metallic magnetic materials. *Left:* can containing a bandpass-filter coil for carrier telephony, the core of which consists of nickel-iron wires and the jacket of dust-core material. The can is necessary on account of the small effective permeability of the dust-core material. Quality factor $Q = 220$ at 60 kc/s, volume 210 cm³. *Right:* coil with core and jacket of "Ferroxcube", $Q = 600$ at 60 kc/s, volume 44 cm³.

Glass is a material of fundamental importance for the manufacture of incandescent lamps and radio valves. Apart from the obvious physical properties, such as the softening point, coefficient of expansion and specific gravity, there are many other properties to be considered, such as the dielectric constant, dielectric losses in high-frequency electric fields, electric conductivity, spectral transmission, etc. Considering the large number of ingredients from which glass is mostly made, and consequently the numerous possible variations in composition, it would seem to be an almost impossible task to gain a clear insight into the effect of the composition of a glass upon its physical properties. Yet in recent times considerable success has been attained in this direction, particularly due to the better theoretical knowledge acquired as a result, i.a., of the work done by *Zachariasen*. It is now possible to form an idea of the internal structure of glasses and from that to predict their properties. Even though the reality proves to be more complicated than the theoretical model, the latter anyhow points the way to approximating the correct relations of the phenomena and for investigating those relations which are most promising for gaining the insight desired.

In cooperation with the glass works the following practical results have been obtained: soft glasses with small dielectric losses, glasses which have small dielectric losses independent of frequency, glasses which under the influence of X-rays do not discolour or which, on the other hand, very readily change colour, and glasses which are transparent to ultra-violet radiation.

Metals

Metals form an important and extensive group of materials. The investigation of tungsten and molybdenum and of metals like titanium, zirconium, hafnium and thorium has already been mentioned, as also that of alloys such as chrome-iron.

The investigation of magnet steels has likewise already been referred to. The very important "Ticonal" (an alloy of iron, titanium, cobalt, nickel and aluminium) with a high value of the product $(BH)_{max}$ and great coercive strength, was further improved by making it anisotropic by cooling in a magnetic field or by some other means.

The nickel-iron with anisotropic structure obtained by rolling, for use in loading coils, has also been mentioned elsewhere. Extensive investigations have been carried out as to the manner in which the texture of this material could be influenced by a suitable thermal and mechanical treatment.

Another group of products is formed by the welding rods, the coating of which has been the subject of particular study. Special mention is to be made of the new method of contact welding, for which purpose the rods are given a special coating with a high iron content.

It is partly in connection with this that extensive investigations have been carried out into the diffusion of gases in metals. This subject had become of real importance as soon as the manufacture of water-cooled metal transmitting valves and metal X-ray tubes was begun, and it proved to be of particular interest in connection with welding problems, especially as regards the penetration of hydrogen into the metal of the welding bead, causing porosity, cracks and fractures. In this connection mention is also to be made of more fundamental research, as for instance that concerning ageing phenomena such as occur, inter alia, in welding.

The absorption of gases by metals has particularly also been studied in the case of metals like zirconium and titanium, which are used as getters in X-ray tubes and transmitting valves. Incidentally reference is to be made here to the coating of anodes of transmitting valves with zirconium to give them a black surface with good heat-radiating properties and also to reduce secondary-electron emission. Investigations of a more technical nature in the field of metals led to new methods of shaping and improved methods for

determining the properties of samples of metal, an example of the former being the so called "lost wax" casting of metallic parts of instruments and apparatus which would be too costly if the ordinary methods of processing were followed, whilst an example of the latter is to be found in the microhardness meter.

This section on metals will be concluded by a short reference to investigations concerning oxidation and corrosion. Interesting aspects are presented by the method of hardening alloys by internal oxidation. A study of corrosion at high temperatures revealed that small quantities of molybdenum oxide may have a harmful effect. Investigations into the non-oxidizing property at high temperatures and the solderability of non-oxidizing metals were particularly of importance in connection with the development of the hot-air engine.

Semi-conductors

Between the insulators of an inorganic and an organic nature, such as glass, porcelain, amber, polystyrene, on the one hand and the metals as good conductors for electric current on the other hand, there is an important group of substances which as compared with metals show little conductivity, e.g. iron oxides, copper oxide, selenium, germanium, etc. These are of interest because they possess typical properties lacking in the other two groups.

Their principal property is a large, negative, temperature coefficient of their resistance, a property found in practically all representatives of this group. Some of them also show photo-conductivity, in that when exposed to light their resistance is reduced. Others, when provided in a certain way with electrodes, appear to have rectifying properties, which means to say that when the current flows in one direction the resistance is many times, say 1000 or more times less than in the other direction. There are also certain materials whose resistance appears to be dependent upon the electric potential gradient.

Although these properties have been known for a long time it is only during the last 10 to 15 years that it has been found possible to utilize them. Obviously a high temperature coefficient of the resistance offers many possibilities. Mention may be made, for instance, of the measuring of temperature and of radiation; furthermore, owing to the negative sign of the coefficient, there is the possibility of suppression of current peaks and application in delay devices, regulating apparatus, etc.

A peculiar feature of these materials is that often their specific resistivity is not a property inherent in the material, as is the case, for instance, with a metal, but that it depends to a large extent upon the antecedents, such as the temperature and the gas atmosphere applied in the preparation, the presence or not of impurities and many other factors.

In the field of oxidic semi-conductors some extensive crystal-chemical investigations have led to the production of semi-conducting materials answering high requirements of reproducibility and stability. Such a material is being marketed in various forms as "N.T.C. resistors" (negative temperature coefficient resistors).

Much attention was paid also to selenium. The resistance of selenium can be varied by the addition of foreign substances. This investigation has led to the production of rectifiers suitable for low voltages, e.g. 18 V, and, by connecting them in parallel, for high currents. On the other hand very small rectifiers are being made for currents of a few milliamperes, as used in the amplifying technique, for example, for modulator cells, which play an important part in carrier-telephony, and in series-connected small cells for rectifying high voltages. By a certain treatment of the surface of the selenium and choice of the electrode material it is possible, to a certain extent, to determine the voltages which are to be applied per rectifying unit.

During the years 1940-1945 there has been renewed interest in the crystal rectifier, a component used in the early days of radio. It appears that in the range of ultra-short waves, which has become of such importance, this element in a more perfected form offers advantages over rectifying valves. Extensive investigations are proceeding in the laboratory also in this field and have already led to a rapidly increasing manufacture of various types of crystal rectifiers.

There are indications that the possibilities of applications in the field of semi-conductors are not by any means exhausted, and that, by the very reason of their abnormal properties, we are only at the beginning of a development of great importance for electrotechnical engineering.

Products

This review will be concluded by referring to some products which have particularly come into existence in the last ten to fifteen years as a result of laboratory work. The reader is not to expect a catalogue with detailed descriptions. Many of the

products which will be mentioned have already been described more or less fully in recent volumes of this journal, whilst a description of some others will shortly be appearing in our columns. Suffice it, therefore, to mention in a few words only the most important products.

Sources of light

In the field of lighting we have already referred to the "TL" lamps, which as such have already long passed the laboratory stage of development. Further investigations in connection with these lamps are being carried out in regard to their colour rendering, for instance with new phosphors used in their manufacture. Particular mention is to be made of the application of "TL" lamps for irradiating plants. Since 1949 the Philips laboratory has had a large horticultural glasshouse at its disposal, in which various plants are being cultivated, and also a dozen smaller glasshouses illuminated exclusively with artificial light. With the large variety of phosphors available it is possible to investigate how the growth and flowering of plants are affected by the spectral energy distribution and the intensity and duration of radiation.

Further, there is the low-pressure mercury lamp without any fluorescent material but made with a glass which is transparent to ultra-violet radiation (2537 Å), viz. the so-called germicidal lamp. Other special sources of light are the flashlamp for the illumination of *Wilson* chambers and for other photographic instantaneous exposures, and a point source of light in the form of a gas-discharge lamp the light of which can be modulated up to high frequencies.

67078.

Glasshouse for biological investigations.

Radio, television, etc.

In the field of radio conditions had already reached such a stable state before 1940 that the normal development of valves and apparatus no longer belonged to the work of the laboratory, so that attention could be directed towards new developments. In the first place mention is to be made of the new valves for ultra-short waves (decimetric and centimetric waves). Development of these valves had already reached an advanced stage abroad during the war years, so that in 1945 Philips had a great deal to catch up with. Extensive investigations had already been carried out with radio valves in the metric wave range. Special tubes were developed based upon the velocity modulation of

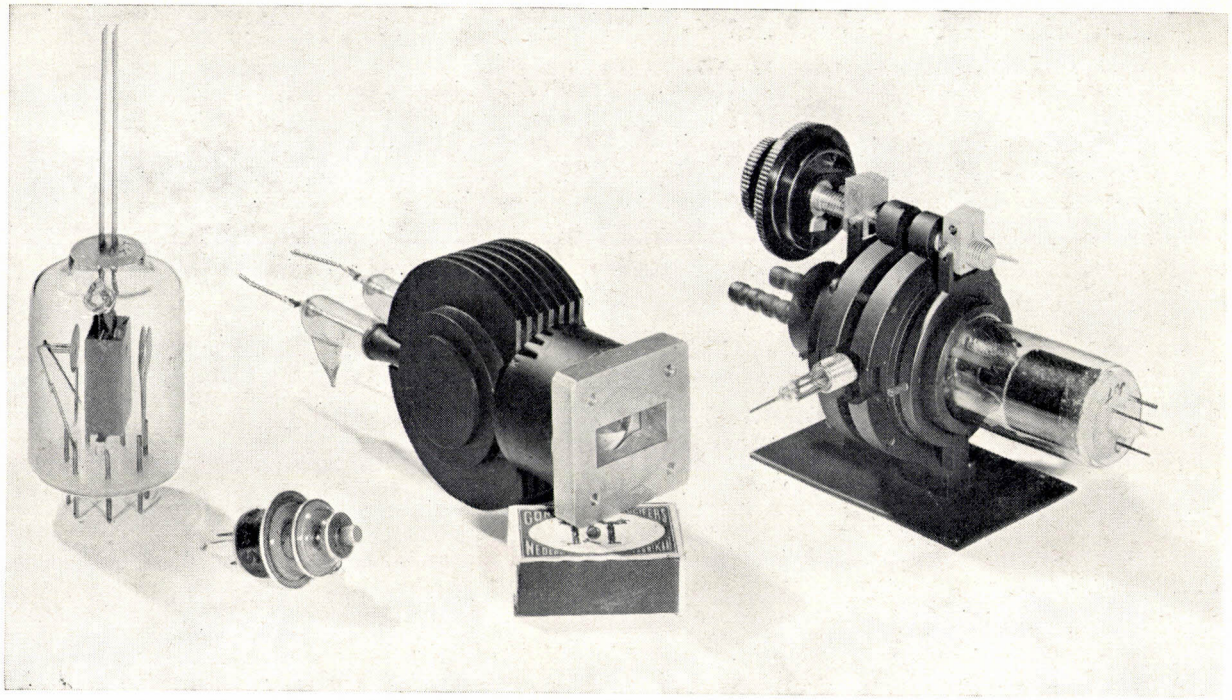
Cucumbers cultivated entirely under artificial light. The light is supplied by "TL" lamps. The temperature inside is prevented from rising too high by causing tap water to flow down the pane of glass in front of the lamps. Half-way down between the "TL" lamps is a TUV lamp, the ultra-violet radiation from which prevents the growth of algae on the glass.

electrons (the velocity-modulation valve, the klystron, the multireflection tube), employing a number of Philips' own inventions.

An important improvement of heated cathodes is the L-cathode, which contains a reservoir filled with barium oxide and enclosed by a porous wall of tungsten acting as the source of electron emission. With the aid of this cathode a disc-seal triode can be built for wavelengths down to 8 cm with a clearance of only 40 μ between cathode and grid. This new type of cathode offers important possibilities for many other applications as well.

An intensive study is being made of the special circuits for generating ultra-short waves, by employing cavity resonators, and of modern means for conducting high-frequency energy with the aid of wave guides.

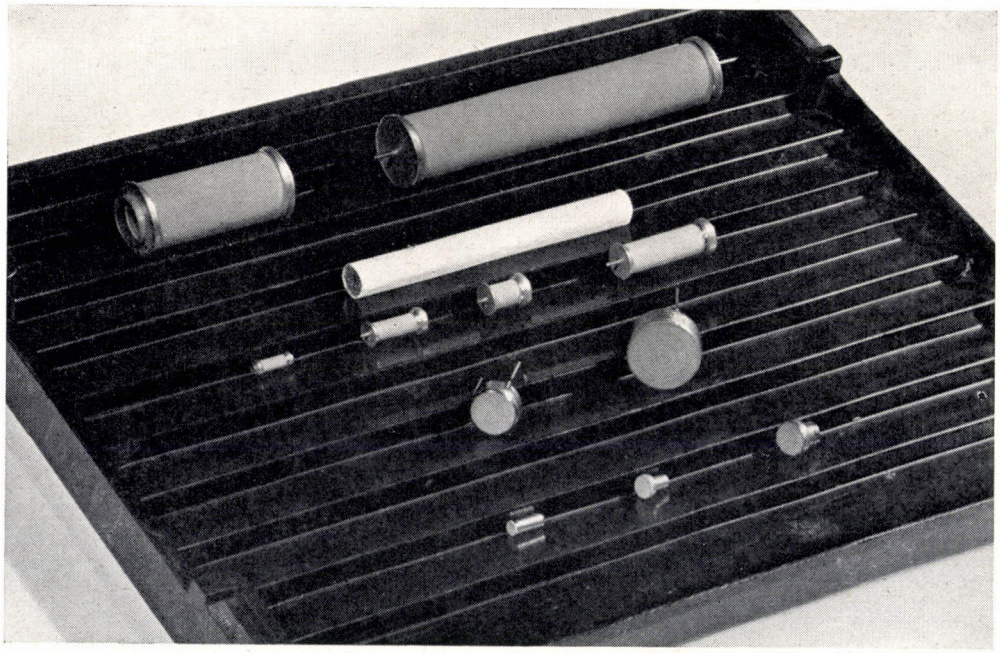
After 1945 the development of television had to be taken in hand again in the laboratory on account of the greater interest being shown in it everywhere.



67086

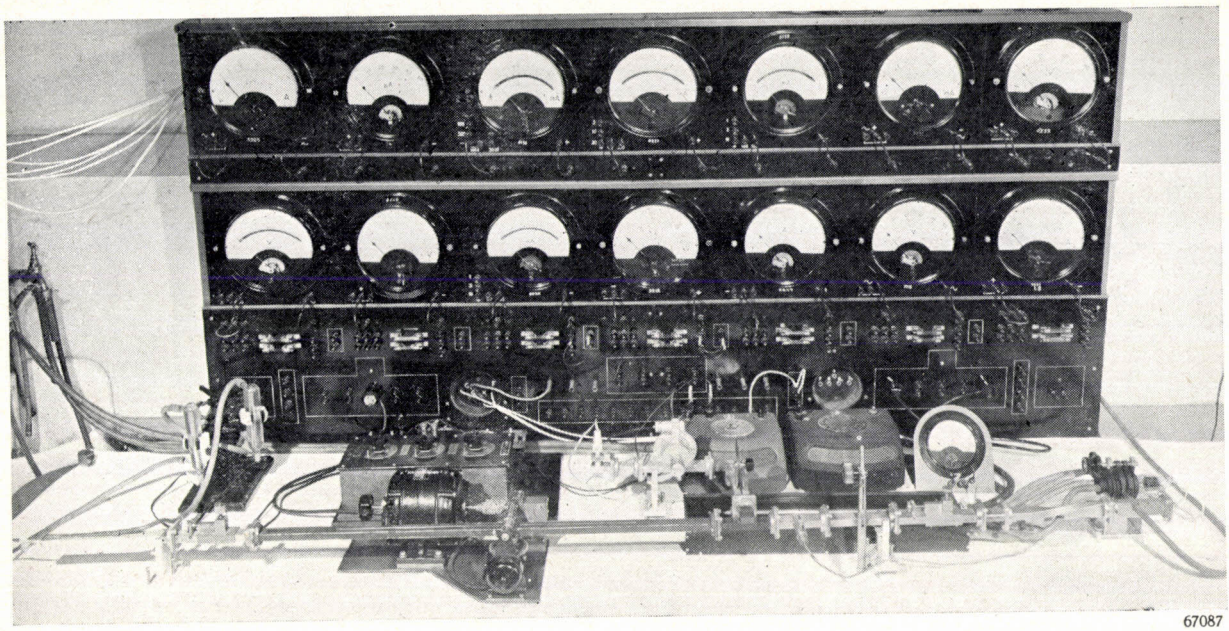
Tubes for centimetric waves. From left to right: a multireflection tube with a continuous output of 20 W on a wavelength of 12 cm; a receiving tube of the "lighthouse" shape for amplification of wide frequency bands (gain 10 at a bandwidth of 50 Mc/s, noise figure 10 db at a wavelength of 10 cm); a magnetron yielding pulses of 1000 kW at a wavelength of 3 cm; a velocity-modulation valve with a continuous output of 100 W at a wavelength of 3 cm.

At first a system of 567 lines was used, with 25 frames per second. In the audio channel frequency modulation was applied. As a result of many discussions on the question of the number of lines the transmitting apparatus was designed so that it could quickly be changed over to different numbers of lines. This proved to be of great value for judging the relative merits of various systems. When, later on, the Netherlands Television Committee advised the adoption of a system with 625 lines



67074

Some L cathodes (with a cigarette for comparison of size). The L cathode has a porous, metal, emitting surface and can withstand heavy loads, say 100 A per cm^2.

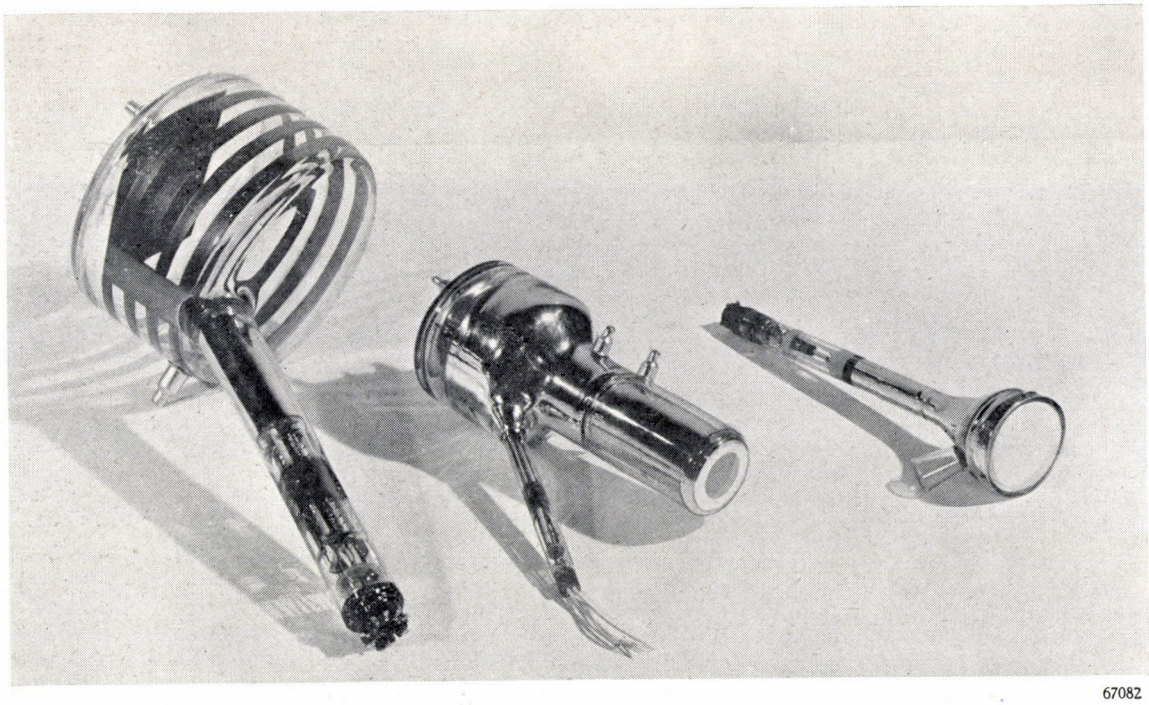


67087

Set-up for taking measurements of a velocity-modulation valve (on the extreme right) at a wavelength of 3 cm. Here wave guides are used for conducting the energy.

and 25 frames per second the installation was made suitable for that system.

Further, attention was devoted both to the problems encountered at the receiving end and to those arising in the transmission. As far as reception is concerned, based upon the results of experiments already mentioned a system of projection television was worked out with the aid of a *Schmidt* optical system, using very small projection tubes with the screen coated with phosphors suitable for the purpose. Special precautions had to be taken against discoloration of the glass of these tubes under the influence of cathode rays and X-rays. In the optical system proper, use was made of a correction plate made of gelatin, for which a separate method of manufacture had to be developed.



67082

Left: two pick-up tubes for television (an iconoscope and an image iconoscope). Right: a picture tube for projection television.

Of course attention had also to be paid to direct-view reception. One of the most important results of the work done in this connection was a considerable improvement in regard to flicker of the image by employing phosphors having a long persistence.

As regards TV transmission the most important advance made was in the construction of new pick-up tubes based on the principle of the image iconoscope. Further improvements lay in the television cameras and in the studio lighting.

Except for some short interruptions there have been regular experimental television transmissions from Eindhoven ever since 1946, for which purpose a studio, a control room and dressing rooms for artists were fitted out in the laboratory building. These experimental transmissions have helped much towards arousing the interest now being shown in this youngest branch of technical development both in the Netherlands and in adjoining countries.

Next in order to radio and television is the work done in the field of telecommunications. Ever since 1935 there has existed a department in the laboratory (organically not belonging to it but as part of the A.F. telephony department of the works) where apparatus for carrier telephony, i.a. a 17-channel system, and for A.F. telegraphy have been developed. About 1940, in cooperation with the laboratory, a modern carrier-telephony system for 48 channels was developed, in which the new core material for coils, "Ferroxcube", plays an important part. This system is already being used on an extensive scale in the Netherlands and in Switzerland, whilst it is to be installed also in Denmark in the near future. In this connection a word of grateful recognition is due to the Netherlands P.T.T. officials for their close cooperation. In 1945 the carrier-wave department was transferred from Eindhoven to the Philips' Telecommunication Industry at Hilversum, whilst a research group for telecommunications was established in the laboratory. Development work in Hilversum is proceeding further in the direction of the improvement and reduction in size of electrical and mechanical parts, with the aid of results reached in the laboratory.

Special carrier systems have been developed which are suitable also for short distances. Further, by developing such things as repeaters, a super-group through-filter and other filters, the laboratory has already contributed much towards a system now in course of development at Hilversum for incorporating some hundreds of channels in a coaxial cable.

In this laboratory work is continuing on new methods of modulation; for instance, a simplified form of pulse-code modulation has been studied, the quantum modulation, which makes it possible for conversations to be carried satisfactorily over long distances in spite of a high noise level. In addition, the possibilities of applying new materials and components for telecommunication apparatus are being studied.

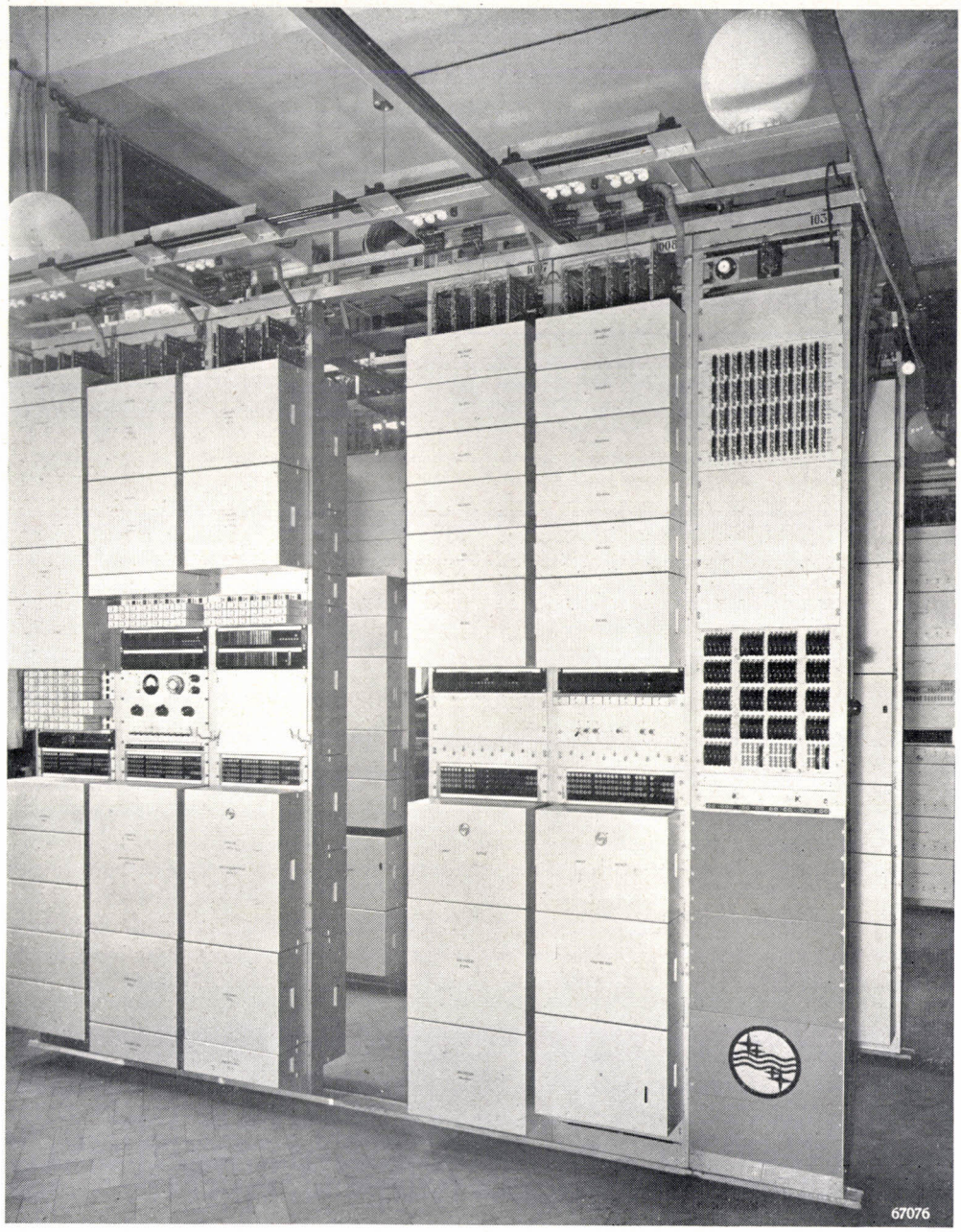
In this connection mention is to be made of the new switching tubes with ribbon-shaped electron beam, which promise interesting applications in the field of telecommunications and electric computers.

Special mention is to be made of the system of facsimile transmission worked out in the laboratory, by means of which documents can be transmitted and photographically recorded at the rate of 5000 cm^2/minute.

Of importance for meteorology is a newly developed radio sonde, in which extremely small radio valves with low current consumption are employed and with which experiments are now being carried out in conjunction with the Royal Netherlands Meteorological Institute at De Bilt.

Dr. A.F. Philips in front of the camera of the Eindhoven television transmitter.

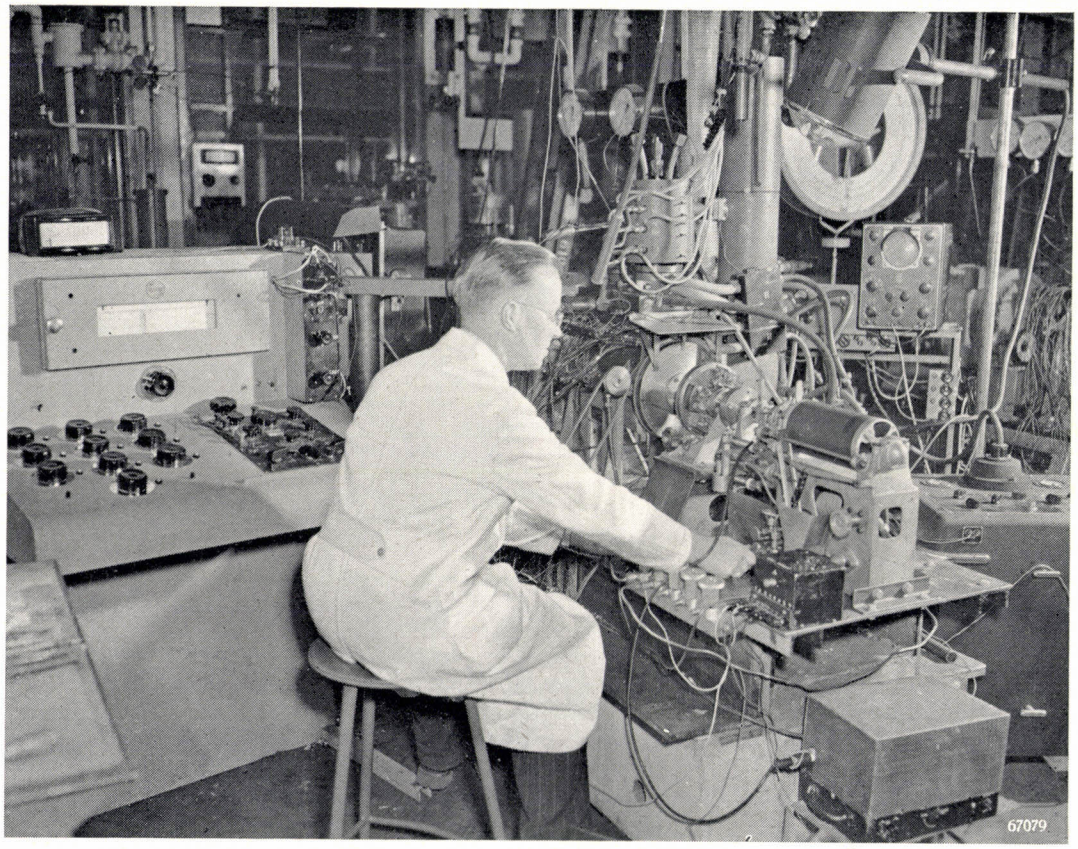
A product which certainly has some connection with radio, considering the need of a supply source in places where no electricity is available, but which represents the outcome of a development of its the piston of one cylinder is caused to act as the transfer piston for another, is of importance in connection with future possibilities the consequences of which cannot yet be fully predicted.



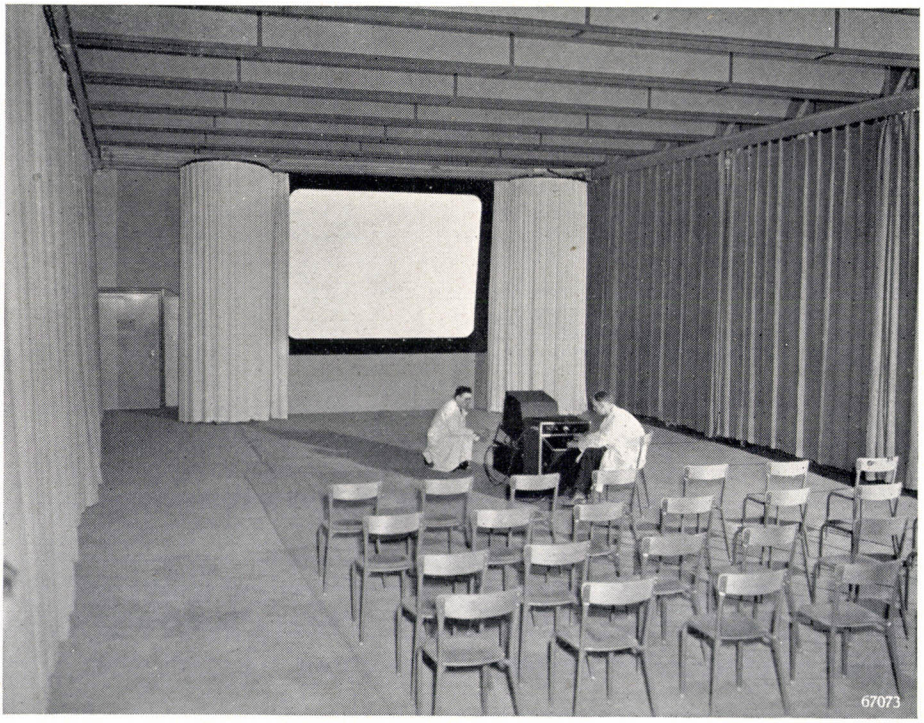
Experimental installation for carrier telephony.

own, is the air engine already referred to in this review. From a close study of the heat exchange between flowing gases and metallic walls and of materials suitable therefor, and further of the thermo-dynamics of the cycle, it has become possible to build a light high-speed motor with good thermal efficiency. A new principle, of application for systems with more than one cylinder, according to which

In the field of acoustics there are quite a number of new products to be mentioned. First of all the improved sound reproduction by means of stereophony, which has led also to the construction of small microphones. The ideal microphone, with small dimensions compared with the wavelength, also for high audio-frequencies, appeared to be most closely approximated by the condenser microphone,



Taking measurements in the testing of an experimental air engine. Energy input by electric heating, energy output via a loaded dynamo. Connected to the desk on the left are thermo couples measuring the temperature at various points (inside and outside). A *Farnboro* pressure indicator records the pressure in the cylinder as a function of time; the phase of this diagram is checked with the aid of a capacitive pressure indicator (connected to an oscillograph) and a stroboscope.



Hall for acoustical investigations (i.a. stereophony) and large-screen television (almost 3 m × 4 m).

which can now be made in a size of about 20 mm, and for special purposes even as small as 7 mm.

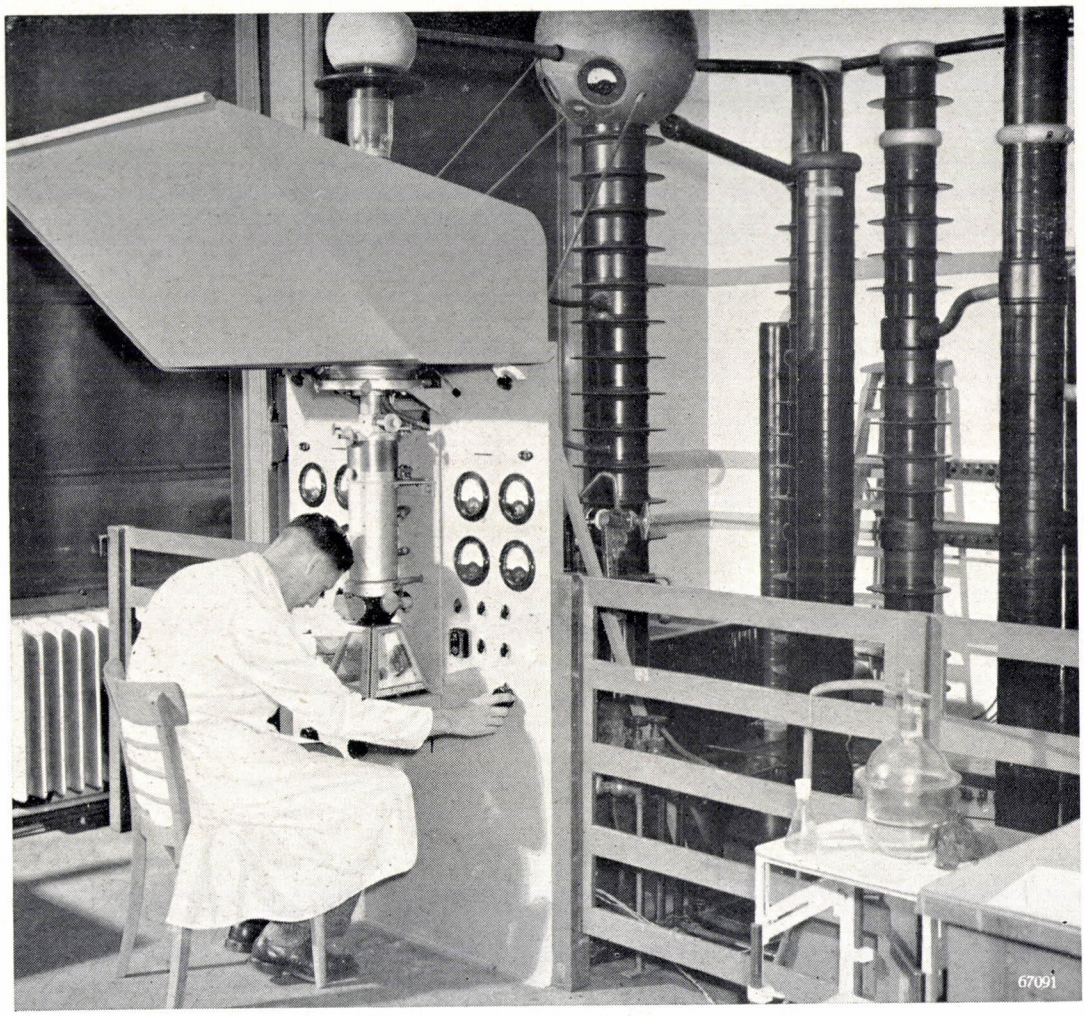
In the Philips laboratory the necessary attention has also been paid to the development, started in Denmark, of the magnetophone for recording and reproducing sound. This development obliged the manufacturers of gramophone records at last to relinquish the demand that it should be possible to play their records both electrically and mechanically, as a result of which considerable improvements became possible in the field of electrical gramophone reproduction by means of microgroove records. Thus the gramophone record may retain its position as an instrument of reproduction.

Physiological study and the resultant connection with the medical world led to the construction of hearing aids.

Experience gained in the amplification of signals of very low frequency, close to the limit of audibility, led further to the designing of a vector cardiograph, some specimens of which have been given to medical experts for testing.

Miscellaneous

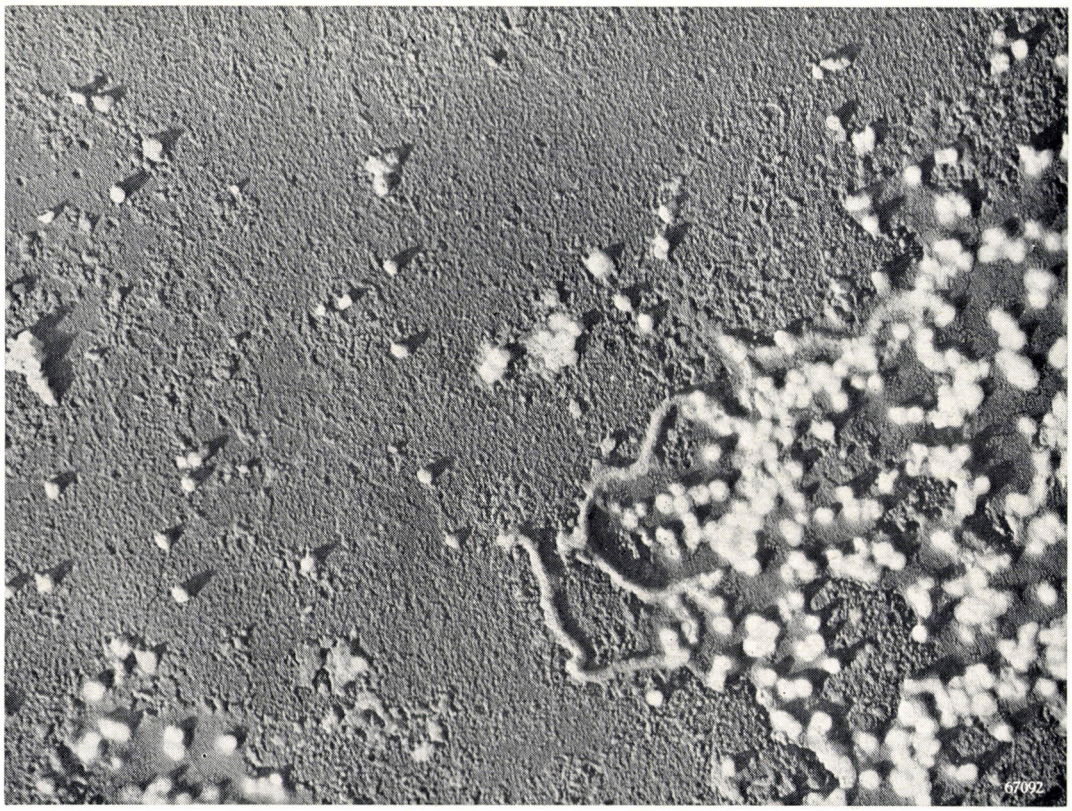
Among the X-ray tubes there is a new one for contact therapy and, further, new high-powered tubes for X-ray diffraction with accessory apparatus.



Experimental electron microscope with an accelerating voltage of 400,000 V.

With the experience gained in the field of television it has been possible to construct an X-ray picture intensifier, by means of which the intensity of the screen picture can be increased to such an extent as to allow of screening patients in a room not blacked out. A number of these intensifiers are now being tried out in Philips' Medical Department.

As an offshoot, so to speak, of the development of cathode-ray tubes, aided by the experience acquired with high tensions and in collaboration



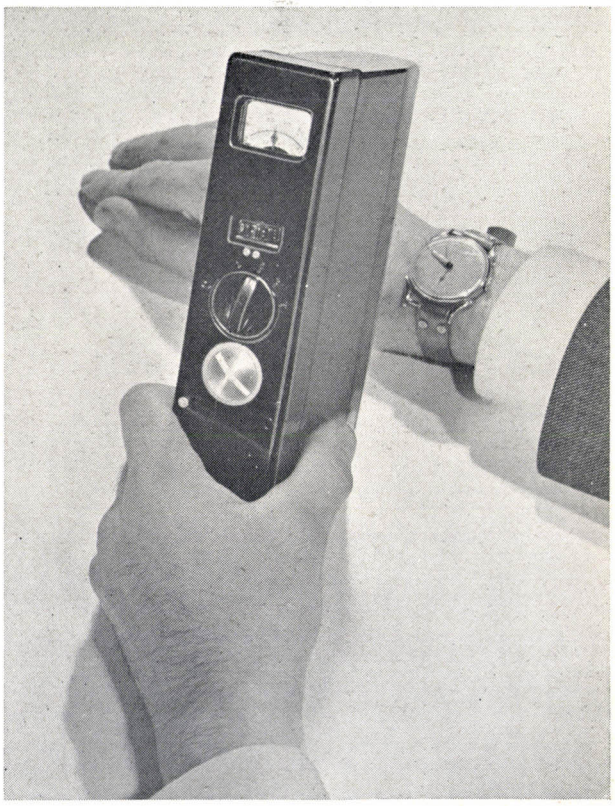
Microgram of influenza virus (specimen not entirely purified) magnified 21,000 times by the electron microscope. This virus occurs in the shape of small pellets as well as in a thread-like form.

with the Technical-Physical Laboratory of the Technical University at Delft, an electron microscope with a magnetic electron-optical system was constructed. An apparatus for 100 kV has been marketed, and in the laboratory an experimental apparatus for 400 kV has also been built. Recently, following upon previous experiments by Burgers, an emission-electron microscope has been developed, with which phenomena at the surface of metals (recrystallization) can be observed with a magnification of 1000.

For the detection and measurement of X-rays and other kinds of rays, radiation counter tubes with accessory apparatus have been made, and a special electrometer with vibrating capacitor has been developed, which can be connected, for instance, to a dosimeter for X-rays.

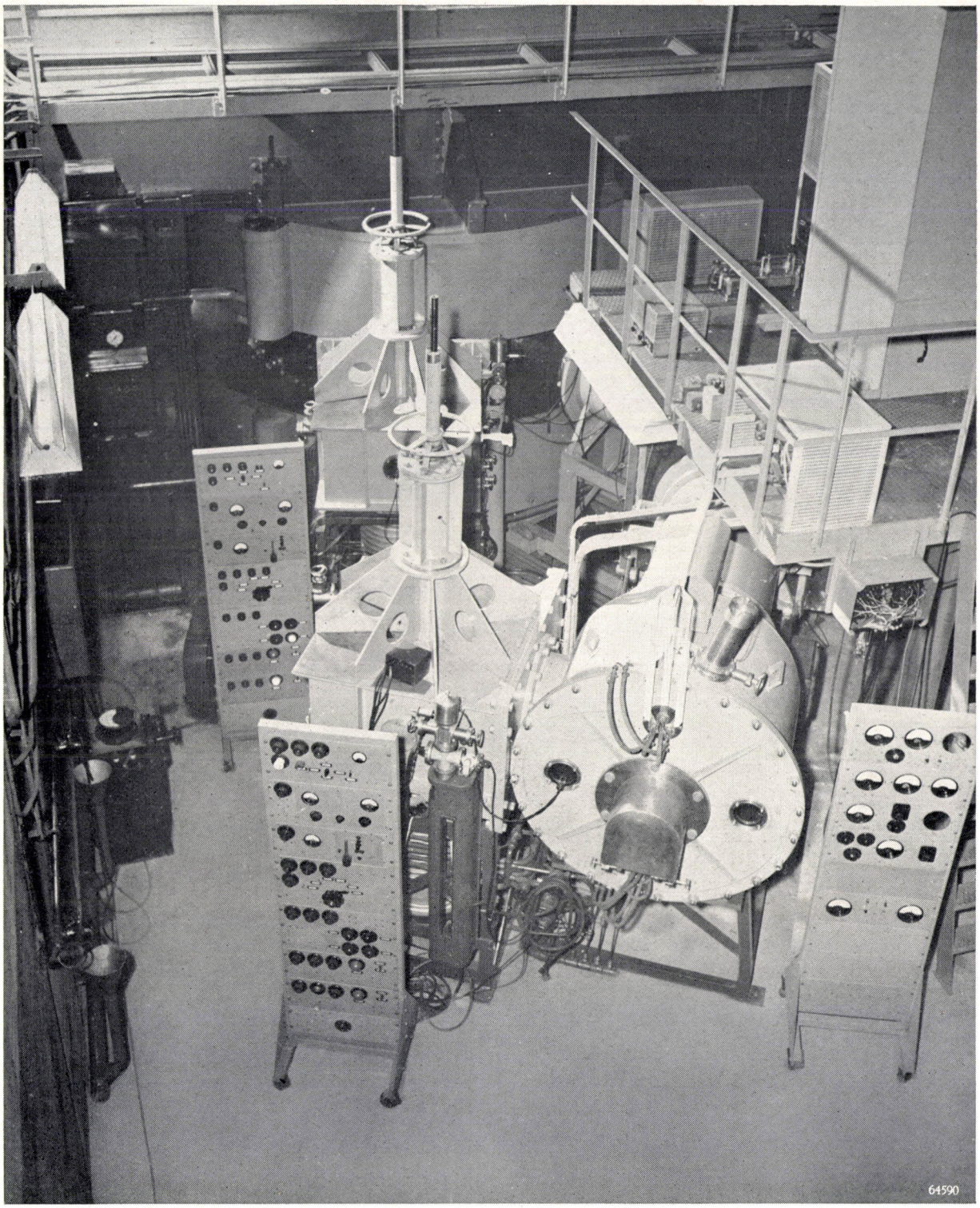
Finally, nuclear-physical research has led to the construction of neutron generators with ion-accelerating tube (maximum voltage 1.2 MV) and also a synchrocyclotron, with which, for instance, deuterons can be accelerated to an energy of 28 MeV; this synchrocyclotron has been installed at Amsterdam.

Further, an experimental betatron with air coils has been developed, with which electrons can be accelerated to an energy of 9 MeV.



67065

Portable radiation counter for detecting beta and gamma rays, fed entirely from built-in batteries. The counter reacts perceptibility to the rays from the luminescent paint on the hands of a watch.



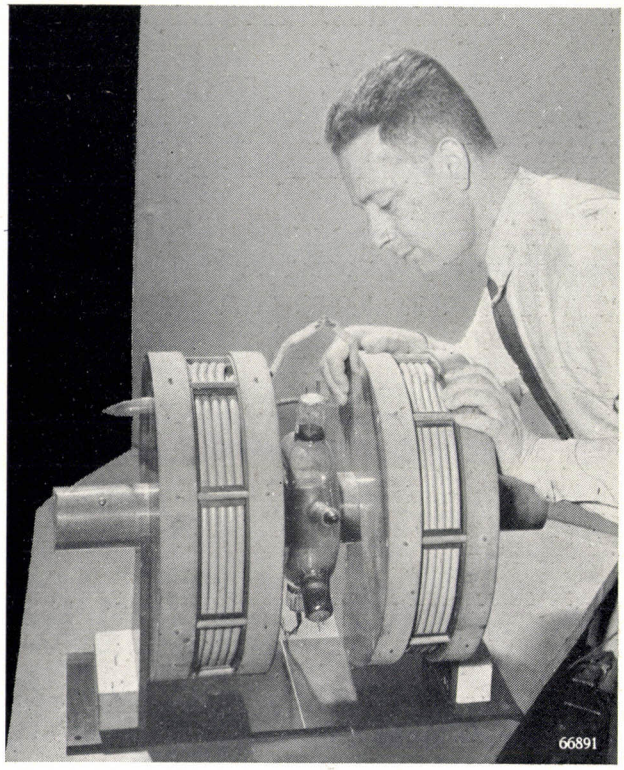
The synchrocyclotron installed in the Institute for Nuclear Research at Amsterdam. This cyclotron produces continuously deuterons with an energy of 28 million electronvolts. In the background the electromagnet, between the poles of which, where there is a maximum flux density of 1.38 Wb/m^2, are the two Dees (not visible in the photograph) inside which the deuterons describe spiral orbits. Between the Dees is an alternating voltage with a peak value of 15 kV and a frequency of about 10.7 Mc/s. The modulator (in the foreground on the right) varies the frequency periodically by an amount of 4%, so as to keep at least some of the deuterons to the desired orbits.

Conclusion

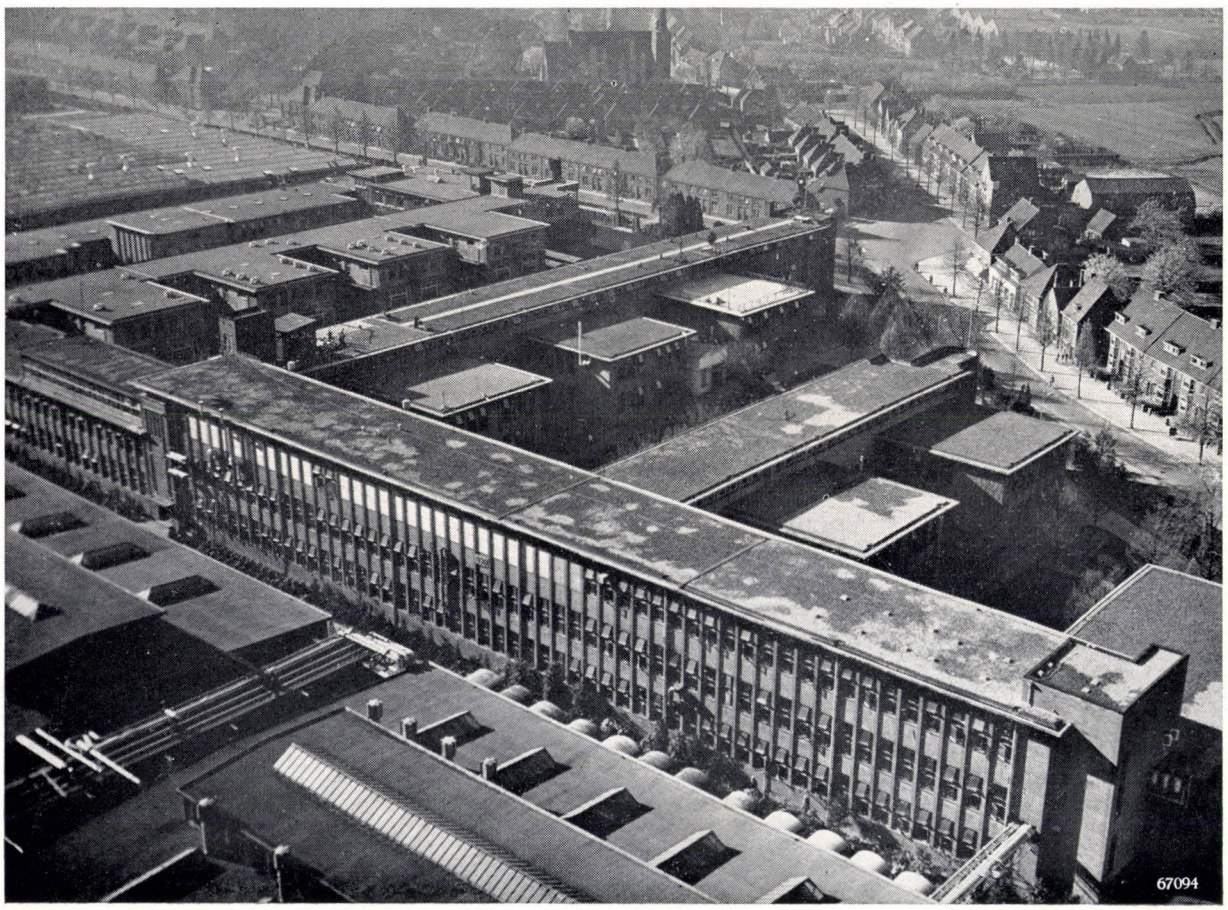
In the foregoing an attempt has been made to give a review of the activities of the Physical Research Laboratory of Philips' Industries in the course of the nearly forty years during which it has formed part of the Concern now celebrating the sixtieth anniversary of its foundation. It goes without saying that these activities could not have been achieved without the build-up of an extensive organization for the equipment of the laboratory and the workshops. In many cases these workshops have made substantial contributions to the realization of various projects.

The writer is fully aware that many sides of the development of the research work have been inadequately treated in this review. Due to lack of space, several subjects have received but scanty attention or have not been mentioned at all. However, the reader will be able to supplement this summary by referring to the Abstracts of Scientific Publications regularly appearing in Philips Technical Review and to the subject index given with the last number of each volume.

W. de Groot.



A betatron, with which electrons can be given an energy of 9 million electronvolts. The electrons travel round in the annular glass tube seen in the middle. The particular feature about the rest of the construction is that no iron yoke is used for the magnetic field.



BOOKS WRITTEN BY SCIENTISTS OF THE PHILIPS' RESEARCH LABORATORY *)

G. Holst, Electrische lichtbronnen en hare eigenschappen (Electric sources of light and their properties; in Dutch only), Haarlem, de Erven Boon, 1920.

A. Bouwers, Physica en techniek der Röntgenstralen (Physics and technics of X-rays; in Dutch only), Deventer, Kluwer, 1927.

A. E. van Arkel and J. H. de Boer, Chemische binding als electrostatisch verschijnsel (Chemical binding as electrostatic phenomenon; in Dutch), Amsterdam, Centen, 1930 (Supplement 1937).
id. id. Chemische Bindung als elektrostatische Erscheinung, Leipzig, Hirzel, 1931.
id. id. La valence et l'électrostatique, Paris, Alcan, 1936.

J. H. de Boer, Electron emission and adsorption phenomena, Oxford, University Press, 1935.
id. Elektronenemission und Adsorptionserscheinungen, Leipzig, Barth, 1937.

W. Uyterhoeven, Elektrische Gasentladungslampen, Berlin, Springer, 1938.

A. Bouwers, Elektrische Höchstspannungen, Berlin, Springer, 1938.

A. E. van Arkel, Moleculen en kristallen (Molecules and crystals; in Dutch), The Hague, Van Stockum, 1941.

W. G. Burgers, Rekristallisation, verformter Zustand und Erholung, Handbuch der Metallphysik III, Leipzig, Akademische Verlagsgesellschaft, 1941.

H. Bruining, Die Sekundär-elektronenemission fester Körper, Berlin, Springer, 1942.

P. J. Bouma, Physical aspects of colour, 1947 (also in Dutch, French and German) edited by N.V. Philips' Gloeilampenfabrieken, Technical & Scientific Literature Department.

E. J. W. Verwey and J. Th. G. Overbeek, Theory of the stability of lyophobic colloids, Amsterdam, Elsevier Publishing Co., 1948.

J. L. Snoek, New developments in ferromagnetic materials, Amsterdam, Elsevier Publishing Co., 1948.

F. A. Kröger, Some aspects of the luminescence of solids, Amsterdam, Elsevier Publishing Co., 1948.

J. M. Stevels, Progress in the theory of the physical properties of glass, Amsterdam, Elsevier Publishing Co., 1948.

H. Bremmer, Terrestrial radio waves, Amsterdam, Elsevier Publishing Co., 1949.

Balth. van der Pol and H. Bremmer, Operational Calculus, based on the two-sided Laplace integral, Cambridge, University Press, 1950.

W. Elenbaas, The high-pressure mercury vapour discharge, Amsterdam, Noord-Hollandse Uitgeversmaatschappij, 1951.

*) This list does not include books the contents of which are not directly related to the laboratory work, a number of contributions to manuals, and suchlike.

VOL. 13 No. 3, pp. 49-80 SEPTEMBER 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

NEW ELECTRONIC TUBES EMPLOYED AS SWITCHES IN COMMUNICATION ENGINEERING

I. CONTACT TUBES

by J. L. H. JONKER and Z. van GELDER. 621.385.832:621.318.57:621.39

The development of telephone exchanges has followed a logical course from manually operated to semi-automatic and thence to the present-day fully automatic exchanges, where subscriber connections are made via mechanical switches operated by electromagnets.

The next phase in the development will probably be the replacement of these electromagnetic relays, selectors, etc. by electronic tubes. One of the advantages of the latter, compared with mechanical switches, is that they have much less inertia, so that they can perform their functions more quickly and thus return earlier to their state of rest ready for the next duty. This means a saving in switching material.

Although the development of special electronic tubes for replacing relays and the like has scarcely begun, laboratory work has already led to the development, in a technically useful form, of a number of types, about which something will be said in this and a following article.

Apart from automatic telephony, these "contact tubes" will also find application in other fields where switching has to be done quickly — either synchronously or not — as is the case in certain systems for multiplex connections, for code modulation, for colour television, etc.

For making and breaking connections in line telephony wide use is made of switches in the form of electromagnetic relays. The simplest types are represented in *fig. 1*. The contacts of relays are subject to corrosion and therefore need cleaning regularly. Another important factor with these relays is the switching speed. This is limited by the inertia in the appearance and disappearance of the magnetic field, or by the inertia of moving parts.

It can be understood that in the search for switches not requiring any attention and having less inertia thought should begiven to electronic tubes, whose switching speed may be 10^2 to 10^4 times higher. Conventional amplifying valves in which the intensity of a stream of electrons is controlled by one or more grids are sometimes used as relays in telephony in certain cases, but there is no question of these valves — designed for quite different purposes — being able to perform the function of, for instance, a relay with make-and-break contacts. Even if the very simplest relay (fig. 1*a*) used in telephony were to be considered for replacement by a normal radio valve there would be the difficulty that the anode alternating voltage is in antiphase with the grid voltage. When transmitting pulses these would arrive either with one polarity or with

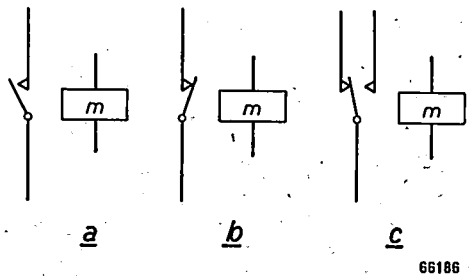


Fig. 1. *a*) Relay with a make contact, *b*) relay with a break contact, both drawn in the position where the magnetic coil *m* is not excited. *c*) Relay with a switch contact (two-way switch; the other two relays are single-way switches).

the other, according to whether the varying number of stages in the connection is even or odd. Only with rather complicated measures might it be possible to overcome this difficulty. Such measures, however, are not needed when an electronic switch is available which, just as is the case with an ordinary switch, would produce no phase difference between the input and output voltages. Presently it will be seen that such a manner of making electronic contact is possible by employing secondary emission, which is in fact done in all the tubes that are to be dealt with in these articles [1]).

of a certain contact), the types of tubes to be dealt with here can be divided into two groups:

1) those where the stream of electrons is passed or blocked by a control grid — these are contact tubes, corresponding to relays with one or more make contacts — and

2) those where an electron beam is deflected (as in a cathode-ray tube) — these are commutator tubes, corresponding to relays with one or more make and break contacts or to multi-way switches.

In this article we shall deal only with tubes of

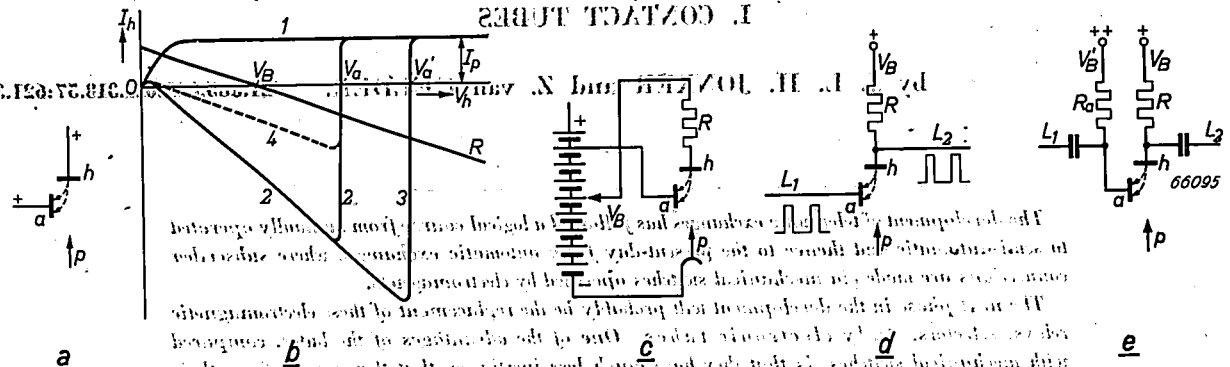


Fig. 2. *a*) A stream of primary electrons *p* is directed onto an electrode *h*, from which secondary electrons are released which proceed towards a collecting electrode *a* at a positive potential.

b) The current I_h from the electrode *h* as a function of the voltage V_h at the same electrode, *1*: for the case where *h* is not emitting secondary electrons, *2*: with secondary emission from *h* and a voltage V_a on the electrode *a*, *3*: the same as *2* but with a higher voltage (V_a') on the electrode *a*, *4*: the same as *2* but with weaker secondary emission.

c) The electrode *h* is connected across a resistor *R* to a point of positive potential V_B, so that the point of intersection of the line *R* on one of the characteristics — see (*b*) — becomes the working point.

d) The collecting electrode *a* functions as input electrode and *h* as output electrode. Positive pulses on the telephone line L_1 are transmitted to the line L_2 as long as the primary electron current reaches *h*.

e) Circuit for transmitting an arbitrary alternating voltage from line L_1 to line L_2. The input electrode *a* receives a positive bias via the resistor R_a.

These tubes, it is to be emphasized, are to be regarded as still being in the experimental stage. Nevertheless in their shape and dimensions (they are of about the same size as a normal radio valve) and, of course, in their electrical properties — to be discussed later — they show that account is being taken of practical requirements right from the outset.

According to the manner in which the stream of electrons emitted by the heated cathode is directed to or diverted away from a certain spot in the tube (corresponding to the closed or open position

Contact-making with the aid of secondary emission

The principle of effecting an electronic contact with the aid of secondary emission [2]) is illustrated in fig. 2. A stream of primary electrons *p* is directed towards an electrode *h* (fig. 2*a*) at a positive potential V_h. Assuming that this electrode is incapable of emitting secondary electrons, then the current I_h in the supply lead of the electrode *h*, which depends on the voltage V_h, follows the curve

[1]) These tubes have partly also been dealt with by J. L. H. Jonker in "Valves with a ribbon-shaped electron beam: contact valve; switch valve; selector valve; counting valve", Philips Research Reports 5, 6-22, 1950 (No. 1).

[2]) It has already been incidentally described in this Journal (Philips Techn. Rev. 11, 141, 1949, fig. 12) as being used in a circuit called an "anode follower".

I_1 in fig. 2*b*; I_h soon reaches a saturation value given by the intensity, I_p, of the stream of primary electrons. If, on the other hand, the electrode h has been processed in such a way as to be able to give a high secondary emission, and in the valve there is an electrode a (fig. 2*a*) at a certain positive potential V_a, then at low positive values of V_h secondary electrons will be emitted and these will proceed towards the electrode a having a higher potential. As soon as the number of secondary electrons released per second from h exceeds the number of primary electrons impinging upon h, I_h becomes negative (see the descending part of the curve *2* in fig. 2*b*, the so-called dynatron characteristic). When V_h is made approximately equal to V_a the secondary electrons can no longer be taken up by a. They then return to h, so that in this range ($V_h \approx V_a$), I_h rapidly increases until the curve *1* is again reached. If V_a is given a higher value (V_a') a characteristic is obtained like that of curve *3*.

When the electrode h is connected via a resistor R to a direct-voltage source V_B (fig. 2*c*), then V_h is as represented by the straight line R (fig. 2*b*) as a function of the current I_h through R. The point where this line R intersects the curve *2* indicates the values of the current and voltage at the electrode h when the voltage at the other electrode a is V_a. If the values of R and V_B are chosen such that the point of intercection lies on the steep part of the curve *2* — as represented in fig. 2*b* — then the potential V_h corresponding to this point is practically equal to that corresponding to the point where the curve *2* intersects the abscissa, thus about V_a. Imagining V_a to be variable, then V_h follows the variations of V_a, without changing polarity. In other words, there is, as it were, contact between the electrodes h and a, which we may call the output electrode, and the input electrode respectively.

When, for instance, a telephone line L_1 (fig. 2*d*) carrying positive pulses is connected to the electrode a and another line L_2 is connected to h, the pulses are transmitted from L_1 to L_2 as long as primary electrons are impinging upon h. For the transmission of an arbitrary alternating voltage (for instance, with audio frequencies) we proceed as indicated in fig. 2*e*; here the potential of a consists of a constant bias, derived via a resistor R_a from a direct-voltage source V_B, and of the alternating voltage derived, say, from the line L_1 connected to a across a blocking capacitor, in the same way as L_2 is connected to h. As long as primary electrons are impinging upon h, the voltage at h follows that at a and thus L_2 has the same alternating voltage as L_1. To break the contact in the cases of figures 2*d* and 2*e* it suffices to apply to a control grid in the vicinity of the cathode a negative voltage which is sufficient to suppress the primary stream of electrons.

The "voltage drop" in the contact, $V_a - V_h$, is small and depends only slightly on the magnitude of the secondary emission, provided it has been arranged that the line R intersects the characteristic approximately in the middle of the steep part: practically the same point of intersection is then found for an electrode h with weaker secondary emission (curve *4* in fig. 2*b*).

The contact described with reference to fig. 2*d* is unilateral, since unless other measures are taken V_a does not follow any fluctuations of V_h. In communication engineering use is very often made of separate channels for the outgoing and the return connections. For such a unidirectional channel a unilateral contact therefore suffices, and this condition is satisfied in most of the tubes to be discussed in this article. For those cases, however, where a bilateral contact is necessary another tube has been developed as described in the next section.

With the circuit of fig. 2*e* a bilateral contact is nevertheless possible within certain limits. If, for instance, V_h is increased by a voltage from L_2 then equilibrium is not restored until V_a is increased by the same amount: if V_a were to lag behind the rising V_h then secondary electrons could no longer pass from h to a and the current through R_a would decrease, which means that V_a does after all increase. The same argument can be followed to show that V_a will decrease as V_h drops.

V_a is not, however, unrestricted in following the fluctuations of V_h. In the first place V_a cannot follow V_h when the latter exceeds the value V_B', because this is the highest potential that a can reach (that is why it is favourable to choose a high value for V_B', higher than V_B). As V_h decreases another limit is reached on account of the secondary emission diminishing with the decrease of V_h until h is no longer capable of supplying the larger current through R_a that would be needed for the same decrease of V_a.

If a contact tube with bilateral action is needed then a tube working on the principle described may serve, but, generally speaking, a tube of the kind described below, which has been designed for bilateral action, will be more suitable.

Symmetrical contact tube

A bilateral contact, identical in both directions, is obtained when the device of fig. 2*e* is made entirely symmetrical. To achieve this both the electrodes h and a (denoted in fig. 3 by h_1 and h_2) are processed for high secondary emission. They are placed symmetrically with respect to the primary stream

of electrons and both are connected via a resistor R to a point of positive potential. This tube operates as follows.

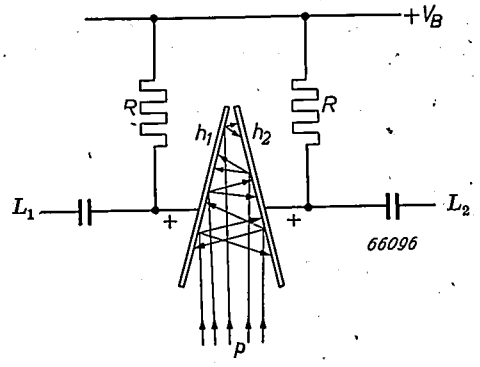


Fig. 3. Entirely symmetrical, bilateral, electronic contact. The primary electrons p impinge upon two symmetrically placed electrodes (h_1 and h_2), both with secondary emission properties and each connected across a resistor R to a point of positive potential V_B. L_1-L_2 external circuit.

We shall first consider only the secondary electrons emitted by the electrode h_1; the potentials of h_1 and h_2 will be denoted by V_1 and V_2 respectively. As long as V_2 is greater than V_1 all these electrons will be attracted towards the electrode h_2. This corresponds to the horizontal part of the curve *1* in *fig. 4*. When the potential difference is reversed ($V_2 < V_1$) then the secondary electrons leaving h_1 may travel more or less in a direction opposed to the electrical field, as their initial velocity is much higher than that of the electrons leaving a heated cathode (see the curved left-hand part of the characteristic *1*). A similar characteristic applies for the secondary electrons emitted by the electrode h_2 (curve *2* in fig. 4). The current flowing in the external circuit (L_1-L_2 in fig. 3) is found by adding the ordinates of the curves *1* and *2*. This gives the curve *3*, which is the current-voltage characteristic of the electronic contact. The corresponding characteristic of an ideal contact would be the I-axis, that of a contact with a low resistance the dotted line *4*. The characteristic *3* closely approximates *4*, being practically linear round about the origin, so that the contact can transmit in both directions, for instance, alternating voltages at audio frequencies without any appreciable distortion. The angle between the tangent to *3* at the origin and the I-axis is a measure of the resistance of the contact, which is inversely proportional to the intensity I_p of the primary stream of electrons. With $I_p = 1$ to 2 mA a contact resistance in the order of 1000 ohms can be reached.

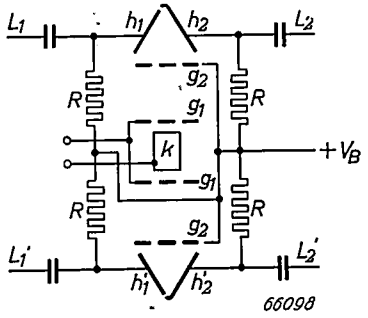


Fig. 5. Bilateral contact tube according to the principle of fig. 3, in double-pole execution. k is the common cathode, g_1 the control grid, g_2 an auxiliary grid at a positive potential, h_1-h_2 and h_1'-h_2' the secondary emission electrodes to which are connected, for example, the lines L_1-L_2 and L_1'-L_2'.

The electrodes h_2 and h_2' are slightly lengthened and bent over in order to prevent electrons reaching the bulb and there causing undesired secondary emission.

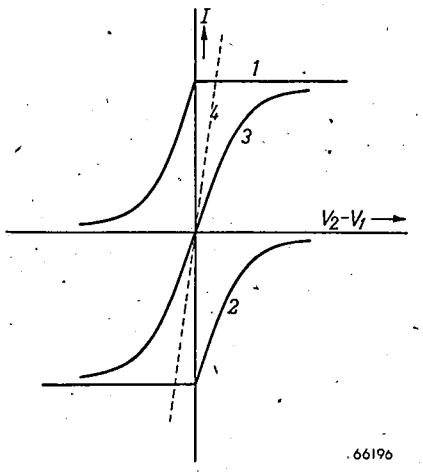


Fig. 4. Plotted as functions of the potential difference V_2—V_1 for the tube of fig. 3: *1* the current strength of the secondary electrons from h_1, *2* that of those from h_2, *3* the total current found by adding the ordinates of *1* and *2*. For comparison the dotted line *4* gives the current-voltage characteristic of a normal contact of low resistance.

The symmetrical contact tube described can be improved in some respects simply by applying a grid g_2 (see *fig. 5*, representing a bipolar contact tube) of such dimensions and at such a positive potential that a part of the secondary emission from the electrodes h_1 and h_2, respectively h_1' and h_2', is drawn away which just corresponds to the current of the primary stream of electrons. The current then flowing through the resistors R is very small. This means that the average potential of the electrodes h is raised (almost equal to V_B), so that the secondary emission coefficient is greater and the current to be transmitted may be greater. When the primary electron stream is switched off the potential of h rises only to the slightly higher value V_B; thus when the contact is closed or opened (passing or suppressing the electron stream respectively) only insignificant voltage pulses arise in the circuits connected to the electrodes h. The

advantages mentioned are still obtained if the resistances are chosen so high as to form only a small load for the circuits.

Fig. 6 is a photograph of the tube last described.

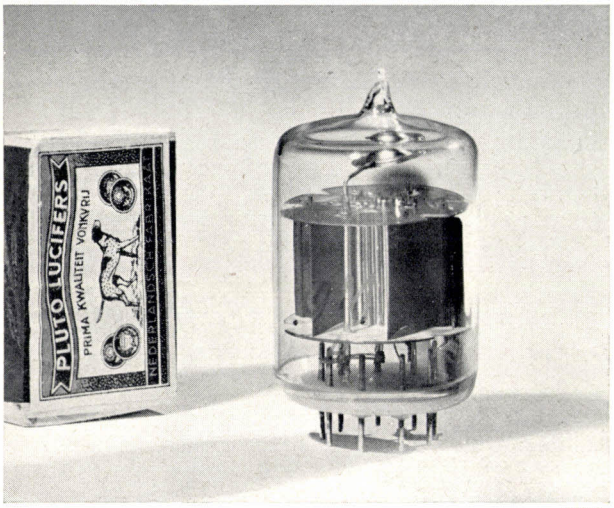


Fig. 6. Photograph of the tube schematically represented in fig. 5.

Limited switching possibilities with tubes of the type discussed

In telephony there are many cases where the requirements demanded of the relays cannot be met, or only partly, with the type of tube described above. It is often desired to connect a fairly large number of relay contacts in series and at least a few of them in parallel. With electromagnetic relays this requirement is easily satisfied. With tubes of the type described, however, this is not so easy when more than three of them are connected in series. This will be briefly explained before discussing the solution that has been found for this problem. In the case of parallel connection, which we shall not go into here, no special difficulties arise.

In *fig. 7* we have a diagram representing the series connection of three contact tubes (for transmission in one direction) as shown in fig. 2. It is presupposed

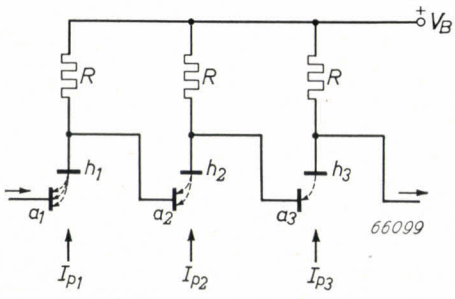


Fig. 7. Three unilateral electronic contacts according to fig. 2*d* connected in series.

that the resistances R are of such a value that the currents flowing through them are negligible in comparison with the primary electron currents, which are assumed here to be equal in intensity ($I_{p1} = I_{p2} = I_{p3}, = I_p$). The secondary-emission electrode h_3 of the third tube will then assume such a potential that its secondary emission coefficient, δ_3, is equal to 1. There then flows to the collecting electrode a_3 a stream of secondary electrons, I_{sec3}, equal to I_{p3}. This current reaches the electrode h_2 of the second tube together with the primary current I_{p2}, so that h_2 must emit a secondary electron current equal to $I_{sec3} + I_{p2} = = 2I_p$. Since the intensity of the primary electron current reaching it is only $I_{p2} = I_p$, h_2 must assume a potential at which the secondary emission coefficient δ_2 is equal to at least 2. In the same way it can be deduced that the electrode h_1 of the first tube will assume a potential at which its secondary emission coefficient δ_1 is equal to at least 3 and, with n tubes connected in series, equal to n. Now for a substance with good secondary emission like caesium, with a primary electron energy of 200 eV the secondary emission coefficient is at most about 5. Taking into account the finite value of R, the δ's must be even slightly greater than n, for the currents through these resistors must also be compensated. Thus the number of electronic contacts that can be connected in series is limited in practice to only three.

It is, however, possible for many more contacts to be connected in series if the two functions hitherto performed by the electrode a, namely the function of input electrode and that of collecting electrode, are separated. The contact system then consists of three electrodes, with the input electrode in the form of a grid to which hardly any current flows.

Contact system with input electrode in the form of a grid

Fig. 8 is a schematic representation of a unilateral contact system with three electrodes, where the primary electrons impinge upon an electrode h with secondary emission and forming one of the contact electrodes (the output electrode); the input electrode in the form of a grid is now denoted by r. By far the majority of the secondary electrons coming from h pass through the meshes of the grid and reach a third, collecting electrode a at a fixed and higher potential. As is the case in fig. 2, h follows the potential variations of r (in this case it is more correct to say that it follows the potential variations in the plane of r). The small current flowing to r can be compensated for the greater

part by arranging for this electrode to emit per second about just as many secondary electrons as there are primary electrons impinging upon it ($\delta \approx 1$); such is the case if the surface of the metal is not contaminated.

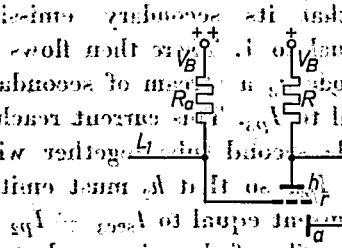


Fig. 8. Contact system with input electrode in the form of a grid. p primary electrons, h secondary-emission output electrode, r input electrode in the form of a grid, a collecting electrode at a fixed, high potential. L_1 incoming line, L_2 outgoing line.

When a number of these contacts are connected in series there is much less trouble from the difficulty discussed above, as practically no current flows towards the electrodes r. About ten contacts can then quite well be connected in series, a number which is ample for very many practical cases.

The voltage difference across the closed contact can be made very small (even negative) owing to the fact that the electrode a at high potential can be made to bring about in the plane of the grid r a potential which is higher than that of the grid itself.

This contact system with three electrodes has been applied in tubes in which the primary stream of electrons is guided in the desired direction by electrostatic deflection. Further particulars of this will be given in another article to follow.

Summary. This is the first of two articles in which some experimental tubes are described which are intended to function as switches. These contact tubes have the advantage over electromagnetic relays in that the switching speed is much greater and no cleaning of the contacts is required.

Contrary to the case of ordinary electronic valves, where the output signal is in antiphase with the input signal, these new tubes are characterized by the two signals being in phase, thereby avoiding many complications. This property is due to a secondary-emission output electrode being employed, which follows the potential of the input electrode. The "contact" thus existing, as it were, between these electrodes can be broken by suppressing the primary beam of electrons with the aid of a control grid. The contact is unilateral, the output electrode following the potential variations of the input electrode (with the same phase), whereas the reverse is not usually the case. In telephony such a unilateral contact often suffices. Nevertheless a symmetrical contact tube, for a bilateral contact, has also been developed. The current-voltage characteristic of the contact is practically linear round about the origin, the contact resistance is in the order of 1000 ohms.

In some cases it is a drawback that no more than three of these contacts can be connected in series, but this can be largely overcome by arranging for the functions of input and collecting electrodes to be performed by two separate electrodes ("contact system with input electrode in the form of a grid"). As many as about ten contacts can then be connected in series.

FLICKER IN TELEVISION PICTURES

by J. HAANTJES and F. W. de VRIJER.

621.397.3:535.7:612.843.53

In a television picture flicker is sometimes experienced. It is therefore important to investigate how flicker can be eliminated. With a suitably chosen luminescent material on the screen of the cathode-ray tube it appears to be possible to reduce flicker considerably, so much so that under normal conditions this effect is entirely absent.

The object of motion-picture projection and television is to reproduce moving scenes as naturally as possible. The fact that this can indeed be technically achieved is due to a fundamental deficiency of the human eye, namely its inability to perceive as separate stimuli a rapid succession of optical impressions. The result is that a series of pictures with only small differences and quickly following one upon the other appear as one continuously changing picture. Of course the separate pictures must follow each other at a sufficiently high frequency. But even when this condition is satisfied an unpleasant effect may arise — the picture flickers. Only when the frequency is raised much higher does this effect disappear.

For film projection the picture frequency has been standardized at 24 pictures per second. The illumination of the film picture is intermittent, being blacked out during the change over from one picture to the next, so that each time a stationary picture appears on the screen. Were it not for the special measures that are taken, at this picture frequency the flicker effect observed would be considerable. The frequency of the projection is therefore raised artificially by interrupting the illumination once more during the projection of each separate picture (the light and the dark intervals are made equal in length). In this way a frequency of 48 pictures per second is reached — two successive pictures are the same — and, as everyone is aware, no trouble is experienced from flicker in a cinema.

For television, too, a series of pictures are made to follow each other in rapid succession, but here the situation is more complicated than in the case of film projection. The picture screen, which is covered with a layer of a fluorescent substance (phosphor), is scanned line for line by a beam of electrons, thereby causing the phosphor to luminesce also line for line. It is desired to keep the frequency at which the "frame" is completely scanned as low as possible, because the bandwidth of the video signal is directly proportional to that frequency. Further, it is advantageous (at least it was felt to be so during the development of television) to match the picture frequency to the mains frequency, and that is why in European countries it is customary to use a picture frequency of 25 c/s, whilst in the U.S.A. it is 30 c/s (in both cases half the mains frequency). Although such frequencies are sufficiently high for reproducing a continuous movement, they are too low to avoid flicker. Consequently interlacing is generally applied, which means that, instead of all the horizontal lines being scanned by the electron beam in the natural sequence, for instance first all the even lines are scanned and then all the odd ones, and so on. Thus twice as many frames are scanned per second, though each has only half the total number of lines, and so frame frequencies are reached of 50 and 60 c/s respectively.

The question now arising is whether such a frequency is high enough to avoid flicker. Owing to the entirely different build-up of the picture the experience gained in motion-picture projection does not suffice to answer this question. Much experimentation has therefore been directed towards the solution of this problem especially for television. In the present article some experiments will be described which have been conducted in the Philips Laboratories at Eindhoven and which were demonstrated in the spring of 1950 on the occasion of a visit to Eindhoven of the 11th committee of the "Comité Consultatif International de Radio" (C.C.I.R.). An extensive report on these experiments has been published elsewhere [1]).

Before proceeding to deal with these experiments it is well to consider briefly those properties of the eye upon which the flicker effect is based.

1) J. Haantjes and F. W. de Vrijer, Flicker in Television Pictures, Wireless Engineer 28, 40-42, Febr. 1951.

Physiology of flicker

When the eye is suddenly exposed to the light from a light source (a lamp, or an illuminated screen) of constant luminance the signal transmitted by the optic nerve to the brain will be as represented by the curve in *fig. 1*. At the instant, t_1, that the

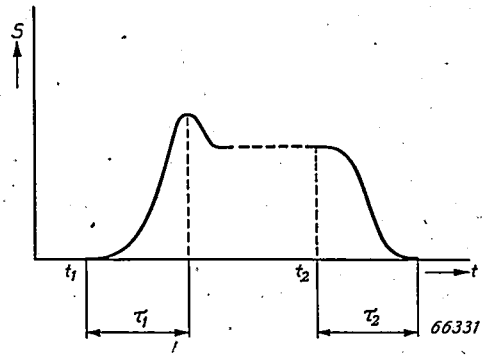


Fig. 1. The signal S, transmitted by the optic nerve to the brain of an observer whose eye at the moment t_1 is suddenly exposed to an impression of light, rises to a maximum value in a finite time τ_1, then falls slightly and after that remains constant. Upon the light source being extinguished at the moment t_2 the strength of the signal gradually diminishes to zero in a period of time τ_2. The intervals τ_1 and τ_2 are in the order of 0.1 to 0.2 sec (for normal luminances); they also depend upon the conditions under which observations are made and, further, differ as between one person and another.

light falls on the eye the observer does not immediately perceive anything, but after a short time (in the order of 0.01 sec) the signal gains in strength and ultimately reaches its maximum after an interval τ_1. Then the stimulus diminishes slightly in intensity but later remains constant. Upon the light source being extinguished at the instant t_2 the signal still remains constant for a short while (again about 0.01 sec) but then diminishes practically exponentially to zero in a period of time τ_2. It may therefore be said that the eye perceives an optical impression with a certain inertia. This inertia is characterized by the intervals τ_1 and τ_2, which are practically equal and from now on will be denoted by the one symbol τ. Experiments teach [2]) that τ depends upon the luminance B of the light source and the colour of the light, whilst also the solid angle under which the light source is seen and the place where the light falls upon the retina are of influence. The value of τ is in the order of 0.1 sec.

Tests carried out by Hartline [3]), among others, show that in addition to the normal receptors behaving as described above, in the retina there are also other receptors reacting particularly to "on" and "off" stimuli. The latter receptors pass a signal to the brain only when the optical impression begins and when it ends, but not during the exposure. There are even some receptors responding only to extinction of the light. Such receptors are naturally of the utmost importance for the perceptibility of flicker. Not much is yet known on this subject except that these receptors occur mainly on the periphery of the retina. This is one of the explanations for the fact — dealt with farther on — that flicker becomes more noticeable when the light falls upon the edge of the retina.

When flashes of light are caused to strike the eye at such a frequency that the interval between two successive flashes is much shorter than τ then they will be observed as one constant impression of light; the brightness of the flashing light source is then perceived as the average brightness of that source over a full cycle of the alternation (Talbot's law).

Talbot's law can be demonstrated with a simple experiment: in front of a light source, say an incandescent lamp, a disk is set up which can be rotated at a variable speed and from which one or more sectors have been removed, so that the light from the lamp is cast for part of each cycle upon a white screen placed in front of the disk; the screen serves as light source for an observer. Talbot's law appears to apply for frequencies higher than about 15 c/s; hence for these frequencies the observed brightness of the screen does not depend upon the speed of the disk.

When the frequency is round about 15 c/s the law is no longer satisfied: probably owing to a resonance phenomenon, the eye then perceives a greater luminance. At such frequencies (thus in the case where the cycle is in the order of τ; as a matter of fact to some extent also at higher frequencies) there is considerable flicker effect, whilst at still lower frequencies each flash is observed as such separately.

When investigating more closely how flicker depends upon the frequency we find the following.

For a given luminance B of the screen there appears to be a critical frequency f_c above which the eye no longer perceives any flicker. When the frequency of the alternation between light and dark is lower than f_c flicker is perceptible; the lower the frequency (to a certain limit), the more unpleasant is the sensation. The higher the level of the screen luminance, the higher is the critical frequency, there being an almost linear

---

[2]) See, e.g., Y. Le Grand, Optique physiologique, Tome II, Editions de la "Revue d'Optique", Paris 1948.

[3]) H. K. Hartline, J. Opt. Soc. Am. **30**, 239, 1940.

relation between f_c and the logarithm of B;

$$f_c = a \log B + b. \quad \ldots\ldots \quad (1a)$$

This is the Ferry-Porter law.

So far we have not considered the luminance of the surroundings, which also has its effect upon the result. It appears, however, that when the luminance of the surroundings is small compared with that of the screen it has practically no influence upon the results of the tests.

When the luminance of the screen is not too low ($B > 0.1$ cd/m^2) the coefficient a lies between 10 and 20 sec^{-1}, so that from formula (1a) it follows that the critical frequency rises 10 to 20 c/s when B is increased by a factor 10.

Flicker is naturally a subjective phenomenon, the results of the tests varying considerably for different persons and under different conditions. When there are influences distracting the observer's attention flicker is generally less noticed. Consequently we can only speak of averaged results.

For television practice, where the frequency is fixed, one may introduce the term critical luminance. If the luminance of the screen (at given a frequency) is less than this critical luminance B_c no flicker is noticed, but if it is greater than B_c then flicker will indeed be noticed. For this case we can write the Ferry-Porter law as:

$$\log B_c = \frac{f - b}{a}. \quad \ldots\ldots \quad (1b)$$

The lower the frame frequency the smaller is B_c, and hence the frequency sets a limit to the permissible luminance of a television picture.

The coefficients a and b depend upon the manner in which the luminance of the screen varies with time, particularly upon the ratio in length of the light and dark intervals. This can be shown by varying the size of the sector in the rotating disk. This phenomenon is demonstrated in *fig. 2* (constructed from data given by Engstrom [4]). It is seen that with a given frequency the critical luminance increases with the part of the total cycle covered by the light interval, as is understandable considering that then the case of continuous illumination is more closely approached. As far as television is concerned it may therefore be concluded that the finite decay time of the fluorescent screen will have a favourable effect upon flicker.

In motion-picture projection a luminance of about 35 cd/m^2 occurs in the brightest parts of the picture. As explained in the subscript to fig. 2, from the graphs it can be directly determined what frequency will be high enough to avoid flicker in the case of equality between the light and dark intervals of the illumination. It is seen that a frequency of 48 c/s is indeed sufficient.

Also the solid angle under which the screen is observed appears to influence the coefficients a and b. This is intimately related to the fact already mentioned that flicker is more noticeable the more of the light from the screen (or a part thereof) falls upon the periphery of the retina. This may easily be verified with the aid of a gas-discharge lamp

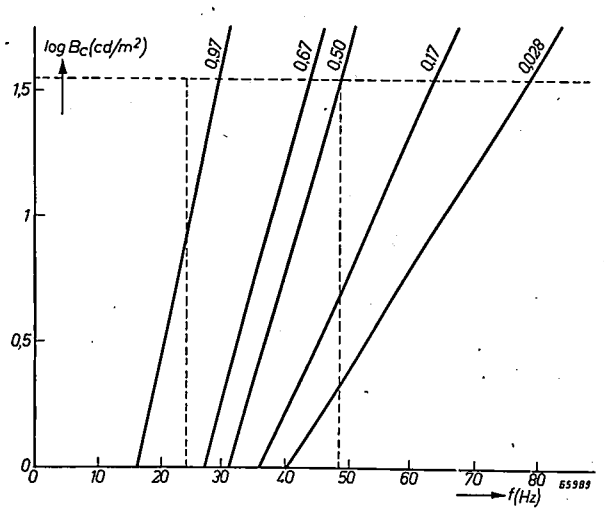


Fig. 2. With a given distance between the observer and an intermittently illuminated screen the logarithm of the luminance at which the observer just perceives flicker (the critical luminance B_c) varies linearly with the frequency f of the optical signal. The coefficients in this Ferry - Porter law depend, inter alia, on the manner in which the light and dark periods alternate, and in particular on the ratio of the lengths of these periods. The various straight lines in the diagram apply for the ratios of light to dark periods indicated. In film projection the intervals are equal and thus the relation between the critical brightness B_c and the frequency is given by the line indicated by the parameter value 0.50. In this case the desired screen luminance (in the lightest parts of the picture) is 35 cd/m^2 (horizontal broken line). With a picture frequency of 24 c/s (projection without additional interruption of the light) this brightness is far above the critical luminance (vertical broken line on the left), but with double that frequency it comes just below B_c (vertical broken line on the right) and thus no flicker is then observed. (Constructed from data given by E. W. Engstrom [4].)

(say a "TL" lamp) working on an alternating voltage. When the eye is directed straight onto the lamp the light is not seen to flicker (or scarcely so), but when the eye is turned upon some point in the vicinity of the lamp some flicker is noticed.

The colour of the light has very little effect upon the perceptibility of flicker, at least upon the critical frequency. This fact even makes it possible to compare the luminances of differently

[4]) E. W. Engstrom, A study of television image characteristics, Proc. Inst. Radio Eng. 23, 295-310, 1935.

coloured light sources. By definition it is said that two light sources have equal luminance when they show the same critical frequency.

The fact that flicker is practically independent of colour is made use of also in the flicker photometer, with which the brightness is compared of differently coloured planes viewed in rapid succession. The luminance scale found with the flicker photometer for the various colours (or rather chromaticities) agrees well with the luminance scale defined with the aid of the international luminosity curve [5]).

This is not the place to enter into a deeper discussion of the subjects touched upon above, and reference is therefore made to the extensive literature available (see, for instance, the publication cited in footnote [2])).

Experiments with television pictures

The experiments which yield us the data for determining, e.g., the minimum frequency for motion-picture projection (as already seen, with the luminance of 35 cd/m^2 a frequency of 48 c/s suffices) have to be extended when we come to consider these phenomena for the case of television. This is because, as intimated in the foregoing, the build-up of the television picture is quite different from that of the film picture.

Furthermore, in the case of television a much greater luminance is required, since it is desired to be able to view the picture in surroundings which are not blacked out or only slightly dimmed. This makes it necessary for the luminance to be more than one hundred cd/m^2. The fact that it is nevertheless possible to reach satisfactory results with a frame frequency of 50 c/s is due to the finite decay time of the fluorescent substance used for the screen.

As already described several times in this journal [6]), the intensity of the light emitted by the phosphor does not disappear immediately after the electron beam has passed, but rather decreases almost exponentially with time. If the decay time is long (say 5 times longer) compared with the time it takes for the electron beam to scan one line, then at any moment several picture lines will be luminescent. Such a band of light has not the same intensity everywhere: as the beam scans the screen in horizontal lines from top to bottom the uppermost lines of the band will be less luminescent than the lowermost ones. If the decay time of the phosphor is of the same order as the "line time" then the "band" has a width of only one line, and there will even be a difference in luminance in the horizontal direction. In practice phosphors are used with a decay time between some hundreds of times the line time and about one line line time. The situation is made still more complicated by the fact that often a mixture of two (or more) phosphors has to be used to give a screen a white luminescence. One of the components may give, e.g., blue light and have a short decay time, while the other component gives yellow light and has a long decay time.

It has been seen that a phosphor with a long decay time is favourable as regards flicker: the decay time of the phosphor in the case of television takes the place, as it were, of the ratio in the lengths of the light and dark intervals in film projection. Some experiments have been carried out to investigate this effect of the phosphor. At the same time the dependency of flicker upon the distance between the observer and the screen was determined, thus its dependency upon the solid angle from which the screen is seen, and also its dependency upon the picture frequency. Tests were carried out with ten observers in a room where the luminance of the surroundings was low compared with that of the picture; in this case the ambient luminance has little effect upon the results. Each observer was allowed a few minutes for visual adaptation before the test began. Only the average of their observations is given here. The maximum spreading between the critical luminances observed amounts to a factor 4; none of the observers gave a luminance below 70% of the average, whilst only in one exceptional case was the critical luminance given three times higher than the average (by an observer who was apparently very insensitive to flicker).

It is to be noted that all the tests described here were carried out with a fixed number of lines, viz. 625.

A study was made of the relationship between the distance from the observer to the screen and the critical luminance as perceived by the observer when using different mixtures of phosphors, at a frame frequency of 50 c/s. The luminance was defined as an average in time (Talbot's law), but not as an average in position. The latter is of no consequence when taking measurements with a television screen on which only the frame is scanned, thus without a picture, but it does count

---

[5]) See R. W. Pohl, Einführung in die Physik, Band III, Optik; Berlin, Springer-Verlag, 1943.

[6]) See, e.g., F. A. Kröger and W. de Groot, The influence of temperature on the fluorescence of solids, Philips Techn. Rev. **12**, 6-14, 1950 (No. 1). There the decay time of a phosphor is defined as the time in which after interruption of the excitation the luminous intensity drops to 1/e of its original value.

when measurements are taken with a normal television picture. In the latter case a distinction has to be made between the average luminance and the high-light luminance (in the white parts) of the picture. In *fig. 3* curves are given showing

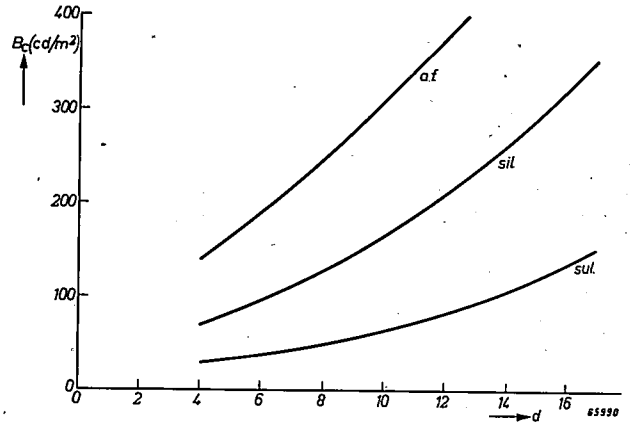


Fig. 3. Critical high-light luminance B_c of a television picture as a function of the viewing distance d (expressed with the screen height as unit). The three curves relate to the three luminescent mixtures mentioned in the text. The frame frequency is 50 c/s.

the relation between this high-light luminance and the viewing distance, it being borne in mind that the average luminance of the picture may be a factor 5 lower. In point of fact a distinction should also be made between stationary and moving pictures. In the latter case less flicker is noticed, probably owing to the observer's attention being distracted.

As the graphs show, the critical luminance increases with the viewing distance. This is probably due to the difference in sensitivity to flicker in the middle of the retina and on its periphery, as already mentioned.

The three curves relate to three different fluorescent mixtures, all giving a white luminescence. The bottom curve gives the results for a mixture of a yellow and a blue luminescent sulphide. Such a mixture is commonly used in direct-view picture tubes. The decay time of sulphides is always short (about 0.1 msec) and it is plausible that with such short decay times flicker will be most noticeable. The middle curve gives the measurements taken with a screen coated with a mixture of a yellow and a blue luminescent silicate. The yellow component has a decay time of about 5 msec, while that of the blue component is much shorter; yet the curve shows that for the mixture as a whole the critical luminance is much higher, being raised by a factor of about 2.5, regardless of the viewing distance.

With a mixture in which both the yellow and the blue components have a long decay time a still better result could be attained. This is illustrated in *fig. 4*, showing how the gain G in critical luminance (compared with a sulphide phosphor) of a single phosphor component (which in general will not give a white luminescence) with a decay time δ depends upon this decay time. Unfortunately a suitable blue phophor with long decay time has not yet been found.

The top curve in fig. 3 relates to a phosphor mixture (anti-flicker mixture) made up from three components, one of which has a decay time of about 13 msec. The critical luminances with this mixture are a factor 4.5 higher than those obtained with the sulphide mixture normally applied for direct-view picture tubes.

On the other hand, of course, the decay time must not be excessive, because otherwise rapidly moving objects in the picture would leave "trails" behind them. Further, the difference in decay time of the components of a phosphor mixture may be the cause of pictures of moving objects showing a coloured edge. These effects can, indeed, sometimes be seen but most of our observers did not notice them.

For projection-television tubes mostly the silicate mixture to which the middle curve relates is used. As regards flicker the results are then the

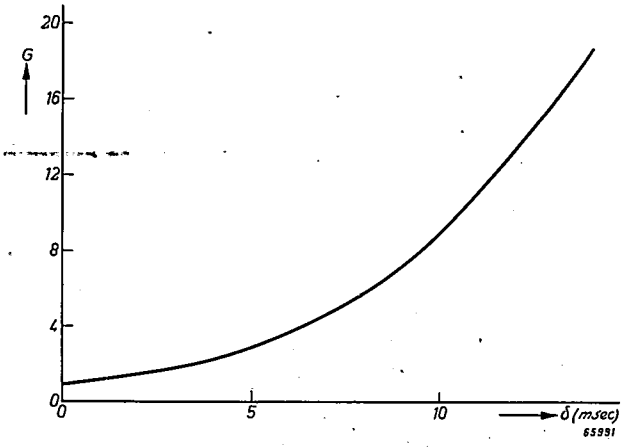


Fig. 4. The gain G in critical luminance, compared with that of a normal sulphide screen, as a function of the decay time δ of the phosphor. This diagram gives the gain for single phosphors, which therefore will not generally give white luminescence.

same as in the case of direct-view tubes, provided, of course, one is looking at the projected picture; the picture on the screen of the projection tube itself flickers considerably, owing to the very high luminance.

From the foregoing it may be gathered that with a suitable choice of the luminescent substance the permissible luminance can be raised considerably without any annoying flicker occurring.

This result is particularly of importance when we come to consider the question whether it is advisable to depart from the frame frequency of 50 c/s (with interlacing) as used in Europe and to adopt the American frequency of 60 c/s. When the critical luminance for a given phosphor and with a given viewing distance is compared for the two frequencies it appears that the gain in critical luminance that is to be reached by using the higher frequencies amounts to a factor 4.2 in the case of a sulphide mixture and a factor 5.5 when using a silicate mixture.

With the "anti-flicker mixture", however, about the same results are obtained at a frequency of 50 c/s as with a normal sulphide mixture at a frequency of 60 c/s. At a viewing distance 6 times the screen height the permissible picture luminance in the white parts is then 200 cd/m^2. Such a value is more than sufficient, even when the picture is viewed in surroundings which have not been blacked out.

Summary. The flicker of television pictures can be compared with the flicker of motion-picture projection. In both cases the perceptibility of flicker depends upon the frequency at which the pictures follow each other, the luminance of the picture and some other factors. In the case of motion-picture projection an important factor is the ratio of the intervals during which the illumination is on and off, and taking the place of this in the case of television is the decay time(s) of the phosphor mixture used for the screen of the tube. This article first deals with the origin of flicker and the properties of the eye giving rise to this phenomenon. Some laws of relationship are mentioned. Finally some experiments with television pictures are discussed, from which is appears that with a suitable choice of the fluorescent substance and when working with the European frame frequency of 50 c/s a high-light luminance of 200 cd/m^2 is permissible without causing troubleseome flicker. This value is ample for good viewing of the picture in surroundings which are not blacked out.

MEASURING THE DIELECTRIC CONSTANT AND THE LOSS ANGLE OF SOLIDS AT 3000 Mc/s

by M. GEVERS.

621.317.335.3.029.63/.64:
621.317.374.029.63/.64

The dielectric properties of solids at frequencies of some thousands of megacycles per second are interesting both from the practical point of view — since decimetric and centimetric radio waves have now found important practical applications — and from the theoretical-physical point of view, for they help in giving an insight into the structure of a solid. The method of measuring the dielectric constant and the loss angle is not new in principle but calls for an entirely new technique on account of the high frequencies occurring.

Electromagnetic waves with a wavelength of about 10 cm (frequency 3000 Mc/s) have the property of being readily focused, so that they are highly suitable for radio communications between two fixed points and for radar. Moreover, in that wave range an enormous number of channels are available and therefore there will not so readily be that "congestion" which has to be contended with nowadays on the longer wavelengths. It was a long time, however, before the technique of the decimetric and centimetric waves, or the so-called ultra-high and super-frequencies, was sufficiently mastered. The fact that this may now be said to be the case — for a number of years already — is due in part to the results of the study that has been made of the properties of various constructional materials at these frequencies.

The variation of the dielectric constant ε [1]) and of the loss angle δ as a function of the frequency has particularly been studied. When a dielectric is placed in an alternating field with an r.m.s. value E and frequency f the amount of heat generated in the dielectric per second is proportional to $E^2 f \varepsilon \tan \delta$. From the proportionality with f it follows that ε and $\tan \delta$ become particularly important at high frequencies. The amount of heat generated in a dielectric is an important point of consideration, for instance, in the choice of the kind of glass to be used for the bulb of a transmitting valve; if the right kind is not chosen the glass may melt owing to excessive dielectric losses! Other examples are to be found in the various kinds of insulators in cavity resonators, in coaxial cables and suchlike, the dielectric losses of which cause undesired damping and therefore have to be limited to the minimum. For these insulators certain plastics or materials of a ceramic nature are usually employed.

The variation of $\tan \delta$ as a function of the frequency is often erratic, so that in order to avoid unpleasant surprises the measurements have to be taken at the frequency at which the insulating material is to be used.

In this connection it is to be pointed out that knowledge of the dielectric properties of as many materials as possible in the widest possible frequency range is a matter of interest not only to the technician but also to the physicist, since it contributes towards the gaining of a better insight into the structure of solids [2]).

The method employed for measuring ε and $\tan \delta$ of solid dielectrics is a resonance method. When a rod (or disk) of the dielectric under investigation is placed in the electric field of a resonant circuit the resonance curve of that circuit undergoes a change in two respects, namely a displacement and a broadening. From the displacement it is possible to deduce the dielectric constant, and from the displacement and broadening together the loss angle can be determined [3]).

This method, in itself well-known, will first be explained under the supposition that the frequency

[1]) By ε is understood here the real part of the complex dielectric constant.

[2]) See, e.g., M. Gevers and F. K. du Pré, A remarkable property of technical solid dielectrics, Philips Techn. Rev. **9**, 91-96, 1947, dealt with more extensively in: M. Gevers, Philips Res. Rep. **1**, 197-224, 279-313, 361-379 and 447-463, 1946.

[3]) In addition to the method described here, for determining the dielectric properties of solids at high frequencies the Philips Laboratories at Eindhoven also apply a method employing wave guides. The latter method is less accurate than that described here but more suitable for still higher frequencies (10,000 Mc/s); for its description, see: H. G. Beljers and W. J. van de Lindt, Dielectric measurements with two magic tees on shorted wave guides, Philips Res. Rep. **6**, 96-104, 1951 (No. 2), and for the results of measurements taken with various kinds of glass, see: J. M. Stevels, Some experiments and theories on the power factor of glasses as a function of their composition, Philips Res. Rep. **5**, 23-36, 1950 (No. 1) and **6**, 34-53, 1951 (No. 1).

at which the measurements are taken is not very high (say 1 Mc/s), so that a normal resonant circuit can be employed consisting of lumped inductance and capacitance. Then the modifications will be discussed which are necessary when for a frequency of, say, 3000 Mc/s the circuit has to be replaced by a cavity resonator [4]).

Measuring with a normal resonant circuit

In *fig. 1a* a resonant circuit is represented consisting of a coil L and an air-capacitor C. The circuit is coupled via a small capacitance C' to an oscillator O with variable frequency, and a voltmeter V is

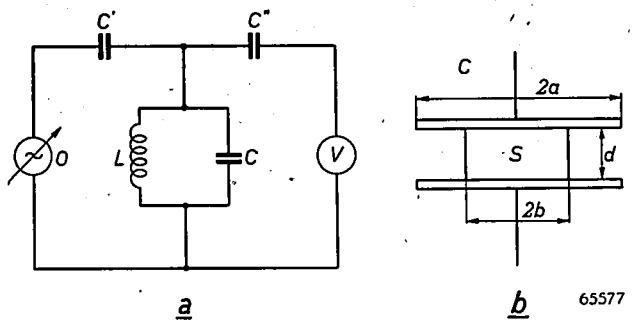


Fig. 1. *a*) A normal resonant circuit (L-C circuit) which can be used for measuring the dielectric constant and the loss angle at frequencies which are not excessively high. O oscillator with variable frequency, V voltmeter, C' and C'' coupling capacitors with small capacitance.
b) A disk S of the dielectric under investigation is placed between the two electrodes of the air capacitor C.

connected to the circuit via another small capacitance C''. The couplings via C' and C'' are so loose that on the one hand any variations in the circuit have no perceptible influence upon the oscillator and, on the other hand, the voltmeter does not cause any appreciable damping of the circuit.

A disk of the dielectric under investigation can be placed between the electrodes of the capacitor C [5]). Before doing so, however, one determines with the "empty" circuit the resonant frequency f_0 at which the voltmeter shows the greatest deflection and the difference Δf_0 between the two frequencies either side of f_0 at which the voltage across the circuit is $1/\sqrt{2}$ times the maximum value. The quality factor Q_0 of the empty circuit then follows from the equation:

$$Q_0 = \frac{f_0}{\Delta f_0}.$$

[4]) F. Horner c.s., J. Inst. El. Engrs **93** III, 53-68, 1946.
[5]) When measuring with a normal resonant circuit one will not actually proceed in this way, it being more usual to make from the dielectric under investigation a separate capacitor shunted across to the air capacitor. The procedure suggested in the text has only been chosen because it resembles more closely what will presently be described for the cavity resonator.

The same measurements are then taken after having inserted between the capacitor electrodes a disk of the dielectric having exactly known dimensions (fig. 1*b*). Denoting the values then found for the resonant frequency and the width of the resonance curve by f_1 and Δf_1 respectively, the circuit quality is then

$$Q_1 = \frac{f_1}{\Delta f_1}.$$

Upon the dielectric being introduced the capacitance is increased from C_0 to C_1, in accordance with the formula:

$$\frac{C_1 - C_0}{C_0} = \frac{f_0^2 - f_1^2}{f_1^2}. \quad \ldots\ldots \quad (1)$$

As we shall presently see, from the variation in capacitance it is easy to calculate ε, and from ε and the change in quality, $\tan\delta$ can be derived, taking into account the dimensions of the disk and the air capacitor, which occur in the formulae as a sort of "form factor". For these calculations it will be assumed here that the field in the capacitor is perfectly uniform (thus disregarding edge effects) and also, for the sake of simplicity, that the capacitor electrodes — two in number — and the disk are circular (with radius a and b respectively, see fig. 1*b*), that the thickness d of the disk equals the distance between the capacitor electrodes and that the "empty" capacitor is entirely free of losses.

Calculation of the dielectric constant

A capacitor with a homogeneous medium between its two electrodes (area A, distance d) has a capacitance C which is given by the equation

$$C = \varepsilon \frac{A}{d},$$

where the dielectric constant ε is the product of the dielectric constant ε_0 of the vacuum and the relative dielectric constant ε_r of the medium. In the Giorgi system of units employed here $\varepsilon_0 = 10^7/4\pi c^2$, with c representing the velocity of light in vacuo $= 2.998 \times 10^8$ m/s, so that $\varepsilon_0 = 8.86 \times 10^{-12}$ F/m.

The capacitance C_0 of the "empty" capacitor ($\varepsilon_r = 1$) of the resonant circuit is therefore:

$$C_0 = \varepsilon_0 \frac{\pi a^2}{d}.$$

When there is inserted between the capacitor electrodes the disk with thickness d and the relative

dielectric constant ε_r (to be determined), the capacitance is:

$$C_1 = \varepsilon_0 \frac{\pi a^2 - \pi b^2}{d} + \varepsilon_0 \cdot \varepsilon_r \frac{\pi b^2}{d} = \varepsilon_0 \frac{\pi}{d} (a^2 - b^2 + \varepsilon_r b^2).$$

Thus

$$\frac{C_1 - C_0}{C_0} = \left(\frac{b}{a}\right)^2 (\varepsilon_r - 1).$$

In combination with equation (1) this gives:

$$\varepsilon_r = 1 + \left(\frac{a}{b}\right)^2 \left\{ \left(\frac{f_0}{f_1}\right)^2 - 1 \right\}, \quad \ldots \quad (2)$$

in which $(a/b)^2$ might be termed the "form factor of the dielectric constant" (or rather of $\varepsilon_r - 1$).

From the graph in *fig. 2* the value of ε_r can be found, for a given value of the parameter b/a, as a function of the ratio of the resonant frequencies f_1 and f_0.

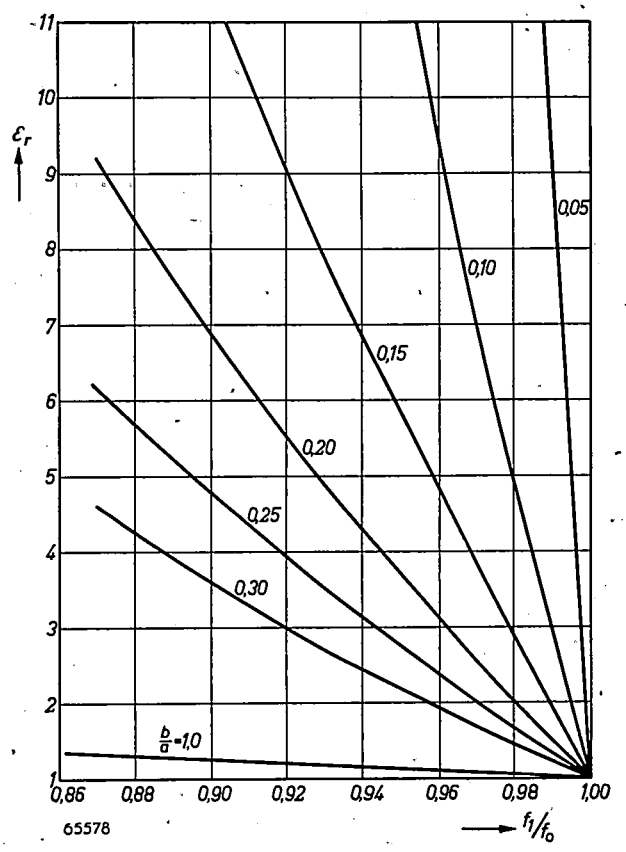


Fig. 2. The relative dielectric constant ε_r as a function of f_1/f_0, with b/a as parameter, according to the formula (2), which applies for the configuration of fig. 1.

Calculation of tan δ

The losses of a resonant circuit are inversely proportional to the quality of the circuit. As is readily understood, tan δ of the material introduced is proportional to $1/Q_1 - 1/Q_0'$, where Q_1 is the quality factor of the circuit with the dielectric in question and Q_0' is the quality factor of the circuit with an equally large disk of an imaginary dielectric having the same dielectric constant ε_r but with zero losses — the quantities Q_1 and Q_0' being measured at the resonant frequency f_1 of the circuit in both cases. Following the method described, the quality Q_1 was indeed measured at the frequency f_1 but the quality Q_0 of the empty circuit was measured at the frequency f_0. To evaluate Q_0' from the value Q_0 it should be realized that in the case of a circuit like the one considered here, where the losses are due only to the resistance R of the coil, the quality is proportional to the frequency f and inversely proportional to R. Now at high frequencies R is proportional to $\sqrt{f}$ and thus the quality is proportional to $f/\sqrt{f} = \sqrt{f}$. Hence $Q_0' = \sqrt{f_1/f_0} \times Q_0$, and

$$\tan \delta = F' \left(\frac{1}{Q_1} - \frac{1}{Q_0'}\right) \quad \ldots \quad (3a)$$

$$= F' \left(\frac{1}{Q_1} - \sqrt{\frac{f_0}{f_1}} \cdot \frac{1}{Q_0}\right), \quad \ldots \quad (3b)$$

where F' is the "form factor of tan δ", which is equal to unity if the disk completely fills the space between the capacitor electrodes. Generally, however, (when $b < a$) there is not only a part of the capacitor with the capacitance

$$C_d = \varepsilon_0 \varepsilon_r \frac{\pi b^2}{d}$$

in which losses occur, but also a loss-free part with air as a dielectric and having the capacitance

$$C_a = \varepsilon_0 \frac{\pi a^2 - \pi b^2}{d}.$$

To find tan δ of the solid dielectric, $1/Q_1 - 1/Q_0'$ must be multiplied by $(C_d + C_a)/C_d$:

$$F'' = \frac{C_d + C_a}{C_d} = \frac{\varepsilon_r b^2 + a^2 - b^2}{\varepsilon_r b^2},$$

so that equation (3b) ultimately assumes the form:

$$\tan \delta = \frac{\left(\frac{a}{b}\right)^2 + \varepsilon_r - 1}{\varepsilon_r} \left(\frac{1}{Q_1} - \sqrt{\frac{f_0}{f_1}} \cdot \frac{1}{Q_0}\right).$$

Measuring with a cavity resonator

When these measurements have to be carried out at frequencies of some thousands of megacycles per second we can no longer use a resonant circuit composed of lumped inductance and capacitance. It is most convenient to use a cavity resonator,

which in our case can be a metal cylinder with a bottom and a cover. For the rest the procedure is as described above: one determines the resonant frequency and the difference between the two frequencies at which the electric (or the magnetic) field strength in the cavity resonator is $1/\sqrt{2}$ times as large as that at resonance, first with the empty resonator and then after introducing the dielectric under investigation, preferably in the form of a

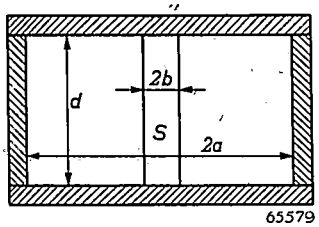


Fig. 3. A cavity resonator composed of a cylinder, bottom and cover (diameter $2a$, height d), with a rod S (diameter $2b$) of the dielectric under investigation placed inside it.

rod (*fig. 3*). From the measured values of f_0, f_1, Δf_0 and Δf_1 we can again derive ε_r and $\tan\delta$, though in this case the calculation is rather more complicated because the field in the cavity resonator is not even approximately homogeneous.

For the results of this calculation, as also for the corresponding graphs and the manner in which they are used, the reader is referred to the Appendix to this article. It may be mentioned here, however, that for $\tan\delta$ a formula is found which can be written in the form of eq. (3), viz:

$$\tan\delta = F''\left(\frac{1}{Q_1}-\frac{1}{Q_0''}\right), \quad \ldots\ldots \quad (4)$$

where F'' is a factor of no consequence at the moment, Q_1 is the quality factor of the cavity resonator with the dielectric introduced, and Q_0'' is a quantity which, like Q_0' in (3a), is proportional to the quality factor Q_0 of the empty resonator (or resonant circuit). The smaller the value of $\tan\delta$ to be measured accurately, the higher Q_0 has to be, so that it is of importance to consider the question of how high a quality can be reached with a cavity resonator.

The quality factor of the cavity resonator

The losses in a cavity resonator which determine its quality are the dissipative losses in the walls. The magnitude of these losses is related to the distribution of current along the walls — which, of course, differs with the infinitely large number of modes of oscillation that are possible in a cavity resonator — and to the depth of penetration of the current into the wall.

For the mode used, to which we shall revert farther on, *fig. 4* depicts the distribution of current in an arbitrary meridian plane of the empty cavity resonator. The density of current is the maximum in the cover and the bottom along a circle with a radius somewhat smaller than that of the cavity resonator. The density of current in the side wall is not muchless. From this it may already be concluded that the contact resistance between the cylindrical part and the bottom or the cover should be very small.

The quality Q_0 of the empty cavity resonator is calculated from the formula:

$$Q_0 = \frac{a}{\left(1+\frac{a}{d}\right)\Delta_0}, \quad \ldots\ldots \quad (5)$$

where Δ_0 is the depth of penetration at the resonant frequency f_0, i.e. the depth to which, at that frequency, the amplitude of the current density has decreased to 1/e of the current density at the surface of the wall. For the penetration depth Δ at the frequency f the formula

$$\Delta = \sqrt{\frac{\varrho}{\pi\mu f}} \quad \ldots\ldots\ldots \quad (6)$$

applies, where ϱ is the specific resistance and μ is the permeability of the conductor. For non-magnetic materials μ is practically equal to the permeability of the vacuum, viz. $4\pi\cdot10^{-7}$ H/m.

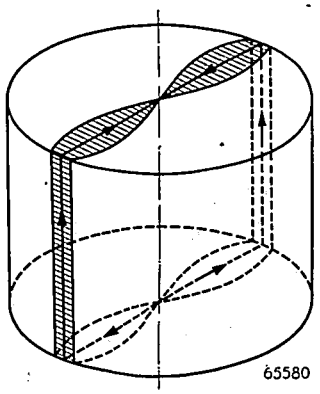


Fig. 4. The width of the hatched strip is proportional to the local current density in an arbitrary meridinal cross section of the (empty) cavity resonator (E_{010} mode, see Appendix). The arrow points indicate the direction of the current at a certain moment.

When ϱ is made as small as possible (by plating the inner walls of the cavity resonator with silver), and bearing in mind that, as we shall presently see, with the mode employed there is a fixed relationship between the radius a and the resonant frequency f_0 ($f_0 = 229.49/2a$ Mc/s, with a expressed

in metres), it appears that in the formula for Q_0 only the height d of the cavity resonator is available. From this equation (5) it is seen that increasing the value of d improves the quality, but there is a limit set to the height, because if it is made too large, then other, undesired, modes may easily occur in the cavity resonator. Besides, if d is greater than a any futher increase of d would, as a matter of fact, yield only a slight improvement in quality, and moreover a tall cavity resonator with the desired extremely high quality is far more difficult to make than a low one.

Our cavity resonator has the dimensions $a =$ $= 33.5$ mm ($f_0 = 3428$ Mc/s, with corresponding wavelength λ_0 in free space 8.75. cm) and $d =$ $= 40$ mm, whilst the theoretical quality factor is $Q_0 = 17{,}560$.

Unless special precautions are taken, the actual quality obtained is usually not more than 70 to 80% of the theoretical value, as has been shown in literature [6]). This has in fact even given rise to doubts about the correctness of the suppositions upon which formula (6) is based. However, we have found that this difference between practice and theory is due to irregularities on the inner surface of the wall causing the current to follow a longer path. Usually a cavity resonator is made by turning out a metal cylinder on a lathe, in the process of which the point of the cutter leaves more or less deep grooves in the surface at right angles to the general direction of the current (*fig. 5a*). Unless the depth of these grooves is small compared with the depth of penetration Δ the current flowing through the thin layer follows the irregularities and thus its path is considerably lengthened. It appears that polishing the surface does not improve matters much. Now from formula (6) it follows that under the given conditions the depth of penetration is about 1 μ, so that great care is certainly essential to avoid the remaining irregularities being many times greater than the penetration depth. This has been achieved in the following way [7]).

First the brass cylinder, its bottom and cover are lined electrolytically with a layer of silver 200 μ thick. On a vibration-free precision lathe this layer of silver is then turned with a diamond having a semi-circular profile (radius of curvature about 100 μ), whilst the pitch (about 1 μ) and the shaving thickness are chosen extremely small. Since this pitch is in the order of the wavelength of light (0.4 to 0.8 μ) the surface thus turned acts as a sort of Rowland grating, showing beautiful interference rings under incident light. From the pitch and the radius of curvature it follows that

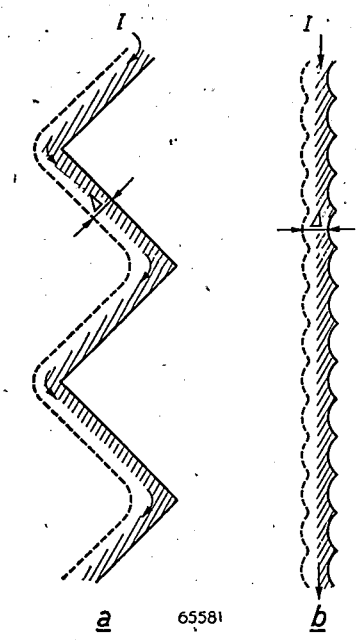


Fig. 5. Inner surface of the cavity resonator, (*a*) with sharp cutting grooves deeper than the penetration depth Δ of the high-frequency current I, (*b*) with rounded-off grooves shallower than Δ. In the case (*a*) the path of the current is lengthened considerably, whereas in the case (*b*) there is only a negligible lengthening.
In both figures the size of Δ (about 1 μ) is the same. In (*b*) the arcs have been drawn with a relatively far too large curvature (radius of curvature of the diamond cutter used was about 100 μ), so that the irregularities are relatively much exaggerated.

the depth of the irregularities is about 10^{-3} $\mu =$ $= 10$ Å The condition at the surface is then as illustrated (for a proportionately much too small radius of curvature) in fig. 5*b*, in which case the difference to be expected from the theoretical value of Q_0 is very small. This has been confirmed by measurement: $Q_0 = 17{,}190$, which is 98% of the theoretical value! *Fig. 6* is a photograph of this cavity resonator.

The bottom and the cover make such a perfect fit with the cylindrical part that the three parts laid loosely together yield the same value of Q_0 as when the bottom and cover are tightly screwed together.

The rod of the dielectric must also be turned or ground smooth and of equal thickness throughout, since its diameter must be exactly known and any irregularities in the surface would disturb the field

[6]) In the article quoted in footnote [4]) mention is made of a measured quality factor amounting to 72% of the theoretical value. See also E. Maxwell, Conductivity of metallic surfaces at microwave frequencies, J. appl. Phys. **18**, 629-638, 1947.

[7]) Thanks to the expert craftsmanship of Mr. L. Leblans, of Philips Research Laboratories.

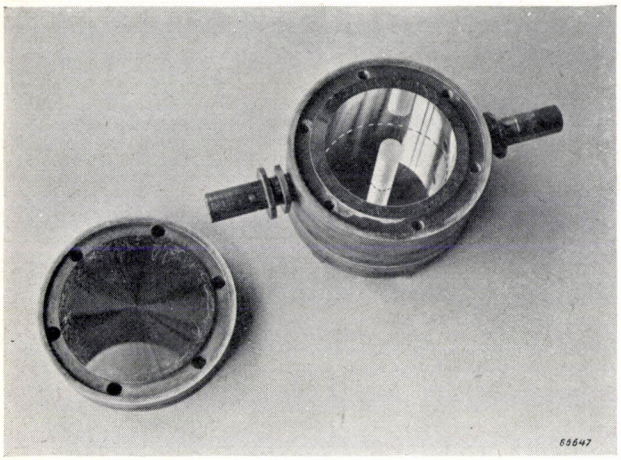


Fig. 6. The opened cavity resonator with a rod of polystyrene inside it. For the sake of clarity the boundaries between the bottom and the side of the resonator and between the rod and its reflection in the bottom have been marked in the photograph with dotted lines.

distribution, in which case the formulae given in the Appendix would apply less strictly. The length d_1 of the rod must equal the height d_0 of the cavity resonator less a certain tolerance; if d_1 were only a mere fraction of a millimetre larger than d_0 the cover would no longer fit on the cylinder, the contact resistance at the top would thereby be considerably increased and the value of Q_1 would become far too low. The centering of the rod in the resonator need not be more accurate than is possible to attain by eye.

The cavity resonator described here need be used only for materials with very low losses or in cases where the utmost accuracy is demanded. In other cases one can manage quite well with a less expensive resonator not processed with such high precision ($Q_0 \approx 9000$).

Accessory apparatus

In addition to the cavity resonator described the apparatus (*fig. 7*) comprises a calibrated oscillator and a device for measuring in a relative unit the magnetic field strength in the cavity resonator. These two components are coupled to the resonator.

As oscillator use is made of a reflex klystron (*fig. 8*), a combination of a kind of velocity-modulation tube [8]) with a cavity resonator, the dimensions of which govern to some extent the oscillator frequency. With the aid of three micrometer screws — two for coarse tuning and one for fine tuning — the frequency can be varied between 3090 and 3260 Mc/s [9]), thus within the range of about 10 cm waves as frequently used for radar. The frequency can be read from calibration curves.

[8]) Tubes of this type have been briefly discussed in the first pages of an article by F. Coeterier, Philips Techn. Rev. **8**, 257-266, 1946, and at more length in, i.a., R. H. Varian and S. F. Varian, A high-frequency oscillator and amplifier, J. app. Phys. **10**, 321-327, 1939.

[9]) The resonant frequency $f_0 = 3428$ Mc/s of the empty cavity resonator falls outside this range. This frequency and Q_0 have been determined once for all with the cavity resonator after the introduction of a rod of a material having a known dielectric constant and a known, very low, value of tan δ.

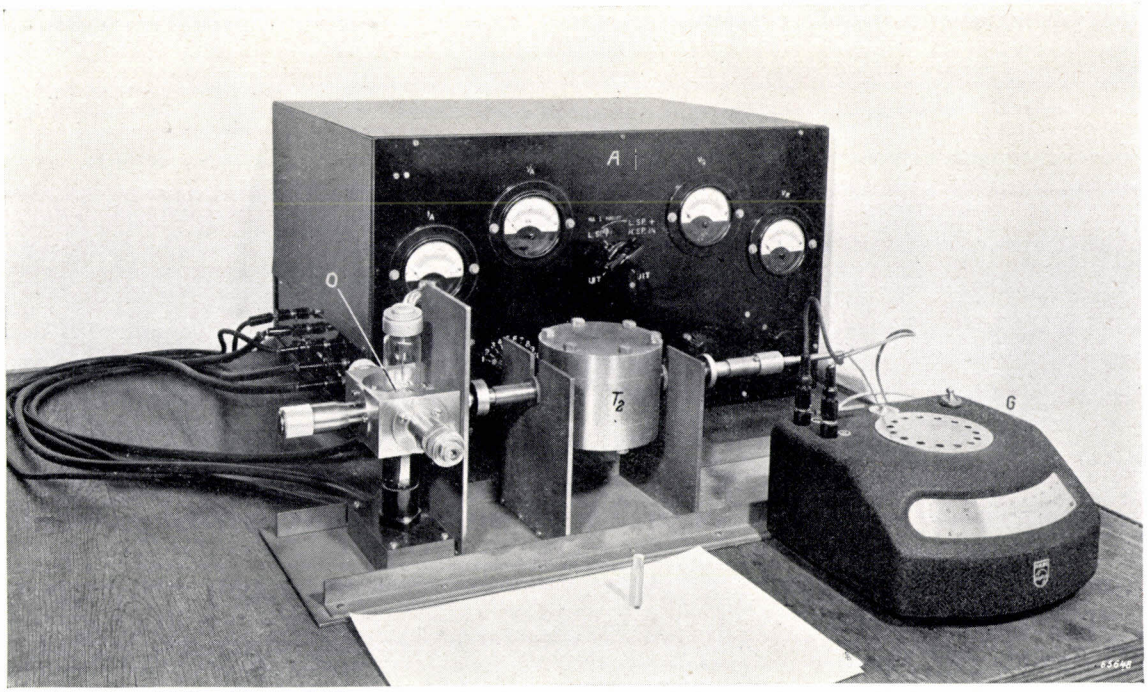


Fig. 7. Photograph of the measuring apparatus. *O* oscillator, T_2 cavity resonator, *G* galvanometer, *A* supply unit. In the foreground is a small rod of quartz prepared for insertion in the cavity resonator.

The cavity resonator used for measuring is coupled both to the resonator of the reflex klystron and to the detector. Both these couplings are made in the usual way with coupling loops and short lengths of coaxial cable. To prevent resonances occurring in the latter they are filled with a material having high dielectric losses. The couplings can be varied by rotating the loops about their horizontal axis.

The dimensions $2b$ and d of the rod are measured with a micrometer ($2b$ at different places). The diameter $2a$ of the cavity resonator must not be measured with a micrometer as otherwise the surface would be damaged. It can, however, be found by optical means, and also from the measured value of the resonant frequency (see footnote [9])); the difference in the results obtained from these two methods is less than 0.01%.

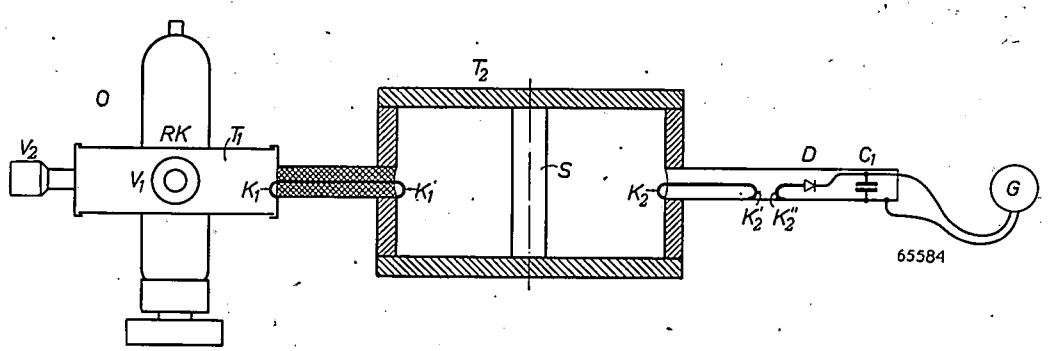


Fig. 8. Schematic representation of the measuring apparatus. *O* oscillator, with reflex klystron *RK* having a cavity resonator T_1 which can be tuned with the knobs V_1 (coarse) and V_2 (fine). T_2 is the measuring cavity resonator, with rod *S* of the material under investigation. The coupling loops K_1 and K_1' couple T_1 to T_2; the loops K_2, K_2' and K_2'' couple T_2 to the detector *D*. C_1 is a smoothing capacitor. *G* the galvanometer.

The detector is a silicon crystal. This supplies a rectified current which, after being smoothed by a capacitor, flows through a sensitive galvanometer (full deflection at 1 μA, resistance 1000 ohms). The characteristic of the crystal is such that the deflection of the meter is practically proportional to the square of the magnetic flux embraced by the coupling loop K_2. When the quality factor is being measured the oscillator is therefore detuned until the deflection is half the value at resonance.

Accuracy

The accuracy with which the values of ε_r and of tan δ can be measured depends upon the accuracy and the constancy of the frequency calibration, the exactness of the dimensions given and the extent to which the cavity resonator and the rod deviate from the rotationally symmetrical shape.

With the aid of a frequency meter (accuracy about 1 : 10^6) the oscillator is first calibrated at different positions of the knobs for coarse tuning and a fixed position of the knob for fine tuning. Then, for a number of positions of the first-mentioned knobs, the detuning which can be brought about with the fine-tuning knob is calibrated (accuracy 1 : 200). It has been found that in a period of two years the oscillator calibrations showed practically no change.

In general, with the apparatus described, an accuracy of $\pm$ 1% can be reached for the dielectric constant, whilst for tan δ the accuracy is $\pm$ 1% $\pm 0.2\times10^{-4}$ (the latter figure is the absolute error.)

The table below gives a few of the results of a number of measurements, carried out at a frequency of approximately 3100 M/cs (wavelength between 9.7 and 9.5 cm). It is particularly to be noted that very small values of tan δ, such as those of polythene and quartz glass, can be accurately measured. For comparison the values of tan δ at 1.5 Mc/s are also shown.

Frequency in Mc/s	3100		1.5
	ε_r	tan $\delta\cdot10^4$	tan $\delta\cdot10^4$
Some samples of glass			
A hard glass	4.20	64.8	36.4
Two kinds of lead glass	6.30	61.7	20.2
	7.15	44.4	8.9
Quartz glass	4.20	0.8	0.8
Some samples of plastics			
Polystyrene	2.52	2.6	2.4
Polythene	2.25	2.0	2.0
"Plexiglass" or "Perspex"	2.6	64.1	130

Appendix

Field distribution in the cavity resonator

By a suitable method of exciting the cavity resonator (not made too tall — see above) it is

ensured that of all the possible modes only the E_{010} mode [10]) occurs, the resonant frequency of which depends solely upon the radius a of the cylinder. The first nought in the index following E means that the electric and the magnetic fields in the cavity resonator are independent of the

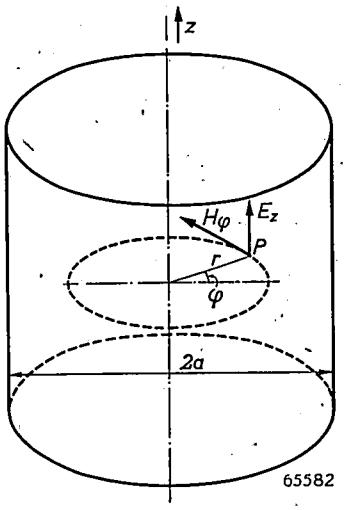


Fig. 9. Cylindrical cavity resonator with cylinder coordinates r, φ and z. In the case of the E_{010} mode, at any point P the quantities E_z and H_φ are the only components of the electric and magnetic fields respectively.

cylinder coordinate φ (see *fig. 9*), and hence that the fields are rotationally symmetrical. The second nought indicates that the fields are independent of the coordinate z. The figure 1 in the index will be explained later.

The E_{010} mode can be described with one component of the electric field and one component of the magnetic field, each of which depends upon only one space coordinate, r. These components are E_z, which runs parallel to the z-axis, and H_φ, which runs tangentially (fig. 9). H_φ and E_z of the empty cavity resonator are represented by the following equations:

$$\left.\begin{aligned} H_\varphi &= A\, J_1(k_0 r) \cos \omega_0 t, \\ E_z &= \frac{k_0}{j\omega_0 \varepsilon_0} A\, J_0(k_0 r) \cos \omega_0 t. \end{aligned}\right\} \quad . \quad . \quad (7)$$

where A is a proportionality factor depending upon the power of the oscillator and the coupling between the oscillator and the cavity resonator, and $k_0 = 2\pi/\lambda_0$. The factors J_0 and J_1 are Bessel functions of the first kind, respectively of the order 0 and 1, whilst $\cos \omega_0 t$ denotes the dependency upon the time t; ω_0 is the angular frequency at which the empty cavity resonator resonates.

[10]) For "E-waves", also called "TM-waves", see, e.g., W. Opechowski, Electromagnetic waves in wave guides, Philips Techn. Rev. **10**, 46-54, 1948, and in particular pp 50 and 51.

The boundary condition is that the electric field strength is zero at the wall of the cavity resonator (the wall assumed to be of infinite conductivity), so that $E_z = 0$ for $r = a$. The first root of $J_0(k_0 a) = 0$ is $k_0 a = 2.4048$, thus the first resonant wavelength $\lambda_0 = 2\pi a/2.4048 = 2.6125a$ and the first resonant frequency $f_0 = c/\lambda_0 = 229.49/2a$ Mc/s, a being expressed in metres. (The figure 1 in the index of E_{010} denotes that this first resonance is meant.) Thus, for instance, with the diameter of 67 mm chosen by us the resonant frequency $f_0 = 3428$ Mc/s, as already mentioned.

The graphs in *fig. 10* show how the amplitudes $E_{z\max}$ and $H_{\varphi\max}$ of the field quantities vary with the distance r from the axis of the cavity resonator.

From the calculation of the inhomogeneous field distribution in the cavity resonator with a rod placed inside it, the following (approximate) formula is found for ε_r:

$$\varepsilon_r = 1 + \frac{M-1}{1 + \frac{k_1'^2 b^2}{8} M} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where $k_1' = 2\pi f_1/c$ and M is an expression in which Bessel functions with the arguments $k_1'a$ and $k_1'b$ occur (see the formula (12) below).

Calculation of ε_r

The deduction of the formula (8) is in broad lines as follows: For the inside of the dielectric, thus for $0 \leqq r \leqq b$, we have:

$$\left.\begin{aligned} H_\varphi &= A\, J_1(k_1'' r) \cos \omega_1 t, \\ E_z &= \frac{k_1''}{j\omega_1 \varepsilon_0 \varepsilon_r} A\, J_0(k_1'' r) \cos \omega_1 t. \end{aligned}\right\} \quad . \quad . \quad . \quad (9)$$

These equations differ from (7) in that k_0 is replaced by a quantity $k_1'' = k_1'\sqrt{\varepsilon_r} = 2\pi f_1 \sqrt{\varepsilon_r}/c$ and ω_0 by the resonant angular frequency ω_1, whilst in the denominator of the expression for E_z a factor ε_r has been added. (In these equations it has been assumed that H_φ and E_z are not affected by the losses in the material, an assumption which is justified except in the case of very poor dielectrics.)

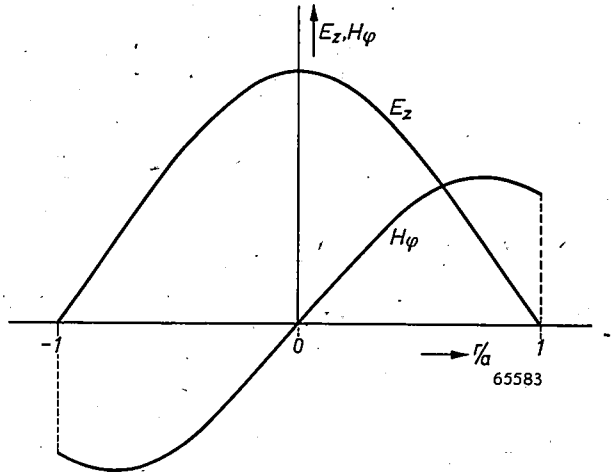


Fig. 10. Distribution of $E_{z\max}$ and $H_{\varphi\max}$ along a diameter of the cavity resonator in the E_{010} mode.

In the remaining part of the cavity resonator ($b \leqq r \leqq a$) the following equations apply:

$$\left.\begin{aligned} H_\varphi &= \\ E_z &= \frac{k_1'}{j\omega_1\varepsilon_0} \end{aligned}\right\} \begin{aligned} &\{B\,J_1(k_1'r) + C\,Y_1(k_1'r)\}\cos\omega_1 t, \\ &\{B\,J_0(k_1'r) + C\,Y_0(k_1'r)\}\cos\omega_1 t, \end{aligned} \quad . \quad (10)$$

where Y_0 and Y_1 are Bessel functions of the second kind, respectively of the order 0 and 1. The constants B and C are proportional to A and can be eliminated with the aid of the boundary conditions. These are as follows: for $r = a$ again E_z must equal 0, and at the boundary between the solid dielectric and air E_z and H_φ must be continuous, which means to say that for $r = b$ the equations (9) must yield the same values for E_z and H_φ as those yielded by the eqs (10).

From the first boundary condition it follows that:

$$B\,J_0(k_1'a) + C\,Y_0(k_1'a) = 0,$$

and from the second boundary condition, in so far as the continuity of E_z is concerned,

$$\frac{k_1''}{\varepsilon_r} A\,J_0(k_1''b) = k_1'\{BJ_0(k_1'b) + CY_0(k_1'b)\},$$

and as regards the continuity of H_φ,

$$AJ_1(k_1''b) = BJ_1(k_1'b) + CY_1(k_1'b).$$

By eliminating B/A and C/A from these three equations we find:

$$\varepsilon_r = \frac{k_1'' \cdot J_0(k_1''b) \cdot J_1(k_1'b)}{k_1' \cdot J_1(k_1''b) \cdot J_0(k_1'b)} M, \quad . \quad . \quad . \quad (11)$$

in which

$$M = \frac{\dfrac{Y_0(k_1'a)}{J_0(k_1'a)} - \dfrac{Y_1(k_1'b)}{J_1(k_1'b)}}{\dfrac{Y_0(k_1'a)}{J_0(k_1'a)} - \dfrac{Y_0(k_1'b)}{J_0(k_1'b)}}. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

In (11) ε_r occurs implicitly in the fraction preceding M, namely in the factor k_1''. If, however, $b\sqrt{\varepsilon_r}$ is very small compared with a, this fraction can be approximated by

$$1 - \frac{(k_1'b)^2}{8} \cdot (\varepsilon_r - 1),$$

so that (11) is then simplified into:

$$\varepsilon_r \approx \left\{1 - \frac{(k_1'b)^2}{8}(\varepsilon_r - 1)\right\} \cdot M.$$

Solving this equation for ε_r leads to eq. (8).

The formula (8) holds to a good approximation only as long as $b\sqrt{\varepsilon_r} \ll a$ (e.g. $< 0.2\,a$). Most of the solid dielectrics that can be considered for use at frequencies in the order of 1000 Mc/s have a relative dielectric constant between 2 and 10. From the condition just mentioned it follows that the diameter of the rod must be smaller than 1/7th to 1/16th of the diameter of the cavity resonator.

The rod has to be much thinner still when taking measurements with materials such as barium titanate, where ε_r may be in the order of 1000 [11]). The cavity resonator described could be made suitable for such cases by drilling a small hole in the centre of the cover and the botton and inserting the dielectric in the form of a wire (thinner than 1 mm) through those holes.

[11]) G. H. Jonker and J. H. van Santen, The ferro-electricity of titanates, Philips Techn. Rev. **11**, 183-192, 1949.

In *fig. 11* a series of curves are given representing ε_r as a function of $k_1'a$, according to eq. (8), with b/a as parameter. Since $k_1'a = 2\pi a f_1/c$, the abscissa is directly proportional to the resonant frequency f_1 of the cavity resonator with the rod in it. For a certain cavity resonator, with given a, the constant $2a\pi/c$ can be incorporated in the scale units.

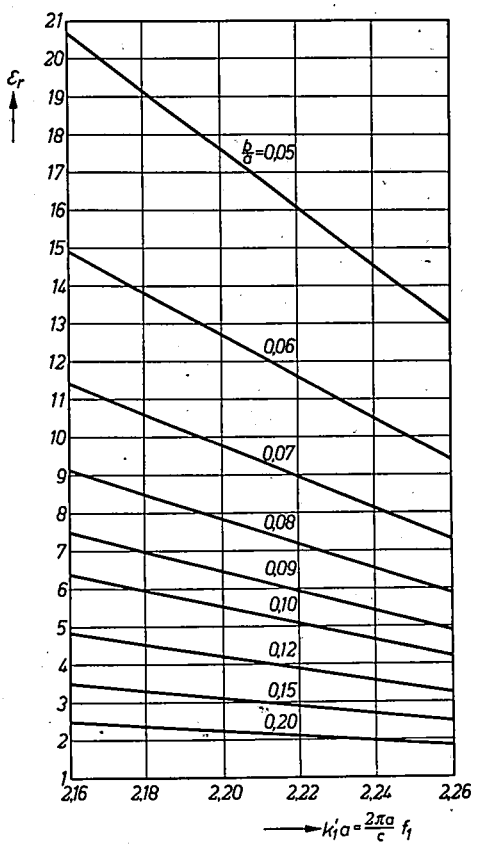


Fig. 11. The relative dielectric constant ε_r as a function of $2\pi a f_1/c$, with b/a as parameter, according to formula (8); configuration as in fig. 3.

The calculations for $\tan\delta$ — which will not be given here — yield the following formula (cf. formula (4)):

$$\tan\delta = \frac{\left(\frac{a}{b}\right)^2 D + \varepsilon_r - 1}{\varepsilon_r\left[1 + \left\{\frac{J_1(k_1''b)}{J_0(k_1''b)}\right\}^2\right]} \cdot \left(\frac{1}{Q_1} - \frac{1}{Q_0''}\right), \quad (13)$$

where D stands for the expression:

$$D = \left[\frac{2}{\pi k_1'a} \cdot \frac{1}{Y_0(k_1'a)\,J_0(k_1'b) - Y_0(k_1'b)\,J_0(k_1'a)}\right], \quad \ldots \quad (14)$$

which is represented in *fig. 12* as a function of $k_1'a$ ($= 2\pi a f_1/c$) with b/a as parameter. *Fig. 13* is a graph representing as a function of $k_1''b (= 2\pi b\sqrt{\varepsilon_r} f_1/c)$ the term $(J_1/J_0)^2$ occurring in the formula (13).

By analogy with the quantity Q_0' occurring in the L-C circuit, the quantity Q_0'' occurring in eq. (13) is understood to be the quality factor at the frequency f_1 if there were to be introduced in the cavity resonator a rod of a hypothetical loss-free dielectric of the same dimensions and with the same dielectric constant as the rod of the material under investigation. Calculations yield for $1/Q_0''$:

$$\frac{1}{Q_0''}=\sqrt{\frac{f_0}{f_1}}\cdot\frac{a}{a+d}\cdot\left\{1+\frac{\frac{ad}{b^2}D}{\left(\frac{a}{b}\right)^2 D+\varepsilon_r-1}\right\}\cdot\frac{1}{Q_0}. \quad (15)$$

This form is more complex than the relationship between $1/Q_0'$ and $1/Q_0$ for the normal resonant circuit, due to the fact that in the case of the circuit the dissipative losses in the capacitor electrodes can be ignored, whereas in the case of the cavity resonator owing to the introduction of the dielectric the current distribution is changed and thus also the dissipative losses are different.

Recapitulating, the procedure is as follows: From graphs like those in figs 11 and 12 we read the relative dielectric constant ε_r and the quantity D, respectively, for the measured resonant frequency f_1 and the value chosen for b/a. With D, the measured values f_0, f_1, $Q_0 = f_0/\Delta f_0$ and the dimensions a, b and d, we calculate $1/Q_0''$ according to the formula (15), and from the graph of fig. 13 we read the quantity $(J_1/J_0)^2$ for the calculated value of $k_1''b = 2\pi b\sqrt{\varepsilon_r} f_1/c$. We then have all the data required to calculate $\tan\delta$ from eq. (13).

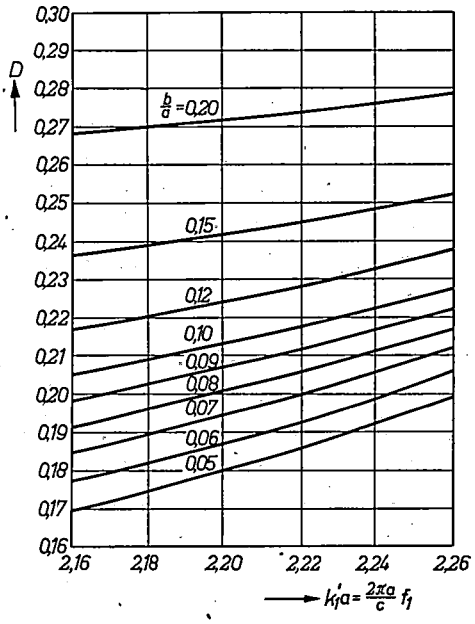


Fig. 12. The quantity D, according to (14), as a function of $2\pi a f_1/c$, with b/a as parameter.

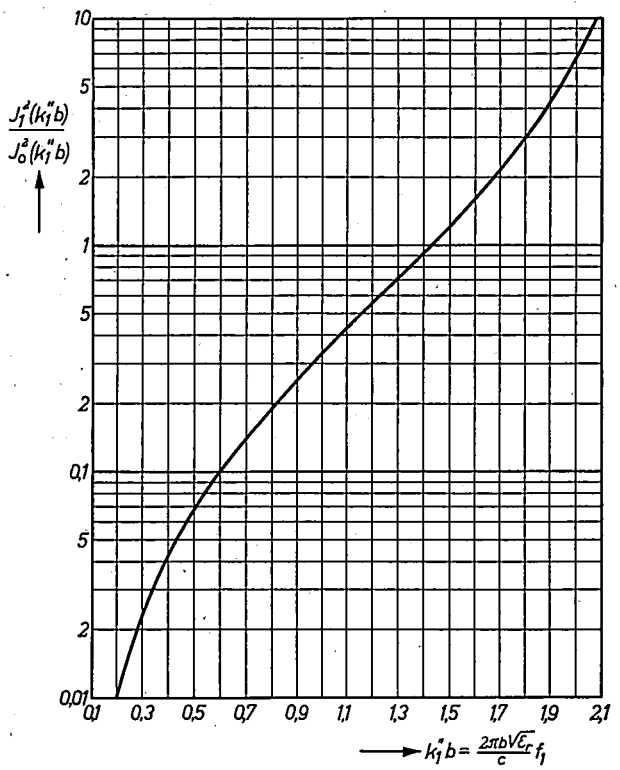


Fig. 13. The quantity $J_1^2(k_1''b)/J_0^2(k_1''b)$ occurring in formula (13), as a function of $k_1''b$.

Summary. The relative dielectric constant ε_r and the loss angle δ of a solid dielectric are determined from the detuning and the reduction of the quality factor to which a resonant circuit is subject when the dielectric is introduced in the electric field of the circuit. Owing to the high frequency at which the measurements are taken (about 3000 Mc/s) a cavity resonator is employed instead of a normal resonant circuit. This is cylindrical in shape and resonates in the E_{010} mode. For accurately measuring $\tan\delta$ of materials with very low losses it is necessary that the quality of the cavity resonator prior to the introduction of the dielectric should be extremely high. The quality factors hitherto reached were never higher than 70 to 80% of the theoretical value. This is ascribed to the fact that the irregularities left in the inner surface of the cavity resonator after the mechanical processing are greater than the penetration depth of the current (1 μ) and thus cause the path followed by the current to be considerably lengthened. By extremely careful finishing of this surface it has now been possible to reduce the depth of these irregularities to the order of 10^{-3} μ, thereby raising the quality factor (17,190) to 98% of the theoretical value.

As oscillator a reflex klystron is used, the frequency of which is variable within a certain range (3090-3260 Mc/s). The calibration has proved to be highly constant. As detector a silicon crystal is used, connected to a galvanometer.

A table is given showing some measured values of ε_r and $\tan\delta$ for a number of materials at a frequency of over 3000 Mc/s. The accuracy of the measurements for ε_r is $\pm$ 1% and that for $\tan\delta$ is $\pm$ 1% $\pm$ 0.2×10^{-4}. In an appendix formulae and graphs are given with the aid of which ε_r and $\tan\delta$ can be determined from the quantities measured.

SPECIAL X-RAY TUBES

by B. COMBÉE and P. J. M. BOTDEN. 621.386.1 : 616-073.75 : 615.849

As in most branches of technical engineering, also in the field of the application of X-rays there is to be seen side by side a development of fundamentally new possibilities — of those of recent date we would mention only the use of the electron image amplifier and of modern electron accelerators — and a further development of existing instrumentarium. As far as the latter development is concerned, in many constructions, in particular of X-ray tubes for special applications, there is a gradual quantitative progress to be observed, which is the fruit of continued research in regard to the materials used and the shaping of the products.

Reviewing the course of development of X-ray tubes up to the present day, a parallel is to be found with the development of the incandescent lamp. In both cases there was first a period of several decades marked by a wrestling with technical problems which ultimately led to the mastery of the physical and technological fundamentals of the product. This was followed by a period in which the field of application was widened by the development of a number of modifications of stabilized prototypes in order to meet the needs of special applications. In the case of the incandescent lamp this led to hundreds of different kinds of "special lamps" being produced, and although we cannot speak of hundreds there are nevertheless quite a respectable number of special X-ray tubes too, differing greatly in appearance and also in their properties. To determine roughly the scope covered by the latter we could take as the mean a normal, universal tube for X-ray diagnostics and a normal 250 kV therapy tube, and take as the extremes the tubes for diffraction radiographs with hours of exposure and the tubes for instantaneous X-ray photographs with an exposure of 10^{-6} sec, the miniature tubes for dental apparatus and tubes for examining materials with tensions of 2 million volts. (Here we leave out of consideration the generation of extremely hard X-rays, for example, by means of linear accelerators, betatrons or synchrotrons and suchlike, to which the conventional conception of an "X-ray tube" no longer applies.)

In this article a few examples will be taken from the store of modern X-ray tubes and it will be shown how, by the application of the materials now available and by the use of modern methods, it has been possible so to improve the construction of X-ray tubes as to make them more suitable for certain special purposes. First we shall discuss a diagnostic tube, the "Rotalix" O-55, and then two therapeutic tubes, the new CT tube for contact therapy and the ET tube for endotherapy.

The "Rotalix" O-55

The principle of the rotating anode, which has been applied in the "Rotalix" tube and was first put into practice by Bouwers [1]), has repeatedly been dealt with in this journal [2]). It is recalled that the motive for causing the anode to rotate lies in the increased specific loading capacity of the focal spot. To take an X-ray photograph of a part of the body that is in motion the exposure time has to be as short as possible, so that a great intensity of X-rays is desired, thus a large power on the anode. On the other hand, the focus, from which the X-rays emerge and in which the power from the bombardment by the stream of electrons is for the greater part converted into heat, has to be kept very small so as to limit the kinetic blurring of the X-ray image. Owing to the rotation of the anode the area bombarded by the electrons during the exposure is increased without enlarging the focus, so that for a given permissible focus temperature the focus can be more heavily loaded.

At first this improvement was applied only in X-ray tubes for diagnostic apparatus that was required to meet the highest demands. To put it more concretely, the "Rotalix" tubes hitherto produced were high-power tubes intended to be D.C. fed. The latter fact results from the former since large diagnostic installations are always D.C. fed, on the one hand because this is the most

[1]) A. Bouwers, Physica **10**, 125, 1930.
[2]) J. H. van der Tuuk, Philips Techn. Rev. **3**, 292-298, 1938 and **8**, 33-41, 1946.

rational method of generating X-rays [3]), and on the other hand because the additional cost of the high-tension valves etc. weighs relatively little in the case of large installations.

It having come to be the common practice to use the rotating anode for large X-ray diagnostic apparatus, it was only natural to introduce this refinement also in simpler installations working at a lower rating and with a self-rectified X-ray tube. Thus the "Rotalix" O-55 was designed, which is illustrated in *figs 1* and *2*.

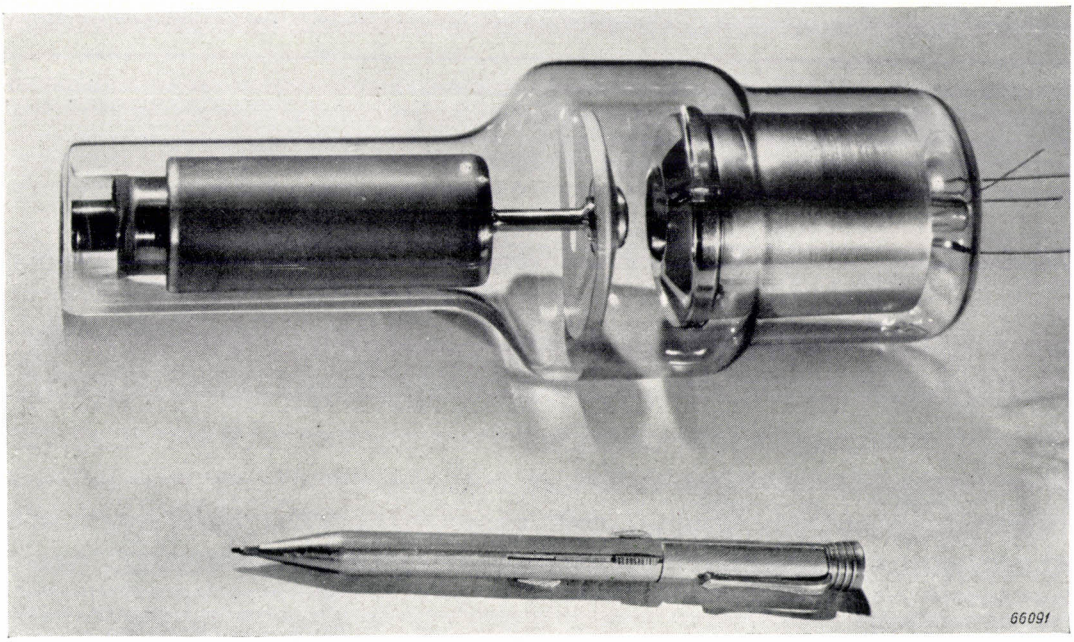


Fig. 1. The "Rotalix" O-55.

Before proceeding to discuss this tube (or at least the details in which it differs from the older types of "Rotalix" tubes) it has to be considered what purpose is served in making the anode of a self-rectified X-ray tube rotate.

Confining our considerations to exposures in the order of 0.1 sec, for which the question of the specific loading capacity of the focal spot is of most importance, it may be said that when working with direct voltage or with full-wave rectified, non-smoothed alternating voltage (*fig. 3a*), the loading capacity of a rotating-anode tube is 8 to 9 times that of a stationary-anode tube. In the case of self-rectified operation (fig. 3*b*) one would at first sight expect the improvement in this respect to be much less, since between two successive current peaks the focus has an opportunity to cool down; to put it in other words, as the anode is rotated there is always some part of the focal spot track that is not bombarded at all by the electrons. Roughly speaking, it might therefore be supposed that the increase in the loading capacity due to rotation of the anode would amount to no more than a factor 4.

A closer investigation, however, shows that this argument is not quite correct. In the above-mentioned gain of 8 to 9, as applying to D.C. operation, it is assumed that the maximum permissible focus temperature is not influenced by the rotation of the anode. This is indeed the case when the tube is D.C. operated; the focus temperature is then only limited by the mechanical stresses resulting in a gradual roughening of the anode surface (cracking). In the case of self-rectified operation, however, account has also to be taken of the risk of backfiring. As a consequence of this, in the case of a stationary anode, the maximum permissible focal temperature is much lower than when the tube is D.C. operated: in the negative half-cycle,

[3]) In self-rectified operation, for the same maximum permissible current (peak value) — determined mainly by the mains — the intensity of radiation obtained is only half that when operating with two-phase rectified voltage. For high powers self-rectified operation is also less suitable owing to the asymmetry of the voltage variation (heavy voltage drop in the half-cycle where the X-ray tube passes current). Finally the yield of X-rays per watt is greater when using a rectified, smoothed voltage than in the case of self-rectified operation.

i.e. when the anode is negative and the filament positive, the focus must not be allowed to become so hot as to emit electrons to any appreciable extent, because these would impinge upon the filament with considerable energy and soon destroy

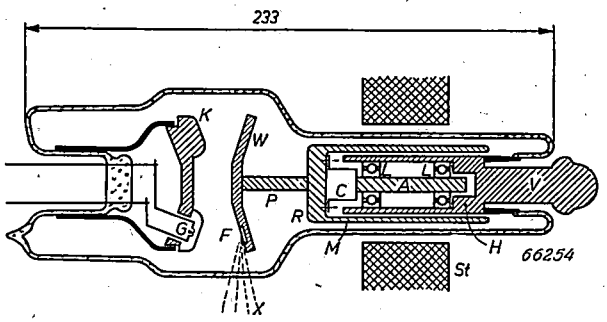


Fig. 2. Cross section of the "Rotalix" O-55. The filament G, placed in a slot in the cathode plate K, emits the electrons, which (during the positive half-cycle) are accelerated by the tube voltage and impinge upon the rotating anode, made in the form of a disk of tungsten (W), in the focus F. Under the electron bombardment this disk becomes red hot and a large part of the heat generated is carried off by radiation. Cooling by radiation is further promoted by the blackened copper sleeve M of the rotor R, to which the disk is connected by the thin molybdenum pin P. The rotor is driven by the rotating field of the stator winding St placed round the anode neck of the X-ray tube. The spindle A, connected to the rotor via a thin cylindrical part C, rotates in ball-bearings L lubricated with lead and fixed in the anode support H. This support is fused by means of a "Kovar" ring into the hard-glass wall of the tube and with the extension V forms the pole for connecting the anode to the high tension. X is the effective cone of X-rays.

it [4]). It is true that in the 0.005 sec between the current peak in the positive half-cycle and the beginning of the negative half-cycle the focus cools down some hundreds of degrees, but the temperature limit below which thermionic emission becomes harmless lies so much lower than the limit set by the surface roughening of the anode that, in the case of self-rectified operation, the permissible loading capacity of the focal spot is only about half that in the case of operation with a full-wave rectified voltage. When the anode is made to rotate instead of being stationary the problem of backfiring at once assumes a different aspect. Those parts of the anode which are intensively heated in the positive half-cycle are no longer facing the filament in the negative half-cycle, since meanwhile the anode has turned (it is caused to rotate at the rate of almost 50 rev./sec). Consequently any electrons emitted by the anode in the positive half-cycle do not impinge upon the filament but can be intercepted by parts of the tube where they can do no harm. Thus it is possible to work with a higher temperature in the negative half-cycle than is permitted when the anode is stationary, and this favourable factor, added to the fact that in the time interval of 0.005 sec following the peak load the part of the anode in question has much better opportunity to cool down than when it is in the stationary state [5]), allows of practically the same gain in loading capacity being attainable with alternating-voltage as with direct-voltage supply.

Of course, even in this favourable situation as regards backfiring every precaution must be taken to limit or render harmless any electron emission from the anode. In the process of manufacture steps are taken to keep the anode as free as possible of impurities, particularly those with a low work function, which would most readily give rise to thermionic emission. Furthermore, every tube is specially tested for the occurrence of anode emission. To provide against the consequences of an increase in anode emission during the lifetime of the tube, the glass wall — which, next to the filament, is most sensitive to electron bombardment — is protected by giving the cathode plate a much larger diameter than in the case of the normal "Rotalix" tubes (see fig. 2). The configuration of anode and cathode is made so favourable, as regards the maximum local field strengths, that the tube can safely be operated with an alternating voltage of 100 kV_{peak}; when working with a rectified voltage, which of course is also possible, one can go up to 110 kV_{peak}.

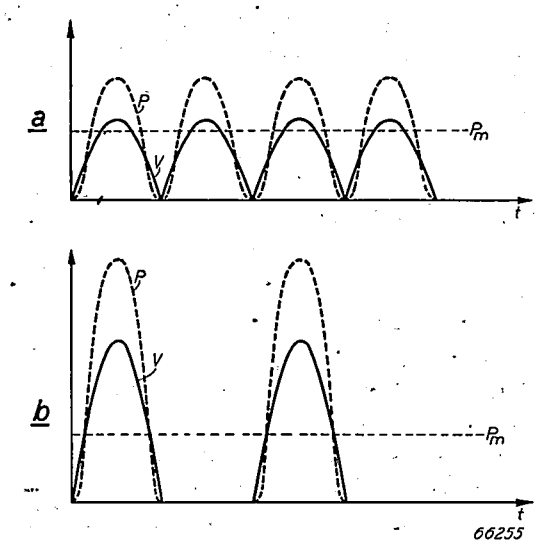


Fig. 3. Variation, as a function of the time t, of the instantaneous values of the voltage V and the power P of an X-ray tube: *a*) when operating with full-wave rectified alternating voltage;
b) in self-rectified operation with the same mean power P_m.

[4]) For a more detailed discussion of backfiring see, e.g., J. H. van der Tuuk, Philips Techn. Rev. 6, 309-315, in particular p. 311.

[5]) In 0.005 sec it has rotated through an arc equal to several times the width of the focus, so that after 0.002 sec it will already be out of the range of the electron beam. For a more detailed discussion of the cooling problem see W. J. Oosterkamp, Philips Res. Rep. 3, 303, 1938.

In a "Rotalix" tube for D.C. operation the anode is rotated only during an exposure. For screening, which is done with the same tube but with a much lower load on the focus, no purpose is served by rotating the anode and it is therefore kept stationary, to save the bearings from unnecessary wear. With the "Rotalix" tube for self-rectified operation, however, it is a different matter, for in this case the rotation of the anode performs an important function in counteracting backfiring, so that the anode must always rotate, even when screening. To make this possible the bearing construction has been modified, compared with that in the old type of tube. In particular the anode support (*H* in fig. 2) carrying the ball races for the rotating anode is now made entirely of iron; formerly it was made of copper, with an iron jacket round it for concentrating as far as possible in the rotor of the anode the magnetic field of the stator winding round the X-ray tube. The new construction has greater mechanical strength and thus easily permits of continuous rotation of the anode, even between successive screenings and radiographing, so that the operating apparatus can be simplified. It is true that the conduction of heat through the iron anode support to the oil in the shield, in which the whole tube is contained, is less satisfactory than with a copper support. However, the necessary low temperature of the bearings is maintained by the thermal resistance provided by a thin-walled cylindrical part connecting the chrome-iron spindle to the rotor sleeve of the anode (see *C* in fig. 2).

Since the "Rotalix" O-55 is intended for small installations it was desired to make the dimensions both of the tube and of the shield much smaller than those of the older type of "Rotalix" tubes. As far as the tube is concerned, the tungsten disk of the anode has been given a smaller diameter (55 instead of 75 mm), thus sacrificing a small part (factor 1.2) of the aforementioned gain in loading capacity of the focal spot. Further, the rotor sleeve is narrower, so that the stator winding to be passed round the tube at this part could be given a much smaller external diameter, whilst, finally, the rotor and thus the whole tube is made slightly shorter. As a result also the shield could be made much smaller (*fig. 4*).

The surface and wall temperature of the shield determine the power that it is permissible to generate in the X-ray tube under continuous operation. Owing to a highly heat-resistant material being used for the insulators, and especially for male and female plugs with which the tube is connected to the high tension supply, the oil and the wall of the shield may be permitted to reach very much higher temperatures than before, viz 100 °C to 110 °C compared with 60 °C to 70 °C, so that the apparatus can be operated with a continuous power of 210 watts (the tube itself can withstand 350 watts continuously).

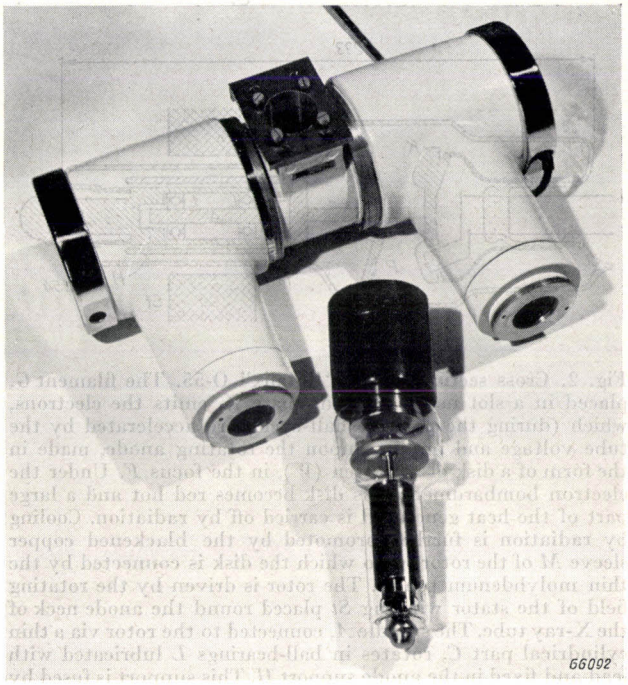


Fig. 4. Shield of the "Rotalix" O-55. In the foreground the tube itself, fitted with a "Philite" cap enveloping the cathode connection.

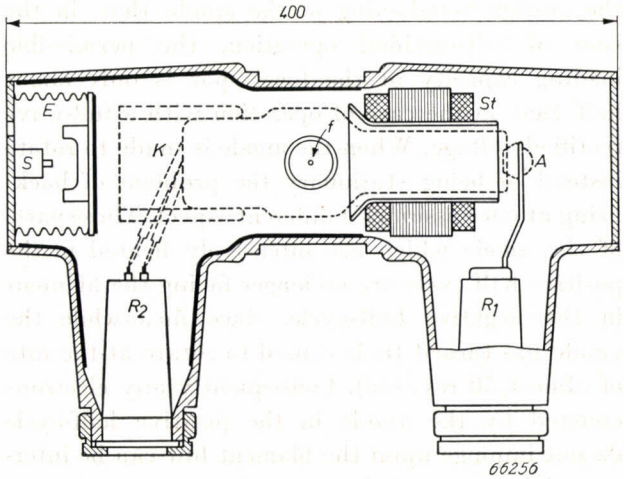


Fig. 5. Cross section of the shield of the "Rotalix" O-55. The tube is outlined in broken lines; the focus lies exactly in the middle, at *f*. *A* anode pole, *K* cathode pole, with female plugs R_1, R_2 for connecting the high-tension cables. *St* stator winding. *E* expansion chamber, taking up the thermal expansion of the oil, with safety switch *S* switching off the tube voltage as soon as the mean temperature of the oil exceeds a certain limit.

The shield of an X-ray tube immersed in oil is always fitted with an expansion chamber, for instance in the form of bellows, which allows for thermal expansion of the oil without any appreciable increase of pressure. In view of the higher temperature reached by the oil, for the "Rotalix" O-55 the bellows had to be made relatively much larger. The space for the bellows was easily found in the shield at the cathode end of the tube (*fig. 5*); the cathode half of the tube is always shorter than the anode half, but for the sake of good manoeuvrability it is nevertheless desirable to make the shield symmetrical, with the X-ray focus exactly in the centre.

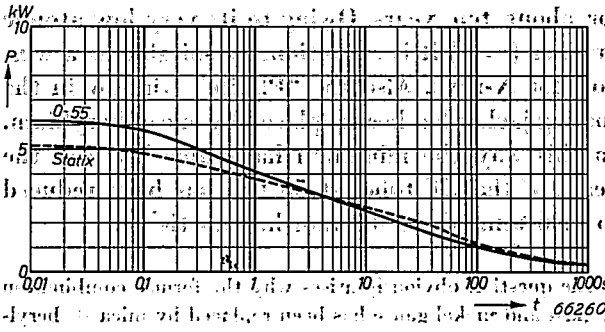


Fig. 6. Permissible load P (in kW) of the "Rotalix" O-55, when operated with 50 c/s alternating voltage, as a function of the time t in sec. Dotted curve: loading capacity of a comparable X-ray tube with stationary anode ("Statix"). As the graph shows, both tubes can be about equally loaded, but the (apparent) focus of the "Rotalix" O-55 is 1 mm × 1 mm as compared with that of the "Statix" measuring 3.1 mm × 3.1 mm, so that the specific loading capacity of the focal spot of the "Rotalix" is about 9 times greater.

Finally, *fig. 6* gives the loading capacity of the O-55 type of tube as a function of the exposure time (at 50 c/s). For comparison the permissible load is also given for a similar tube with stationary anode ("Statix"), the focus of which is three times as long and as wide as that of the O-55.

A new X-ray tube for contact therapy

When malignant growths are being treated with X-rays it is of the utmost importance to make sure that the healthy tissue around the tumour is not damaged, or at least no more than the healthy tissue can stand. The most favourable situation is when the tumour is on the surface of the skin, for then the X-rays do not have to pass through healthy tissues before reaching the tumour and it is only the tissue behind it that one must try to save. This is achieved to a large extent by applying the method of contact therapy.

For contact therapy, which has already been fully described in this journal [6]), soft X-rays are used, which are almost entirely absorbed in the tumour, and the focus is placed at a distance of only a few centimetres from the skin. With this short distance advantage is taken of the quadratic reduction of the intensity with the distance, since the relative differences in distance from focus to tumour and from focus to the underlying tissue are then large and the dose on the underlying tissue, even without absorption, is therefore relatively small.

One of the most important properties of the X-ray tubes designed for contact therapy is the "inherent filtration" of the tube. It is desired to use soft rays, which are for the greater part absorbed in a thin layer of tissue, but for this very reason such rays are easily absorbed before they ever leave the X-ray tube! The fact is that the rays have to pass through at least one window and usually two (or even more) windows: one in the wall of the X-ray tube proper forming part of the high-vacuum envelope, and another in the earthed metal shield round the tube. It is just the softest components of the spectrum emitted by the anode that these windows (the "inherent filter") absorb most.

This can be expressed in more exact terms. In passing through any material (filter) the X-radiation always becomes harder. As a measure for the hardness the half-value layer H is used, this being the thickness of a layer of aluminium or some other specified material which reduces the intensity of the radiation to half its original value. It may therefore be said that owing to the presence of the inherent filter the half-value layer of the useful radiation of an X-ray tube cannot be reduced below a certain limit. In the case of the X-ray tube for contact therapy described earlier in this journal [6]) the minimum half-value layer, for a tube voltage of 50 kV, was $H = 0.30$ mm Al, or about 4 mm human tissue. The focus could be brought to a distance of 2 cm from the skin.

The small value of H mentioned was reached by a very unconventional design of the tube, which is represented diagrammatically in *fig. 7*. The ring-shaped filament lies between the anode and the patient, so that the effective X-ray beam passes out through this ring. On its way the beam passes through two windows, one in the cathode can — which in this tube is earthed — and one in a metal jacket placed round the tube to provide for cooling of the cathode can by means of a forced current

[6]) H. A. G. Hazeu, J. M. Ledeboer and J. H. van der Tuuk, An X-ray apparatus for contact therapy, Philips Techn. Rev. 8, 8-15, 1946.

of air. The first window in the old type of tube was made of glass, 0.25 mm thick, and as protection against possible electron bombardment a thin gauze of nickel was placed in front of it (not shown in the drawing). The second window, in the jacket, was of "Philite".

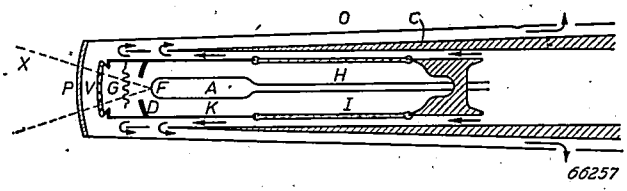


Fig. 7. Principle of the construction of the CT tube (X-ray tube for contact therapy). *K* earthed cathode can with ring-shaped filament *G* and focusing ring *D*. *A* anode with focus *F*. *I* insulation between cathode can and anode holder *H*. *V* window in the cathode can. *O* cooling jacket. *C* insulating cylinder for guiding the cooling air (see arrows). *P* window in the cooling jacket. *X* effective X-ray beam.

For most cases where contact therapy is to be applied the half-value layer of 4 mm body tissue is small enough, but for the treatment of some very superficial affections still softer rays are desired (these are often referred to as "Grenz" rays, because they lie at the "Grenz" (boundary) of the spectral range of the medically useful X-rays). For the latter cases it is necessary to try to make the inherent filter of the X-ray tube even smaller than that of the construction described.

A means of reducing the inherent filter lies in the application of what is known as Lindemann glass (a glass of lithium beryllium borate) as material for the tube window. This kind of glass, however, has the drawback that in course of time it disintegrates, especially in humid surroundings. Moreover it is highly fragile and — what applies to glass windows in general but especially so to windows of Lindemann glass — when the glass is fused into the tube wall it is inevitable that rather large differences appear in the thickness of the glass locally and as between one tube and another. The local differences cause the distribution of the dosage rate over the irradiated field of the skin to be uneven. The individual differences have to be compensated by means of separately chosen extra filters, in order to avoid too great variations in output as between one tube and another. These filters necessarily involve an undesired loss in the dosage rate.

In recent years it has been found possible to reduce the inherent filtration of the CT tube in another way. The window in the cathode can is now being made of a mica disk only about 20 microns thick, the disk being 1.5 cm in diameter and secured vacuum-tight in the opening of the cathode can by a newly developed technique [7]). When fused in, the disk remains flat and of even thickness all round, so that a uniform distribution of the dosage is obtained without any trouble, whilst the desired uniformity of all tubes does not involve any loss. This extremely thin disk of mica stands up against atmospheric pressure, but, like the glass formerly used, it cannot withstand bombardment by electrons. In the new construction it is therefore protected by a disk of beryllium 0.2 mm thick. This light metal, which formerly was used mainly as an alloy component in copper for contact springs and suchlike and for nuclear-physical research, has now been used in X-ray apparatus for about ten years. Owing to its very low atomic number (= 4) it is exceptionally suitable as a window for X-rays. Also the "Philite" window in the cooling jacket has been replaced by one of beryllium. In this way the minimum half-value layer of the beam of the CT tube, at 50 kV, has been reduced to 0.06 mm Al, or 0.8 mm human tissue.

The question obviously arises why the former combination of glass and nickel gauze has been replaced by mica + beryllium and not by beryllium alone. Such is in itself possible, but the solution chosen has great advantages. For the vacuum-sealing of a beryllium window the disks would have to be thicker than those which are now being used. Since beryllium is a very expensive material this makes a considerable difference in cost. Furthermore the treatment of the beryllium would then involve much more work and take longer than is the case with thin foils. Such treatment was considered undesirable, not only on account of the additional cost it entails, but more so because of the known toxic effects that beryllium has upon anyone coming too much in contact with it. Finally, the vacuum-tight fusing-on of a beryllium window is a much more difficult operation (and causes more wastage) than that of a mica disk.

The advantage of the reduction of the inherent filter of the CT tube lies particularly in the fact that even with a tube voltage of only 10 kV a reasonable dosage rate [8]) can still be obtained,

[7]) The use of mica as a window for X-rays has already been described in this journal in connection with a radiation counter tube: J. Bleeksma, H. J. Di Giovanni and G. Kloos, X-ray spectrometer with Geiger counter for measuring powder diffraction patterns, Philips Techn. Rev. **10**, 10, 1948. By that time J. Bleeksma had already initiated and developed the use of mica windows for X-ray tubes proper.

[8]) The dosage (expressed in roentgen units) is measured by the amount of energy absorbed in a certain volume of air. Since the absorption coefficient for the radiation varies greatly with the hardness of the rays there is no simple relation between the number of roentgen units per unit of time and the power entering the measuring volume. The dose per unit of time does not therefore signify an intensity of radiation (watts/cm^2), and for that reason the term "dosage rate" has been introduced; this quantity has the dimension watts/gram.

namely about 6400 roentgen units per minute for 2 mA tube current and with a focus-skin distance of 2 cm. (The therapeutic total doses required are in the order of 6000 r, the permissible dosage rate is 0.3 r per week.) The half-value layer with this low voltage is only 0.3 mm human tissue! With a tube voltage of 50 kV the dosage rate at the same current and the same distance is 160,000 r/min, as against 8000 r/min in the case of the old type of tube.

In many cases the medical practitioner will not, of course, want to work with the smallest half-value layer. For thicker tumours harder rays are needed, and the CT apparatus should therefore be able to supply also these harder rays. The window has consequently been designed as indicated in *fig. 8*. The ring carrying the beryllium window

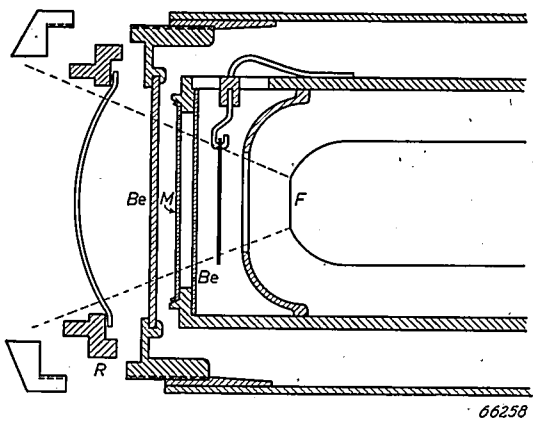


Fig. 8. New construction of the windows of the CT tube, with ring *R* for additional filters as required. *M* mica window, *Be* beryllium windows.

of the cooling jacket has been made so that a second ring can be screwed onto it, by means of which various additional filters can be placed in the path of the X-rays, in the form of aluminium foils. An additional filter of 0.1 mm Al increases the half-value layer from 0.3 mm tissue at 10 kV without additional filter, to 0.5 mm; at 50 kV the corresponding half-value-layers are 0.8 and 1.3 mm. Thus it is possible to vary the dosage distribution in the irradiated tissue within wide limits. For cases where a fixed irradiation technique has been established, based on the old CT apparatus, the radiation properties of the new tube can be made practically identical to those of the old one by placing in front an additional filter of "Philite" admixed with a certain quantity of metal oxide powder.

X-ray tube for endotherapy

It has already been pointed out that in contact therapy the conditions for a cure are more favourable than is the case with "depth therapy", where only too often nothing comes of the hoped-for success owing to the healthy tissue being seriously damaged. With some tumours not seated in the outer skin the advantages of contact therapy are still available to the practitioner, notably in cases of growths lying in cavities of the body, in particular the frequently occurring tumours in the uterus. For this purpose X-ray tubes have been developed with the focus at the end of a long, thin, earthed anode (immersion anode) which can be inserted in the cavity of the body, so that the source of the rays is brought in the immediate vicinity of the tumour (sometimes there may still be several centimetres of healthy tissue between the anode and the tumour). This is referred to as "endotherapy". Owing to the size and the position of the growths to be irradiated, for this application not too soft rays are desired, corresponding to tube voltages of up to say 60 kV or still better up to about 100 kV, a rather heavy filter being used.

Formerly it was quite common, and still at the present day it is a frequent practice where the situation is more favourable for it or where economic considerations dictate it, to irradiate growths in the uterus not with X-ray tubes but with radium in the form of needles or suchlike inserted *in* the tumour and left there for a few days until the desired total dosage has been administered.

It is to be pointed out that the X-ray tubes referred to here have not been developed solely with the object of replacing radium in certain cases. They make it possible for the irradiation to be carried out in a few minutes instead of days. This is of importance for a modification of "endotherapy": the irradiation of growths that may be seated deep in the body but have been made accessible from the outside by surgical operation; of course in such a case only a very short time is available for the irradiation.

For endotherapy it is particularly essential that the tube should be easy to handle, and therefore light in weight and of small size, that the anode should be very thin and of considerable length, and that the dosage rate should be high.

The endotherapy (ET) tube constructed by us fully answers these requirements. The tube with shield is illustrated in *figs 9* and *10*.

As regards easy handling, a striking advance has been made upon former designs by departing from the conventional mounting in air and placing the glass tube in oil in a sealed jacket. The principle of this construction has been followed for a long time in other X-ray tubes and in this connection has frequently been mentioned in this journal (see the foregoing, and particularly the article quoted in footnote [4]), so that we need

not go into it any further here. In passing we would, however, point out another constructional feature helping to make the tube easier to handle: the cathode of the tube is connected to the high tension by means of a rubber cable with a conical end fitting into a female plug of synthetic resin attached to the tube (fig. 10). This simple plug construction, which for a considerable time has been applied by C. H. F. Müller A.G. of Hamburg for all sorts of X-ray tubes, makes it possible for the shield, for a given voltage, to be made much shorter than when applying the former combination of male and female plugs, both of synthetic resin. As a result of these expedients the total weight of the endotherapy tube with shield, which for the best design was formerly about 25 kg, has now been reduced to 3.5 kg, whilst the tube can be used with voltages up to 100 kV.

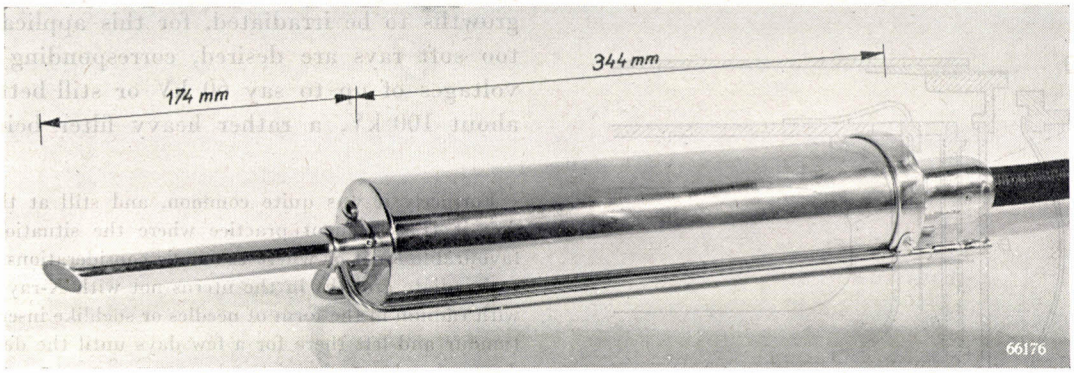


Fig. 9. ET tube (X-ray tube for endotherapy). The focus is on the oblique target at the end of the earthed anode pipe protruding from the left-hand end of the cylindrical shield. On the right the high tension cable which is connected to the cathode and the connections for the cooling water.

The other requirements mentioned, viz. a long and thin immersion anode and high dosage rate, call for rather more attention. The beam of electrons directed upon the focus must not be allowed to touch the walls of the hollow anode pipe, which means to say that it must already be extremely narrow at the entrance to the anode and must not diverge too much as it passes through. Consequently the longer and the thinner the pipe the more attention has to be paid to the problem of focusing the electron beam.

In our case it was definitely decided that the focusing would have to be done electrostatically, that is to say by giving the electrodes a suitable shape, and not by means of magnetic fields, since in the latter case there is the inconvenience that a variation of the tube voltage involves a change in the dimensions of the focus.

Electrostatically focused, narrow electron beams have already been known for some time, in their application for cathode-ray tubes where the focusing is brought about with the aid of auxiliary electrodes placed between cathode and anode and to which suitable voltages are applied, if necessary variable from the outside. For an X-ray tube, where we have to do with much higher voltages, such a method is not practicable. Here the focusing has to be brought about by a suitable shaping of the cathode, the filament mounted in it and the anode. Roughly speaking, the object is achieved with the aid of a beaker-shaped cathode with its open end facing the hollow anode and with the filament coil mounted near the bottom of the beaker. The quality of the beam depends greatly, however, upon the details of this geometrical configuration, as for instance the depth of the cathode beaker.

To make the anode as thin and as long as possible we took advantage of extensive investigations in regard to the shaping of the electrodes carried out by A. Kuntke (of C. H. F. Müller A.G., Hamburg), the results of which have been successfully applied also for various other kinds of X-ray tubes with a similar anode. With the aid of a model of the electrode configuration made to scale and immersed in a tank filled with an electrolyte, measurements were taken to determine the potential field [9]). *Fig. 11* shows as an example the field

[9]) For a further description of this method see, e.g., G. Hepp, Measurements of potential by means of the electrolytic tank. Philips Techn. Rev. **4**, 223-230, 1939.

found for one of the configurations investigated. The part of the field lying between the heavily drawn equipotential line (2%) and the entrance to the anode can be regarded as a certain type of electronic lens the focusing action of which upon any electron beams entering it can be determined by geometric-optical methods. Since the directions of incidence of the rays entering this lens could

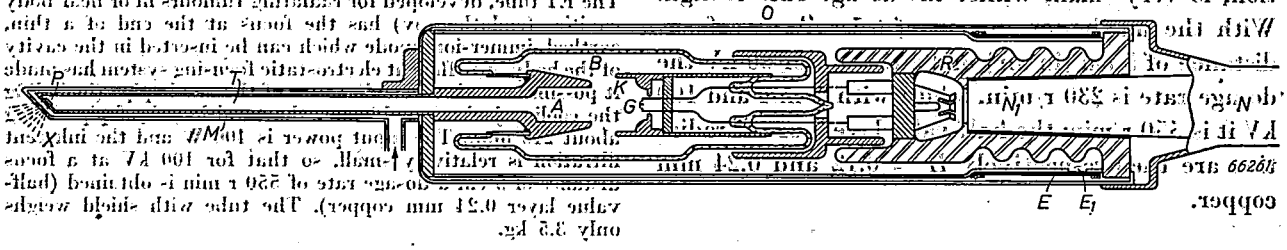


Fig. 10. Cross section of the ET tube with shield. B evacuated tube, A hollow immersion anode with pipe T and target P upon which the electrons impinge. The X-rays, X, pass out through the target. K cathode beaker with filament G; R female plug, connected to the cathode, which is mounted, together with the tube, in the shield O filled with oil. N high-tension cable with conical end N_1. M cooling-water jacket for cooling the immersion anode. E rubber tubing allowing for thermal expansion of the oil, via the opening E_1.

not be constructed from the equipotential lines with sufficient accuracy, they were determined by imitating the narrow first part of the field by a rubber membrane and the electrons by small balls which are caused to roll over the membrane from the point where the filament would be situated [10]). From this mechanical analogue it immediately becomes evident that the spiralized filament in particular has to be constructed and mounted with a high degree of precision.

The focusing reached with the configuration ultimately chosen as the result of these investigations is such that with a total length of pipe of about 210 mm the diameter of the focus can be less than 6 mm. With the cooling-water jacket placed round the anode pipe we thus arrived at an overall diameter of 18 mm, which is considerably smaller than that of former constructions.

It goes without saying that the power of 400 watts previously mentioned cannot be dissipated on the focus inside the patient's body without adequate means of carrying off the heat. Hence the water cooling referred to above. The system of supply and discharge of the water is clearly shown in fig. 9. The water flows along the anode in a narrow jacket and along the target affixed at an angle to the end of the anode, this target being bombarded by the electron beam (fig. 10). The X-rays have to penetrate through the target (hence the name transmission target) and in order to avoid too great a loss of radiation by absorption the target is made of a very thin foil of copper. Copper is chosen for its good mechanical properties and its relatively low atomic number. For a good yield of X-rays in the focus, however, a high atomic number is needed and therefore an extremely thin layer of gold has been applied to the target on the inside. The flat head of the cooling jacket, through which the X-rays have to pass, consists of a very thin foil of copper, which makes the jacket watertight, whilst its outer face is covered with a thin plate of a special kind of synthetic resin absorbing the soft components of the secondary radiation by the

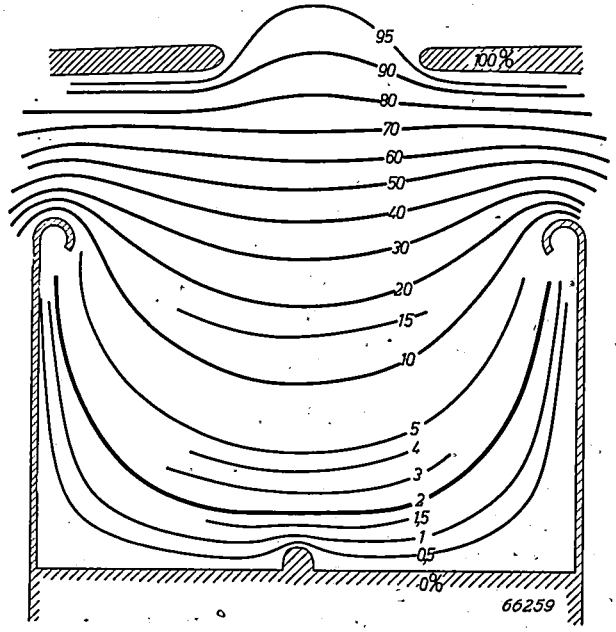


Fig. 11. Equipotential lines of the electrostatic field between cathode and anode, for the configuration of these electrodes as drawn. The lines have been constructed from measurements taken with a model in a tank filled with an electrolyte. 0% is the cathode potential, 100% the anode potential.

[10]) See P. H. J. A. Kleynen, The motion of an electron in two-dimensional electrostatic fields, Philips Techn. Rev. 2, 338-345, 1937.

X-rays in the copper. These soft components — mainly the CuK-alpha radiation with a wavelength of about 1.5 Å — are undesired because when irradiating a tumour slightly below the inner wall of the cavity in the body they give a high extra dosage rate on that wall.

Owing to this construction of the target and the cooling jacket the inherent filtration of the tube, compared with similar tubes of a different construction, is very small, whilst the dosage rate is high. With the maximum current of 6.5 mA at a focus distance of 5 cm (in air) and a tension of 60 kV the dosage rate is 230 r/min, whilst with 4 mA and 100 kV it is 550 r/min; the half-value layers of the radiation are then respectively $H = 0.12$ and 0.24 mm copper.

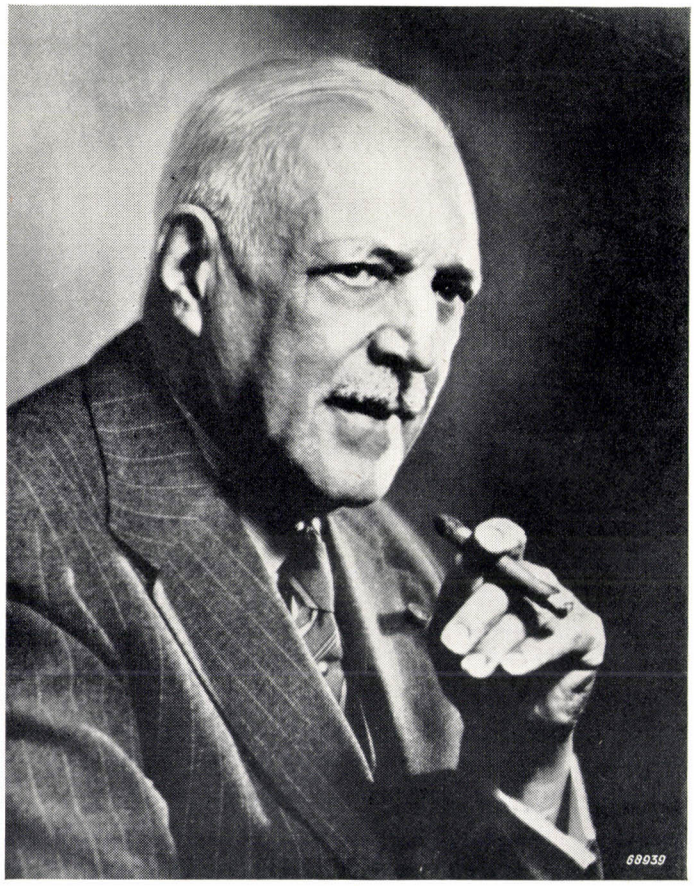
Summary. Characteristic details of three newly developed types of X-ray tubes are described. The "Rotalix" O-55 is a diagnostic tube with rotating anode especially constructed for self-rectified operation. Owing to its high loading capacity with small focus dimensions (about 6 kW during 1/10 sec on a focus of 1 mm × 1 mm) it is possible to obtain with a modest apparatus much better radiographs than hitherto. The new tube for contact therapy (CT tube) is provided with a window construction consisting of a vacuum-tight fused-in mica disk with a foil of beryllium in front of it. This gives the tube such a small inherent filtration that the minimum half-value layer of the radiation is thereby reduced to 0.3 mm human tissue for a tube voltage of 10 kV (dosage rate at 2 cm focus-skin distance 6400 r/min) and to 0.8 mm for 50 kV (160,000 r/min). The ET tube, developed for radiating tumours in or near body cavities (endotherapy) has the focus at the end of a thin, earthed, immersion anode which can be inserted in the cavity of the body. An efficient electrostatic focusing system has made it possible to reduce the diameter of the anode pipe (including the cooling jacket) to 18 mm, the total length of pipe being about 210 mm.. The input power is 400 W and the inherent filtration is relatively small, so that for 100 kV at a focus distance of 5 cm a dosage rate of 550 r/min is obtained (half-value layer 0.24 mm copper). The tube with shield weighs only 3.5 kg.

VOL. 13 No. 4 pp. 81-108 OCTOBER 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS



IN MEMORIAM
Dr. h.c. A. F. PHILIPS

IN the early hours of Sunday, Oct. 7th, 1951, Anton Frederik Philips, born at Zaltbommel on the 14th of March, 1874, and one of the founders of the Philips concern, passed away at the age of seventy seven.

He was originally intended by his father, B. F. D. Philips, himself a banker, to occupy a position in the banking business. In 1895, at the age of 20, he was requested to assist his 16 years older brother, Gerard Leonard Frederik Philips, in developing the Philips lamp factory, founded in 1891, by undertaking the commercial part of the business, and not only did he succeed in bringing the young concern through its initial difficulties but he also very soon became, at his brother's side, the pushing power of this industry.

He was deeply convinced of the necessity of research, and the erection of a new building for the Physical Laboratory — founded in 1914 — was one of his first deeds after his brother resigned in 1922 and the management of the concern fell entirely into his hands. Although not a technician he on several occasions functioned as chairman of technical discussions in the laboratory when the development of important new technical products was to be promoted.

In 1939 Dr. Philips resigned as managing director, but in the years 1940-1945, in the U.S.A., he was still active in the interests of the concern and its overseas activities.

He did not assist personally at the festivities on the occasion of the 60-years' jubilee of the concern, May 1951, on the grounds of ill-health, but television made it possible for him to follow the events at home, which he did with lively interest. To the very last all questions concerned with the future of the Philips industries had his full attention.

The Philips personnel all over the world, and especially the staffs of the laboratories, deeply regret the passing of this remarkable man to whose indomitable initiative they are proud to owe their enviable sphere of activities.

THE EDITORS, PHILIPS TECHNICAL REVIEW

NEW ELECTRONIC TUBES EMPLOYED AS SWITCHES IN COMMUNICATION ENGINEERING

II. SWITCH TUBES

by J. L. H. JONKER and Z. van GELDER.

621.385.832:621.318.57:621.39

The authors continue their discussion of some experimental tubes designed for performing switching functions, in particular of tubes with which switching is brought about by the deflection of an electron beam. Among others, a tube has been constructed which can serve in this way as a multi-way switch.

Electromagnetic relays, such as are frequently used, inter alia, in line telephony, exist in a large variety. As far as the contacts are concerned, for instance, there are to be distinguished make contacts, break contacts and change-over contacts (see fig. 1 in a previous article[1])), which can be operated in various numbers and combinations by a single excitation coil. Exclusively make and break contacts are one-way switches (mono- or multi-polar), but there are also multi-way switches, such as relays with change-over contacts, selectors and line finders, which connect one contact point to one of two or more other points. A distinction can also be made according to the switching speed: as a rule the switching speed has to be as high as possible, but in certain cases some delay is essential in order to ensure that the relay comes into action after another has performed its function.

In article I an electron tube was discussed which can act as relay with two contacts closed or opened simultaneously (bipolar symmetrical contact tube). Here it will be shown how tubes can be constructed for more complicated relaying functions.

In the tube just referred to, the control-grid voltage determines whether the electrons emitted by the hot cathode can reach a more positive electrode or not; this corresponds to the closing or opening of the contacts. Wider switching possibilities are afforded when the electrons are concentrated to a beam which, as in the case of a cathode-ray tube for an oscilloscope or for television reception, can be directed by deflection and thus caused to act, as it were, as a switch arm. This type of tube has been given the name of switch tube. Such tubes have already been described in various publications and patents, and suggestions have been made for applying this principle for multifarious switching operations[2]).

A technical application is found in what is known as the distributor, a tube used in multiplex telephony systems for connecting a line (or radio link) to different channels in turn. In this tube the electron beam is deflected in such a way as to make contact in rapid succession with a number of electrodes arranged in a circle. For a multiplex connection two such tubes are needed, working synchronously.

The deflectable-beam type of switch tubes so far known have a number of serious drawbacks: owing to their large dimensions (they are much larger than a normal amplifying tube) they take up much space between the other elements and their size alone already makes them rather expensive; the high working voltage (some thousands of volts) is an undesired complication; the maximum permissible beam current (in the order of 100 μA) is low and as a consequence the resistance of the external circuit is high, so that the time constant, which is usually required to be very small, is rather large.

When in a tube like the distributor the beam current is modulated with the aid of a control grid the output voltage is in anti-phase with the input voltage, and in many cases this is an objection.

---

[1]) J. L. H. Jonker and Z. van Gelder, New electronic tubes employed as switches in communication engineering, I. Contact tubes, Philips Techn. Rev. **13**, 49-54, 1951 (No. 3), here further referred to as I.

[2]) A similar tube was described in: Tj. Douma and P. Zijlstra, Recording the characteristics of transmitting valves, Philips Techn. Rev. **4**, 56-60, 1939 (figs. 2, 3 and 4).

As explained in article I, however, by means of secondary emission — which is not employed in the tube referred to in note [2]) — it can be so arranged that the voltage variations of the output electrode are in phase with those of the corresponding input electrode. Instead of bearing the character of a conducting switch arm, owing to the secondary emission, the beam then functions rather as a movable element capable of effecting contact between two adjacent electrodes.

The aim should therefore be to produce cathode-ray tubes — provided with secondary-emission contacts — in the size of normal amplifying tubes and suitable for the voltages and currents commonly used with them (e.g. some hundreds of volts and a few mA). A good step in the right direction was the introduction of the ribbon-shaped electron beam, which will now first be discussed.

Ribbon-shaped electron beam

In a contact tube such as described in article I there is no difficulty in producing a primary stream of electrons which, given a fairly large cross section, has an intensity of 1 to 2 mA and which can be passed through or blocked as desired with the aid of a control grid. The accelerating voltage required need be no more than say 300 V.

It is a different matter, however, when an electrostatically deflectable beam is required, which has to serve two or more contacts and, in order to avoid cross-talk, has to be focused on one contact at a time. A similar problem is encountered in cathode-ray tubes for oscilloscopes and for television receivers: for the sake of a sharply defined picture the spot on the fluorescent screen has to be small. Owing to the mutual repulsion of the electrons this is possible only when the current density in the beam is small (i.e. with low current strength) and the electrons have a high velocity (i.e. a high voltage). That is why oscilloscope and television tubes are made for currents in the order of 10 to 100 μA and for voltages of 1000 to 25,000 V. Since the picture field has to be rather large these tubes are made with the familiar elongated shape.

An important factor arising in the problem of focusing is the space charge in the electron beam. The greater the space charge the more difficult it is to keep the beam focused, owing to the mutual repulsion of the electrons. In the formulae for calculating the space charge in a beam a number of quantities always appear in the following relation [3]):

$$\frac{I}{A}\,\frac{l^2}{V^{3/2}}, \qquad \ldots\ldots\ldots \quad (1)$$

where I is the current strength, A the cross section of the beam, l the length of the beam between the last and the penultimate electrodes, and V the potential in the space between those two electrodes. The smaller the factor given by (1), the smaller is the space charge. This already points to one way of arriving at tubes having the desired higher currents and lower voltages, namely a reduction of l, which in our case is permissible since with switch tubes there is no need for the beam to cover any large field. Naturally this reduction in the length of the beam is also favourable for producing a smaller tube.

The reduction of l, however, is not sufficient to allow of the current I being raised to some milliampères and the voltage V being lowered to, say, 300 V without making the expression (1) and thus the space charge intolerably large. Some other improvement is therefore needed, which is found by departing from the usual circular cross section of the beam and giving it a flat rectangular shape, i.e. in the form of a ribbon. This is possible in view of the fact that in the case of a switch tube the beam has to be deflected in one direction only, so that high definition is required only in that direction. (This definition is necessary on account of the large number of contact systems which in a tube required to act, for instance, as a distributor have to be mounted side by side within a limited space.) This is contrary to the case of a tube used for an oscilloscope or television receiver, where high definition is required in two directions of deflection at right angles to each other, which in itself makes it necessary for the cross section of the beam to be circular (or square).

The ribbon-shaped electron beam has several advantages over a beam of circular cross section.

If we give the two beams the same definition in the direction of deflection, that is to say if the short side of the rectangle, d, is made equal to the diameter of the circular beam *(fig. 1)*, then with the same current density the current in the ribbon-shaped beam will obviously be much greater than that in the circular beam.

Further, calculations prove that — again with equal current density — the adverse effect of the space charge (mutual repulsion of the electrons) is so much less in the case of a broad beam that the

[3]) K. R. Spangenberg, Vacuum tubes, MacGraw Hill Book Co., New York, 1948 (pp 443 et seq.).

strength of current can be increased two-fold to four-fold.

Thanks to these two advantages, with a ribbon-shaped beam currents of some milliamps can be obtained and the voltages need not be higher than 200 to 300 V.

A third important advantage of the ribbon-shaped beam is the simpler build-up of the electrode system. The apertures in the electrodes, forming together the "gun" which focuses the electrons, need only be accurately aligned in one direction (at right angles to the flat side of the beam). Thus it is possible for the electrode system to be mounted between supporting plates of mica, as in normal amplifying valves, which greatly simplifies manufacture.

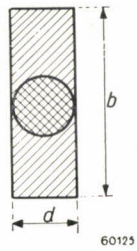


Fig. 1. Cross section of a circular electron beam (diameter d) and of a ribbon-shaped beam ($d \times b$). The latter beam is deflected to the left or to the right.

In another respect, too, advantage can be taken of the technique in construction of normal valves, by following the same principles for keeping the capacitances between the electrodes small. In the case of switch tubes this helps in obtaining a small time constant (short switching time).

Two further advantages of the ribbon-shaped beam are the following:

1. The focal length of a slit-shaped electronic lens, given the same potentials and field strengths, is only half that of a circular lens. Consequently the flat beam can be made shorter than the circular one, so that lower potentials are required.
2. In order to avoid cross-talk, the beam must be strictly focused on the one, desired contact system. This implies that the cathode must be sufficiently far removed from the lens to reduce the size of the "image" of the cathode to less than that of the contact system. Consequently the field strength at the cathode becomes rather small and possibly not sufficient current density can be obtained. Owing to the comparatively great width of the ribbon-shaped beam, which may be regarded as a large number of narrow beams in parallel, the density of current may be small without making the total beam current too small.

Tubes with one or more make and break contacts or change-over contacts

The tube illustrated in *figs 2* and *3* has two contact systems each comprising three electrodes (h-r-a and h'-r'-a', cf. fig. 8 in article I). With the aid of the deflecting plates d_1-d_2 an electron beam

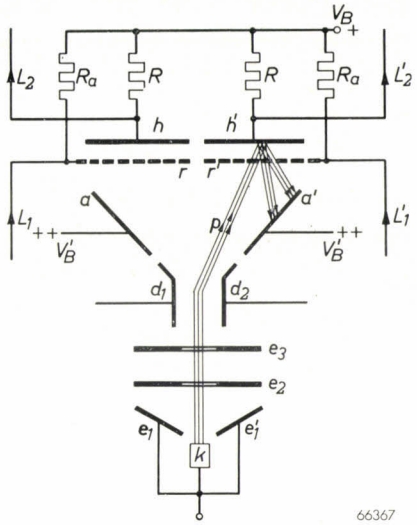


Fig. 2. Cross section of a switch tube with a make and a break contact. The cathode k supplies the electrons from which a ribbon-shaped beam p is formed by the electrodes e_1-e_1', e_2 and e_3. With the aid of the deflecting plates d_1-d_2, this beam can be focused onto h or onto h'. h-r-a and h'-r'-a' are contact systems according to fig. 8 of article I. A connection is made either between the lines L_1 and L_2 or between L_1' and L_2'.

can be directed upon either of the contact systems as required, thus closing either the contact h-r or the contact h'-r'. In this way the tube can replace a relay having a make and a break contact, or, if r is connected to r', a relay with a change-over contact. Owing to the grid-shaped input electrode

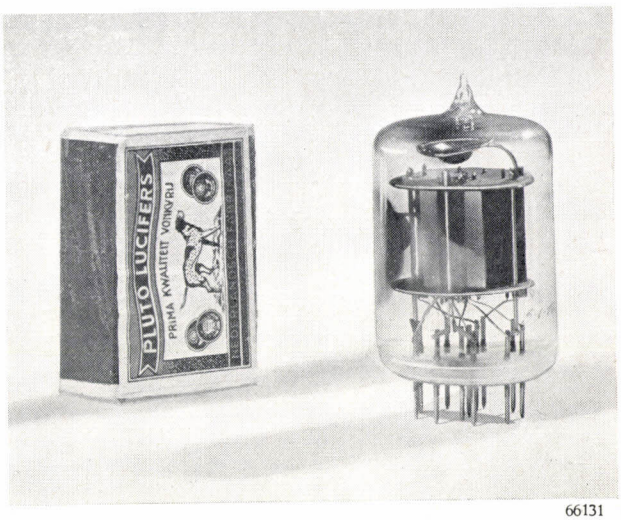


Fig. 3. Photograph of the tube represented in fig. 2.

used in the construction of the contact systems, as explained in article I, about 10 contacts can quite well be connected in series.

For the tube with only two contact systems it is not necessary that the beam should be particularly narrow in the direction of deflection, but even so it is advantageous to use a ribbon-shaped beam: the smaller the thickness d of the beam (in the direction of deflection), the more may the deflecting voltage (between d_1 and d_2) vary without causing the beam to diverge from the contact system upon which it is focused.

As shown in fig. 8 of article I, r was connected via a resistance to a point of a higher potential (V_B') than h. This is favourable for effecting a bilateral contact, such as explained for a similar case with reference to fig. 2*e* in article I. If a unilateral contact suffices then it is an advantage, as represented in fig. 2 (page 83), to feed r and r' with the same low voltage (say 70 V) as that applied to h and h', because when the beam is directed upon h (or h') this does not in itself cause any change in potential at h (or h'). This means that in the case of telephone-current transmission the switching-over of the beam does not give rise to any click, and in the transmission of pulses there is no false signal.

If the width b of the ribbon-shaped beam is made large enough this beam can be made to cover a second, a third or more pairs of contact systems simultaneously, so that the tube can perform the function of a multipolar relay (with two or more make and break or change-over contacts).

By connecting the secondary-emission electrode h to the deflecting plates of one or more additional tubes it can be so arranged that the voltage variation brought about at h when this is struck by the beam is just sufficient to deflect the beam in the following tubes onto another contact system. This is equivalent to a relay contact closing or breaking the circuit of one or more other relays. This will be reverted to later.

Selector tube

The second type to be discussed is the selector tube, i.e. a switch tube which can connect one electrode to any other one of a large group of electrodes.

Fig. 4 is a sketch representing the electrode system. According to the amplitude of the voltage between d_1 and d_2 one of the output electrodes h_1, h_2, h_3, . . . is struck by the beam and owing to secondary emission that electrode assumes the constant potential of a screen grid r, which forms the one electrode to which each of the output electrodes can be "connected". The slots in the screen grid r and the small thickness of the ribbon-shaped beam avoid the possibility of more than one output electrode h being in contact with r at the same time.

In front of the screen grid r is a suppressor grid g_3, which is at zero potential; the wires of this grid are parallel to the plane of the drawing of fig. 4. This grid prevents any undesired transmission of secondary electrons to the deflecting plates. A screen s, likewise at zero potential, behind the electrodes h prevents electrons reaching the glass bulb.

As an application of the selector tube may be mentioned its use as a distributor in multiplex systems, reference to which has been made in the introduction. A sawtooth voltage between the deflecting plates causes the beam to scan the contacts periodically. By means of a simple circuit, which cannot be gone into in detail here, a selector tube acting as distributor can be synchronized with another. The pulses of constant amplitude on the output electrodes h trigger in turn a repeater valve in each of the channels [4]).

A selector tube with six, or even with twelve contacts need be no larger than a normal radio valve *(fig. 5)*.

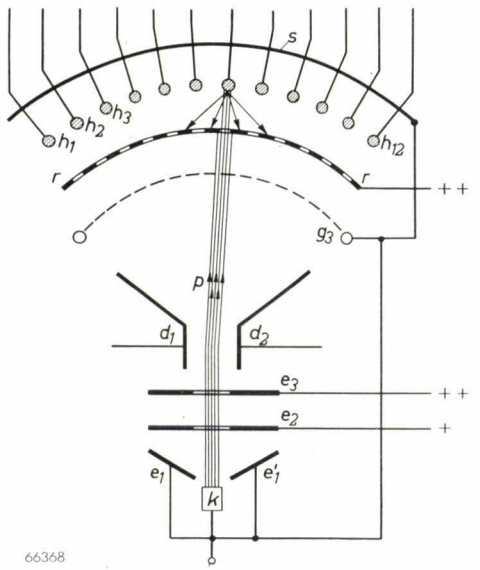
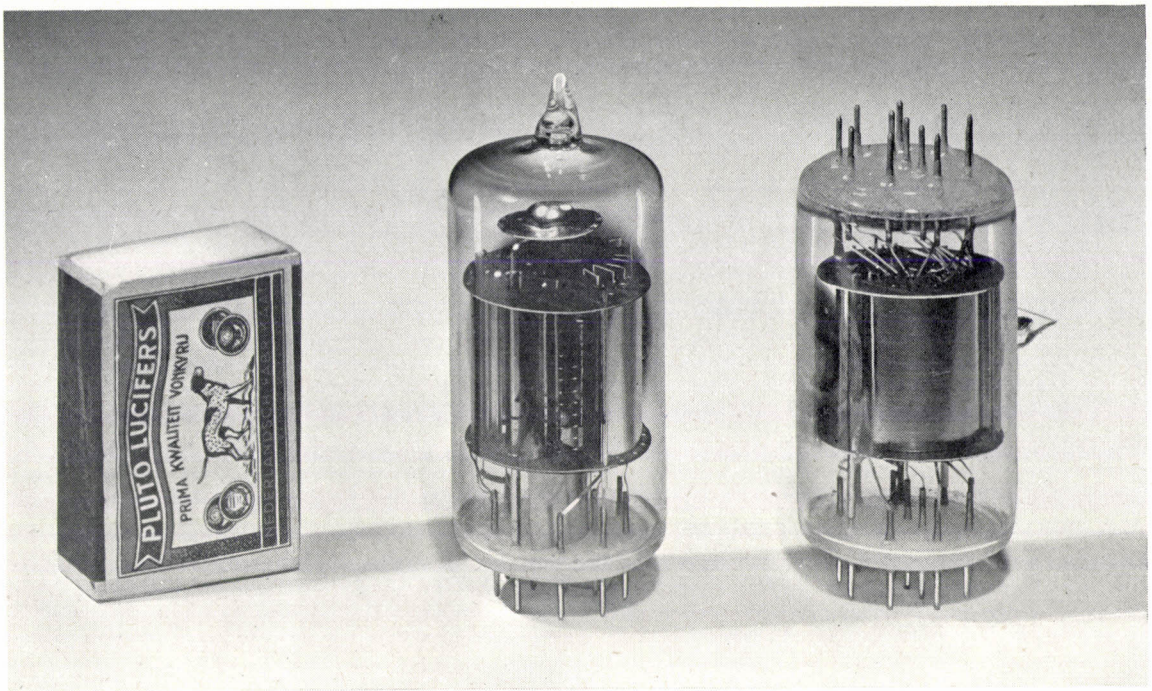


Fig. 4. Selector tube. The ribbon-shaped beam of primary electrons p is focused onto one of a large number of electrodes h_1, h_2, h_3, ... which emit secondary electrons, whereby the electrode struck by the beam assumes the constant and high potential of the screen grid r. g_3 is a suppressor grid and s a screen, both at cathode potential, preventing electrons from reaching undesired places. The rest of the letters have the same meaning as in fig. 2.

[4]) Cf. the channel-gate circuits described in: C. J. H. A. Staal, An installation for multiplex-pulse modulation, Philips Techn. Rev. **11**, 133-144, 1949.



66132

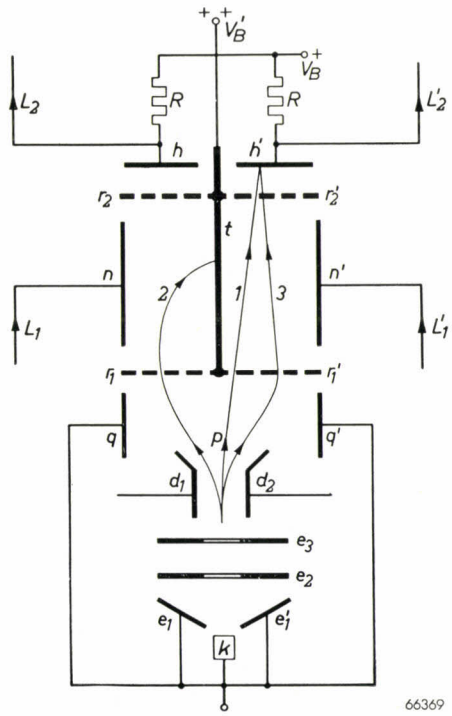
Fig. 5. Two selector tubes according to fig. 4. On the left one with six and on the right one with twelve contacts h.

Signalling tube

In telephony signalling is understood to comprise everything connected with the transmission of signals excepting the conversations themselves, i.e. signals for selecting or dialling, ringing, buzzing for "engaged", etc. These signals are in the form of pulses.

The contact tubes in question can, of course, also serve for transmitting signals with this special form of modulation, but nevertheless a switch tube has been designed with a system of electrodes as sketched in *fig. 6*, which is particularly suitable for this purpose. *Fig. 7* is a photograph of two signalling tubes, one of which is fitted with a "Noval" base.

This tube, too, works with a ribbon-shaped beam, which can be deflected to the left or to the right with the aid of the deflecting plates d_1 and d_2. The beam then enters a space enveloped by the grids r_1 (r_1') and r_2 (r_2'), the partition t and the plate n (n'); the grids and the partition have the same constant high potential V_B' (say 200 V); n and n' are the input electrodes receiving the signal pulses (with positive polarity). In fig. 6 the line *1* represents the beam deflected to the right in the case where the plate n' receives a signal with a positive voltage approximately equal to V_B'. After passing through the grid r_1' the primary electrons come into a space which is practically free of any electric field, so they shoot straight on and strike the electrode h'.



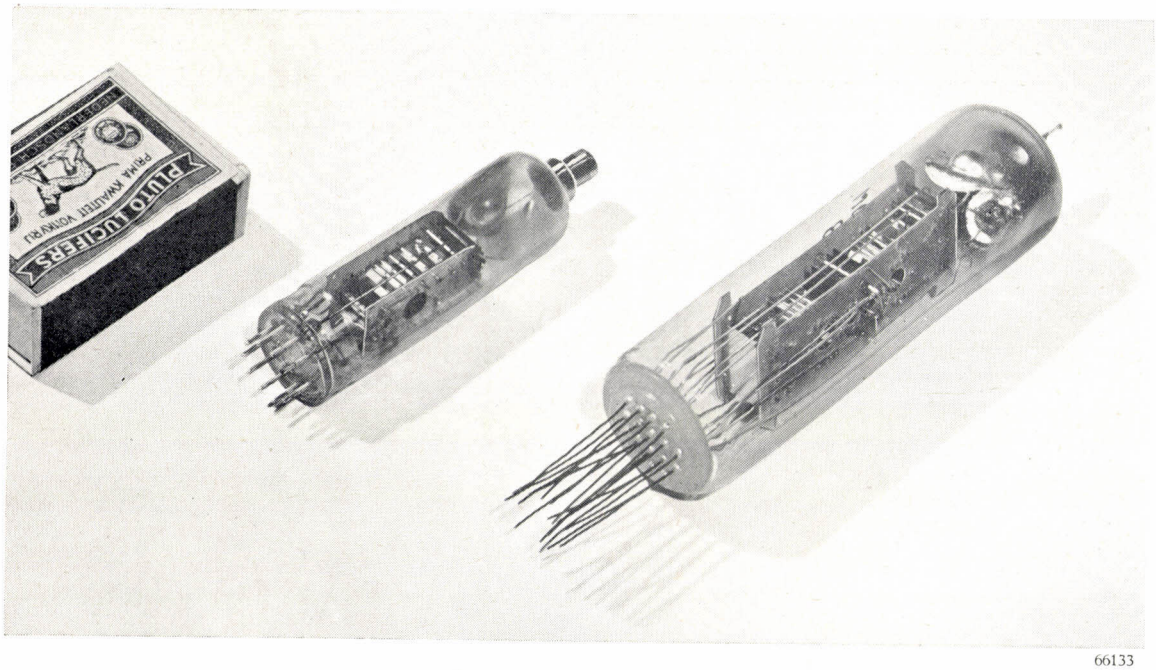
66369

Fig. 6. Cross section of a signalling tube. For the meaning of k, e_1-e_1', e_2, e_3, d_1-d_2, p, h and h', see fig. 2. The plates q-q' at cathode potential prevent the beam from being deflected too far. The collector t and the grids r_1-r_1' and r_2-r_2' are at the high potential V_B'. *1* is the path followed by the primary beam when this is deflected to the right and the plate n' (input electrode) has approximately the potential V_B'; as the beam strikes the output contact h' the latter assumes the potential V_B' of the grid r_2', and the same is the case when owing to a higher deflecting voltage the beam follows the path *3*. The beam (deflected to the left) follows the path *2* if the incoming line L_1 and the input electrode n are at a low potential. The beam does not then reach h and the outgoing line L_2 does not undergo any change in potential.

The latter is thus caused to emit secondary electrons, which pass over to t and r_2', and h' thereby assumes the high potential of r_2', which forms the output signal.

In the absence of the input signal (or if the potential of the input electrode drops below a certain value) the situation is as indicated by *2* in fig. 6 (for the beam deflected to the left): owing to the low potential of n the beam is diverted away from n and strikes against the partition t, without reaching h.

fore, is analogous to the case of the selector tube, and essentially different from that of the switch tubes previously discussed where the voltage at the output electrode h follows that of the input electrode a. This implies that the signalling tube cannot transmit any arbitrary modulation of the input signal, but that on the other hand it is excellently suited for imparting pulses the amplitude of which is adjustable with the size of V_B' and independent of the amplitude of the input pulses.



66133

Fig. 7. On the left a signalling tube according to fig. 6 with "Noval" base. On the right a double signalling tube incorporating two of the electrode systems illustrated in fig. 6 (but with one common cathode).

On either side is a plate, q and q', at cathode potential. These plates prevent the beam from being deflected too far, so that in the event of the deflecting voltage being higher than necessary the primary electrons follow a path as represented by *3* in fig. 6. Thus, owing to the effect of the plates q and q' the deflecting voltage between the plates d_1 and d_2 is not at all critical and, moreover, the dimensions of the deflecting system can be kept small.

The plates n and n' have the character of control electrodes, rather than that of contacts. Here the "closing of the contact" means that the electrode h or h', which previously had the potential V_B (say 70 V), assumes the potential V_B' (say 200V), regardless of the voltage at n (or n'), except that it must have exceeded a certain value. This, therefore, is analogous to the case of the selector tube, and essentially different from that of the switch tubes previously discussed where the voltage at the output electrode h follows that of the input electrode a. This implies that the signalling tube cannot transmit any arbitrary modulation of the input signal, but that on the other hand it is excellently suited for imparting pulses the amplitude of which is adjustable with the size of V_B' and independent of the amplitude of the input pulses.

Contrary to the case of other switch tubes, there is practically no limit to the number of signalling tubes that can be connected in series.

This is to be explained as follows. A series connection of the contacts of three tubes according to fig. 2 is represented in *fig. 8a*, with an equivalent circuit in fig. 8*b*. When the three contacts are closed the resistances R_i of each contact are in series, and although R_i may be small with respect to R, a limit is therefore nevertheless set to the number of contacts that can be connected in series (maximum about 10). When, however, a number of signalling tubes are connected in series (fig. 8*c*) the equivalent circuit is as shown in fig. 8*d*: here a "closed contact" is represented by the connecting of an output electrode h via the contact resistance $R_i (\ll R)$ to the voltage source V_B', so that the resistances R_i are not connected in series and consequently any number of contacts of these tubes can be connected in series.

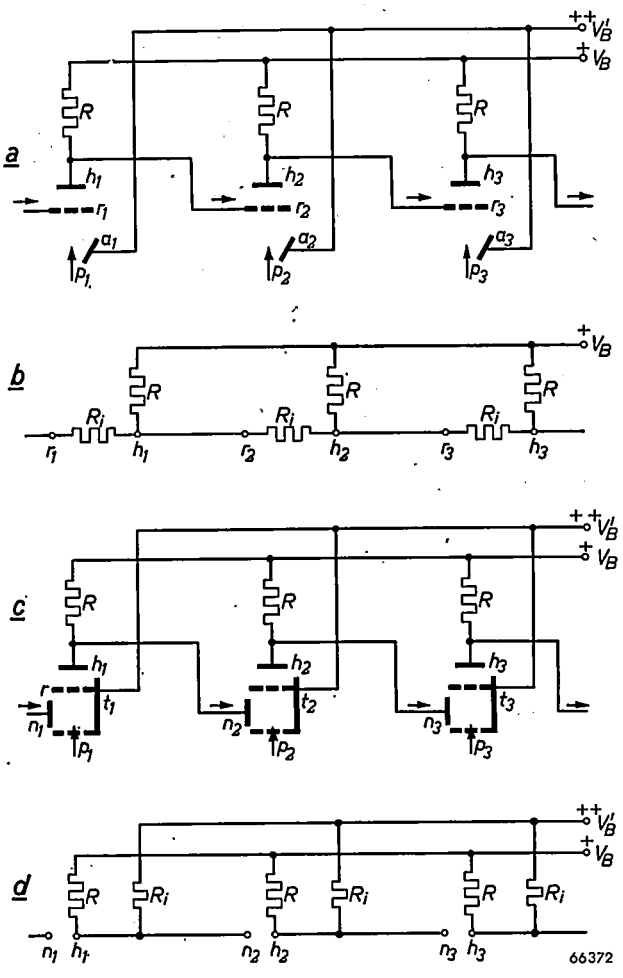


Fig. 8. *a*) Series connection of the contacts of three tubes according to fig. 2 (only the left-hand halves are drawn here). *b*) Equivalent circuit for the case where the three contacts are closed; the contact resistances R_i ($\ll R$) are thereby brought into series. *c*) Series connection with tubes according to fig. 6 (only the left-hand halves drawn). *d*) Equivalent circuit of (*c*) with the three contacts closed; here the contact resistances R_i ($\ll R$) are not in series.

Some switching-technical problems

Intentional delayed action

In the introduction it was already pointed out that sometimes it is necessary that certain (electro-magnetic) relays, simultaneously activated, should close their contacts in a predetermined sequence. For this purpose special constructions are employed with which a certain intentional delay is obtained, adjustable or fixed.

To attain the same object with switch tubes one may proceed as follows. In an auxiliary tube *(fig. 9)* a stream of electrons p_2 is generated and directed as a beam upon an anode a_2. The deflecting plates of the auxiliary tube are connected to the corresponding plates of the tube whose action is to be delayed; this tube may be of the type shown in fig. 2 or fig. 6. The anode a_2 is connected via a resistor to a point of positive potential and via a capacitor C to the deflecting plate d_1. When the deflecting voltage of the other plate (d_2) is raised to a level whereby the beam is made to diverge from anode a_2 (the beam in the other tube then still strikes the electrode h) the potential of a_2 suddenly rises and as a consequence also the potential of d_1 is increased (both in the auxiliary tube and in the switch tube), thereby retarding the deflection of the beam in both tubes. This retardation is determined by the capacitance C and the resistances through which the charges can leak away.

If necessary a variety of electrode systems can be built into the auxiliary tube, with different RC times, so as to provide for different delay times. In principle it is also possible to mount the anode a_2 in the switch tube whose action is to be delayed and to arrange for a part of the (widened) beam to be directed upon it.

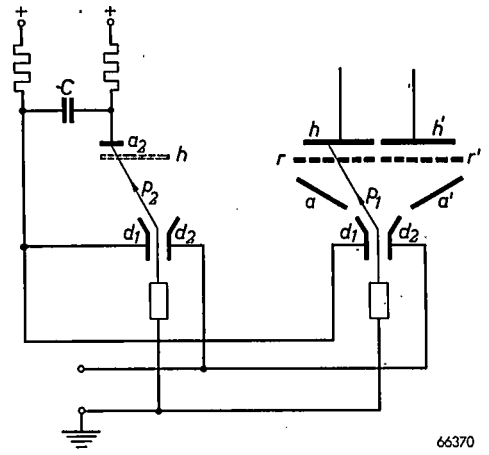


Fig. 9. On the right a switch tube according to fig. 2, the action of which is given a certain delay by means of the auxiliary tube on the left. When the voltage at the two right-hand deflecting plates d_2 is raised to such a level that the beams p_1 and p_2 are deflected to the right, instead of to the left, then upon p_2 leaving the anode a_2 the potentials of a_2 and (via the capacitor C) the two plates d_1 are suddenly increased, thereby retarding the deflection of the beams in both tubes. In the sketch of the auxiliary tube the output electrode h of the main tube has been drawn in broken lines so as to indicate that p_1 is still directed upon h when p_2 is just leaving the anode a_2.

Replacement of several relays by one switch tube

Fig. 10a represents a certain combination of three monopolar relays which can be replaced by a single switch tube according to fig. 6. The deflecting plates d_1-d_2, directing the beam to the left or to the right, take the place of the coil of relay *1*, whilst the plates n-n' in fig. 6 replace the coils of the relays *2* and *3* (the "contact" is only brought about when the potential of the latter electrodes is sufficiently positive).

The relay circuit can also be replaced by the tube represented in fig. 2, with r, r', h and h' fed from one and the same voltage source. Relay *1* again corresponds to the switching-over of the beam. To raise the potential, for instance of h', it is not sufficient to focus the beam on this electrode, but also the potential of r' has to be raised to the level desired for h'. The analogy of this in the relay circuit according to fig. 10*a* is that, in order to connect the lowermost point with the higher point on the right, not only relay *1* but also relay *3* has to be activated.

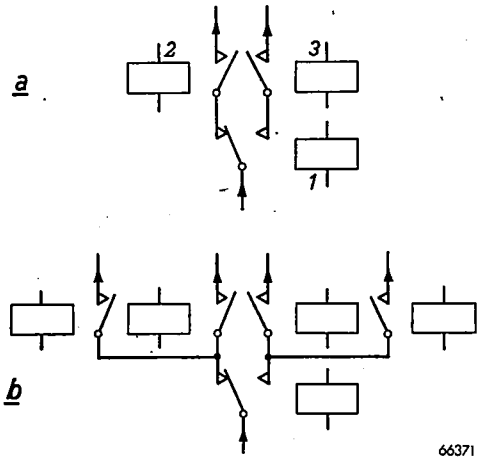


Fig. 10. The relay system *a* can be replaced by a tube according to fig. 6, and the system *b* by a similar tube with two contact systems side by side.

When there are two such groups of relays it can be arranged for the current passing through the coil of relay *1* of one group to be switched on and off by the contact of relay *2* or *3* of the other group. With two tubes according to fig. 6, for instance, a corresponding circuit can be built up in the manner already described, by using the output voltage of an electrode h of one tube as deflecting voltage between the plates d_1-d_2 of the other tube.

If the beam is made wide enough to cover two contact systems placed side by side then one such a double tube can replace the five simple relays according to fig. 10*b*. If the whole of the electrode system were to be doubled (except for the cathode, which then has to produce two beams each serving two pairs of contact systems) then in this way, in a circuit similar to that of fig. 10*b*, even one tube could perform the functions of ten monopolar relays! It is obvious that this means a considerable saving in space.

It will be clear that the possibilities are not confined to the examples given here. Practice will have to show how many and what combinations can best be incorporated in one tube.

Summary. In this second article on "electronic switching" some experimental tubes are dealt with in which, as in the case of those already described, "contacts" are brought about by means of secondary emission (so that the output signal is in phase with the input signal), but with the beam of primary electrons guided in the desired direction by electrostatic deflection, thus giving these tubes the character of a change-over switch. The use of a ribbon-shaped beam, having a cross section large in one direction but small in the other, has a number of advantages: the strength of the beam current can be raised to several milliamps, the voltages required need not be higher than 200 to 300 volts, and the dimensions of the tubes can be kept small (about the size of normal radio receiving valves).

There are dealt with in turn a tube with one or more make and break contacts (or change-over contacts), a selector tube with 6 and one with 12 contacts, and a signalling tube (which can transmit only pulses; the amplitude of the pulses transmitted is independent of that of the input pulses, provided that the latter exceed a certain value). If desired an adjustable delay can be introduced in these tubes. Some examples are given to explain how one single switch tube can replace a number of monopolar relays.

NEW VIEWS ON OXIDIC SEMI-CONDUCTORS AND ZINC-SULPHIDE PHOSPHORS

by E. J. W. VERWEY and F. A. KRÖGER.

537.311.3:535.371

Just as a few decades ago great advances in the electrotechnical industry followed the introduction of electron tubes such as the diode and triode, the cathode-ray tube and the thyratron in electric engineering, so now great possibilities can be invisaged by the application of specific properties of solids. All kinds of new materials with special properties or combinations of properties are being introduced: Semi-conductors play a large part in blocking-layer rectifiers, as basic material for resistors, photo-conductive layers, etc.; fluorescent materials are being used on a large scale in illumination engineering, television, radar, etc.

In view of the great practical importance of these materials there is every reason to study them intensively also from the theoretical aspect and to attempt to gain a deeper insight into their behaviour.

It has long been known that important physical properties of solids are often governed by certain "impurities". As examples may be mentioned the conductivity of semi-conductors and the optical properties of fluorescent substances. In many cases the mechanism whereby the "impurities" exercise their influence is still obscure, while in other cases although some idea can be formed of this mechanism it is extremely difficult to prepare a substance with the desired content of impurities.

With regard to some points, however, a deeper insight has recently been gained, and new possibilities for a better mastery of the preparation of materials with desired properties have thereby been gained.

In this article semi-conductors will be dealt with first. Following upon a discussion of some ideas of older date, it will briefly be explained how new ideas have led to the "method of controlled valency". Then some new views will be discussed which have been gained by consideration of the zinc-sulphide phosphors, from which it will appear that there is a great analogy in the lines of thought followed in both these fields.

Semi-conductors

Semi-conducting elements

The first materials whose behaviour under the influence of impurities came to be fairly well understood were the semi-conducting elements silicon and germanium, thanks to extensive investigations carried out by, among others, Lark-Horowitz and Johnson[1]). It appears that when an element in the third group of the periodic system, for instance boron, aluminium, gallium or indium, is admixed with Si or Ge a semi-conductor is usually obtained which behaves as if its conductance were due to positive charges (a "hole" semi-conductor). If the impurity is an element in the fifth group of the periodic system, such as nitrogen, phosphorus, antimony or arsenic, usually a semi-conductor is obtained whose conductance is due to negative charges.

This behaviour is to be explained by the obvious supposition that the foreign atoms occupy a normal place in the lattice. Considering, for instance, the case of germanium containing antimony, it can be imagined that four of the five valency electrons of the Sb atom take part in the chemical bonding with the four closest neighbours. (Ge and Si both have four valency electrons and crystallize into the so-called diamond structure, where each atom has four immediate neighbours.) The fifth electron is, as it were, superfluous and can be regarded as an electron which in a medium (germanium) with dielectric constant ε moves in the Coulomb field of a positive Sb^+ ion.

Owing to the high value of ε it requires only an energy of some hundredths of an electron-volt to release such an electron from the Sb^+ ion, after which it is comparatively free to move about in the crystal. Thus, even in very small quantities such an impurity causes a considerable increase in the conductivity of the crystal.

The above-mentioned hole-conductance occurring in the case of contamination with an element of the third group of the periodic system can be explained in a similar manner. The replacement of a Ge or Si atom by, for instance, Al, gives rise to a deficit in electrons. It can be assumed that

[1]) See, for instance, K. Lark-Horowitz, El. Eng. **68**, 1047, 1949; W. Shockley, Electrons and holes in semiconductors with applications to transistor electronics, Van Nostrand, New York, 1950.

an electron from a Ge atom in the immediate vicinity of the Al atom has the tendency to form, together with the Al atom, a negative ion. Of course it is a matter of indifference which of the immediate Ge neighbours releases an electron. In the position from which at a certain moment an electron is missing there is said to be an electron "hole".

This "hole" behaves as a positively charged particle which, in a medium with dielectric constant ε, moves in the Coulomb field of the negative Al^- ion. As in the first case considered, the binding energy of this "positive charge carrier" is very small. After being released from the Al^- ion it can move about in the crystal practically unhindered, so that again there is a contribution towards the conductivity.

Apart from this superficially sketched theory of the origin of conductivity in Si and Ge there are also detailed theories by which it is possible to explain the dependency of conductivity upon temperature and upon the concentration of the impurity as found experimentally. By measuring the conductivity of materials to which known quantities of foreign atoms have been added, it is possible to calculate, with the aid of these theories, the specific resistance of ideal pure Si or Ge. It is then found that at room temperature this specific resistance should be respectively 10^6 and 10^2 Ωcm. When these values are compared with those measured with samples of Ge or Si which have been purified by the most rigorous methods, it appears that the specific resistance of the materials prepared in this way is still, respectively, a factor 10^4 and a factor 2 too small. From this it can therefore be concluded that these materials still contain a certain small amount of impurities which cannot be removed; this concentration is so slight that it is usually impossible even to determine their nature.

Non-stoechiometric semi-conductors

In addition to the semi-conducting elements mentioned there is a large group of semi-conducting compounds, many of which belong to the metal oxides or metal sulphides. Of the oxidic semi-conductors the non-stoechiometric semi-conductors have been known longest [2]). Although the electrical properties of these semi-conductors are not governed by impurities in the sense of "foreign atoms", a few words will be devoted to them here. The fact is that the mechanism of conductance in these substances is analogous to that in semi-conductors prepared on the basis of the method of controlled valency to be described below.

The non-stoechiometric semi-conductors owe their conductivity to a departure from the simple, stoechiometric, relation of the component elements. This deviation may consist in the occurrence of an excess of metal or an excess of oxygen. Generally speaking, an excess of ions of one kind is accompanied by lattice flaws; further, non-stoechiometric compounds are generally found in compounds of metals with varying valency. For example, in the case of Cu_2O with an excess of oxygen there must be a number of Cu^+ places unoccupied, whilst, for the electrical neutrality of the lattice, an equivalent number of Cu^+ ions must have changed into the bivalent state. Schematically the situation may be represented by $(Cu^+_{2-2\delta}\ Cu^{2+}_{\delta}\ \square_{\delta})\ O^{2-}$, where δ is the excess of oxygen in atomic percentage and $\square$ represents a vacant site in the lattice. Given sufficiently low temperatures it can be expected that these Cu^{2+} ions will be found next to the vacant Cu^+ sites, since, for Cu^+ ions in the immediate neighbourhood of the vacant Cu^+ sites, less energy is required to remove an electron from a Cu^+ ion, on account of the positive charge being lacking (see *fig. 1a*). At higher temperatures an electron can also be released from one of the many ions farther removed from the vacant site. This electron may then again convert the original Cu^{2+} ion into a Cu^+ ion. It is then said that the electron "hole" is removed from the immediate vicinity of the empty place in the lattice through thermal dissociation. Provided it is removed far enough the electron "hole" is able to move about in the lattice practically unhindered, since the various positions at a distance from the empty place in the lattice correspond approximately to one and the same energy [3]) (see fig. 1*b*). It is in this manner that the non-stoechiometric Cu_2O receives its conductivity, and it is also seen that the conductivity increases greatly with rising temperature.

The practical utility of non-stoechiometric semi-conductors is strictly limited. In the first place it is almost impossible to produce a homogeneous, stable, non-stoechiometric semi-conductor with a particular conductivity value. In the second place the deviations from the stoechiometric state which can be reached are as a rule small. This implies also that the conductivity is limited in its variability. These limitations arise from the lack of stability in a lattice with many lattice flaws.

[2]) See, for instance, E. J. W. Verwey, Philips Techn. Rev. 9, 47, 1949.

[3]) J. H. de Boer and E. J. W. Verwey, Proc. Phys. Soc. 49, extra part, 59, 1937.

It will be shown how it has been found possible to produce oxidic semi-conductors in a manner whereby these difficulties are avoided and the conductivity is brought about, as in the case discussed above, by ions of the same element but with different valency occupying crystallographically equivalent places in the lattice.

The better conducting nickel oxide obtained in this way is of the type of non-stoechiometric semi-conductors discussed above and has the same drawbacks: it is not of practical utility owing to lack of stability. But an oxide containing Ni^{3+} in addition to Ni^{2+} ions can now be obtained in quite a different way, by admixing a certain quantity

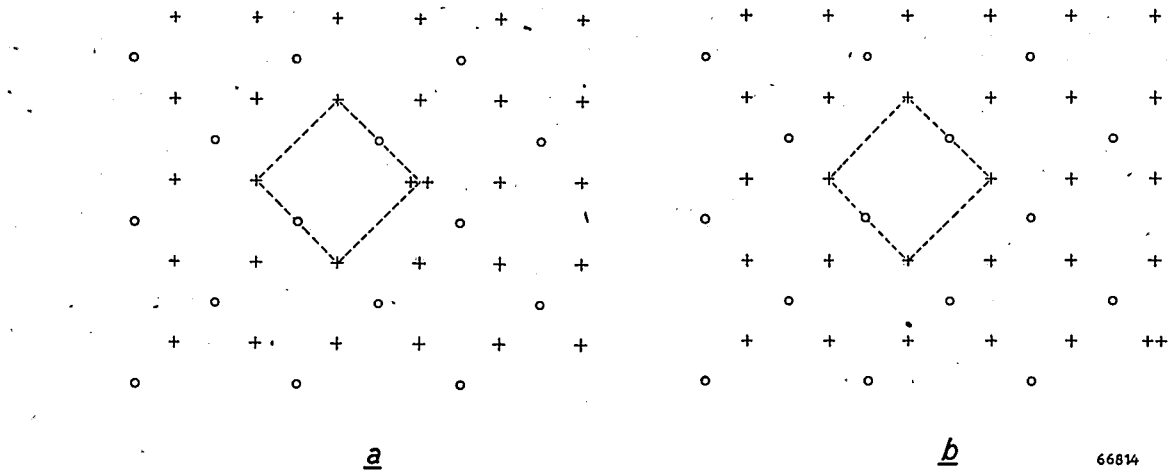


Fig. 1. Two-dimensional model of an M_2O lattice with an excess of oxygen. The regular pattern characterizes a lattice in which the number of positive ions (denoted by +) is twice the number of negative ions (which obviously carry a double charge and are denoted by O). The vacant lattice site (surrounded by the broken line) expresses that a positive ion is missing. To ensure the electric neutrality of the lattice the number of double-charged positive ions (++) must be equal to the number of vacant sites (hence Cu^{2+} ions in the case of Cu_2O). Given a sufficiently low temperature, the electron "hole" corresponding to such an ion lies in the immediate vicinity of the vacant site in the lattice (*a*). Once the electron "hole" is removed from the vacant site in the lattice at higher temperatures, it is practically free to move about in the whole of the lattice (*b*).

The method of controlled valency

The new method of producing semi-conductors, indicated by Verwey, Haaijman and Romeyn [4]), will be described by taking the case of nickel oxide as an example.

When NiO is prepared by heating $NiCO_3$ in an oxidizing atmosphere (say air) at about 1200 °C, the oxide is obtained in the form of a green crystalline mass with a high specific resistance (up to 10^{12} Ωcm). This resistance is too high for practical purposes; better conductivity can be expected if in addition to the bivalent Ni ions present in the NiO compound there are also Ni^{3+} ions in the crystal lattice, and this can be arranged: when $NiCO_3$ is dissociated in air at a low temperature (about 500 °C) a nickel oxide is obtained which has a small excess of oxygen and thus contains some trivalent nickel ions. The presence of these Ni^{3+} ions is disclosed by the black colour of the oxide.

of Li_2O with the NiO (in practice the two carbonates are mixed) and again heating this mixture to 1200 °C in air. The product is a black oxide with low specific resistance. This oxide is a homogeneous crystalline mass, the composition of which is found by chemical analysis to be

$$(Li_\delta Ni^{2+}_{1-2\delta}\, Ni^{3+}_\delta)\, O^{2-},$$

the amount of Ni^{3+} being equivalent to the amount of Li added as impurity, at least as long as the latter amount is kept within certain limits. X-ray diffraction shows that the crystal structure of this black oxide is the same as that of NiO; only the dimensions of the elementary cell have become somewhat smaller. From the above formula it follows that in the preparation of this oxide an amount of oxygen must have been absorbed from the air, whilst a certain amount of Li_2O is taken up in the lattice of the NiO, so that the final product is again an oxide with a minimum of vacant sites in the lattice. The reaction taking place may be written as follows:

[4]) E. J. W. Verwey, P. W. Haaijman, F. C. Romeyn and G. W. van Oosterhout, Philips Res. Rep. 5, 173, 1950 (No. 3).

$$\tfrac{1}{2}\,\delta\,Li_2^+\,O^{2-} + (1-\delta)\,Ni^{2+}O^{2-} + \tfrac{1}{4}\,\delta\,O_2^{2-} \rightarrow (Li_\delta^+\,Ni_{1-2\delta}^{2+}\,Ni_\delta^{3+})\,O^{2-}.$$

For $\delta = 0.1$ the specific resistance is about 1 Ωcm, a factor 10^{12} lower than the specific resistance of pure NiO (see *fig. 2)*.

The artifice applied here in using lithium has a very general significance, as may be gathered from the following.

If the changing of the valency of some Ni ions were to be brought about by an excess of oxygen, then at the same time there would be some unoccupied Ni sites in the lattice. If we were to aim only at introducing Li_2O into the NiO lattice this would necessarily be accompanied by unoccupied O sites (or interstitial Li ions). Actually in our case neither of these two processes takes place separately. The process which does take place may be regarded as a combination of the two, giving the above-mentioned reaction and without leaving any vacant sites in the lattice.

This leads to the conclusion that the apparent preference for a situation without vacant sites in the lattice can be utilized for bringing about desired changes in valency by the admixture of suitable impurities. This is what the authors [4]) call the method of controlled valency.

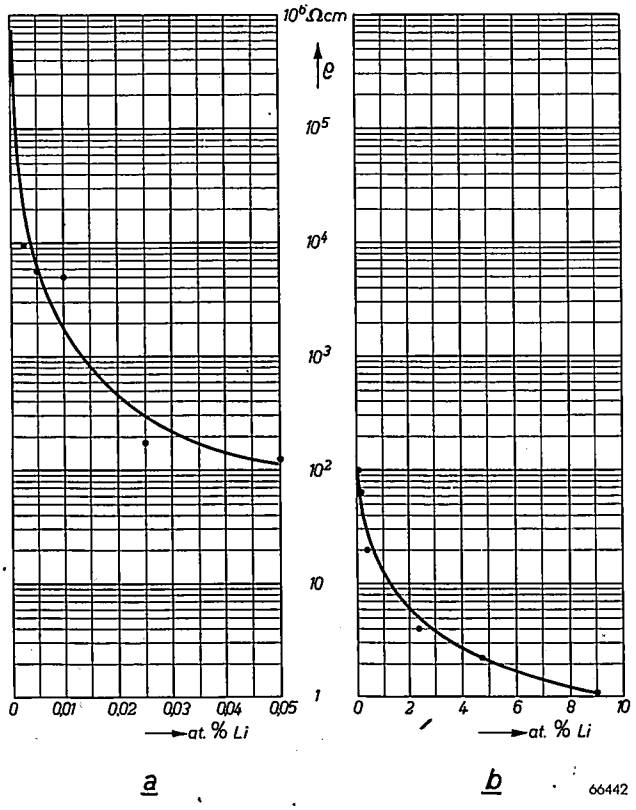


Fig. 2. Specific resistance of $(Li_\delta Ni_{1-\delta})O$ as a function of the Li content. The atomic percentage of Li varies between *a*) 0% and 0.05%, *b*) 0.05% and 9%.

For this method to be successful, certain conditions have to be fulfilled, which in our case of NiO mixed with Li may be formulated as follows. In the first place the Li must have a more stable valency than Ni. In the second place the ion radius of Li^+ must not differ much from the radius of Ni^{2+}, so as to permit the incorporation of Li^+ions in the NiO lattice. These two conditions are indeed satisfied. The energy required to release an electron from an Li^+ion, so as to change that ion into an Li^{2+}ion, amounts to 75.3 electron-volts, which is much greater than the energy required to turn a bivalent Ni ion into a trivalent one, namely 35 eV. The radii of the two ions Li^+ and Ni^{2+} are both 0.78 Å.

The expression given above for the reaction is based upon the assumption that a deficit of electrons in NiO leads to the formation of trivalent Ni ions and not to the formation of monovalent O ions. A closer theoretical study proves this assumption to be justified.

Practice has already shown that the method of controlled valency can be applied to a large number of compounds if only the suitable impurity is chosen. This makes it possible for the conductivity of these compounds to be adjusted within certain, wide, limits. Usually the maximum soluble quantities of the impurity are much smaller than in the case of NiO, where as many as 30% of the sites of metal ions can be occupied by Li^+ ions. In all these compounds a number of cations of the original lattice are replaced by foreign ions of a different charge, which, given an equivalent number of cations in the basic material, bring about a change in valency, so that the electrical neutrality of the whole crystal is preserved. The foreign ion may also occupy a site in the lattice which is not crystallographically identical to the sites occupied by the ions with variable valency. This is illustrated by the example:

$$(Sr_{1-\delta}\,La_\delta)\,(Ti_{1-\delta}^{4+}\,Ti_\delta^{3+})\,O_3.$$

In discussing the phosphors an example will be taken where the valency of the cations is changed by replacing some of the anions by foreign anions with a different charge.

Of the semi-conductors produced by the method of controlled valency there is one which has already found practical application as a resistor with negative temperature coefficient.

Zinc-sulphide phosphors

We shall now pass on from semi-conductors to zinc-sulphide phosphors, but first we would

briefly recall what has hitherto been known about the nature of the centres of fluorescence where the fluorescence is assumed to originate.

Purely stoechiometric ZnS does not show any visible fluorescence when irradiated (excitation) with ultra-violet rays or bombarded by electrons. The excitation energy is then either imparted to the crystal lattice in the form of heat through some process of dissipation or else radiation takes place in the ultra-violet. The incorporation of impurities in the ZnS lattice enables the ZnS crystal, after excitation, to radiate a large part of the excitation energy in the form of light. Hence the name "activators" sometimes given to these impurities [5]), which may consist, for instance, of Ag, Cu, Au, or Zn. By mentioning Zn as an impurity in a ZnS lattice we have taken the conception of an impurity very broadly. When speaking of ZnS as having Zn as an "impurity" or being activated by Zn we mean that the ZnS has an excess of Zn.

As to the form in which the activators are absorbed, the general belief was that they are taken up in the lattice as atoms, occupying interstitial sites. According to other investigators they turn into an ion by releasing an electron, which in the ionic form occupies a normal cation site. In the latter case it was assumed that an equivalent number of vacant anion sites are formed which can just be occupied by the electrons released.

It will now be explained how more recent investigations have led to these ideas being revised.

ZnS phosphors are made by heating precipitated ZnS in a suitable atmosphere, for instance H_2S. In order to promote crystallization a so-called flux is added, which is subsequently removed by washing. Then, according to the colour of the fluorescence required, very small quantities of the elements Ag, Cu, Au are added as activators. In the case of a product having an excess of Zn one sometimes speaks of self-activated ZnS. It has now been found that traces of Cl, Br or I play an essential part in the formation of centres of fluorescence: a fluorescent material can only be produced if the flux contains one of these halogens or the atmosphere contains HCl or HBr [6]).

For instance, pure ZnS heated in an atmosphere of H_2S and HCl shows a blue fluorescence, the intensity of which increases with the HCl content of the atmosphere (see *fig. 3*). Thus it appears that Cl is very closely related to the formation of the centres. It is fair to assume that a number of S^{2-} ions are replaced by Cl ions, so that at the same time an equivalent number of Zn^{2+} ions change into the monovalent state. It can indeed be proved that the number of centres introduced into the non-activated ZnS is about equivalent to the amount of Cl absorbed in the lattice.

The concentration of the centres can be determined from the phenomenon of the "saturation" of the fluorescence. This phenomenon implies that when employing an exciting beam of electrons of great current density (an electron beam is used for the excitation because it is difficult to produce a sufficiently intensive ultra-violet radiation) the intensity of the fluorescence no longer increases linearly with the electron current but approaches a constant value. This saturation intensity is proportional to the concentration of the centres.

Similarly, also in the case of the phosphors activated with Ag, Cu or Au it can be assumed that the activator is taken up in the form of its halide and not, as was formerly believed, in the atomic form. Since the colour of the fluorescence is determined exclusively by the activator used,

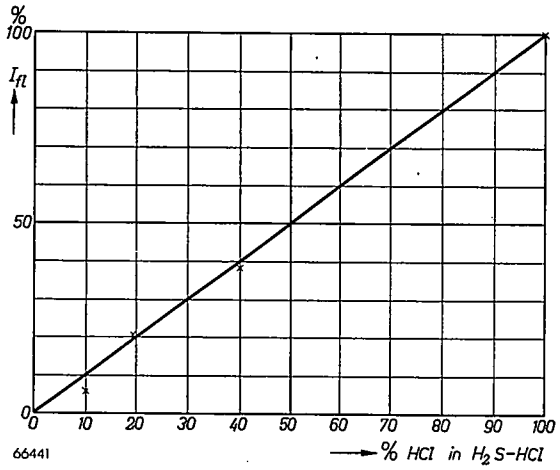


Fig. 3. Intensity of the blue fluorescence of ZnS.ZnCl as a function of the composition of the atmosphere in which the product is prepared. The curves relate to fluorescence after excitation by rays of a wavelength of 2537 Å at −180 °C.

and not at all by the halogen used in the preparation of the phosphor, it can be assumed that the halogen ions are not present in the centres of fluorescence but that, in analogy with the semi-conductors, the halogen ions serve exclusively to keep the crystal electrically neutral (compensation of charge) and to prevent the formation of vacant sites in the lattice. This may also be expressed by saying that in this case we have to do with the ordinary formation of a mixed crystal ZnS-AgCl. The correctness of this assumption is proved by the following.

[5]) See, for instance, F. A. Kröger and W. de Groot, Philips Techn. Rev. **12**, 6, 1950 (No. 1).

[6]) F. A. Kröger and J. E. Hellingman, J. Electrochem. Soc. **93**, 157, 1948. See also footnote [7]).

It is also possible to prepare fluorescent zinc-sulphides in a manner whereby the charge-compensating function is taken over by a trivalent cation. In the neutralization of the charge introduced by the activator ($Zn^{2+} \rightarrow$ for instance Ag^{+}) the replacement of an S^{2-} ion by a Cl^{-} ion serves in the same way as the replacement of a Zn^{2+} ion by a trivalent cation[7]). For instance, ZnS containing as impurity exclusively Al or Ga shows blue fluorescence in a frequency band absolutely identical to the band in which the frequencies of the fluorescent light from self-activated ZnS lie. Also products containing Ag, Cu or Au together with Al, Y, Yb, La or Gd show fluorescence in the blue, green or yellowish green, which again is absolutely the same as the fluorescence from the corresponding phosphors prepared with halogen ions instead of trivalent cations [8]).

Thus it is seen that the fluorescence depends solely upon the activator used and not upon whether halogen ions or trivalent cations are used. Neither has the choice of any special halogen ions or trivalent cations any influence.

This applies with a certain restriction. Some trivalent cations, such as Ga, In, Sc, Pr or Ce, are in themselves capable of performing the function of activators. As a consequence, in the fluorescence from phosphors prepared with these elements there appear new fluorescence bands in addition to those ascribed to the actual activator. This complication will not be dealt with here [8]).

From the foregoing it can be concluded that the halogen ions or trivalent cations take no part in the actual process of fluorescence and, therefore, cannot be considered as belonging to the centres. Apparently the centres of fluorescence consist of a monovalent activator ion with its surroundings of sulphur ions.

On the basis of extensive experiments, the description and discussion of which would lead us too far here, it can be plausibly concluded that the action of the activator ion lies mainly in its different (smaller) charge reducing the ionization energy of a neighbouring S^{2-} ion. To gain a more complete insight into the process, however, account has to be taken not only of the electrostatic forces but also of the polarization effects, repulsive forces and van der Waals' attractive forces. These are in fact responsible for the differences in fluorescence found when ZnS is activated with different elements.

These new conceptions not only form an important contribution towards the theory of fluorescence but at the same time present the possibility of producing entirely new phosphors with the aid of trivalent cations, such as would not be possible by the incorporation of Cl^{-}.

Summary. Following upon a brief discussion of the semiconducting elements silicon and germanium and of non-stoechiometric oxidic semi-conductors, the "method of controlled valency" is described by taking nickel oxide as an example. This method consists in bringing about desired changes in valency by the admixing of suitable "impurities". There is a great similarity in the mechanism of conductance between non-stoechiometric semiconductors and other semi-conductors prepared by the method of controlled valency. Compared with the former, the latter group however possess great advantages, which are to be ascribed to the lack of thermal stability of non-stoechiometric compounds. This lack of stability is due to the fact that in non-stoechiometric compounds there are always vacant sites in the lattice.

Somewhat similar considerations apply to zinc-sulphide phosphors. After the part played by halogen ions in the formation of these substances had been realized, a better picture could be formed of the fluorescent centres: the centres are formed by an activator ion surrounded by sulphur ions. The accuracy of the hypothesis developed finds confirmation in the possibility of having the part played by the halogen ions taken over by trivalent cations. This opens the way to the production of entirely new phosphors.

[7]) The possibility cannot be excluded that very small quantities of activator are taken up in the lattice without the presence of Cl, Br or I or of trivalent cations. The concentrations of the activator which can be introduced into the crystal when one of these ions is used in the preparation of the product are, however, so much greater that this fact may safely be ignored in our considerations.

[8]) F. A. Kröger and J. Dikhof, Physica **16**, 297, 1950.

COLLIMATING X-RAYS IN BEAMS OF VERY SMALL DIVERGENCE AND HIGH INTENSITY

by J. A. LELY and T. W. van RIJSSEL. 537.531:535.312:539.26:621.386

When X-rays fall upon a polished surface at a very small angle total reflection may take place. Use has been made of this principle in constructing an "X-ray lens" and even an X-ray microscope. It also enables one to make — as an aid in the X-ray diffractional examination of proteins and suchlike substances — a simple collimating system producing an X-ray beam with very small divergence, a relatively high intensity and very little scattered radiation.

X-ray diffraction

The examination of crystalline substances with the aid of X-ray diffraction is indispensable for the identification or chemical analysis of such substances and for the investigation of crystal structure, texture and many other propertes. Several articles on this subject have already appeared in this Review[1]), so that there is no need to go into the theory of X-ray diffraction here. We would recall, however, the well-known Bragg's equation, which reads:

$$2\,d \sin \Theta = n\lambda, \quad \ldots\ldots \quad (1)$$

where d is the distance between a system of parallel lattice planes in a crystal, Θ the angle between the incident ray and those planes, λ the wavelength of the rays and n a whole number giving the order of the reflection. Reflection takes place only when the angle Θ satisfies the condition given in the formula; a ray striking the lattice planes at any other angle is not reflected. It is to be noted that in this connection it is always the angle between the incident ray and the reflecting lattice plane which is used, and not the angle between the direction of the ray and the normal to the reflecting plane, as is usually taken in optics. The scattering angle of the X-rays in the case of reflection is therefore equal to 2Θ.

When employing the CuKα rays emitted with a wavelength of 1.54 Å by a copper anticathode (copper is the material frequently used for X-ray diffraction work; molybdenum and tungsten are also used, but then the rays are of a shorter wavelength) and with spacings in the order of a few Å such as occur in normal crystals, one finds for the scattering angles of the reflected rays values lying between approximately 10° and 180°. Particularly the angles between 10° and 80° are of importance.

Divergence of the X-ray beam

Disregarding special focusing systems, in order to obtain diffraction angles which are defined with sufficient accuracy it is as a rule essential to have a narrow X-ray beam. This is obtained, for instance, by confining the rays emerging from the focus of the X-ray tube within certain limits fixed by means of two diaphragms set up some distance apart. The narrower the diaphragms, the less is the divergence of the X-ray beam, but at the same time also its intensity is reduced. When applying the Debye-Scherrer method, whereby the diffraction of the X-ray by a finely distributed crystalline powder is investigated (all kinds of lattice planes in the separate crystals then occur partly in the reflecting state), one finds that the reflections are rather weak. If in that case an accurate definition of the diffraction angle is desired, then the exposure has to be rather long (some hours) to be able to record all the reflections. Recording is mostly done on a photographic plate or film, but nowadays a Geiger-Müller counter is increasingly used.

Scattered radiation

In addition to the X-rays limited by the diaphragms and reflected by the lattice planes in the crystal, some scattered rays also fall on the photographic plate or film.

This scattering is caused by the edges of the diaphragms and by the molecules of the air. It can be suppressed for the greater part by taking suitable measures (rounding off the edges of the diaphragms, enclosing the primary X-ray beam in front of and behind the specimen in a narrow cylinder). With perfected cameras (such as described, for instance, in the second article quoted in footnote[1])) it is possible to work with angles

[1]) See, e.g., W. G. Burgers, The use of X-rays and cathode rays in chemical and metallographic investigations, Philips Techn. Rev. 5, 157-166, 1940; W. Parrish and E. Cisney, An improved X-ray diffraction camera, Philips Techn. Rev. 10, 157-167, 1948.

as small as 3°. When using CuKα rays these angles correspond to a separation of the lattice planes by about 15 Å. With still smaller angles there is a troublesome background of scattered radiation.

This scattering gives rise to difficulties when it is desired to study "crystalline" substances whose lattice planes are spaced far apart, up to some hundreds of Å, as is the case with proteins and other organic substances. The scattering angles of the reflected rays are then very small (less than one degree) and the intensity of the reflected rays from these organic substances is also extremely small, so much so that with a normal camera it is impossible to observe these reflections. A similar difficulty is encountered when it is desired to study the small-angle scattering of X-rays by a very fine powder as such. From this small-angle scattering, information can be obtained on the distribution of the grain sizes and on the mutual distances of the grains. The intensity of the scattered image decreases rapidly with increasing scattering, so that the image becomes of importance mainly with scattering angles smaller than 1°. But this image, which is still very weak, can be observed only if the scattering caused by the edges of the diaphragms and by the molecules of the air is suppressed as far as possible.

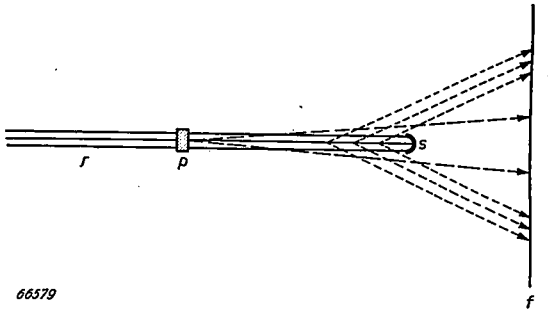


Fig. 1. The scattering of X-rays by the molecules of the air (see short-dash lines) takes place mainly over large angles. When the lead bead s intercepting the primary beam r is placed a few centimetres in front of the film f, practically none of the air-scattered rays fall within the area of small angles of diffraction, while the solid angle from which these scattered rays emerge is also small. The rays diffracted by the specimen p are represented by the long-dash lines.

For the investigation of this small-angle scattering special measures have to be taken, with the object of minimizing the intensity of the scattered radiation at such small angles of diffraction and increasing the intensity of the primary beam. In this case, it should be noted, the scattering by the air is of no consequence. The molecules of the air tend to scatter the X-rays at large angles to the direction of the primary beam. The concave lead bead with which the primary beam is cut off for these investigations, so that it shall not blacken the film, is placed a few centimetres in front of the film, the rays scattered by the molecules of the air then striking the film only in the region of the larger scattering angles. At the same time the solid angle from which the scattered rays emerge is very small *(fig. 1)*. Thus in our case the scattering at the edges of the diaphragms is the most important factor.

Various collimating systems

To reduce this edge scattering as far as possible, various systems have been developed. For instance, there is the system whereby a third slit diaphragm is added to the series of two diaphragms already

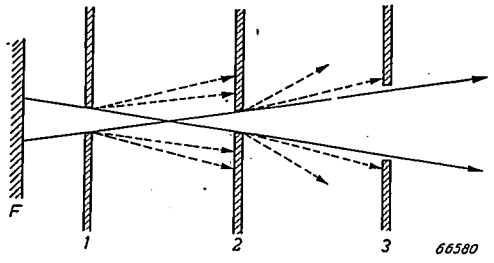


Fig. 2. Extended slit-diaphragm collimator. Of the rays emitted by the focus F only a narrow beam is allowed to pass through the apertures *1* and *2*. The edges of these apertures cause scattering. Therefore a third diaphragm, *3*, is added, usually placed at the same distance from *2* as that between *1* and *2*. The width of this third aperture is such that the primary beam can just pass through without touching the edges The rays scattered by the edges of aperture *2* are intercepted by the third diaphragm. The width of the apertures *1* and *2* is usually 100 μ to 200 μ, whilst the distances between the diaphragms are of the order of 10 to 20 cm.

mentioned (see *fig. 2)*. Such a system requires very precise alignment (the slits are narrow and spaced far apart; the edges of the third slit must just be clear of the primary beam) and yields only a small intensity. The third diaphragm, moreover, makes the system longer and as a consequence the intensity at the specimen is much less still than that obtained with only two diaphragms.

A collimator based upon an entirely different principle is the planar crystal monochromator, the working of which is graphically explained in *fig. 3*. As its name implies, this collimator produces a monochromatic beam, this being an advantage on account of the different wavelengths produced by the X-ray tube. In addition to the CuKα rays a copper anticathode yields also CuKβ and CuKγ rays, etc. Furthermore, there is continuous radiation, particularly at the smaller wavelengths.

It is only with monochromatic radiation that sharply defined diffraction lines can be obtained. With slit-diaphragm systems, both for small and for normal angles of diffraction, it is therefore always necessary to have a filter in the primary beam

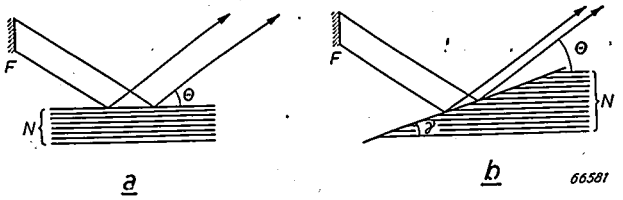


Fig. 3. Planar crystal monochromator. *a)* The rays falling from the focus F onto the crystal are reflected if their angle of incidence Θ with the lattice planes N satisfies Bragg's law. All reflected rays are (theoretically) parallel; the width of the beam is determined by the width of the focus. *b)* Often the crystal is ground in such a way that the lattice planes make an angle γ with the surface which is only a little smaller than Θ; the reflected beam is then extremely narrow and a greater "brightness" is obtained. The depth to which the X-rays penetrate into the crystal is very small, of the order of a few lattice spacings. Here these distances are greatly exaggerated.

so as to make it monochromatic (i.e. to filter out the Kβ rays, etc; the continuous radiation cannot be suppressed in this way). Naturally such a filter involves additional loss of intensity. With the crystal monochromator no filter is needed, and moreover a larger part of the beam emerging from the focus is utilized. Compared with the slit-diaphragm system, however, this crystal monochromator still does not give any appreciable increase in intensity, since about 80% of the intensity is lost in the reflection at the crystal.

By grinding the crystal concave and then bending it (bent crystal monochromator) a much larger part of the beam emerging from the focus can be made effective. But neither the planar nor the curved crystal monochromator is of any practical use for angles of diffraction smaller than about 10′, owing to the irregularities in the crystal (mosaic structure).

Collimation by means of total reflection

In the Philips Laboratory at Eindhoven a collimator has been developed with which an X-ray beam with a divergence of about 5′ can be produced in a very simple manner. This collimator yields an intensity about three times greater than that obtainable with a slit-diaphragm system for the same divergence. There is little scattering, the beam passed through is partly monochromatic and the construction is very sturdy. It is based upon the principle of the total reflection of X-rays.

For all solids the index of refraction for X-rays is less than 1 (contrary to that for light rays). As a result, when passing from air to, say, glass, the X-rays can be totally reflected provided their angle of incidence is smaller than a certain critical angle η. As we shall see presently, the indices of refraction for X-rays by the various materials differ very little from unity, and this means that η is very small, so that only rays falling at a very small angle are reflected; all others are almost completely absorbed (disregarding here possible Bragg reflections).

Let us turn again to the simple slit-diaphragm system consisting of two apertures *(fig. 4a)*. Of the rays emerging from point A at the anticathode of the X-ray tube only those which lie in the cone with boundary rays p and q are effectively used. Rays such as p' and q' passing through the first aperture do not reach the second one and are therefore lost. This accounts for the low intensity of the beam passing through a slit-diaphragm system. The effective aperture angle as seen from any point on the surface of the anti-cathode is determined by the width of the last slit and its distance from the focus.

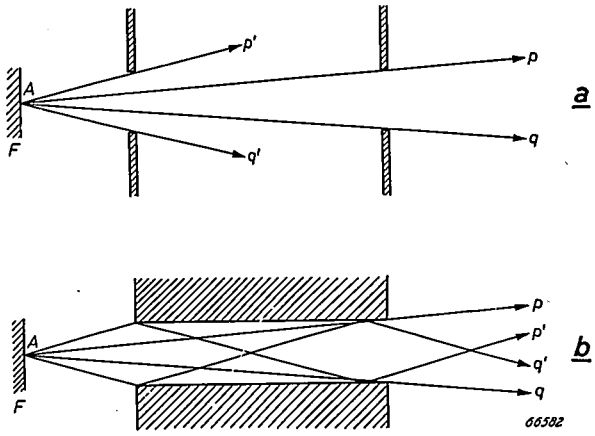


Fig. 4. *a)* With a slit-diaphragm system only those rays are passed through which lie in the solid angles formed by the points of the focus and the aperture of the last diaphragm, so that only part of the beam emitted by the focus is effectively used. *b)* When the collimator consists of two totally reflecting plates full use is made of all the rays lying in the solid angle formed by a point of the focus and the first aperture. (It is assumed that in the situation drawn total reflection is possible.)

The rays p' and q' can, however, be utilized by substituting for the two slit-diaphragms two polished glass plates (fig. 4*b*), from which these rays are totally reflected. (Here we are speaking of bilaterally limited beams and not of all-round limited beams; for the latter circular diaphragms would have to be used instead of slit diaphragms, and a cylindrical hollow tube instead of plates.)

Such a collimator has been described by Nähring[2]). Thus the whole of the beam emerging from the point A passes through the collimator. Compared with a slit-diaphragm system with the same divergence the total-reflection collimator therefore gives a more intensive beam. It has yet another advantage, in that scattering of the rays at the edges of the slits is entirely avoided; since the rays are reflected from the plates there is no absorption and thus no scattering, not even at the edges at the end of the collimator.

The quality of this collimator, however, is bound to two limits. In the first place the angle of divergence of the incoming rays cannot be chosen haphazardly; only those rays from A striking the reflecting surface at an angle smaller than the critical angle η are passed through. The second limitation is that the divergence of the emerging beam cannot be made very small; it will amount to 2η, if the collimator is of sufficient length (a least just as long as a split-diaphragm collimator with the same divergence). If the collimator is made even longer this does not have any further influence upon the divergence of the emerging beam, as is the case with a slit-diaphragm collimator. Now the critical angles of the materials suitable for the collimator are in the order of 10′ to 15′, which means to say that with the total-reflection collimator beams cannot be obtained with a divergence less than 20′ to 30′. For investigations where this divergence is small enough this collimator offers the above-mentioned advantages over the slit-diaphragm system, but for the examination of proteins and suchlike, where a still smaller divergence (of a few minutes of arc) is required, the collimator is of no use.

However, by slightly modifying the construction it is possible to give the emerging beam a much smaller divergence, while still retaining the advantages of higher intensity and the minimum of scattering. This is achieved by mounting the polished plates at a small angle with respect to one another instead of parallel *(fig. 5)*. Thus the collimator is given a wedge shape, the rays entering through the narrow opening and emerging through the wider one. Say that the angle made by the two plates to the plane of symmetry is β. A ray entering at an angle δ to this plane and striking the wall is reflected at an angle $\alpha - 2\beta$ to the plane (we are considering only the absolute value of the angle). After n reflections the angle made by the ray with the plane of symmetry is

$$a_n = a - 2n\beta \,. \quad \ldots\ldots \quad (2)$$

Thus the path followed by the ray will be more and more in the direction of the plane of symmetry. Since this applies to each ray, the divergence of the emerging beam will be smaller than that found with parallel plates. In this way a beam can be obtained with a divergence of about 5′ without sacrificing any of its intensity.

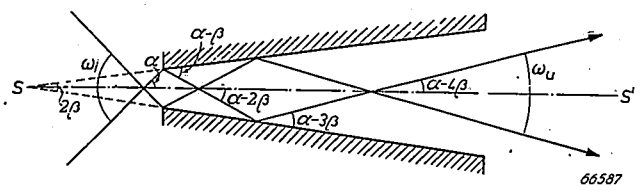


Fig. 5. The wedge collimator. Each of the two reflecting surfaces makes an angle β with the plane of symmetry ss'. A ray having an angle of incidence a with respect to this plane (and striking the wall) will be reflected if $a<\eta+\beta$ (where η represents the critical angle of the material) and will be absorbed if $a>\eta+\beta$. After having been reflected n times a ray entering with an angle of incidence a to the plane of symmetry will have only the angle $a - 2n\beta$ to that plane. With a suitable choice of the dimensions of the wedge (determining n) and the angle of inclination β, the rays entering within the aperture angle $\omega_i = 2(\eta+\beta)$ will leave the wedge with a much smaller aperture angle ω_u. Here two rays are drawn which have the largest possible angle of incidence $(\alpha-\beta=\eta)$. The angles are strongly exaggerated in this diagram.

Before considering further the path of the rays it will be useful to discuss briefly the phenomenon of total reflection of X-rays.

Total reflection of X-rays

The condition for total reflection is:

$$\sin i_1 > n_2/n_1 \,, \quad \ldots\ldots \quad (3)$$

where i_1 is the angle made between the incident ray and the normal to the boundary plane of the two media, n_1 and n_2 are the indices of refraction of the media $(n_1 > n_2)$, whilst the direction followed by the ray is from medium *1* to medium *2*. If the angle made with the surface is, say, $\varphi = 90° - i_1$ then the condition for X-rays passing from air to a substance with an index of refraction $1 - \delta$ is:

$$\cos\varphi > 1-\delta \,. \quad \ldots\ldots \quad (4)$$

Thus the critical angle η, the largest angle at which total reflection is still possible, is given by:

$$\cos\eta = 1-\delta \,,$$

or, since $\delta \ll 1$, to a very close approximation, by:

$$\eta = \sqrt{2\delta} \,. \quad \ldots\ldots \quad (5)$$

[2]) E. Nähring, Phys. Z. **31**, 401-418, 1930, in particular p. 405.

Since $\sqrt{2\delta}$ is very small, viz. less than 1.5×10^{-2} for all solids, total reflection of X-rays can be obtained only when they strike the surface at a very small angle ($\varphi < 50'$).

Following this reasoning one would expect that the coefficient of reflection R for rays having an incident angle smaller than the critical angle would be equal to unity, and for rays with an incident angle larger than η equal to zero.

Actually the variation in the coefficient of reflection is more gradual, as indicated in *fig. 6*.

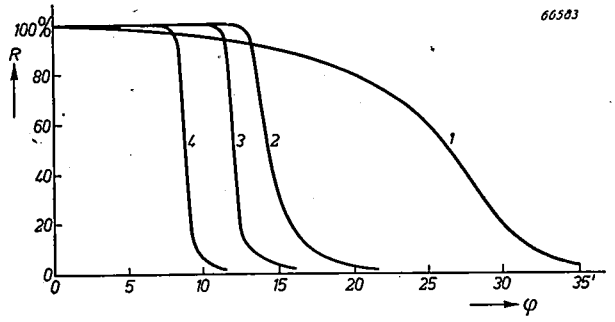


Fig. 6. Reflection curves for different materials. The reflection coefficient R is plotted against the angle of incidence φ (in minutes of arc). The lighter the material, the less it absorbs the X-rays and the smaller is the critical angle (the point where the curve bends), whilst also the curve is steeper. *1* silver, *2* glass not containing lead, *3* plastic material, *4* beryllium. These curves apply for CuKα rays of 1.54 Å wavelength.

For angles of incidence slightly smaller than the critical angle the coefficient R is not yet equal to 1, whilst with angles larger than η reflection still takes place, though to a smaller extent. This is related to the absorbent properties of the material: the lower the absorption coefficient, the steeper are the curves and the smaller is the critical angle.

The curves drawn in fig. 6 apply for a wavelength of 1.54 Å. For larger wavelengths the curves are displaced to the right and are less steep; the critical angle appears to be proportional to the wavelength.

Since the indices of refraction differ so little from unity it is not possible to make an X-ray lens. The focal length f of a simple refracting surface with radius r is of the order of r/δ, thus much greater than r. Therefore, in order to give a useful focal length an X-ray lens would have to be built up from a large number of components. If a material with large δ is to be chosen for this "lens" then that material would also have a large absorption coefficient. Consequently the total absorption in the X-ray lens is always far too great for the lens to be of any practical use.

The fact that ordinary light can be concentrated with lenses is due to the indices of refraction of solids for light being in general much higher than unity, while there are materials, such as glass, which practically speaking absorb no light at all.

Apart from collimation, which, as we have seen, can be brought about by total reflection, it has also been found possible to form an image by means of X-rays with the aid of this principle. With one or more spherically or elliptically ground mirrors an "X-ray lens" (actually it is not a lens) has indeed been made [3]), and even an X-ray microscope has been designed, also with concave mirrors [4]). Compared with an optical microscope such a microscope has the advantage that its resolving power is much greater, and compared with an electron microscope, the rays penetrate more deeply into the object being examined, so that also the internal parts of the object can be studied. However, it has not yet been possible to obtain a magnification of more than 30 times, and the aberrations are very considerable.

The path of the rays in the wedge collimator

We shall now consider briefly the relation between the directions of the incident rays and those of the emergent rays. The conditions to be satisfied if the incident rays entering through the narrow opening are to be reflected is that the angle of incidence at one of the reflecting planes must be smaller than η, or, that the angle α between the ray and the plane of symmetry (see fig. 5) shall be given by:

$$\alpha < \eta + \beta . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The effective aperture angle of the incident rays is thus given by:

$$\omega_i = 2\,(\eta + \beta) . \quad . \quad . \quad . \quad . \quad (7)$$

Here it is assumed that the dimensions of the collimator are such that rays with an angle of incidence greater than $\eta+\beta$ cannot pass through the collimator without being reflected. If d_i is the width of the entrance and d_u that of the outlet, whilst l represents the length of the collimator, then it is assumed that the condition

$$\tan \alpha_0 = \frac{d_i + d_u}{2l} < \tan\,(\eta + \beta) ; \quad . \quad (8)$$

is satisfied; α_0 is the maximum angle of incidence at which a ray can pass through the collimator without being reflected.

To calculate the shape and divergence of the emerging beam an artifice is employed *(fig. 7)*. By projecting mirrored images of the wedge from its reflecting surfaces it is possible to draw, for instance, instead of the zig-zag ray $APQRD$ the straight ray AD''. Since in practice the angle β is only a few minutes of arc, the line $DCD'C'D''$ can be regarded as being straight, although it does not

[3]) W. Ehrenberg, J. Opt. Soc. Am. **39**, 741-746, 1949, and P. Kirkpatrick, J. Opt. Soc. Am. **39**, 796-797, 1949.
[4]) P. Kirkpatrick and A. V. Baez, J. Opt. Doc. Am. **38**, 766-774, 1948.

appear to be so in this diagram drawn on such an exaggerated scale.

The rays coming from A which pass through the wedge without being reflected at all have an angle of incidence α such that

$$\tan\alpha < \tan\alpha_0 . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

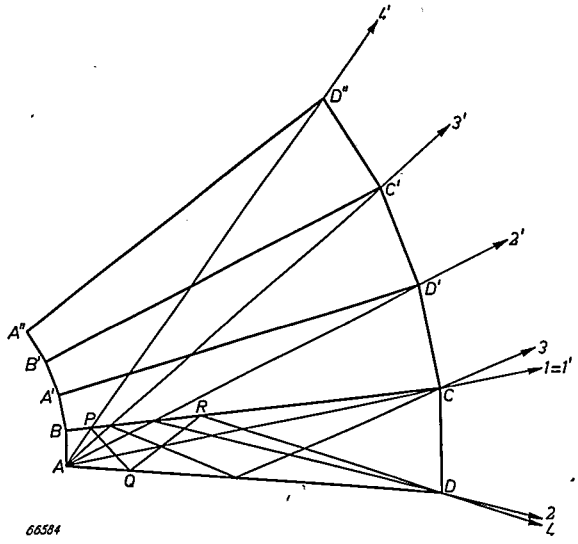


Fig. 7. Projection for calculating the path followed by the rays through the wedge.

A ray (passing through A) will be reflected once if its angle of incidence lies between α_0 and α_1, the latter being given by:

$$\tan\alpha_1 = \frac{d_i + 3d_u}{2l}, \quad . \quad . \quad . \quad . \quad (10)$$

as follows directly from fig. 7. So we may continue and ultimately find that a ray passing through the point A will be reflected n times if its angle of incidence lies between α_{n-1} and α_n, where α_n is given by

$$\tan\alpha_n = \frac{d_i + (2n+1)\,d_u}{2l} . \quad . \quad . \quad . \quad (11)$$

Since all angles are very small, in these formulae the tangents may be replaced by the angles.

In this way it is also possible to follow the rays not entering at A but at some other point in the entrance opening. By calculating the angle of emergence of all these rays with the aid of formula (2) we arrive at the shape and divergence of the emerging beam.

Calculations show that the intensity of the emergent rays is practically the same in any permissible direction of emergence, whilst the distribution of the intensity over the whole width of the wider outlet is uniform. The outlet can easily be made 100 to 200 μ wide, thus giving it a width comparable to that of the slits in a diaphragm system. Although the width of the entrance to the wedge collimator is much smaller than that in a slit-diaphragm system, the intensity of the emergent beam is much greater, owing to the larger solid angle ω_i for the incidence of the rays; in the slit-diaphragm system this angle is equal to the divergence of the emerging beam, whereas in the wedge collimator it is much larger.

The gain in intensity compared with a slit-diaphragm system yielding a beam of the same width and divergence amounts to a factor of 3 to 5.

The wedge also functions somewhat as a monochromator. The effective aperture angle for the incident rays is given by (7), and since β is always much smaller than η it may be assumed, without making any great error, that the effective aperture angle is proportional to the critical angle. Now for all materials the rule is that η is proportional to λ, and so the aperture angle will also be proportional to λ. This is just what is desired: the continuous part of the radiation emitted by the focus, which part has mainly shorter wavelengths and cannot be intercepted by a filter, is attenuated with respect to the monochromatic component, the ratio of this attenuation being equal to that of the wavelengths, i.e. for normal tube voltages a factor of 2 to 3.

The construction of the wedge is determined by the critical angle of the material and the divergence desired. As will presently be shown, it is preferable to choose a material with the steepest possible reflection curve, thus also with a small critical angle. Common window glass with $\eta = 14.5'$ (curve *2* in fig. 6) answers very well; still better would be an organic material (such as plastics) containing only light atoms, but for our purpose this is not sufficiently stable in shape.

As an example the dimensions are given of a wedge collimator made of window glass in this laboratory, the construction of which is outlined in *fig. 8*. The length is 120 mm, the entrance is

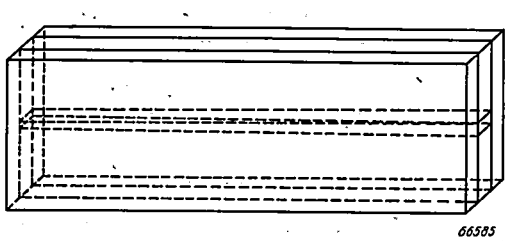


Fig. 8. Construction of the wedge collimator. The reflecting plates are placed at the precise distances and with the precise inclination with the aid of gauges and then firmly cemented between two other plates. The whole is very sturdy and easy to make.

9 μ wide and the outlet 140 μ wide. With CuKα radiation a beam can be obtained with a divergence of 4.8 minutes of arc. According to formula (7) the effective aperture angle for the incident rays is 33′ (since $\beta = 2'$).

Around the emerging primary beam with a divergence of approximately 5′ there appears a "halo" of about 25′ diameter. This halo is not to be ascribed to scattering by diaphragm edges such as arises with a slit-diaphragm collimator, and it is therefore of much lower intensity. It arises from other causes of "scattering".

"Scattering" in the wedge collimator

There are three causes of this scattering in the wedge collimator.

1) As a result of the irregularities in the surfaces of the glass plates some scattering is bound to occur. Assuming that by thorough polishing the differences in thickness of the plates can be reduced to about 20 Å, it may be estimated that for the small angles of incidence with which we are dealing here the scattering due to such irregularities will not be of much consequence. Owing to this effect the scattering halo might have an aperture angle of about 15′; this does not, therefore, account for the 25′ mentioned above.

2) Owing to the "tail" of the reflection curve also rays with angles of incidence larger than η are reflected, thus giving the emergent beam a greater divergence. This effect, of course, is not scattering in the real sense of the word, being rather a regular reflection, but here it is considered as scattering because it likewise contributes towards the formation of a halo around the primary beam. The larger the number of times the ray is reflected, the smaller is the angle the ray makes with one of the polished surfaces. Consequently a ray striking one of the planes for the first time at an angle $\varphi > \eta$, and thereby being reflected, even if for only a small percentage, will be much better reflected the second time because then the angle of incidence will be nearer η, or it may perhaps have become smaller than η. At the next reflection the ray will be still better reflected.

The decrease of the angle of incidence is determined by formula (2). When, as is the case in practice, φ is only a few minutes of arc, all rays with an incident angle $\varphi > \eta$ will be considerably weakened when passing through the collimator, so that with $\varphi > 1.5\,\eta$ the intensity of the emergent rays will be about one million times smaller than that of the incident rays, as a rough calculation shows. Rays with an incident angle φ between η and $1.5\,\eta$ will emerge partly within the angle of divergence of the rays with an incident angle $\varphi < \eta$. This contribution towards the intensity of the emergent beam is an advantage, albeit a very slight one. Part of these rays, however, have a larger divergence and give rise to a halo around the emergent beam. It appears that owing to the "tail" of the reflection curve the aperture angle of this halo will amount to about 8′.

3) The polishing agent, which contains iron, leaves some iron atoms behind in the uppermost layer of the surface, and when these atoms are excited by the incident X-rays they themselves radiate in all directions, thus giving rise to "secondary X-rays". This effect can be made quite perceptible with the aid of various filters introduced in the beam. It is possible that this accounts for the divergence of 25′ for the scattering halo. This halo could be reduced by polishing, for instance, with aluminium oxide.

Use of the wedge collimator

The specimen to be examined is placed immediately behind the wide opening, and the film 10 to 20 cm behind that. A nickel filter is placed in front of the wedge so as to suppress the CuKβ rays.

As an example of a photographic recording *fig. 9* gives a diffraction pattern of wet collagen. The

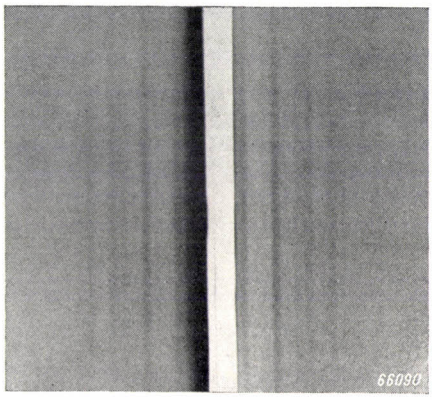


Fig. 9. Diffraction pattern of a wet collagen specimen 1 mm thick, taken with the aid of the wedge collimator (scale of the reproduction 3:1). Focus 1.2 mm × 1.2 mm, tube voltage 30 kV, tube current 30 mA, distance object to film 200 mm, exposure 20 hours. The spacing in wet collagen is 660 Å and the reflections of the second and higher orders are quite perceptible (mutual distance 8′). The reflection of the first order is rendered invisible by the scattering background. The primary beam was intercepted with the aid of a lead bead.

distance of the most important lattice planes of this protein is 660 Å, and thus with CuKα X-rays reflections are found with 8′ between their successive orders. The reflections of the second and higher orders are quite perceptible. The reflection of the first order (angle of diffraction 8′) is not visible

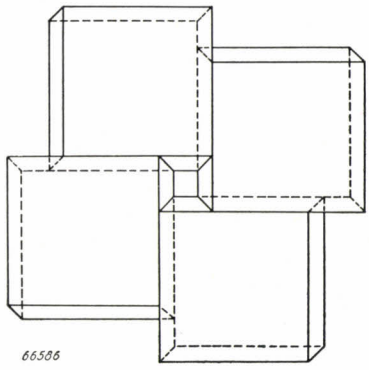


Fig. 10. Sketch representing the construction of a collimator according to the principle described here and limiting the X-ray beam on all sides (seen from the end with the wide opening),

owing to the background of scattered radiation. The exposure for this recording was 20 hours. With a slit-diaphragm system of the same divergence an exposure of 75 hours would have been required.

A collimator according to the principle described here can also be constructed for limiting a beam on all sides. This is of importance for many applications, e.g. for examining single crystals in different directions. The principle of the construction is represented in *fig. 10*. Theoretically, with such a collimator an all-round limited beam can be obtained with an intensity 10 to 30 times greater than a beam of the same diameter and divergence limited by a number of diaphragms with small aperture.

Summary. For studying the structure of certain substances it is necessary to be able to measure the diffraction of X-rays over very small angles ($<1°$). This requires an X-ray beam with very small divergence, of a few minutes of arc. To satisfy this requirement and at the same time obtain sufficient intensity of the beam and a sufficiently weak scattering background, special collimators have been developed, which are briefly described. The authors have constructed a collimator which is very sturdy and quite simple to make. This collimator, with which a divergence of about 5′ can be obtained, is based upon the principle of the total reflection of X-rays when they strike a polished surface at a very small angle. By means of a wedge-shaped arrangement of the reflecting planes, which can be made, for instance, from window glass, a beam of about 30′ divergence can be reduced entirely to the much smaller divergence mentioned, thereby producing a relatively intense beam. Consideration is given to the causes of the scattering occurring, which is found to arise to a smaller extent than is the case with other types of collimators. Finally an example is given of the results that can be achieved with this collimator.

MF 8664

MEASURING RAPIDLY FLUCTUATING GAS TEMPERATURES

by B. H. SCHULTZ.

621.317.39:536.5:536.717:621.4/.5

Petrol engines, diesel engines and suchlike are being used by the million and yet the processes taking place inside them are still not known in great detail. For engines which are still in course of development, such as gas turbines, air engines, etc., this obviously applies all the more. Perhaps the most important detail in respect to which a deeper insight is desired in all these cases is the variation of the gas temperature in the engine cylinders during a cycle of the work process. The method of measuring described here promises to broaden considerably the knowledge we have about this problem.

Fluctuating temperatures in thermal engines

One of the deciding factors for the work process taking place in any kind of thermal engine, be it a steam or a petrol engine, a jet or an air engine, is the temperature of the gases taking part in the process. This may be explained by an example. In the case of the air engine [1]) the gas (in this case air) is presumed to have a constant and uniform high temperature T_w in the hot space and a constant and uniform low temperature T_k in the cold space. Energy is obtained by causing the gas in the hot space to expand and then transferring it via the regenerator to the cold space and compressing it, after which it is transferred back to the hot space via the regenerator, the cycle then starting again. The efficiency of the process in the ideal case (when there are no losses whatever) is:

$$\eta = \frac{T_w - T_k}{T_w} \quad \ldots\ldots\ldots \quad (1)$$

Actually these temperatures T_w and T_k are not exactly constant with time, and the gas temperature also varies from point to point. The walls of the hot and cold spaces are indeed at constant temperatures, but in the course of expansion the gas in the hot space is slightly cooled, especially in the central part of the gas volume, since the heat exchange with the wall cannot be complete in the very short time available (in the case of an engine running at 3000 r.p.m. the time for each expansion is only about 0.01 second). Much the same applies for the gas in the cold space during compression. Thus at any point in the gas space there will be rapid periodical fluctuations of the gas temperature. Equation (1) is then no longer valid. The power and efficiency then have to be calculated from the periodical variation of pressure as influenced by the fluctuating gas temperatures.

[1]) See, e.g., H. Rinia and F. K. du Pré, Air engines, Philips Techn. Rev. 8, 129-136, 1946.

In principle it is possible to calculate these fluctuating gas temperatures from constructional data and the working conditions of the engine, not only of that taken as an example here but also for other engines, but when making such calculations one is faced with considerable difficulties. A method has therefore been developed by means of which fluctuating temperatures can be measured quickly and with the desired accuracy. The novelty of the method to be described here is not to be sought in its principles but in the details of application imposed by the exceptional requirements.

Principle of the method of measurement

The method is based on the use of a resistance thermometer which consists of a thin wire of a suitable material introduced into the gas at the point where the temperature is to be measured, forming, one of the four arms of a bridge. Any variation in the temperature of the filament causes a variation in its resistance, the bridge is thrown out of balance and the diagonal voltage of the bridge serves as a measure for the change in temperature. By applying this voltage to a cathode-ray oscillograph it is possible to record and measure also rapid fluctuations in temperature. In our case, where the fluctuations occur periodically with a cycle equal to the duration of one revolution of the engine (twice that in the case of a four-stroke engine), it is most convenient to synchronize the time base of the oscillograph with the engine. Thus a stationary picture of the temperature variation is obtained on the screen, in so far as the temperature fluctuations occurring are sufficiently reproducible.

Fig. 1 is a photograph of the measuring set-up. Further on the reader will find reproductions of some oscillograms obtained with this set-up, with explanations in the subscript.

Thermal inertia of the thermometer wire

Owing to its thermal inertia the thermometer wire will respond to the gas temperature with the required accuracy only when the engine is running at a sufficiently low speed. As the speed increases, the temperature of the wire lags more and more behind that of the gas, and to a first approximation the situation might be regarded as if there were only a delay in the wire temperature, the form and amplitude of the temperature variation of the wire still being precisely the same as that of the gas. The time lag τ is given by:

$$\tau = \frac{W}{aF} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where W is the thermal capacity of the wire, F its surface area and a the coefficient of heat transfer between gas and wire, which is determined mainly by the geometrical configuration (thickness of the wire).

The coefficient a is defined by the equation:

$$dQ = aF(T_g - T_d)\,dt, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

which expresses that the amount of heat dQ passing over from the gas to the wire in the short time dt is proportional to the surface of the wire and to the difference between the temperatures T_g and T_d of the gas and the wire respectively.

From eq. (3) it is easy to derive eq. (2), since dQ is the heat absorbed by the wire in the time dt, thus being equal to the product of thermal capacity and rise in temperature of the wire [2]): $W dT_d$. Hence eq. (3) may be written in the form:

$$T_g = T_d + \frac{W}{aF}\frac{dT_d}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Denoting W/aF by τ, the right-hand term may be regarded as the MacLaurin development of $T_d(t + \tau)$ broken off after the second term, so that eq. (4) gives expression to the fact that at the instant t the gas temperature T_g is approximately equal to what the temperature of the wire T_d will be at the instant $t + \tau$. Thus, with respect to T_g, T_d lags behind by a period of time τ, which is given by eq. (2).

Owing to the fact that the exact MacLaurin development for $T_d(t + \tau)$ contains terms of the second and higher derivatives which do not occur in the right-hand term of eq. (4), the wire cannot exactly follow sharp changes in the gas temperature but shows rounded-off temperature curves. Such a distortion, however, is of less practical importance than another which will now be dealt with. The remedy that will be discussed greatly reduces both distortions simultaneously.

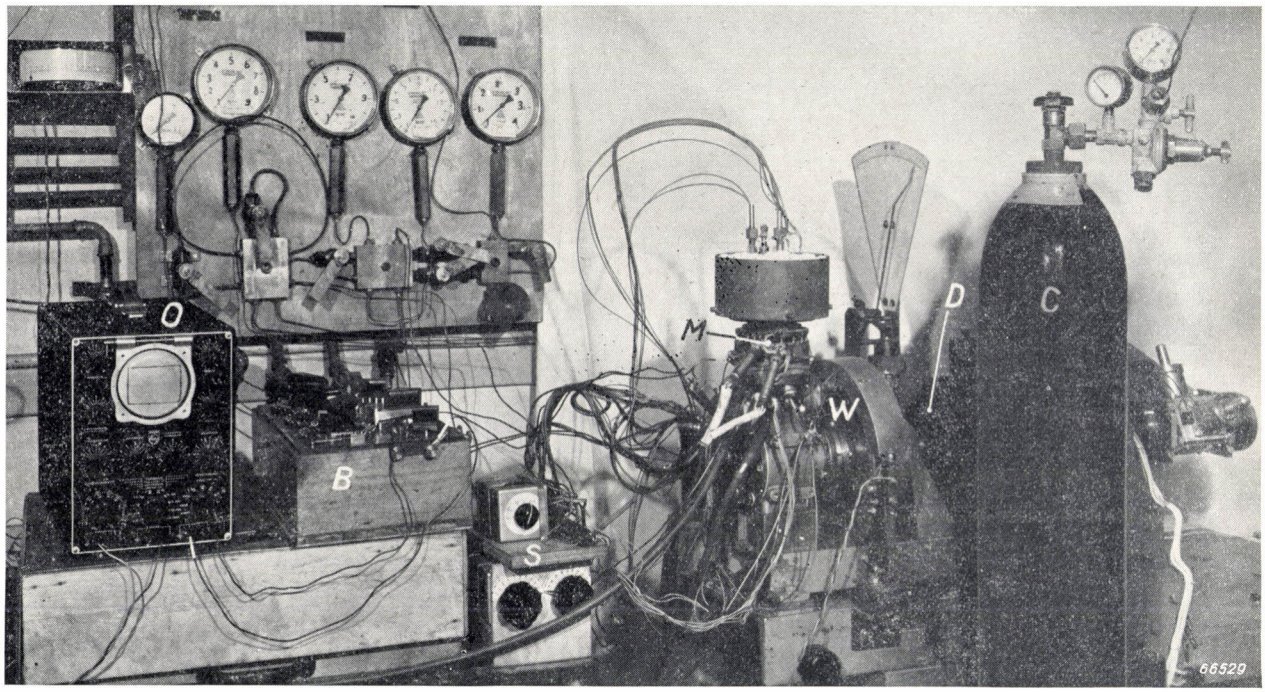


Fig. 1. The complete measuring set-up. M is an air engine in which thermometer wires have been introduced at various places. B are the variable resistors of the bridge circuit. O is the cathode-ray oscillograph. S are switches for connecting one of the wires as required. In this set-up there are coupled to the engine a flywheel W (to which is attached the interrupting mechanism referred to later on) and a brake dynamo D. The pressure inside the engine is adjusted with the aid of the high-pressure cylinder C. The manometers at the top of the picture on the left serve for various static pressure measurements.

[2]) At least if it is permitted to disregard the heat transfer between the wire and the surroundings through radiation or through conduction via the extremities; see below.

When the time lag τ is known, the temperature curve as recorded could be corrected accordingly. Actually, however, this description is still too rough, for account must be taken of the fact that τ is also somewhat dependent upon the velocity and the pressure of the gas flow, and that α, and with it τ, therefore varies during the stroke of the engine. We could take an "average delay", but the varying value of τ naturally implies a certain distortion of the temperature curve which is not easy to correct. It is preferable, therefore, to try to make τ so small that the correction can be entirely ignored, even at the highest engine speeds.

To attain this object it was necessary to make the thermometer wire extremely thin. It is easy to understand that the thinner the wire the better it is able to respond to changes in the gas temperature, but it is well to define this more precisely, because the manner in which α depends upon the thickness of the wire d plays an important part in this case. With a given length of wire the surface area F is proportional to d, whilst the volume and thus the thermal capacity W is proportional to d^2, so that owing to these two factors in eq. (2) the lag τ already decreases proportionately with d. Moreover, as empirically established, the coefficient of heat transfer α changes in proportion to $d^{-0.7}$, so that with decreasing thickness of the wire the delay time is reduced at an even higher rate, viz. in proportion to $d^{1.7}$.

For our thermometer we mostly use Wollaston wire, this being made by coating fine platinum wire with silver and drawing it down further so that the platinum core is reduced, in our case, to a thickness of from 1.5 to 2 microns; after the wire has been mounted the silver is dissolved off by etching, so that only the platinum wire of the thickness mentioned remains. We have also frequently used tungsten wire reduced to such a fine thickness by etching, and which is strong enough to be handled in this form. With these extremely fine wires the coefficient of heat transfer α reaches the exceptionally high value of about 0.3 cal/cm^2 degree sec (in the case of normal configurations where a gas gives off heat to a wall the coefficients α are usually [3]) in the order of 10^{-3} cal/cm^2 degree sec). The time lag τ is from 1 to 1.5 $\times$ 10^{-4} sec, corresponding, at an engine speed of 3000 r.p.m., to a phase angle of from 1.5 to 2.5 degrees, a correction which in our case may reasonably be ignored, so that *a fortiori* the variations of α are of no consequence.

[3]) The difference agrees well with the power —0.7 in the relation between α and d. It has been directly proved by experiments that for our extremely fine filaments the $d^{-0.7}$ law is still valid.

It will now be understood why we do not work with thermo-elements, which are generally used for measuring wall temperatures of cylinders and the like. The joint of a thermo-element, suitably made, may well allow of a somewhat more accurate localization of the measuring point than is possible with a resistance wire, but the thermal inertia of the joint cannot be reduced to the same extent as that of our wires.

Incidentally it may be pointed out that the great increase of α with decreasing thickness of the wire was one of the decisive factors in the development of the modern gas-filled incandescent lamps. The long and thin tungsten wires answering to the desired voltage and the desired output involve in themselves large losses in the form of heat given off to the gas. By coiling the filament it is made equal, as far as heat transfer is concerned, to a wire of a thickness corresponding to the diameter of the coil, so that the thermal losses are reduced to a fraction of what they were originally.

With such fine metal wires the temperature is equalized over the whole cross section of the wire so quickly (within about 10^{-8} sec) that even with the highest engine speeds the temperature distribution over the cross section may be regarded as uniform. Thus it is possible to calibrate the set-up statically. By giving the resistance wires a length at least 2000 times their thickness (i.e. in our case at least 4 mm) it is possible to reduce also the measuring error due to lack of uniformity in the distribution of the temperature over the length of the wire (owing to conduction of heat to or from the electrodes between which the wire is stretched), so that for our purpose it may be ignored.

The temperature of the wire could be influenced to a certain extent also by the (constant) temperature of the walls of the cylinder owing to heat exchange through radiation, but with the intensive exchange of heat through convection between the flow of gas and the very thin wires this effect becomes negligible; there is, therefore, no need to counteract this by screening, not even at the highest temperatures at which measurements have been taken (about 600 °C).

The bridge current must, of course, be sufficiently low to avoid any perceptible heating of the resistance wires; a current of a few milliamps was found to be quite permissible.

Calibration of the oscillograms

Once an oscillogram of the temperature fluctuations at a point in the engine has been recorded, this has to be calibrated. That is to say, from the vertical deflections the absolute temperature values have to be deduced and each point of the curve has to be correlated to the right phase of the cyclic process. These two operations are carried out in one step, in the following way.

Before being mounted in the engine each resistance wire is statically calibrated with the aid of an adjustable furnace: the bridge is balanced for a number of different, known, wire temperatures by adjusting one of the bridge resistors, so that it is known what gas temperature corresponds to each value of this resistance. Mounted on the shaft of the engine under test is a mechanical interrupter (one working with a photocell can also be used) — see fig. 1 — which in one complete revolution, thus in one cycle of the oscillogram, cuts out the bridge current twice, or if necessary more frequently, for a brief moment and thus momentarily reduces all bridge voltages to zero. As a result, in the oscillogram there are two discontinuities where the spot returns for a moment to the level corresponding to the zero point of the diagonal voltage of the bridge (for the bridge voltage to drop quick enough to zero the capacitances of the arms of the bridge must be sufficiently low). The bridge balance is varied by adjusting the resistance previously mentioned in such a way that the curve in the oscillogram is raised (or lowered) just sufficiently to cause one of the discontinuities to disappear. At that moment of interruption — the phase of which in the cyclic process is exactly known — the diagonal voltage is then equal to zero also without the interruption, and from the adjustment required the temperature at that instant can be found. After having applied this procedure to both the jumps we have sufficient data from which to plot a complete phase and temperature scale in the oscillogram — at least provided the deflection of the spot is a linear function of the bridge voltage; if this is not the case within the desired accuracy of a few percent then we must use more interruptions of the curve.

In *fig. 2* we have two temperature curves recorded in the manner described. These have been obtained from measurements taken on an experimental single-cylinder air engine. *Fig. 3* is a schematic cross-sectional drawing of the engine and shows the various points where thermometer wires were mounted. The measuring points from which the oscillograms *a* and *b* were obtained are denoted by the corresponding letters. From the oscillograms the temperature can be derived with an accuracy of about 1 °C and the phase with an accuracy of a few angular degrees, which was more than sufficient for our purpose. From the shape of the curves important conclusions could be drawn as to the instantaneous value of the heat transfer at various points in the heat exchangers, the mixing of the gas in the cylinder and the heat transfer between the gas and the walls of the cylinder.

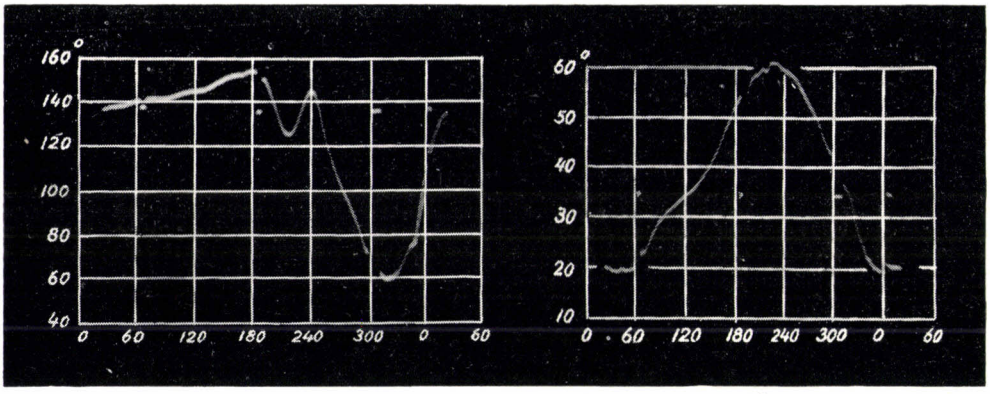


Fig. 2. Oscillograms showing the variations of the gas temperature at two places in an air engine; these points (*a*, *b*) are indicated in fig. 3. The phase angle 0° corresponds to the moment at which the volume of the hot cylinder space is greatest.

Mounting the wires

In conclusion something may be said about the manner in which the wires were mounted in the engine. Each wire is affixed to its electrodes with a minimum quantity of solder; for measuring points where high gas temperatures occur the soldering is done with silver, using a burner with a very fine flame. The electrodes have to be insulated and the opening in the cylinder wall through which they are passed has to be sealed air-tight. A successful method is to use small metal tubes of 1.6 mm internal diameter containing an electrode rod of 1 mm diameter insulated with magnesium oxide pressed in between. Two of such tubes are soldered into borings in a flange. The Wollaston wire is then mounted between the electrodes, etched to remove the silver coating and then statically calibrated in the manner described. Finally the flange with the

electrodes is fixed air-tight over an opening made in the wall of the cylinder (see *fig. 4*).

It may seem astonishing that it should be possible in practice to work with wires no more than 1.5 to 2 microns thick, which are almost invisible to the naked eye. What is perhaps most surprising, however, is that it has also been possible to measure with such a wire the fluctuating temperature of the gas in the thin layer near the surface of the moving piston. For this test the filament was fixed to the piston and thus moved up and down with it (*c* in fig. 3). The electrodes were passed through the body of the piston and connected to the bridge circuit quite simply by means of flexible, coiled wires connected to fixed, insulated electrodes outside the cylinder. This apparently so fragile system proved to be of sufficient mechanical strength to stand up against several millions of cycles of the engine. Only now and then did a thermometer wire come to an untimely end owing to minutely small droplets of lubricating oil floating in the cylinder being deposited on the wire and destroying it through their inertia.

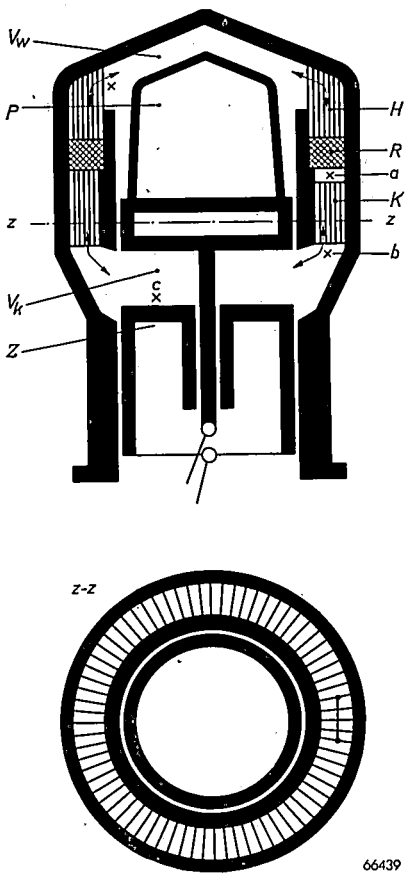


Fig. 3. Simplified cross-sectional drawing of a single-cylinder air engine with displacer piston (see fig. 7 in the article by F. L. van Weenen, The construction of the Philips air engine, Philips Techn. Rev. *9*, 125-134, 1947). V_w is the hot space, V_k the cold space, *H* the heater, *R* the regenerator, *K* the cooler, *Z* the working piston, *P* the displacer piston (which transfers the air expanded in V_w via *H-R-K* to V_k and after compression in V_k back again, via *K-R-H*, to V_w). Thermometer wires were mounted at the points marked with a cross. The orientation of the filaments *a* and *b* is given in the horizontal cross section (*z-z*). These wires yielded the oscillograms of the variation of the gas temperature as reproduced in figs 2*a* and *b*. In order to mount the wire *a* it was necessary to remove a small portion of the cooler, so that the dead space was thereby slightly increased.

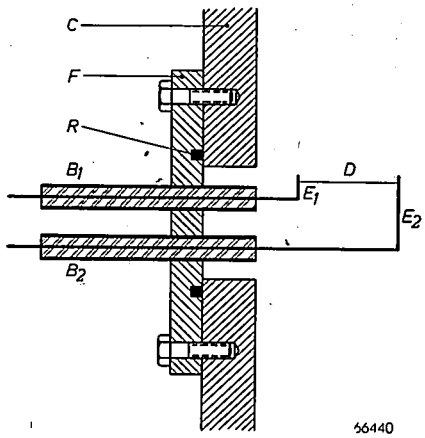


Fig. 4. Insertion of the electrodes E_1 and E_2 for a thermometer wire (*D*) through the wall (*C*) of the engine cylinder. B_1, B_2 metal tubes with internal insulation of magnesium oxide, *F* flange, *R* gasket to make an air-tight seal.

Summary. In the cylinders of high-speed thermal engines, where one cycle lasts only a few hundredths of a seond, very rapid fluctuations take place in the temperature of the gas, which fluctuations are of great importance for the functioning of the engine but are almost impossible to calculate. It has been found a practicable possibility to measure such rapid temperature fluctuations with the aid of an extremely fine metal wire connected as a resistance thermometer to an electric bridge. This wire, of a thickness of from 1.5 to 2 microns, responds to the variations in the temperature of the gas with a time-lag of only 1 to 1.5 × 10^{-4} second. The fluctuations are recorded with the aid of an oscillograph: the oscillograms can be calibrated in a simple manner. The error in the measurement of the temperature is about 1 °C, while the error in the phase angle is at most a few degrees. As an example it is described how measurements were made on an experimental single-cylinder air engine.

VOL. 13 No. 5, pp. 109-144 NOVEMBER 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

A NEW HIGH-PRESSURE MERCURY LAMP WITH FLUORESCENT BULB

by J. L. OUWELTJES, W. ELENBAAS and K. R. LABBERTÉ.

621.327.312 :535.37

In the last decade the use of fluorescent substances has brought about a revolution in illumination engineering, the principal example of which was the low-pressure mercury lamp with fluorescent envelope (the "TL" lamp). Before the triumphal advent of this low-pressure lamp, however, the idea of applying fluorescent substances for the purposes of illumination had already materialized in the high-pressure mercury lamp. It is interesting to give thought to the differences in the problems encountered with these two kinds of lamps. Thanks to the development of new materials and methods, the high-pressure mercury lamp with fluorescent bulb has now also an opportunity to take an important place in the field of illumination, inter alia for industrial and street lighting.

The HP lamp

The principle of the electric discharge in mercury vapour of low pressure was applied several decades ago in tubes for illuminated advertising signs. For lighting purposes, however, this could not be applied because of the blue colour of the light and the very low efficiency of the light source in lumens per watt; the radiation produced by the discharge consists for about 80% of ultra-violet with a wavelength of 2537 Å, only about 5% being in the visible part of the spectrum.

About 1935 a lamp was made which was based on a discharge in mercury vapour under high pressure [1]) (e.g. 5 atm or higher). This meant a great step forward in the direction of applying such discharges for lighting purposes. The high-pressure mercury lamp (HP lamp [2])) has a radiation spectrum as represented in *fig. 1*. A large part of the energy of this lamp is also radiated in the ultra-violet part of the spectrum, the main wavelengths being 2537 Å and 3650 Å, but owing to the high pressure there is at the same time a continuous spectrum extending over the whole of the visible range; again due to the high pressure, the spectral lines in the visible range are also of a relatively greater intensity. The result is that quite a reasonable efficiency is obtained, viz. 35 to 40 lumens/watt. It is true that the colour of this light is still rather bluish and the colour rendering leaves much to be desired, but the high efficiency compared with the incandescent lamp made the use of these high-pressure mercury lamps so attractive that they soon found application in various cases, sometimes in combination with incandescent lamps.

Fig. 2 shows the construction of the HP lamp in its present form. The discharge takes place in a small tube of fused silica containing a certain amount of mercury together with a rare gas under a pressure of a few cm Hg. This rare gas and the provision of an auxiliary electrode enables the lamp to be started with normal mains voltage (220 V~). The heat generated by the discharge causes the mercury to evaporate and the pressure in the tube rises to a few atm (according to the type of tube). The discharge tube is mounted in a bulb of the same shape as that of an incandescent lamp

[1]) W. Elenbaas, Discharges in mercury vapour under high pressure (in Dutch), De Ingenieur **50**, E83-90, 1935; C. Bol, A new mercury lamp (in Dutch), De Ingenieur **50**, E91-92, 1935: W. de Groot The emission and absorption spectra of mercury vapour at very high pressures (up to 300 atm) (in Dutch), De Ingenieur **50**, E92-94, 1935. See also G. Heller, The mercury vapour lamp HP 300, Philips Techn. Rev. **1**, 129-134, 1936.

[2]) Sometimes the HP lamps, together with the SP lamps, which are water-cooled and in which pressures of about 100 atm occur, are denoted as "super-high-pressure" mercury lamps, the name of "high-pressure" lamp then being reserved for lamps with a mercury pressure of about 1 atm (HO lamps).

and filled with nitrogen. The gas jacket thus formed round the discharge tube has a two-fold function. In the first place it protects the current leads fused in the silica, which are heated to a very high temperature, so that they would rapidly oxidize at the end where they come into contact with air. In the second place the nitrogen affords a certain thermal insulation: if the tube were exposed to the open air the thermal dissipation through convection would be so great that upon the current being switched on the tube would only very slowly heat up and the mercury would take a long time to evaporate completely, if at all. (A more complete insulation, by evacuating the bulb — in which case the only dissipation of heat would be by radiation —, is not desirable because then, in order to avoid the wall of the tube reaching too high a temperature, only a smaller power can be allowed to develop in the discharge tube.)

Some other details of the construction are explained in the subscript to fig. 2. We would only mention here — because they will be referred to again later — the strips of molybdenum, serving to carry the current through the fused silica, and the inside frosting of the bulb which hides the inside of the lamp from view and gives the lamp a pleasant appearance when it is not burning while at the same time providing for a not too great surface brightness when the lamp is alight.

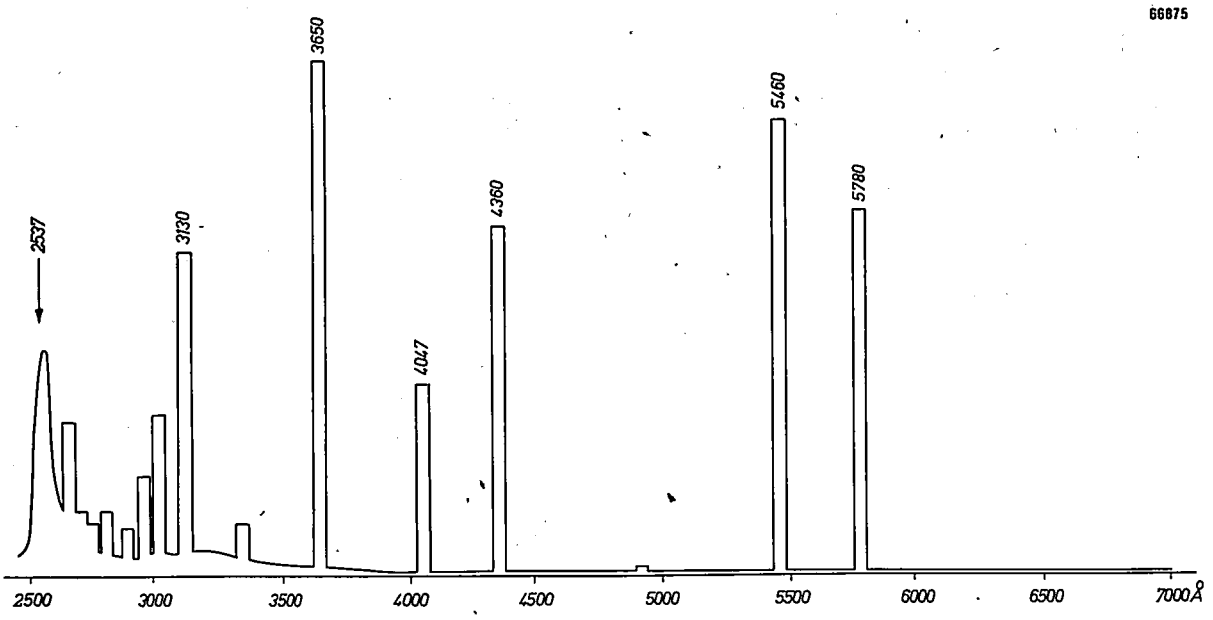


Fig. 1. The spectral energy distribution of the HP lamp (without glass bulb). The spectral lines superposed on the continuum have all been drawn relatively too low and with the same exaggerated width, such that the area covered by each line represents the energy emitted therein. About 1½ times as much energy is radiated in the ultraviolet as in the visible part of the spectrum.

Possibilities for improving the colour rendering of the HP lamp

Considering the high efficiency and other good properties of the HP lamp, there was every inducement to investigate whether the colour rendering could not be so improved as to make these lamps suitable for more universal use. The lack of colour rendering is most noticeable with objects the colour of which is predominantly red: under the HP lamp such objects assume a brownish hue. This can be understood from a look at the spectrum in fig. 1. In the red, i.e. roughly between 6000 Å and 7000 Å, the mercury vapour has no spectral lines of any intensity worth mentioning, and also the continuous spectrum of the HP lamp is weak in that range. The radiation of the HP lamp therefore needs supplementing in the red.

Attempts have been made to attain this by various means: (1) By adding a little cadmium to the mercury vapour. It is true that the cadmium vapour does to a certain extent yield the desired contribution in the red, but this artifice is accompanied by a considerable lowering of the efficiency and a shortening of the life of the lamp. (2) The choke (or low power factor transformer), usually connected in series with the mercury discharge in order to stabilize it, was replaced by a filament mounted together with the discharge tube in one bulb. The incandescent light from the filament, being very rich in the red, supplemented

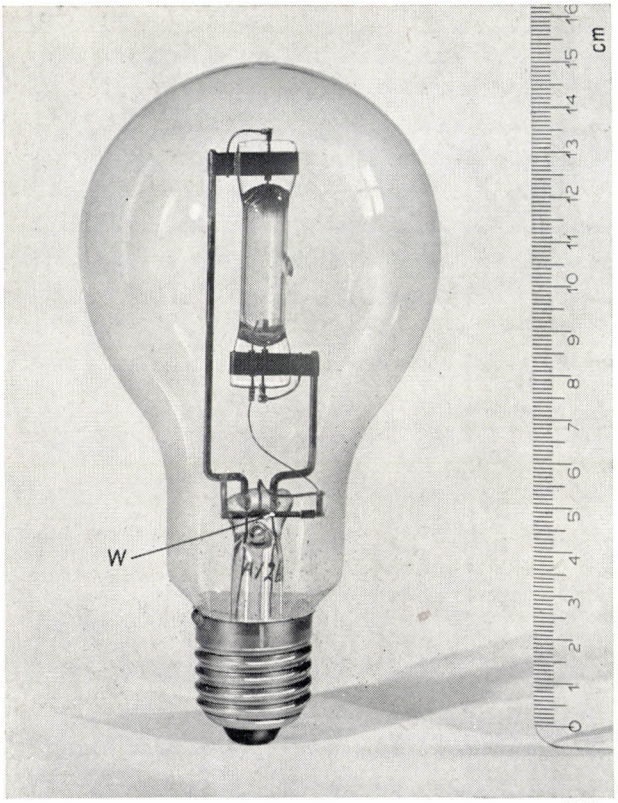


Fig. 2. The HP lamp. Inside, in the centre of the bulb, is the tube of fused silica in which the discharge takes place in mercury vapour of high pressure. The current is fed to the electrodes via leading-in strips of molybdenum (almost concealed by the supports of the discharge tube) sealed airtight in the tube. Beside the lower main electrode is an auxiliary electrode connected to the other main electrode via a current-limiting resistor *W* and serving to facilitate ignition (ignition voltage about 180 V). Usually the bulb is frosted on the inside (here it has been left clear to show the mounting inside). It is filled with nitrogen under a pressure of 50 cm Hg (in the cold state).

the radiation from the mercury discharge so as to give it a white light with good colour rendering. The contribution of light needed from the filament, however, is of such a high percentage that the efficiency of these blended-light lamps, named ML lamps, is greatly inferior (about 20 lm/W) to that of the HP lamp. (3) The most elegant solution is that whereby the otherwise useless ultra-violet radiation from the mercury vapour is converted by means of a fluorescent substance into the red radiation desired. In this way the efficiency in lumens/watt need not by any means suffer but may in principle even be made better than that of the HP lamp. This line of thought led to the construction of the HPL lamp, which is a high-pressure mercury lamp having a bulb coated on the inside with a fluorescent layer.

[3]) E. L. J. Matthews, The blended-light lamp and other mercury lamps with improved colour rendering, Philips Techn. Rev. **5**, 341-347, 1940. J. Funke and P. J. Oranje, The development of blended-light lamps, Philips Techn. Rev. **7**, 34-40, 1942.

Before dealing more closely with this idea and the manner in which it was materialized, further development will be briefly outlined.

The first HPL lamps brought onto the market about 1937 were far from being ideal as regards colour rendering. In those days, however, when the art of producing fluorescent substances (phosphors) was still more or less in its infancy, we had to consider ourselves fortunate in having found a phosphor which could be excited by the radiation of both 2537 Å and 3650 Å and which, moreover, answered fairly well the other requirements set (about which more will be said later). It is to be realized that once thoughts had turned to the use of fluorescent substances it was only obvious to try to make use of them also for the low-pressure mercury lamp for converting the lost radiation of 2537 Å into visible light. Such could not be done at that moment because no phosphors were then known which satisfied the entirely different requirements for that lamp.

As is known, some time later however phosphors were also found which were suitable for the low-pressure lamp, and thus the tubular fluorescent lamps of the "TL" type made their appearance. These very soon became enormously popular, so much so in fact that the previously developed but still immature HPL lamp was relegated to the background. However, investigations with the HPL type of lamp were continued, though for a long time without any great success. It was not until in the course of last year that a phosphor was discovered which has made a decided improvement of the HPL lamp possible. The new type of HPL lamp that is now in production promises to take up a field of application of its own, side by side with the so popular "TL" lamp.

What is required of a phosphor for the HPL lamp and what is to be expected from it

As already indicated, to improve the colour rendering of the HP lamp a phosphor is needed which is excited by radiation of wavelengths in the range of 2537 and 3650 Å and which gives an emission spectrum lying mainly in the red.

Further, the phosphor has to satisfy the following requirements:

1) It must not absorb the visible light of the mercury discharge to any appreciable extent, so as not to affect the efficiency of the lamp. A strong selective absorption is also undesired; apart from the coloured appearance it would give the lamp while it is not burning, which effect in some cases

may be unpleasant, selective absorption might again spoil the colour rendering.

2) Although any phosphor has the tendency to lose its fluorescent power as the temperature increases, this phosphor must still give a reasonably strong fluorescence at a temperature of 150 °C to 200 °C, this being approximately the temperature reached by the bulb that is to be lined with the fluorescent layer. Of course the temperature of the bulb can be reduced somewhat by making the bulb larger, and in fact a step has been taken in that direction (the diameter of the old 120 W HPL 500 lamp was 130 mm, as compared with the 90 mm of the HP 500), but there is a limit to this because the fittings in which the HPL lamps are to be used should not be too large.

source is described by quoting the percentage of the total luminous flux supplied in each of the eight adjacent spectral "blocks" covering together the whole of the visible spectrum.

In *table I* figures are given for the HP lamp and for daylight; it should, however, be noted that these figures represent energy contributions and not contributions of luminous flux. The qualitative differences between the sources of light as considered in each block are, of course, the same in both cases, and by quoting the energies it is possible to give in the same table also the contributions in two ultra-violet blocks. Now these ultra-violet energy contributions of the HP lamp are to be imagined as being transferred to the red blocks 6100 Å - 6600 Å and 6600 Å - 7200 Å, where the radiation

Table I. Block-diagram distribution of the spectrum of the HP lamp and of average daylight. Contrary to the usual custom the relative energy contributions in the blocks are given (slightly rounded off), instead of the relative luminous flux contributions. Also the ultra-violet energy in two spectral sections is given, in the same relative measure as the visible radiation.

	< 3000 Å	3000-4000	4000 No. 1 4200	2 4400	3 4600	4 5100	5 5600	6 6100	7 6600	8 7200
HP 125 W	60	80	13	24	1	2	31	26	1,5	1,5
Average daylight	0	9	6	6	7	18	17	16	14	16

3) Finally there is the very important requirement that the phosphor must not be subject to dissociation under the strong ultra-violet irradiation (the photo-chemical action of which is further promoted by the high temperature). Anticipating what is stated farther on in this article, it may be said here that for many substances which otherwise appeared to be suitable for our purpose this requirement proved to be the stumbling block: the photochemical dissociation occurring with the substances first becomes noticeable in a reduction of the content of red light in the radiation from the lamp; later the fluorescent layer often turns grey, thus absorbing visible light and causing the total luminous flux of the lamp to diminish considerably.

Supposing that a phosphor were found which ideally answers these requirements, the question is what result is to be expected from it as regards the colour rendering of the lamp. This can be calculated in a simple manner. The colour rendering of a lamp is most easily judged with the aid of the "block diagram" devised by Bouma [4]). According to this method the spectral distribution of a light from the HP lamp shows such a marked deficit. In this transference account has to be taken of the depreciation in energy taking place in any process of fluorescence: for one quantum of ultra-violet energy $h\nu_1$ (given a quantum efficiency of 100%) at most one energy quantum $h\nu_2$ of the changed wavelength can be obtained, and since a red quantum contains only about half the energy of one of our ultra-violet quanta ($\nu_2 \approx \frac{1}{2}\nu_1$) in the transference the energy has to be halved. In each of the two red blocks there is then a relative energy contribution of about 35. In reality, of course, a quantum efficiency of 100% cannot be reached, and, moreover, to ensure sufficient transmission of the visible light from the mercury discharge, the phosphor can only be applied in such a thin layer that the ultra-violet radiation is not completely absorbed. The conclusion, therefore, is that in the most favourable case possible of realization the colour rendering of daylight in the red can just about be reached [5]). The improvement that can be attained appears to be indeed quite sufficient for the red colour rendering, provided — and this is the lesson

[4]) P. J. Bouma, Colour reproduction in the use of different sources of "white" light, Philips Techn. Rev. **2**, 1-7, 1937. See also P. M. van Alphen, A photometer for the investigation of the colour rendering of various light sources, Philips Techn. Rev. **4**, 66-72, 1939.

[5]) For this argument we are indebted to Dr. J. Voogd, and for the necessary connections between energy measurements in the ultra-violet and in the visible spectrum we have to thank J. Riemens.

to be learned from the foregoing — the fluorescence is utilized as far as possible to augment the red content of the radiation.

Phosphors tried out for the HPL lamp

The phosphor which until recently had always been used in HPL lamps is zinc cadmium sulphide, activated with copper, (Zn, Cd)S-Cu. The spectral energy distribution of the emission from this phosphor is represented in *fig. 3*, from which

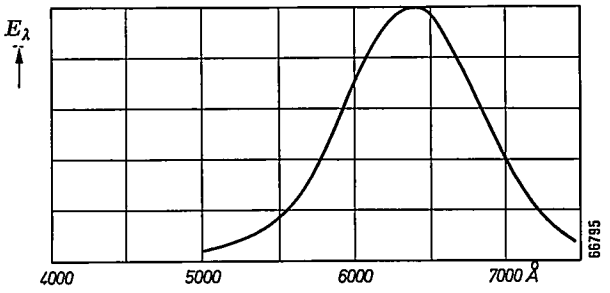


Fig. 3. Spectral energy distribution (in arbitrary units) of the emission of zinc cadmium sulphide activated with copper.

it is seen that there is a good maximum in the orange-red. Further, the "excitation spectrum" is not unfavourable; the phosphor is highly absorbent for all the ultra-violet radiation that is to be considered and it fluoresces with a reasonable quantum efficiency: upon activation with 3650 Å it is about 50% and with 2537 Å about 25%, measured at room temperature. At the working temperature (about 175 °C) these values are only slightly lower. The phosphor is also sufficiently durable during the life of the lamp. A serious drawback, however, is that this zinc-cadmium-sulphide phosphor is highly absorbent also for the blue radiation (an extension of the ultra-violet absorption band, which in itself is indispensable). Instead of a lack of red the light from the burning lamp now has a deficit in blue and is greenish in hue. Consequently under this old HPL lamp all practically white objects assume an unpleasant appearance, and this applies in particular to the human skin.

Various means have been tried in an attempt to remedy this failing of the zinc-cadmium-sulphide phosphor. Attemps were made to reduce the absorption of blue by applying the phosphor to the bulb in the form of a finer powder, but this appeared to be impossible without deteriorating the phosphor. For a long time the zinc-cadmium-sulphide phosphor used in the old HPL lamps was also replaced partly by a blue-fluorescing phosphor (ZnS-Ag), but this solution was not entirely satisfactory and larger bulbs had to be used on account of the admixture being rather sensitive to temperature.

Thus right from the beginning there were reasons for looking for entirely different fluorescent substances suitable for correcting the colour of the light emitted by the HP lamp. This became all the more imperative with the advent of the "TL" lamp, not only because this lamp with its high efficiency and excellent colour rendering made it necessary to lay down more stringent standards, but also because it had thereby been learned that phosphors exist with a quantum efficiency up to about 80%, as compared with the 50% and 25% for the zinc-cadmium-sulphide phosphor.

Of the various phosphors which subsequently became available and which were tried out for use in the HPL lamp four will be discussed here, viz. cadmium borate, calcium phosphate, the group of halophosphates and a group of phosphors with quadrivalent manganese as activator.

Cadmium borate, with divalent manganese as activator, absorbs only the short-wave ultra-violet; thus the emission near 3650 Å is unused. The emission spectrum, however, is very favourable, as is shown by curve *a* in *fig. 4*, and there is no absorption of visible rays, so that with this phosphor quite acceptable lamps could be made, be it with a somewhat larger bulb because the luminescence is considerably weaker even at 150 °C. Unfortunately there was such a great decline in luminescence during the useful life of the lamp that this phosphor could not be considered for practical use.

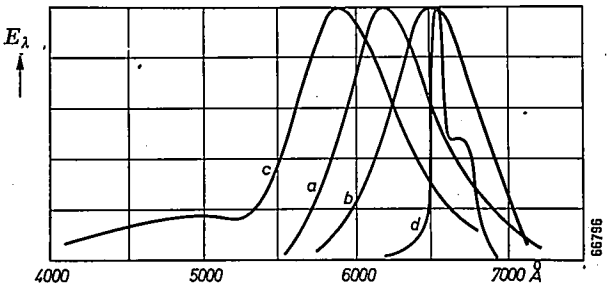


Fig. 4. Spectral energy distribution of four phosphors tried out for their suitability for use in the HPL lamp (maxima all taken to be equally high):
a) Cadmium borate activated with divalent manganese.
b) Calcium phosphate activated with cerium and manganese.
c) Halophosphates activated with antimonium and manganese.
d) Magnesium titanate activated with quadrivalent manganese.

Much the same applies also to calcium phosphate (activated by cerium and manganese). This substance absorbs only the short-wave ultra-violet; in daylight it is white and it shows a beautiful red fluorescence (fig. 4*b*) with a fairly good quantum efficiency

(50%). In regard to temperature-sensitivity this phosphor is even better than cadmium borate, the efficiency remaining absolutely constant up to about 300 °C. But here again no remedy could be found for the rather rapid decline in luminescence during the useful life of the lamp.

The third kind of phosphors mentioned, the halophosphates, are of the general formula $3Ca_3(PO_4)_2CaX_2$, where X stands for chlorine or fluorine. In this case the activators are antimony and manganese. These phosphors have gained much in practical importance through their application in "TL" lamps. Essential for this application is a strong luminescence upon excitation by 2537 Å and sufficient resistance against the dissociating effect of these rays, as also against the effect of the mercury vapour, with which the fluorescent layer in "TL" lamps comes into contact. The halophosphates possess the above-mentioned properties and also retain their fluorescence at high temperatures (this is a property that is not really essential for "TL" lamps because the wall of the "TL" lamp does not reach a temperature higher than 40 °C to 50 °C); further, they are also resistant against dissociation at the high temperatures prevailing in an HPL lamp. They do not, however, answer the specific emission requirements for the HPL lamp: at the best only a spectral distribution can be obtained as drawn in fig. 4*c*, with a maximum at about 5900 Å. This is not sufficient for improving the red content of the radiation from the HP lamp.

Finally there are the substances containing quadrivalent manganese as activator, several of which have been throughly investigated in the Philips Laboratory at Eindhoven [6]). All these show a deep red fluorescence. Magnesium titanate, which belongs to this group of phosphors, appeared to offer the best prospects as regards suitability under the conditions occurring in the HPL lamp. The emission spectrum is given in fig. 4*d*. The excitation spectrum, unfortunately, is less satisfactory, for, though the rays with wavelength 2537 Å are well absorbed, the quantum efficiency of the conversion of those short-wave rays is practically nil. Attempts were made to remedy this evil by causing the ultra-violet HP radiation to fall first upon a layer of some other suitable phosphor (e.g. cadmium borate) which absorbs only the short-wave rays and converts them into red. However this idea has not been put into practical execution because magnesium titanate also has the drawback of its fluorescence being too sensitive to temperature: above 100 °C it diminishes rapidly.

Magnesium arsenate

Although magnesium titanate, belonging to the group of Mn^{4+} phosphors, did not answer the requirements, it was this very same group which yielded the phosphor with which the problem of the HPL lamp was ultimately solved, viz. magnesium arsenate [7]).

Having already become familiar with the problems of preparing Mn^{4+} phosphors and with the characteristics of these substances, it was very soon possible, once attention had been drawn to it, to prepare this new phosphor in a form suitable for application to lighting technique.

The properties of this magnesium arsenate [8]) proved to be highly favourable for its application in the HPL lamp. Its fluorescence is bright red.

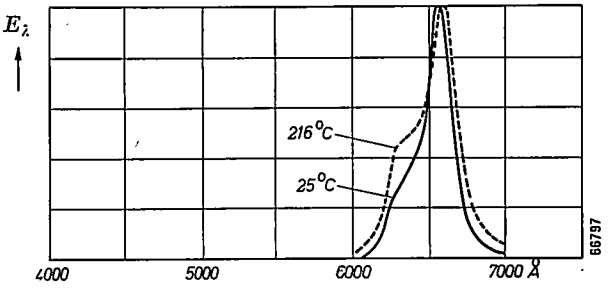


Fig. 5. Spectral energy distribution of magnesium arsenate activated with quadrivalent manganese, at 25 °C and at 216 °C (maxima taken to be equally high).

The spectral distribution of the emission (*fig. 5*) shows a steep peak at 6560 Å. This phosphor absorbs the whole of the ultra-violet radiation spectrum of the HPL lamp and has a quantum efficiency of about 75 % when excited either with 2537 Å or with 3650 Å (and also with intermediate wavelengths). There is a slight absorption of blue light — considering the lack of any sharp demarcation of the absorption bands of solids this absorption of some blue is inevitable if the whole of the long-wave ultra-violet is to be absorbed — but the resultant faintly yellow colour of the magnesium arsenate in daylight is not troublesome, and, contrary to the case with the old zinc-cadmium-sulphide lamps, there is no question of any green coloration of the resultant light from the HPL lamps in which this new phosphor is used. Another very favourable property of magnesium arsenate

[6]) F. A. Kröger, Some Aspects of the Luminescence of Solids, Elsevier Publ. Co., Amsterdam, 1948, in particular pp 57-106.

[7]) M. Travnicek of Graz (Austria) brought this substance to our notice.

[8]) The investigations into the properties of the new phosphor, as also the first investigation into the conditions for obtaining a high efficiency from it, were carried out in the Philips Laboratory by Dr. F. A. Kröger.

is the fact that it shows very little sensitivity to temperature, the quantum efficiency still being about 65% at 300 °C (*fig. 6*).

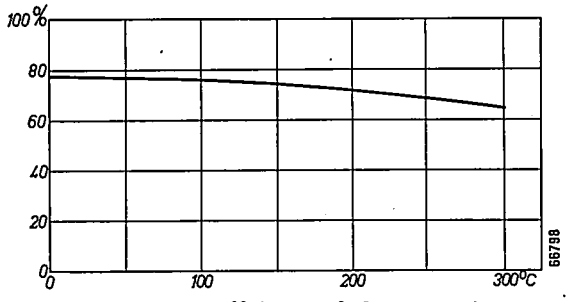


Fig. 6. The quantum efficiency of the magnesium arsenate phosphor upon excitation with 2537 Å, as a function of the temperature.

Chemical behaviour of magnesium arsenate in an HPL lamp

After this summing up of the excellent properties of magnesium arsenate it seems like an anti-climax to have to state that in the beginning this phosphor, like so many others, failed to satisfy the last requirement, namely that of durability. The red content of the light from the first HPL lamps made with magnesium arsenate declined rapidly.

Contrary, however, to the case with other phosphors which came to grief over this stumbling block, a simple means was found of improving the durability of magnesium arsenate. All that was necessary was to replace the nitrogen filling of the bulb of the HPL lamp by carbon dioxide. It is a most remarkable phenomenon that then during the first 20 to 50 burning hours the red content of the light increases, sometimes by a factor of as much as $1^1/_2$, after which it gradually decreases, but after 2000 burning hours it is still about 80% of the peak value.

Though the mechanism by which CO_2 improves the resistance of the magnesium-arsenate phosphor has not yet been fully explained, it was by fairly logical reasoning that the choice of CO_2 came to be made.

Magnesium is one of the basest metals (i.e. difficult to reduce). It was therefore probable that the decline in the fluorescence of the phosphor is not due to any separation of metal (in this case Mg), as was found to be the case with the other phosphors, but that it is to be ascribed to a reduction to a lower valency of the small amount of Mn^{4+} present (for the best yield of light a composition something like 10 $MgO{\cdot}As_2O_5{\cdot}0.01$ Mn is chosen). Chemical analysis confirmed the supposition that the decline in fluorescence of the phosphor is accompanied by a reduction from Mn^{4+} to Mn^{2+}.

Experiments were therefore made with a small addition of oxygen to the nitrogen filling of the bulb of the lamp. It then appeared that (at a bulb temperature of 150 °C to 200 °C) with the addition of 0.3 % to 0.5 % of O_2 and under a nitrogen pressure of 50 cm Hg there was indeed no longer any decline at all in the red content of the light.

However, with such an admixture of oxygen another difficulty arose, viz.: the extremely thin leading-in strips of molybdenum which are fused airtight in the silica for carrying the current through the "pinches" at the ends of the mercury-discharge tube (*fig. 7a*; see also fig. 2) tend to oxidize in an atmosphere containing O_2, thereby expanding and after some time causing the pinches to crack. The higher the temperature at the outward extremity of the molybdenum strip, the more likely is this to occur. With the dimensions of the pinches usual in HP lamps this temperature amounts to about 500 °C, and with a partial oxygen pressure of 0.15 cm Hg (see above) the pinch cracks even after the lamp has been switched on for some tens of hours only.

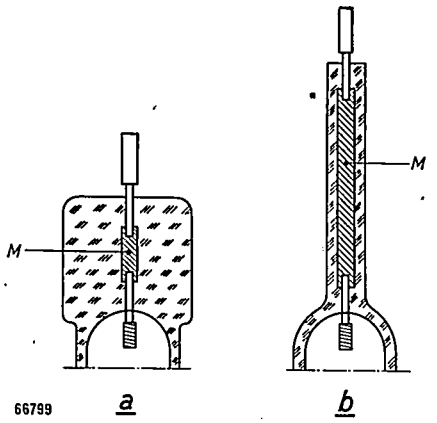


Fig. 7. Construction of the pinch of the discharge tube in the HP and HPL lamps. When the lead is being fused in, it has to be heated up to the softening temperature of the fused silica, i.e. 1600 °C to 1700 °C. Of the few metals capable of withstanding this temperature molybdenum is the most suitable. Since, however, the thermal expansion coefficient of molybdenum is much greater than that of silica, instead of a leading-in wire, use is made of a strip of molybdenum *M*, rolled out extremely thin (13 μ thick) ,which, being readily deformable, adheres to the fused silica in spite of the differences in expansion. *a*) Pinch of normal length (drawn to approximately true size). *b*) The pinch made longer so that the outward extremity of the leading-in strip does not become so hot as in (*a*), thus reducing the rate of oxidation of the molydenum when oxygen is added to the nitrogen filling of the bulb (the round leads are not sealed airtight in the fused silica and hence oxygen can penetrate from the outside of the tube and reach the molybdenum).

This can be remedied by making the pinch longer (fig. 7*b*). Of course oxidation then still takes place, but at a slower rate because the temperature at the danger point is lower. In this way the durability of the lamp is no longer determined by that of the

pinch construction. HPL lamps made in this manner were quite satisfactory, but their manufacture proved to be too costly because the longer pinches could not be made mechanically; furthermore, the O_2 percentage of the filling is very critical, because if it is slightly too low the yield of light from the phosphor diminishes rapidly and if it is a little on the high side then the leading-in strips oxidize too quickly.

It was then thought that perhaps in itself a partial oxygen pressure much lower than 0.15 cm Hg would suffice to maintain the light yield of the phosphor, and that the "high" filling pressure of 0.15 cm Hg, which proves to be fatal for the molybdenum strips, is only needed on account of the fact that the various metal parts in the lamp selectively "catch" the oxygen and more or less retain it ("getter" action). In that case matters might be improved by doing without the oxygen in the filling of the bulb and using instead an inert gas which, under the influence of the high temperature and possibly also under the influence of the ultra-violet radiation, continuously gives off a little oxygen, just sufficient to prevent the reduction of Mn^{4+}, and too little to cause any serious oxidation of the molybdenum strips. CO_2 seemed to be suitable for this purpose and, indeed, it yielded the excellent results already mentioned.

Confirmation of the line of reasoning followed is found in the fact that the desired effect is obtained even with only a very low CO_2 pressure, viz. about 0.5 cm Hg. In the case of dissociation of the carbon dioxide to a degree of say 5%, which is plausible, according to the formula

$$2\ CO_2 \rightarrow 2\ CO + O_2,$$

the partial oxygen pressure is then no more than about 0.01 cm Hg.

With an oxygen filling the metal parts can take up only a certain amount of O_2. The larger the volume of the bulb, the smaller will be the resultant reduction of the O_2 pressure, and it may therefore be expected that then a lower initial pressure suffices in order to keep the O_2 pressure above the minimum required for the phosphor during the desired lifetime. This expectation was fully confirmed by a test whereby the O_2 volume was increased by connecting the bulb of the HPL lamp to a large reservoir via a small tube.

It appears that the CO_2 pressure is not at all critical. One can safely increase the pressure to 50 cm Hg. such as is usual for the nitrogen filling of the HP lamps. The fact that the molybdenum strips do not then oxidize too quickly is apparently due to the partial O_2 pressure not rising proportionately with the higher CO_2 pressure: the degree of dissociation of the carbon dioxide decreases with the increasing pressure, as follows qualitatively directly from the reaction equation given above (the volume on the right is $1^1/_2$ times that on the left).

Construction and properties of the new HPL lamps

HPL lamps are now being made in two wattages, viz. 80 W and 125 W. *Fig. 8* shows these two lamps and, for comparison, a 150 W "Argenta" incandescent lamp. *Table II* gives the principal data for these lamps, for an HP lamp and for daylight.

As follows from the foregoing general considerations, in aiming at the maximum possible increase of the red content of the light emitted by the HP lamp one's efforts are limited by the fact that for complete absorption of the ultra-violet such thick layers of phosphor would have to be applied that the visible radiation from the mercury discharge would be reduced too much: notwithstanding the gain in the red, the total yield of light from the lamp would then be reduced. Now the thickness of the phosphor layer has been so chosen that the gain in the red and the loss over the entire visible spectrum just about compensate each other; in other words, the efficiency of the new HPL lamp is practically equal to that of the HP lamp.

The red content of the light given in the table and serving as a rough measure for the improvement reached in the colour rendering is the sum percentage of the contributions to the luminous flux in the seventh and eighth blocks. For a better

Table II. Properties of the new HPL lamp, compared with those of an HP lamp, an incandescent lamp and daylight.

	HPL 80 W	HPL 125 W	HP 125 W	150 W incand. lamp (2850 °K)	Daylight
Luminous flux (lm)	3000	5000	5000	2000	—
Efficiency (lm/W)	37,5	40	40	13	—
Percentage of red	8,8	7,5	1,3	18	11
Colour point in the colour triangle. $x =$	0,390	0,385	0,319	0,448	0,320
Colour point in the colour triangle. $y =$	0,434	0,425	0,405	0,407	0,330
Luminance (cd/cm²)	5	8	25	3	—
Diameter of bulb (mm)	80	90	90	80	—

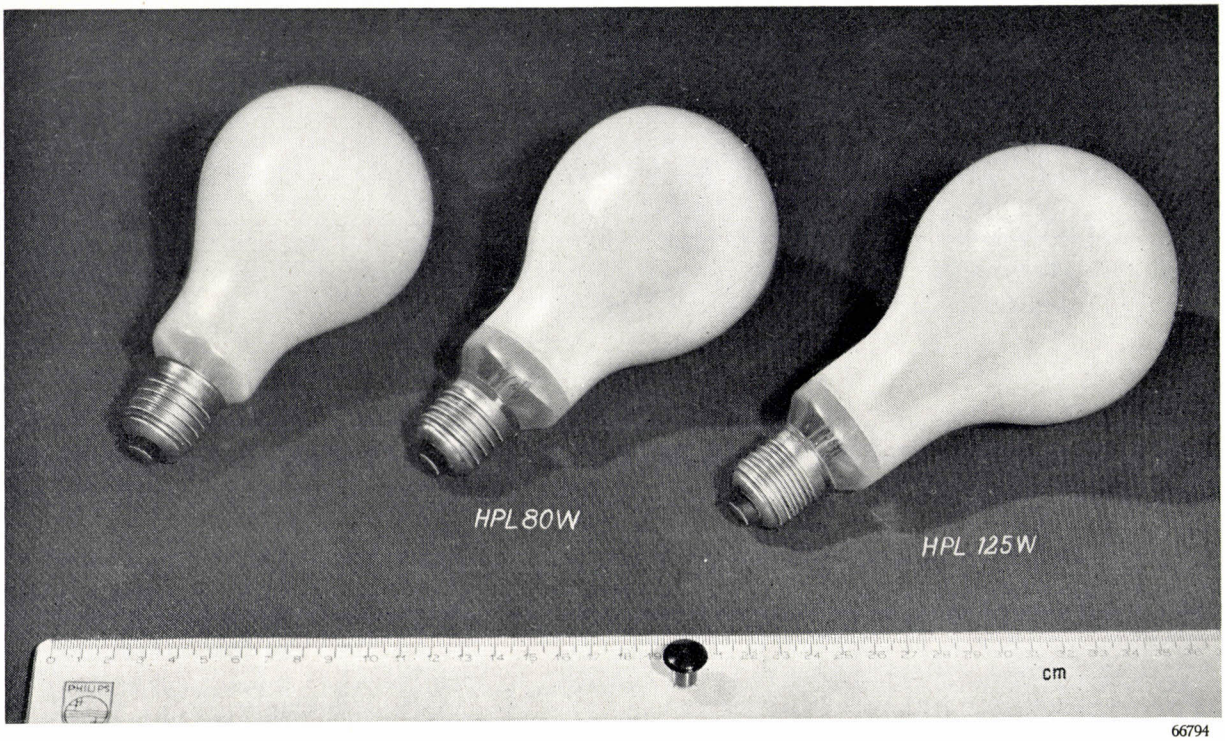


Fig. 8. The new HPL 80 W and HPL 125 W lamps coated on the inside with magnesium-arsenate phosphor. When not alight their colour is practically white. For comparison an "Argenta" 150 W incandescent lamp is also shown.

judgment of the colour rendering *table III* gives the complete block diagram distribution for the radiation of the HPL 80 W and the HPL 125 W, together with the distributions for the HP lamp and for daylight as already given in table I (but there expressed in luminous flux percentages) and also those for the incandescent lamp and for a "TL" lamp ("white" type).

this light. Further, with the aid of inconspicuous fittings the light from the HPL lamp can be just as easily concentrated and directed as that from a frosted or opalised incandescent lamp. For this reason in a number of cases the HPL lamp may be preferred to the "TL" lamp, the application of which for street lighting is now being tried out in various places.

Table III. Block distribution (relative luminous flux contributions) of the spectrum of the new HPL lamp, compared with that of the HP lamp, daylight, an incandescent lamp and a "TL" lamp ("white" type).

	4000	4200	4400	4600	5100	5600	6100	6600	7200
	1	2	3	4	5	6	7	8	
HPL 80 W	0,008	0,37	0,025	0,62	49,8	40,5	8,0	0,76	
HPL 125 W	0,008	0,42	0,04	0,54	47,0	44,4	7,0	0,65	
HP 125 W	0,02	0,76	0,14	1,66	55,1	41,0	1,2	0,09	
Average daylight	0,027	0,23	0,80	10,8	40,8	36,2	10,3	0,73	
150W incand.lamp(2850°K)	0,006	0,060	0,25	5,47	33,4	42,6	16,7	1,57	
"TL" 40 W ("white") . .	0,010	0,35	0,30	4,80	38,0	43,0	13,0	0,43	

An important field of application for the HPL lamp is the lighting of streets and squares in towns, With the high efficiency of these lamps a high level of illumination can be reached in a very economical way and the colour rendering is such that there is no longer any question of the human skin assuming an unpleasant hue under

In this connection something has to be said about the ballasts required for the HP and HPL lamps.

Lamps which are to be installed separately are usually (when a mains voltage of 220 V~ is available) connected in series with a choke (*fig. 9a*). Upon the lamp being switched on, thus while it is in the cold state, a current flows through it

which is about $1^1/_2$ times the normal working current. This starting current is strong enough — with the usual nitrogen or carbon dioxide filling of the bulb under a pressure of about 50 cm Hg — for the discharge tube to be gradually heated and the discharge to reach full strength in 3 to 4 minutes, during which the voltage across the lamp rises and the current drops. The power factor ($\cos \varphi$) of the lamp with choke is about 0.5.

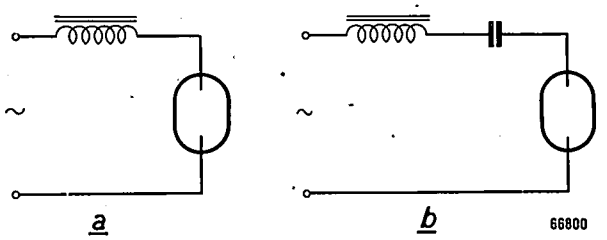


Fig. 9. As ballast for the HP and HPL lamps either a choke (*a*) or a combination of a choke and a capacitor (*b*) can be used. In the latter case the current is about just as much advanced in phase with respect to the voltage as it is lagging in the case (*a*). The starting current immediately after switching on the lamp is much lower in case (*b*) than in case (*a*).

This low power factor is no objection for single lamps, but in street lighting often whole rows of lamps have to be connected to one branch of the distributing network and in such a case a higher $\cos \varphi$ value is desired. This can be obtained by connecting half the number of lamps to a different ballast, viz. a combination of a choke and a capacitor (fig. 9*b*). The current is then about just as much advanced in phase with respect to the voltage as it is lagging in the circuit of fig. 9*a*. This brings the power factor for the whole installation to about 1. In the circuit of fig. 9*b* however the starting current is no stronger than the working current, and therefore, in order to ensure that the lamp will burn at full strength within a reasonable time after switching on, allowing also for severe frost, the discharge tube has to be thermally more insulated, i.e. the pressure of the bulb filling has to be made lower; for such a case a carbon dioxide pressure of about 20 cm Hg is quite suitable.

For this application of the HPL lamp it is therefore very convenient that the pressure of the carbon dioxide is not at all critical for maintaining the luminescence of the magnesium arsenate phosphor.

Summary. The high-pressure mercury lamp (HP) has a very good specific light yield, viz. 35 to 40 lm/W, but the colour rendering under the light of this lamp is poor, mainly due to a deficit in the red. The colour rendering can be greatly improved by converting the useless ultra-violet radiation from the mercury discharge into red light by means of a fluorescent substance. For the HPL lamp developed on this principle the fluorescent substance first used was zinc cadmium sulphide activated with copper. Upon excitation with the ultra-violet rays this substance shows an orange-red fluorescence with a reasonable quantum efficiency, it retains its fluorescence also at the temperature of about 150 °C to which it is exposed when applied to the inside of the bulb of an HP lamp, and while the lamp is alight this substance undergoes only a very gradual dissociation. A drawback however is that it absorbs the blue, so that the resultant colour rendering is nevertheless unsatisfactory. After a number of unsuccessful attempts to find more suitable phosphors, a very good solution has now been reached by employing magnesium arsenate activated with quadrivalent manganese. The quantum efficiency of this bright-red fluorescing phosphor is even appreciably higher than that of the phosphor first mentioned (viz. 75%, upon excitation either with 2537 Å or with 3650 Å) and less sensitive to temperature, whilst the absorption of blue is so slight as to be practically of no effect, so that the colour rendering is quite satisfactory. At first this phosphor was not sufficiently resistant to dissociation, but this has been remedied and excellent resistance obtained by filling the bulb of the HPL lamp with carbon dioxide instead of the nitrogen commonly used for HP lamps. The action of the carbon dioxide in this respect is to be ascribed to a little oxygen being given off which prevents a reduction from quadrivalent to divalent manganese in the phosphor. This effect can also be obtained by the direct admixture of a little oxygen to the normal nitrogen filling, but then there is a prohibitively rapid oxidation of the molybdenum leading-in strips of the discharge tube. The carbon dioxide pressure can be varied between 0.5 and 50 cm Hg, thus making it possible to adjust the thermal insulation of the discharge tube according to the starting current available. This possibility is of practical importance for the application of HPL lamps for street lighting.

THE IMAGE ICONOSCOPE, A CAMERA TUBE FOR TELEVISION

by P. SCHAGEN, H. BRUINING and J. C. FRANCKEN.

621.385.832:621.397.611

The oldest television camera tube, the iconoscope, is now used only for transmitting still pictures (e.g. the signal picture of a certain station) and film pictures. Further development of camera tubes in Europe has followed a course different from that in America. In the U.S.A. the image orthicon has become predominant, whilst in Europe the image iconoscope is widely used. Of the latter there are British and French versions and also one that has been developed in the Philips Laboratory at Eindhoven (Netherlands). This Philips image iconoscope is described here and compared with other camera tubes.

The object of television is to transmit moving pictures via electrical means. This is achieved by "measuring" in succession the brightness of the very large number of picture elements into which the picture to be transmitted is imagined as being divided. This measuring consists in the conversion of the brightnesses into corresponding fluctuations of an electric current which in some way or other govern the signal transmitted.

To be reproduced with a satisfactory definition a picture has to be divided into some hundreds of thousands of elements; let us say, for the sake of convenience, that there are 400,000. Just as in cinema projection, to keep the picture free of flicker the number of pictures produced per second has to be greater than a certain minimum; in Europe the picture-repetition rate is mostly 25 complete pictures per second.

Hence the time available for measuring the brightness of one picture element is only $^1/_{400,000\times 25}$ sec $= 10^{-7}$ sec. Already in 1908 — when 10^7 picture elements per second was beyond the wildest dreams and one had in mind a number something like 160,000 — Campbell Swinton realized that it would be imperative to have a practically inertialess electronic apparatus [1]).

To explain what is meant by this, let us describe the so-called flying-spot method frequently used for televising film pictures.

On the fluorescent screen of a cathode-ray tube (*fig. 1*) a raster of lines is scanned. A lens L_1 focuses this raster on the film picture to be transmitted, and another lens L_2 projects the transmitted light onto a photo-electric cell (a multiplier tube). The latter produces a current varying in strength according to the density of the successively illuminated picture elements on the film. The fluctuating current forms a signal with which the intensity of the electron beam in a cathode-ray tube (picture tube) at the receiving end is modulated. This beam is synchronized with that at the transmitting end. Thus the televised film picture is reproduced on the screen of the picture tube.

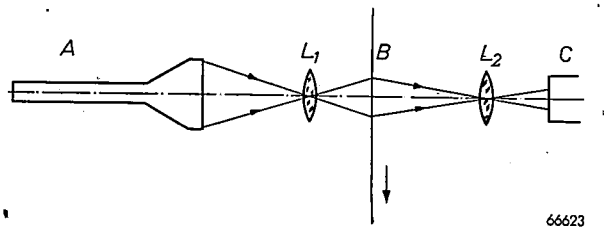


Fig. 1. Flying-spot method for televising film pictures. *A* cathode-ray tube with fluorescent screen on which a raster of lines is scanned. The lens L_1 projects this raster onto the film picture *B* and the transmitted light is concentrated by the lens L_2 in the multiplier tube *C*.

In principle it is possible to televise any scene in a similar way, by projecting a raster of light onto the scene and picking up the scattered light in a photocell. Practice shows, however, that in such a case the photo-electric currents are so small as to be "drowned" in the statistical noise of the photo-electrons, which noise creates at the receiving end the impression of the scene having been televised during a snowstorm! This poor result is due to the fact that only a very small part of the light scattered by the scene in all directions can be picked up in the photocell, such being contrary to the case with the film picture, where practically all the transmitted light can be collected in the photocell.

Fundamental failings of this otherwise fairly simple method are that the fluorescent spot is a weak source of light and that each picture element is illuminated for only 10^{-7} sec, so that one has to manage with the naturally limited amount of light thereby obtained.

[1]) A. A. Campbell Swinton, Distant electric vision, Nature **78**, 151, 18th June 1908, where a remarkable outline is given of the principles of present-day television.

However, another method can be imagined, whereby the scene is illuminated continuously, while for each picture element in succession in the space of time of 10^{-7} sec a signal is transmitted which corresponds to an illumination that was present during the whole of the time ($^1/_{25}$th sec) available per picture. This idea is to be found materialized in all present-day television camera tubes (with the exception of the tubes used for transmitting film pictures according to the method outlined above). With this method there is a continuous accumulation of charge during a frame period, and thus these tubes have come to be known as "storage tubes".

The oldest form of storage tube is the iconoscope, designed by Zworykin (1933), which will presently be dealt with. In the main this article will be devoted to a modern camera tube named the image iconoscope. Some other types will be mentioned in passing.

Classification of modern camera tubes

In the most important camera tubes of modern design there is a plate ("target" or "mosaic") on which is projected an electrical image, consisting of a two-dimensional pattern of electric potentials corresponding in amplitude and position to the luminance in the optical image of the scene to be transmitted. This electrical image is scanned point by point by a focused beam of electrons (the scanning beam), the potentials being thereby reduced to a certain "stabilizing potential" which in some way or other produces an electric signal.

The target is, of course, made of an insulating material, e.g. mica. When an electron beam is directed upon it the rule is that for every surface element, in the stable state, on an average just as many electrons have to be emitted as impinge upon it. This number of electrons may be zero or greater than zero, and it is these two possibilities which, as will be seen, form the basis for the classification of camera tubes into two groups.

Behaviour of an insulator when bombarded with electrons

Let us consider the following case. A beam of electrons (the primary electrons) is focused upon an insulating plate set up in vacuo. In the vicinity of the plate is an electrode, the collector, which has an adjustable positive potential V_{coll}.

When the primary electrons impinge upon a surface element of the plate they release secondary electrons from the material. The secondary-emission coefficient δ, i.e. the average number of secondary electrons released by one primary electron, depends upon the material and the velocity (thus the energy) of the primary electrons at the plate. If V_{coll} is so high that the collector attracts all the secondary electrons towards it then the variation of δ as a function of the energy V_{pr} (expressed in electron-volts) of the primary electrons is as represented in *fig. 2*. In the case of most materials there are two values for V_{pr} where $\delta = 1$; the smaller of the two is denoted by V_1, the larger by V_2. For mica, for instance, these material constants are in the order of 10 volts and some thousands of volts respectively.

As already stated, there are two possible stable states of the surface element, one where no electrons strike or leave the surface, and one where per second a certain number of electrons (>0) impinge upon it and just as many leave it. Which of these two states will be obtained depends upon the value of V_{pr}. If V_{pr} is less than V_1, and consequently δ is smaller than 1, the number of primary electrons striking the surface element per second will be greater than the number of secondary electrons released from it. Thus the potential of the element drops; only when it has reached zero (the cathode potential) — no primary electrons at all then reach the plate — is one of the two stable states of equilibrium obtained.

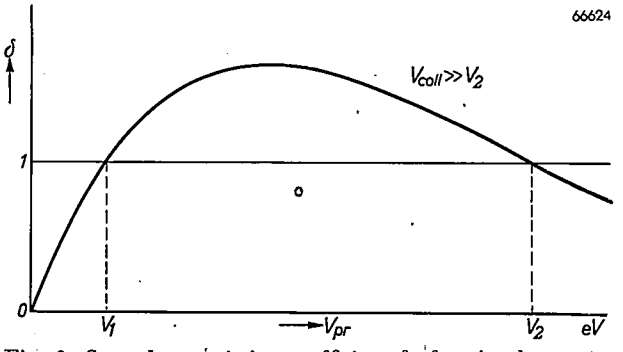


Fig. 2. Secondary-emission coefficient δ of an insulator, as a function of of the energy V_{pr} of the primary electrons. At two values of V_{pr} (V_1 and V_2) δ is equal to 1.
This curve applies when the collector potential is high compared with V_2.

Let us now consider the case where V_{pr} lies somewhere between V_1 and V_2. Then δ is greater than 1, the number of secondary electrons released from the surface per second is greater than the number of primary electrons impinging upon it, and the potential of the surface rises until at V_2 a stable state is reached.

The supposition that all the secondary electrons pass over to the collector only holds if V_{coll} is much higher than V_2. If that is the case then, upon

gradually reducing V_{coll}, the shape of the curve does not change at first until V_{coll} approaches V_2. This is explained by the fact that when the surface element has a potential higher than that of the collector the released secondary electrons have to overcome an electric field before reaching the collector. Whether they succeed in doing so, or whether they fall back

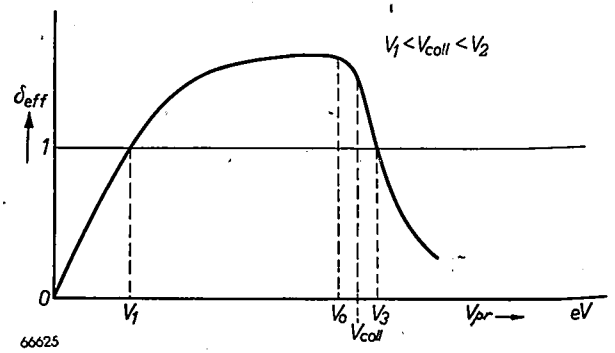


Fig. 3. Effective secondary-emission coefficient δ_{eff} of an insulator, as a function of the energy V_{pr} of the primary electrons when the collector potential V_{coll} is smaller than V_2 (cf fig. 2). $\delta_{eff} = 1$ at $V_{pr} = V_1$ and at $V_{pr} = V_3$, the latter value being a few volts higher than V_{coll}. To the left of V_0 (slightly lower than V_{coll}) the curve is identical to that in fig. 2.

upon the target, depends entirely upon their energy and the geometry of the set-up. The potential of the surface will be stabilized at a value V_3, where the current intensity of the secondary electrons actually reaching the collector (i_{coll}) is equal to the current intensity i_{pr} of the primary beam. The ratio i_{coll}/i_{pr} may be called the effective secondary-emission coefficient, δ_{eff}. For $V_{pr} < V_3$ (but $> V_1$) δ_{eff} is greater than unity, and for $V_{pr} > V_3$ the effective secondary-emission coefficient δ_{eff} is less than unity. As a rule V_3 is slightly higher than V_{coll} (*fig. 3*); in contrast with V_1 and V_2, V_3 is therefore not a material constant.

Thus it is seen that when bombarded with slow electrons ($V_{pr} < V_1$) the surface potential becomes stabilized at zero, and when bombarded with electrons of high velocity it becomes stabilized at the value V_2 (provided $V_{coll} > V_2$) or at V_3 ($\approx V_{coll} < V_2$). For the target of a camera tube however no use is made of the value V_2, for practical reasons; it is strongly influenced by the condition of the surface and thus is too variable from point to point.

It is according to these possibilities that camera tubes are classified as:

1) low velocity tubes, where the target is stabilized at cathode potential (on that account they are also referred to as "cathode potential stabilized tubes", or for short "CPS tubes"), and

2) high velocity tubes, where the target is stabilized at the potential $V_3 \approx V_{coll}$ (e.g. 1000 V).

Among the first belongs the image orthicon, which is the type of tube mainly used in the U.S.A., while belonging to the second group are the iconoscope and the image iconoscope, the latter often being given preference in European television circles. One of the reasons for this preference is related to the large number of lines adopted on the West-European continent (625, and in France 819): with a high electron velocity it is easier to satisfy the high requirements for the the focusing of the scanning beam which are demanded for the definition required for such a large number of lines.

In this article exclusively high velocity tubes will be dealt with, though at the end some comparisons will be made with the image orthicon.

The iconoscope

The iconoscope is the camera tube which at the time gave such an impetus to television[2]). It is schematically represented in *fig. 4*, while in *fig. 5* a photograph is given of the Philips iconoscope, type 5852.

A lens (objective) projects an image of the scene onto a target of thin mica coated on the front with a mosaic of minute, mutually insulated, photo-sensitive elements. On the reverse side is a coating of metal, called the signal plate, forming the output

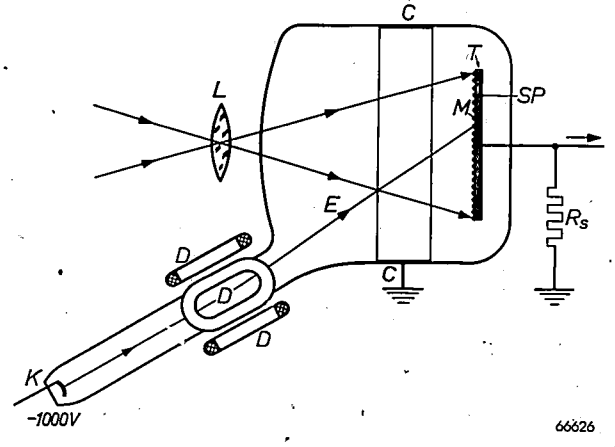


Fig. 4. Iconoscope. L a lens projecting the scene on the mosaic M of the target T. SP signal plate, R_s load resistor, C collector, K cathode, D deflection coils, E scanning beam. The (electrostatic) focusing is not shown.

electrode and externally connected to earth via a resistor. A ring-shaped coating of metal on the inside of the envelope serves as collector and is connected to earth direct.

In an arm of the envelope is an electron gun supplying a beam of electrons which is focused on the target plate. The cathode is at a potential of, say, —1000 V with respect to earth. The beam is made

[2]) See, e.g., Philips techn. Rev. 1, 18-19, 1936.

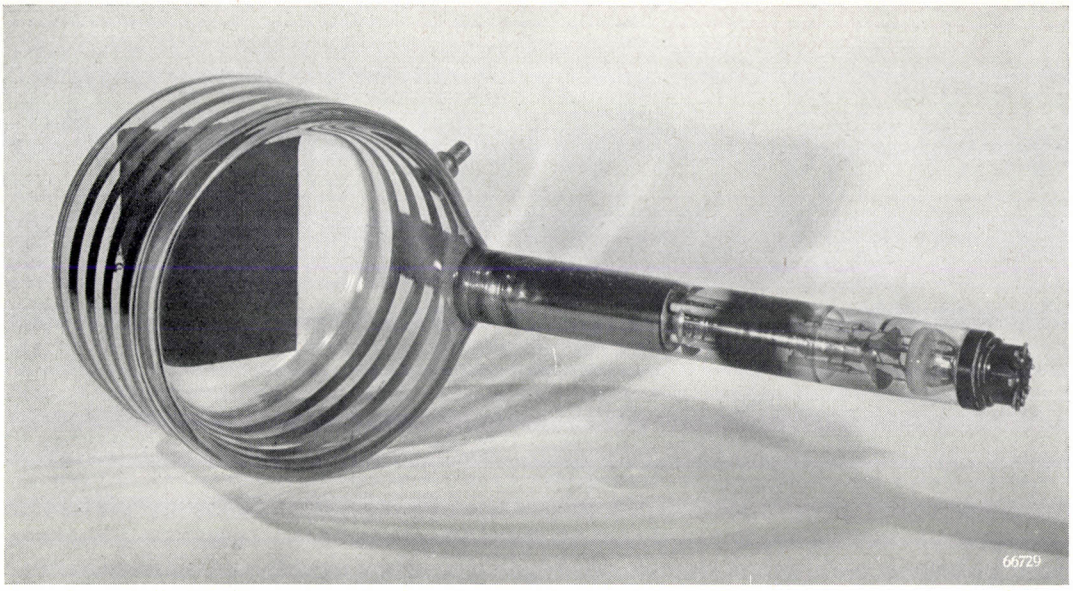


Fig. 5. The Philips iconoscope, type 5852.

to scan the mosaic continuously by means of two pairs of deflection coils. Thus a picture element is formed by the small area of the surface of the mosaic struck by the beam; although this is very small it covers a large number of photo-electric elements.

The action of the iconoscope is sometimes explained in the following (inadequate) way. The incident light causes the photo-electric elements of the mosaic to emit photo-electrons, which are taken up by the collector. Thus a positive electrical image is formed on the mosaic. The photo-electric elements together with the target form as many minute capacitors. As the scanning beam moves across the mosaic the group of capacitors belonging to a certain picture element are discharged. Through the resistor via which the signal plate is earthed there then flows a small current corresponding in intensity to the charge of the picture element, thus corresponding to the local luminance of the optical picture. Thus in the scanning of the electrical image a series of current impulses are generated which together form the video current.

Actually the position is not so simple as this. Such a description does not take into account the part played by secondary emission[3]). In point of fact we have here again the case of an insulator bombarded with electrons, where the collector potential is smaller than V_2 (cf. fig. 3). For the moment photo-emission may be left out of consideration by assuming the iconoscope to be in total darkness.

Not all the secondary electrons reach the collector, firstly because the potential of a bombarded surface element is higher than V_{coll}. If the scanning beam were to be brought to a standstill then, as explained with reference to fig. 3, the element would assume a potential V_3, for instance, 3 volts higher than V_{coll}. Normally, however, the beam is in motion and "rests" on the surface element for such a short space of time (10^{-7} sec) that although the value actually reached (V_{max}, *fig. 6*) is a fraction of a volt lower than V_3, it is still always higher than V_{coll} and thus prevents secondary electrons from reaching the collector. Furthermore, the collector is rather far removed from the target.

The secondary electrons which do not reach the collector fall back on other parts of the mosaic,

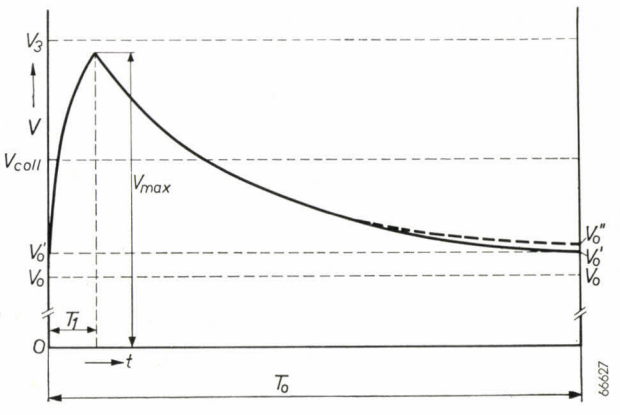


Fig. 6. Curve of the potential V of a picture element on the mosaic of an iconoscope, as a function of the time t. Fully drawn line: mosaic not illuminated; broken line: mosaic illuminated. T_0 = scanning period for the whole image ($^1/_{25}$th sec), T_1 = scanning time for one picture element (10^{-7} sec; in the drawing highly exaggerated). For the meaning of V_{coll} and V_3 see fig. 3, and for V_0, V_0', and V_0'' see the text.

[3]) V. K. Zworykin, G. A. Morton and L. F. Flory, Proc. Inst. Rad. Engrs **25**, 1071-1092, 1937, and W. Heimann and K. Wemheuer, Z. techn. Phys. **19**, 451-454, 1938.

which, as we have already seen, are at a slightly higher potential than the collector. This distribution of the secondary electrons is called the redistribution effect, and it is of essential importance for the action of the iconoscope.

After the surface element in question has been scanned, thereby assuming the potential V_{max}, it will continue to receive secondary electrons originating from other surface elements, until it is scanned by the beam again. Thus its potential V begins to drop (fig. 6) and, if a sufficient length of time were to elapse before the next scanning, would ultimately reach a value V_0 so low that the surface element in question could no longer take up secondary electrons from its surroundings. In practice, however, the surface element is again struck by the beam $^1/_{25}$th sec after the first scanning; the potential it reaches immediately prior to the second scanning is denoted by V_0' (fig. 6).

How is this course of affairs affected when the mosaic is illuminated? An illuminated image element will not emit photo-electrons continuously: from fig. 6 it is seen that during a considerable part of the scanning period the potential V of the element is higher than V_{coll}, and the photo-electrons do not possess sufficient energy to overcome this potential difference. Photo-emission begins, therefore, when — owing to the redistribution effect — the potential V has been sufficiently reduced. Experience teaches even that only during about the last 5% of the scanning period is V low enough for the photo current to reach the point of saturation. Owing to the photo-emission the right-hand part of the V curve of the picture element in question comes to lie somewhat higher (see the broken line in fig. 6), ending with a value $V_0'' > V_0'$, so that in the next scanning the surge from V_0'' to V_{max} is smaller than that from V_0' to V_{max} in the case of a non-illuminated element. It is this difference that gives rise to the signal current. (The illumination causes a change also in V_{max}, but this is small compared with the difference between V_0'' and V_0'.)

The most important features of the iconoscope will now be briefly discussed.

Sensitivity

As already explained, it is due to the redistribution effect that photo-emission can take place, but this is only possible during a fraction of a scanning period. Thus we are still far removed from a continuous photo-emission such as was imagined in the case of an ideal storage tube! This is one of the reasons for the rather low sensitivity of the iconoscope.

A second cause of the lack of sensitivity lies in the mosaic form of the light-sensitive layer. The insulation between the elements does not contribute towards photo-emission, so that a considerable part of the surface of the target is photo-electrically inactive.

Spurious signals

Another drawback attaching to the iconoscope is the fact that, in the case of the non-illuminated tube, the potential surge from V_0' to V_{max} when the element is scanned (fig. 6) varies with the position on the mosaic. Thus a certain signal, a "spurious" signal, is given also when the iconoscope is not illuminated.

The main cause of spurious signals (see the literature quoted in footnote [3])) is that the redistribution does not take place in the same way all over the mosaic, owing to the surroundings of the elements not being the same everywhere. When a certain line is being scanned the part of the mosaic above that line was more recently scanned than the part below it and thus has higher potentials. This is manifested in the picture at the receiving end showing a luminance increasing from top to bottom.

There are similar differences between the left and right sides and between the edge and the middle: the side of the picture where the lines end is reproduced with a higher luminance than the side where they begin, and the edge is lighter than the middle of the picture.

An additional effect is the variation, taking place during the scanning, in the angle at which the beam strikes the mosaic. This angular variation is accompanied by a variation both in the size of the area of the picture element and in the secondary emission coefficient.

When the iconoscope is illuminated the spurious signal is superposed on the picture signal and only if the latter is of a reasonable strength is the spurious signal not very disturbing. It is for this reason that with the iconoscope very high intensities of illumination are needed.

Linearity

The stronger the illumination on a certain part of the mosaic, the higher is the potential V_0'' at that spot just before it is scanned by the beam. This has two consequences: there is slightly less chance of further photo-electrons escaping, and there is a somewhat greater attraction of redistributed secondary electrons. Both these effects result in the amplitude of the signal increasing less than proportionately with the illumination. This non-linearity is rather an advantage than a disadvantage in that it compensates fairly well an inverse non-linearity between the beam current and the control voltage in the picture tube of the receiver. Thus

there is no need to take steps to compensate the latter non-linear effect.

Other features

Apart from the non-linearity there are some other favourable features of the iconoscope, namely that the pictures obtained with it are good both in geometrical configuration and in gradation, and that its installation and operation do not involve any particular technical difficulties.

The image iconoscope

The greatest disadvantage of the iconoscope is its lack of sensitivity, and it is for that reason that attempts have been made to develop camera tubes with greater sensitivity, while still retaining the good picture quality obtained with the iconoscope when the scene is sufficiently illuminated.

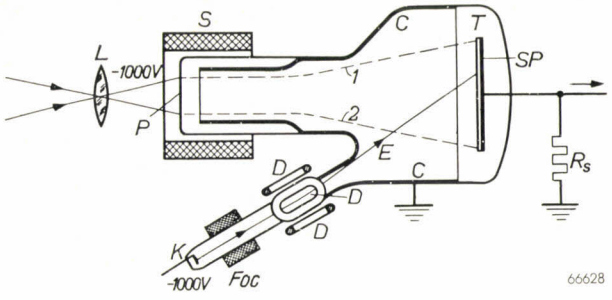


Fig. 7. Image iconoscope. *P* photocathode, *S* coil of the magnetic electron lens, *1* and *2* paths of photo-electrons, *Foc* focusing coil. Other letters have the same meaning as in fig. 4.

A year or two prior to 1940 a more sensitive version of the iconoscope, called the image iconoscope, was developed in the U.K. and in the U.S.A. [4]). Some improvements on this have since been made in the Philips Laboratory at Eindhoven, as will appear in the course of this article.

In the case of the image iconoscope (*fig. 7*) a lens (objective) projects an optical image of the scene to be televised onto a continuous, transparent photocathode. The local density of emission of the photo-electrons corresponds to the local luminance of the optical image. This photo-emission image is focused by an electron lens onto a target consisting in this case of a thin layer of insulating material applied to the signal plate. The metallized inner wall of the envelope serves as collector. An electron gun mounted in an arm of the envelope supplies the beam of electrons scanning the target.

[4]) See, e.g., H. Iams, G. A. Morton and V. K. Zworykin, The image iconoscope, Proc. Inst. Rad. Engrs. **27**, 541-547, 1939.

The differences, compared with the conventional iconoscope, which are mainly responsible for the gain in sensitivity, are the following:

1) The surface of the photocathode is continuous, so that none of its effective area is lost in insulation between the separate photo-electric elements.
2) The stream of photo-electrons reaching the target is reinforced by secondary emission, each photo-electron releasing on an average more than two secondary electrons.
3) The secondary electrons released from the target by the photo-electrons have a much greater energy than the photo-electrons in the ordinary iconoscope, so that secondary emission from a surface element begins immediately after that element has been stabilized by the scanning beam. This means a considerable gain in storage action. The curve for the potential V of a surface element is therefore similar to that in fig. 6, except that in the case where the cathode is illuminated the curve begins to diverge earlier from the curve applying for a non-illuminated cathode, the difference between V_0'' and V_0' thus being greater.

Let us now consider more closely the principal parts of the image iconoscope and also the important question of electron-optical projection. The Philips type of image iconoscope is illustrated in *fig. 8*.

Fig. 8. The Philips image iconoscope, type 5854.

The photocathode

Contrary to ordinary photo-electric cells, an image iconoscope must have a photocathode which is semi-transparent, because the light enters from the outside while the photo-electrons have to emerge on the inside.

In addition to this transparency great sensitivity is needed (at least some tens of μA photo-current per lumen incident light) and moreover the spectral sensitivity must not deviate too much from the relative luminosity curve of the average human eye.

These three requirements greatly restrict the choice of photo-electric material to be used. The photocathode in the Philips image iconoscope consists of a very thin coating of cesium, antimony and oxygen applied to a flat part of the glass envelope. The sensitivity for the light from an incandescent lamp with colour temperature 2600 °K is about 45 μA per lumen. The spectral sensitivity curve, compared with the relative luminosity curve for the normal eye, is slightly displaced towards the blue (*fig. 9*).

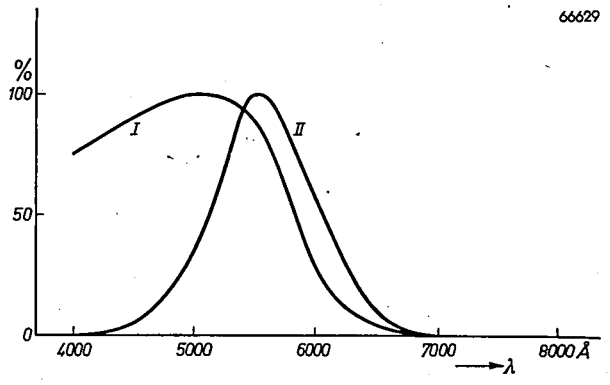


Fig. 9. Relative spectral sensitivity of the type 5854 image iconoscope (curve *I*), compared with the relative luminosity curve (*II*), as functions of the wavelength λ of the light.

Electron-optical image formation

The optical image of the scene is converted into a corresponding photo-emission image on the photocathode. The next step is to produce on the target an electrical image which is a faithful replica of the photo-emission image. This requires that the small beams of photo-electrons emitted from points of the photocathode are focused on corresponding points on the target. For this electron-optical image formation an electron lens is needed. As such use can be made of electric and/or magnetic fields.

For the image iconoscope the choice of the nature of the electron lens is determined mainly by the following considerations:

1) For the emission of photo-electrons to reach the saturation point an accelerating electric field is needed at the photocathode.

2) For a faithful image of a large part of the cathode surface to be produced in a not too long tube, the useful visual field angle of the electron lens — i.e. the angle from which the edges of the image are seen from the centre of the electron lens — must be large.

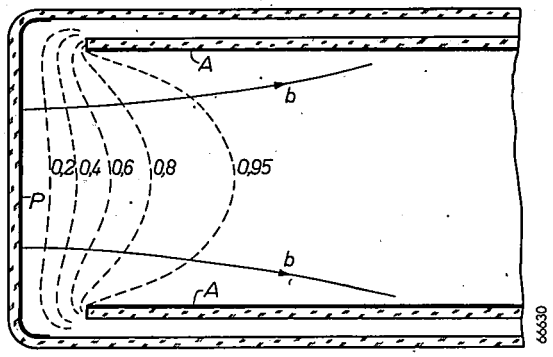


Fig. 10. Between the photocathode *P* and the metal coating *A* (on the inner wall of a glass cylinder) is an accelerating electric field. Broken lines: equipotential planes (the potential of *P* is zero, and that of *A* is taken as 1). *b* are electron paths, in the absence of other fields.

3) The image formed must be sufficiently sharp at all points. A prerequisite is that both the cathode and the target are flat (if the photocathode were curved it would be necessary to employ means of optical correction for the proper projection of a scene upon it; a curved target is very difficult to make and when scanned would cause barrel distortion).

Condition (1) implies that an electric field has to employed. This is obtained by means of a metal cylinder (e.g. the metal coating on the inner wall of a glass tube, *fig. 10*) facing the photocathode and applying a potential difference of, say, 1000 V between these electrodes. Since the cylinder forms, electrically, one whole with the earthed collector, the photocathode is given a potential of —1000 V with respect to earth.

This electric field alone, however, does not suffice. Fig. 10 shows some equipotential planes and some electron paths, from which it is seen that the latter diverge, so that we have here a negative electron lens. Now a negative lens cannot produce a real image from a real object. Therefore, in order to obtain a positive lens, either the shape of the electric field has to be changed in some particular way or a magnetic field has to be added which focuses each electron pencil.

The first alternative comes into conflict with the requirements (2) and (3), for it has not yet been found possible to design an electrostatic lens system which is capable of forming a faithful image of a

sufficiently large part of a flat photocathode on a flat target.

The addition of a magnetic field is therefore indicated. Such a field can be produced by means of a coil placed concentrically around the tube. The coil has to be of such dimensions and in such a position as to minimize aberrations, whilst the magnetic field must not disturb the movement of the scanning beam. As the best solution a fairly long coil was used (see the article quoted in footnote [4]), protruding beyond the end of the tube (see *fig. 11*) far enough for the plane of the photocathode to be intersected practically at right angles by a uniform magnetic flux.

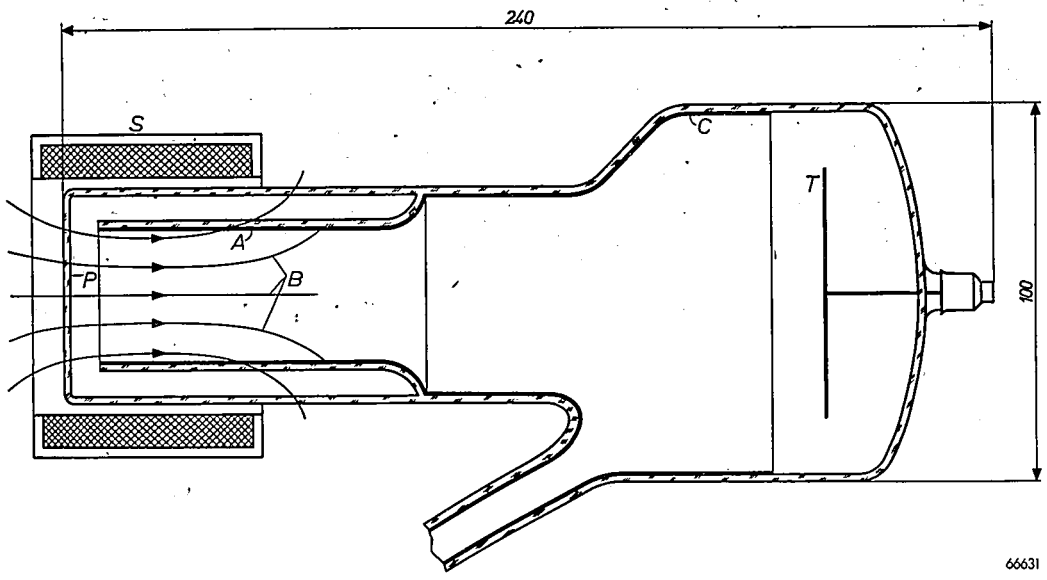


Fig. 11. Formation of the electron-optical image of the photocathode P on the target T with the aid of an electric field (between P and the cylinder A, cf. fig. 10) and a magnetic field. The latter (lines of flux density B) is produced by a focusing coil S.

For a theoretical treatment of the mechanism reference may be made to a publication elsewhere [5]), but the results of this study may be summarized and qualitatively commented upon here.

The movement of the electrons depends not only upon the two fields mentioned but also upon the velocities of the electrons leaving the photocathode. Some of them have zero initial velocity, and the paths they follow are called the principal rays. Generally, however, the electrons leave the cathode with a certain velocity, with the result that they follow a more complex path.

In order to get some idea of the principal rays we have to consider what electrostatic and magnetic forces act upon an electron leaving the cathode with zero velocity (e.g. at the point M in *fig. 12*.)

The cathode being an equipotential plane, the electric lines of force (field strength F) are perpendicular to it. Provided the cathode is at a sufficient depth inside the coil, the same may be said to be approximately the case with the lines of magnetic flux density B. Thus, under the influence of the electric field the electrons following a principal path emerge from the cathode at right angles to its plane; at first they are not subject to any force from the magnetic field.

Very soon, however, there will be a noticeable divergence of the F lines, and a force directed radially outward deflects the path of the electrons outward (fig. 10), thereby causing it to intersect B lines before these diverge. The electron is then subjected to a tangential Lorentzian force, which, with the direction of the B lines as drawn in fig. 11 and for an electron travelling above the centre line, is directed forward. Thus a rotation about the axis of the tube (the z axis) is superposed on the movement of the electron along the lines denoted by b in fig. 10; viewed in the direction of the positive z axis (fig. 12), the electron therefore rotates clockwise. Consequently the principal ray is no longer a planar curve but describes a sort of helix about the z axis. Owing to its having acquired tangential velocity the electron becomes subject to a secondary Lorentzian force directed radially towards

[5]) J. C. Francken and R. Dorrestein, Paraxial image formation in the "magnetic" image iconoscope, Philips Res. Rep. 6, 323-346, 1951 (No. 5).

the z axis and thus counteracting the divergence of the path.

Towards the end of the coil the B lines diverge more with respect to the z axis than does the principal path of the electrons, so that there the electron becomes subject to a Lorentzian force which is tangential, like the first one, but directed backward instead of forward (for an electron above the z axis), because the angle between the B lines and the direction of motion is of opposite sign. Thus the clockwise rotation about the z axis is retarded and even reversed. Then a force comes into action which is directed radially away from the z axis.

Briefly, therefore, the course of a principal ray is as follows: at first the path is approximately parallel to the z axis, then it diverges farther and farther from that axis, thereby turning about the z axis first clockwise and later anti-clockwise in the form of a widening helix.

Although most of the electrons which leave the photocathode have velocities greater than zero and thus do not follow any principal paths, still it is the principal rays which determine the geometry of the electricon-optical image. Each forms the axis of a small electron pencil.

This may be explained as follows. Let us resolve the initial velocity into three components: a radial one (v_r) and a tangential one (v_φ) in the plane of the cathode, and an axial one (v_z) parallel to the z axis (see point N in fig. 12).

Electrons with initial velocity components $v_z = 0$ but v_r and $v_\varphi \neq 0$ follow a path more or less helical about the principal ray but ending, owing to the secondary Lorentzian forces, in the same point on the target as the principal ray. Thus the components v_r and v_φ do not cause any aberrations.

Such is not the case, however, with v_z. This component gives rise to a certain "chromatic" aberration: a point of the photocathode from which electrons emerge with axial velocity does not result in a point being formed on the target but a small circle (scattering circle), the diameter of which is:

$$d = \text{const.}\ \frac{\varepsilon}{F}\,,$$

where ε is defined by $e\varepsilon = \frac{1}{2}m\,v_0^2$ (e = charge, m = mass and v_0 = initial velocity of the electron), and F = electric field strength at the cathode. In the image iconoscope ε is in the order of 1 volt, F about $6 \cdot 10^4$ V/m. Owing to this high value of F the diameter d is so small — thus the image so sharp — that the image iconoscope can quite well be worked with more than 600 scanning lines. In the image orthicon, on the other hand, F is ten times smaller [6]), so that with this type of tube the formation of the electron-optical image is a limiting factor for the number of lines.

---

[6]) H. B. De Vore, Proc. Inst. Rad. Engrs **36**, 335-345, 1948.

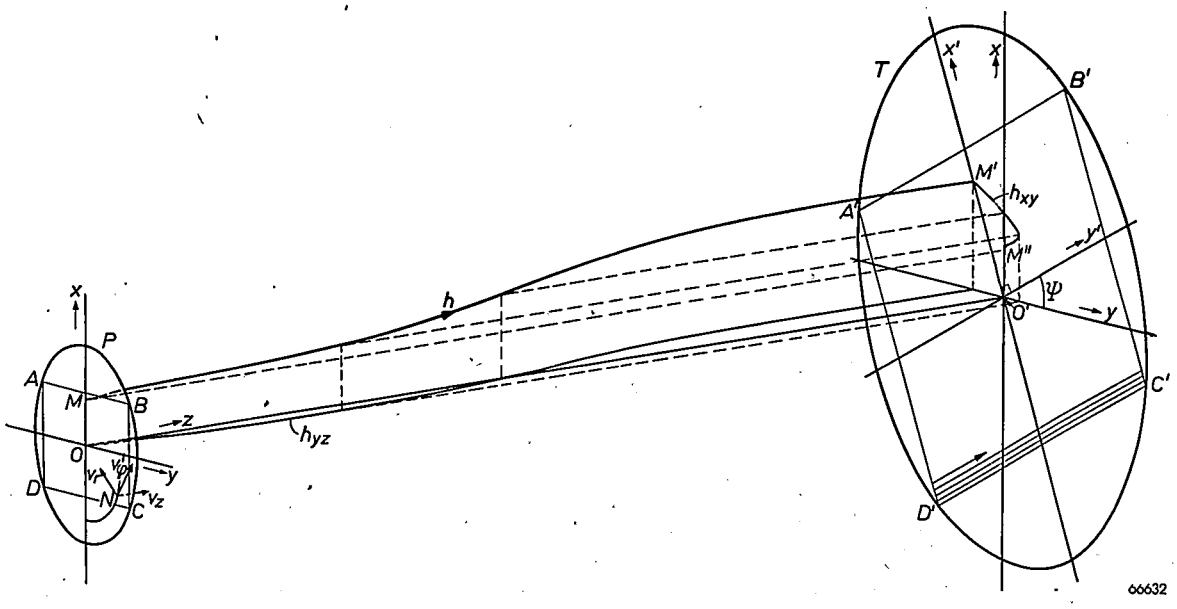


Fig. 12. Formation of the electron-optical image of the photocathode P on the target T: h is a principal path followed by a photo-electron emerging at M with zero initial velocity and reaching the target at M'. M'' is the projection of M on T, h_{xy} the projection of h on T, and h_{yz} the projection of h on the yz plane. At the point N on the photocathode are drawn the three components v_r, v_φ and v_z of the velocity of an electron emitted with a certain initial velocity.

The electrical image on the target is turned over an angle Ψ with respect to the image on the photocathode. $A'B'C'D'$ is the part of the target scanned (a few lines are drawn on it), $ABCD$ the corresponding part of the photocathode. The diagonal $A'C'$ is 75 mm long, while the length of the diagonal AC is normally 20 mm; when other coils are used AC is maximum 27 mm and minimum 10 mm.

Owing to the predominance of the diverging forces acting upon the electrons following the principal path the image on the target is magnified, and owing to the tangential forces the electron image is rotated with respect to the optical image on the photocathode, the angle of rotation Ψ (fig. 12) being about 30 to 40 degrees.

The degree of magnification is related to the effective length of the magnetic field: if the field is lengthened (in the direction of the target) it takes longer for a photo-electron to reach the area where the B lines appreciably diverge, and the secondary Lorentzian force radially directed towards the z axis is active longer, so that this converging force then has greater influence and hence the magnification is smaller. Conversely, the shorter the effective magnetic field, the greater is the magnification.

With our image iconoscope the magnification is normally 3.75, which means to say that the scanned part of the target, which always covers an area of 45 mm × 60 mm, corresponds to an area of 12 mm × 16 mm on the photocathode (the diameter of the active part of the photocathode is 20 mm). By exchanging the coil for another of different dimensions it is also possible, however, to work with a larger or a smaller magnification, thus projecting a smaller or a larger part of the photocathode on the target. The choice as regards the size of the effective photocathode is governed by requirements of an optical, light-technical and camera-technical nature. The limits for the magnification are 2.75 and 7.5 (diameter of the projected part of the cathode respectively 27 mm and 10 mm).

With a magnification greater than 7 to 8, owing to the "chromatic" aberration of the photo-electrons emerging with axial velocity (see above) there is too great a loss in resolving power.

The lower limit of 2.75 is due to various other aberrations, which with a smaller magnification can no longer be sufficiently compensated. As such may be distinguished: field curvature, pin-cushion distortion and so-called S distortion. The first two are known from light-optics [7]). By S distortion is meant the effect of the image of a straight line being projected as a line curved somewhat in the shape of the letter S *(fig. 13)*. This effect, which always occurs to a certain extent with magnetic lenses, is due to the fact that the principal rays at the edge of the image are rotated over a larger angle than those nearer the middle. As is the case with pin-cushion distortion, S distortion is proportional to the third power of the diameter of the active part of the photocathode, thus in the case where the size of the image on the target is constant this distortion is inversely proportional to the third power of the magnification. If the magnification is not too small the S distortion can be sufficiently corrected by electrical means (which we cannot enter into here), but if it is less than 2.75 this is no longer possible. In *fig. 14a* a picture is given showing all three aberrations to a marked extent. The picture in fig. 14*b*, however, has only a scarcely perceptible S distortion, which is not troublesome.

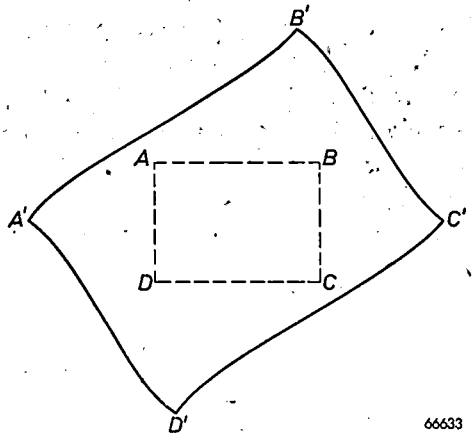


Fig. 13. $ABCD$ is an image on the photocathode, $A'B'C'D'$ the corresponding electrical image on the target. The latter is magnified and turned with respect to $ABCD$ and also shows some S distortion, which always occurs when magnetic lenses are used (straight lines are projected with a slightly S-shaped curve). If the magnification is too small the S distortion becomes so pronounced that it can no longer be sufficiently corrected.

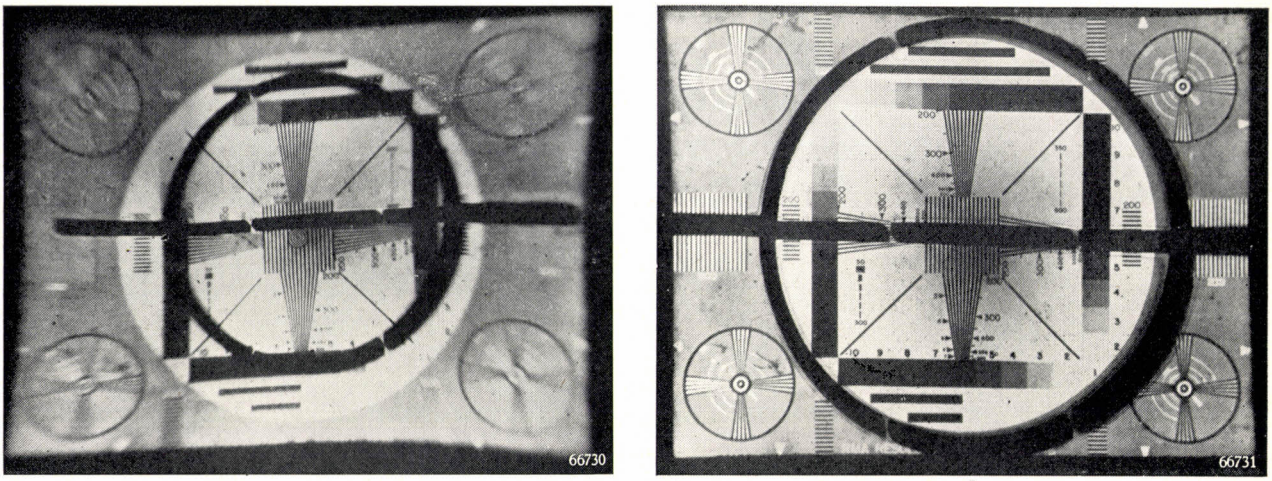
The effective length of the magnetic field can, in principle, be varied in a simpler way than by exchanging the coil, namely by shifting the coil along the tube. In practice, however, the magnification can be only slightly varied in this way: as the coil is shifted farther towards the target a strong field curvature very soon arises (the B lines are then no longer at right angles to the photocathode), whilst if it is slightly shifted in the other direction it touches the objective (to give the camera a large visual field an objective with a short focal length is used, and it is therefore close to the photocathode).

The electron gun

The electron gun supplies the scanning beam. Just as is the case with most picture tubes, in the image iconoscope the beam is focused and deflected with the aid of magnetic fields.

In regard to the sharpness of the scanning, there are two things to be considered. First let us deal with the non-deflected beam. This is focused on the centre of the target, where its diameter must be so small that the lines do not overlap when being scanned. If it is desired to work for instance with 1000 lines then, if the height of the scanned part of the target is 45 mm, the effective diameter of the focus must not be more than 45 μ. This requirement is all the better fulfilled the higher the acceleration voltage is chosen, but, as will be shown in the next section, this voltage should preferably not exceed 1000 V.

[7]) A review of various optical aberrations is to be found, for instance, in: W. de Groot, Philips techn. Rev. 9, 301-308, 1947, in particular pages 304 and 306.



a *b*

Fig. 14. *a*) Picture showing a marked field curvature, pin-cushion distortion and S distortion. *b*) Here there is only a slight S distortion, which can easily be corrected electrically.
These photographs have been taken with the aid of an experimental tube in which a fluorescent screen was used instead of a target. On the photocathode a test pattern was projected as used in television for detecting aberrations and checking the definition and gradation. The heavy black circle and the thick horizontal line in the middle correspond to markings on the photocathode for determining the magnification.

Further, account has to be taken of the fact that in the image iconoscope the electron gun has to be mounted with its axis at an angle to the target. Consequently when the beam is deflected upward or downward the focus is no longer situated on the target. Therefore, to obtain sufficiently sharp scanning also away from the centre, the beam must have a good depth of focus, which means that is has to be extremely narrow. Hence the angle of divergence $2\alpha_f$ (see *fig. 15*) has to be kept very small.

It is, in general, difficult to obtain a fine focus with a very narrow beam on account of the mutual repulsion of the electrons, but fortunately the intensity of the beam current required is very low, in the order of 0.1 μA.

In addition to this space-charge repulsion there is another factor limiting the spot size obtained with a very narrow beam: there is a very fundamental relationship between the angle of divergence $2\alpha_f$ and the current density in the beam. In the case where the space charge is negligible this relationship is:

$$\sin^2 \alpha_f = \frac{V_0}{V} \cdot \frac{j_f}{j_0}, \quad \ldots\ldots \quad (1)$$

where $V_0 = {}^1/_2\, mv_0^2/e$ (with m = mass, v_0 = initial velocity and e = charge of an electron), V = the potential difference traversed by the electrons, j_f = density of the beam current in the focus, and j_0 that at the cathode of the gun.

Equation (1) is derived as follows. In optics there is a law (ascribed to Abbe as well as to Helmholtz or to Lagrange) which says that the brightness of a picture divided by the square of the index of refraction of the space in which the picture is formed is constant (as long as losses through absorption and reflection are negligible). In electron-optics the analogue of this law is expressed in the following formula:

$$\frac{j}{V \sin^2 \alpha} = \text{const.}, \quad \ldots\ldots\ldots \quad (2)$$

where j is the current density and α half the angle of divergence of the electron beam, while V is the potential difference between the point in question and the (imaginary) point where the electron velocity is zero.

The values of these quantities at the cathode are denoted by j_0, α_0 and V_0. The value of α_0 is 90°, because the electrons emerge from the cathode in all directions. (The same holds if there is a limiting diaphragm in the beam, as is the case in the gun of our image iconoscope, since the electron spot is formed by small beams of electrons originating from different

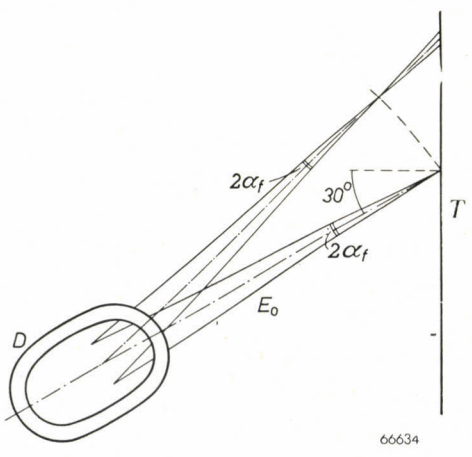


Fig. 15. Assuming that the non-deflected beam E_0 has been focused onto the centre of the target T, when the beam is deflected the focus will no longer be in the plane of T. This gives rise to blurring, the extent of which increases with the angle of divergence $2\alpha_f$.

points of the cathode and each having a half-angle of divergence of 90° at the cathode [8])). Thus the value of the constant in (2) is j_0/V_0.

For the electrons which in the gun of the image iconoscope have traversed a potential difference V the relationship between the half-angle of divergence α_f and the current density j_f in the focus is given by (2) as:

$$\sin^2 \alpha_f = \frac{V_0}{V} \cdot \frac{j_f}{j_0},$$

which is the same as equation (1).

From eq. (1) it is seen that, with given values of V_0, V and j_0, a reduction of α_f leads to a reduction of j_f and thus, for a given value of the beam current intensity, to a larger cross section of the focus. Further it appears that the greater the value of j_0 (thus the heavier the cathode may be loaded) and the higher the value of V, the less is the danger of the focus becoming too large.

What has to be found, therefore, is an optimum value for α_f at which, on the one hand, the focus is not too large and, on the other hand, the sharpness at the edges of the image does not differ too much from that in the middle. With our image iconoscope the position is such that this optimum value of α_f lies at about 3×10^{-3} radians.

This small angle of divergence, combined with a low beam current intensity (about 0.2 μA), has been obtained by placing two diaphragms in the beam. The first, with a narrow aperture, confines the beam within the desired small angle. The second one, with a wider aperture, allows the beam to pass through without hindrance but intercepts the low-velocity secondary electrons formed round the edge of the first diaphragm.

With the focus of 45 μ already mentioned (1000 lines at 45 mm) and a beam current of 0.2 μA, the average current density in the focus is $j_f = 12$ mA/cm². Substituting this in eq. (1), and for V the value at which the secondary emission coefficient of the target is greatest, viz. 1000 V, and for V_0 the value corresponding to the average initial velocity (≈ 0.1 V), we find for the average current density at the cathode of the gun $j_0 \approx 120$ mA/cm². The peak value of the current density is in fact several times greater. Although an ordinary oxide-coated cathode may indeed be continuously loaded with such a current density, it is better to use what is known as an L cathode [9]), since this has a much longer life. It would be quite undesirable if the useful life of a costly tube such as the image iconoscope were to be dependant upon the life of a component like the cathode of the gun.

From the manufacturing point of view the L cathode also has the advantage that it lends itself well for mounting with very close tolerances of the distance between cathode and grid. Thanks to this property there are fewer rejects, and thus less wastage, in manufacture through errors of mounting in the gun.

The glass arm of the envelope containing the electron gun has been kept as narrow as possible (internal diameter 11 mm, external 14 mm), so that also the focusing coil and the deflection coils may be small. The advantages of this are : (1) that the field of the deflection coils in the space facing the target is so weak as not to have any perceptible influence upon the paths followed by the photo-electrons, and (2) little power is needed for excitation of the coils, so that the generators of the deflection currents, mounted together with the tube in the camera, can also be small. This makes the handling of the camera all the easier. A camera with an image iconoscope is shown in *fig. 16*.

In practical use the resolving power of the Philips image iconoscope is found to be 900 to 1000 lines in the middle of the image and about 700 lines at the edges. (These limits are set by the electron gun; the resolving power of the electron-optical projection is very much greater.)

The target

As in the conventional iconoscope, also in the first of our image iconoscopes the target was of mica (of course without mosaic). In the preparation of the photocathode cesium vapour is driven into the tube. This vapour tends to affect the surface of the mica, thereby causing the secondary emission coefficient, δ, as a function of the energy V_{pr} of the primary electrons impinging on it, to follow a curve as represented by I in *fig. 17*. It is true that the peak of this curve lies higher than in the case where the mica has not been in contact with cesium, but then this peak is situated at a point corresponding to a fairly small energy of the impinging electrons. For the electron-optical projection the electric field at the photocathode should be strong, and this, as already stated, is reached with a potential difference of 1000 V. In the electron gun, as we have seen, a high accelerating voltage is necessary in order to obtain a fine focus; for the sake of simplicity a potential difference of 1000 V is likewise chosen here. With an energy of 1000 eV, however, δ remains considerably below the maximum value.

---

[8]) D. B. Langmuir, Proc. Inst. Rad. Engrs **25**, 982, 1937.
[9]) H. J. Lemmens, M. J. Jansen and R. Loosjes, A new thermionic cathode for heavy loads, Philips techn. Rev. **11**, 341-350, 1950.

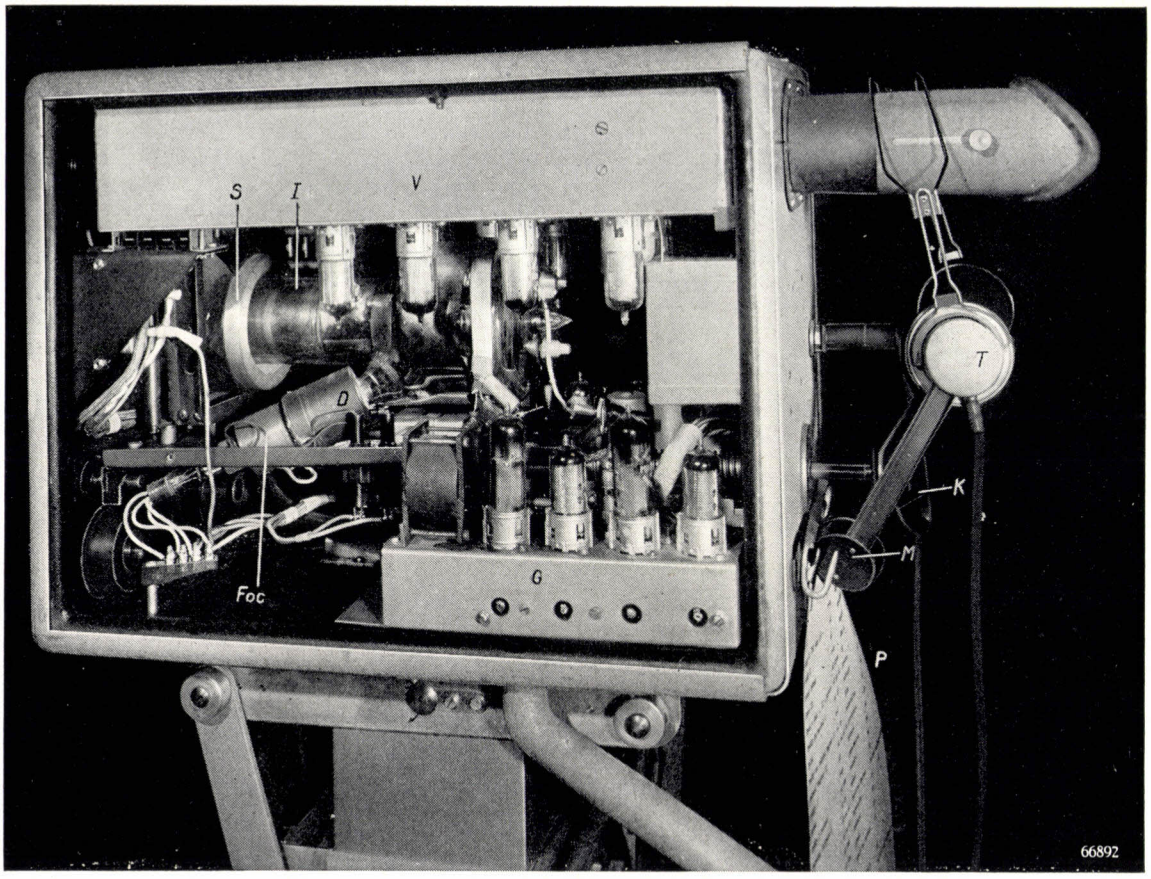


Fig. 16. One of the cameras used for the experimental television broadcasts at Eindhoven. One side panel and a screen have been removed. *I* image iconoscope, type 5854. *S* image coil, *Foc* focusing coil. *D* deflection coils. *G* time-base generator. *V* chassis with monitor picture tube and accessories. *M* microphone and *T* telephone for communication between the operator and the control room. *K* knob for exchanging the objective. *P* playbook.

An improvement has been reached by coating the mica with a thin layer of MgO. The curve $\delta = f(V_{pr})$ now has the shape of curve *II* in fig. 17: at $V_{pr} \approx 1000$ V the coefficient δ is about 4,

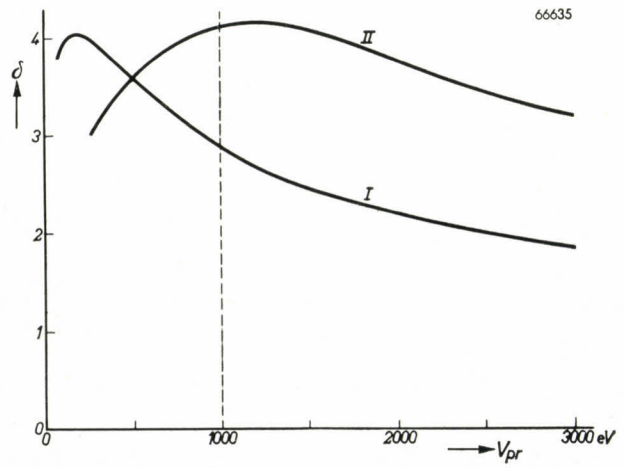


Fig. 17. Secondary-emission coefficient, δ, as a function of the energy V_{pr} of the primary electrons, *I* for mica affected by Cs, *II* for mica coated with MgO. At $V_{pr} \approx 1000$ V *II* is much more favourable than *I*.

which means a considerable gain in sensitivity. Furthermore, owing to the coating of MgO, stains on the mica which cannot be removed and otherwise show up clearly in the picture are thereby made invisible.

Notwithstanding this gain, the sensitivity of the first tubes made was still unsatisfactory. A suitable method of measuring the sensitivity of the target — i.e. the ratio of the signal current to the electron current striking the target — is the following. Except for a narrow vertical strip in the middle, which is kept dark, the photocathode is uniformly illuminated with a variable intensity of light. Each time the scanning beam passes one of the edges of the "black" strip on the target the signal current suddenly changes. The magnitude of this current surge I_s at the transition from "black" to "illuminated" is measured as a function of the photocurrent I_{ph}, which is varied with the strength of illumination.

The result, at first, was that the signal did not rise sufficiently above the noise level of the first

amplifying valve. In order to gain an insight into the factors determining the shape of the signal curve an attempt was made to approach theoretically the problem of the stabilization of the potentials on the target and the production of the signal. Suffice it to say here that these investigations [10]) led to the conclusion that the capacitance of a surface element of the target with respect to the signal plate is an important factor, and that an increase of this capacitance must lead to greater sensitivity.

New tubes were therefore made with a mica sheet of only about 25 μ thickness (also with a layer of MgO, thin compared with the mica) instead of the original sheet thickness of 50 μ.

The success attained from these two measures — applying a layer of MgO and halving the thickness of the mica — is evident when comparing curve *II* in *fig. 18* with curve *I*.

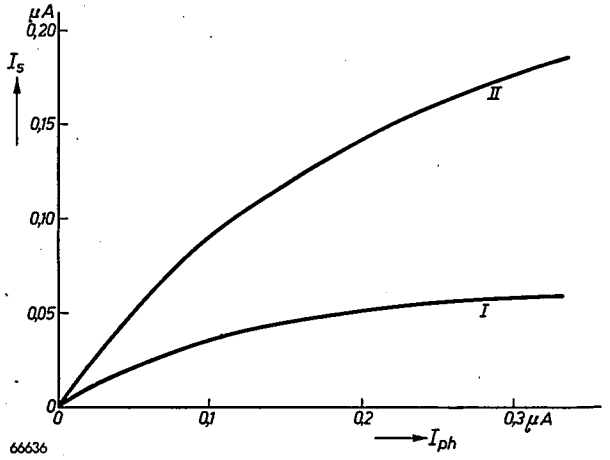


Fig. 18. The current surge I_s in the signal current each time the scanning beam passes from a "black" strip on the target to an "illuminated" part, plotted as a function of the photo-current I_{ph}. Curve *I* applies for a mica sheet 50μ thick without MgO, curve *II* for a sheet 25μ thick coated with MgO.

The reproduced picture of a scene televised under the normal studio lighting, or of an outdoor scene in daylight (even in bad weather), with the image iconoscope last described, is almost free of "noise" and shows excellent gradation.

The scanning beam has to erase, as it were, the electric image on the target point by point. For it to be able to do so sufficiently also on a thinner target, the strength of the beam current had to be increased approximately in proportion to the capacitance of the target. The value of 0.2 μA mentioned corresponds to the mica thickness of 25 μ.

[10]) P. Schagen, On the mechanism of high-velocity target stabilization and the mode of operation of television-camera tubes of the image-iconoscope type, Philips Res. Rep. **6**, 135-152, 1951 (No. 2).

In the image iconoscope spurious signals arise from the same cause as in the case of the conventional iconoscope: the various surface elements of the target are not all in the same position with respect to the scanning beam. In the image iconoscope, however, the situation is more favourable: with the tube described (mica 25 μ thick, beam current 0.2 μA) and with an illumination producing a photo-current of more than 0.1 μA, the spurious signals are so weak that there is hardly any need of compensating measures. In practice a photo-current of 0.1 μA can be obtained with an illumination of the scene of about 1000 lux, when using a non-diaphragmed, normal objective with aperture $f:2$.

Comparison of different types of camera tubes

Let us now compare, briefly, the two main types of camera tubes, the high-velocity and the low-velocity types.

In the first place there is the question of sensitivity. This resolves itself into two factors (disregarding the efficiency of the optical system), viz. the sensitivity of the photocathode (photo-current I_{ph} in relation to the light flux falling on the cathode) and the sensitivity of the scanning mechanism (ratio of signal current I_s to photo-current I_{ph}):

As regards the sensitivity of the photocathode it has already been seen that of the two described representatives of high-velocity tubes — the conventional iconoscope and the image iconoscope — the latter has very much the advantage, owing to the continuity of the photocathode. Among the low-velocity tubes there are likewise types with a mosaic cathode and others with a continuous cathode, the latter including the image orthicon, which as regards photocathode sensitivity is equal to the image iconoscope.

The scanning sensitivity of low-velocity tubes (and without an inportant improvement to be mentioned below) can be directly determined: in these tubes the electrons emitted by the light-sensitive layer are supplemented by electrons from the scanning beam, the scanning sensitivity in this case therefore being exactly 1 μA signal current per μA photo-current. In high-velocity tubes the phenomenon of redistribution complicates matters, but from measurements taken the scanning sensitivity of the ordinary iconoscope can be put at $^1/_{20}$th μA/μA and that of the image iconoscope, with weak to moderate illumination, at about 1 μA/μA (see fig. 18).

Although, therefore, the image iconoscope has about the same scanning sensitivity as the (simple)

low-velocity tube, there is a difference in the $I_s = f(I_{ph})$ curve in favour of the image iconoscope: in the case of this tube, and also the ordinary iconoscope, the said curve is not linear (see fig. 18), whereas in the case of low-velocity tubes it is linear; the non-linear curve is favourable, as explained when dealing with the iconoscope.

There is a means, however, of appreciably increasing the scanning sensitivity of low-velocity tubes. The electrons from the scanning beam which are not taken up by the target and return to the gun can be collected in a multiplier, placed around the gun, which works with secondary emission and thus multiplies them. This is what takes place in the image orthicon, the camera tube commonly employed in the U.S.A. In this way the scanning sensitivity may be raised to a value of 25 to 100 $\mu A/\mu A$, which is of course a valuable property when, for instance, scenes have to be televised in poor light (e.g. sporting events under artificial light). An inherent drawback is, however, that the current of the returning beam can be modulated only up to about 20% and consequently contains a relatively large amount of noise. More electrons return from dark parts than from light ones, so that the noise from the darker parts is both relatively and absolutely the strongest, and often very troublesome.

It has already been explained that in regard to spurious signals the image iconoscope has a decided advantage over the ordinary iconoscope. The image orthicon is free of spurious signals of this nature, but on the other hand it is subject to another interference connected with the fact that the secondary-emitting surfaces of the multiplier do not have exactly the same secondary-emission coefficient over the whole area ("dynode spots").

Electron-optically, high-velocity tubes have undeniably the advantage over those of the other group, in that with electrons of a high velocity it is easier to obtain a scanning beam with a high resolving power, and there is much less trouble from interfering electric and magnetic fields.

Summary. In the introduction television camera tubes are divided into two groups, high-velocity and low-velocity tubes. This article deals only with tubes of the former group, first briefly touching upon the iconoscope and then considering the image iconoscope in more detail.

Discussed in succession are the photocathode, the mechanism for the electron-optical projection, the electron gun and the target as found in the Philips image iconoscope, type 5854. Measured under the illumination from an incandescent lamp with colour temperature 2600 °K, the photocathode (Cs-Sb-O) gives an output of about 45 μA per lumen. For the formation of the electron-optical image an electric and a magnetic field are used, situated in the same space (near the photocathode) and giving a magnified replica of the photocathode image on the target. The electron gun supplies a magnetically focused and magnetically deflected scanning beam (beam-current strength normally 0.2 μA) with such a fine focus and such a focus depth that the resolving power at the middle of the target amounts to 900-1000 lines and at the edges is still 700 lines. The cathode employed in the gun is of the "L" type, which has a long life even under a high specific load. As regards the target it has been found that increasing the capacitance with respect to the signal plate gives greater sensitivity. A very thin mica sheet (25μ) has therefore been used, coated with a thin layer of MgO in order to increase the number of secondary electrons released by the high-velocity electrons impinging on the target ($\delta > 4$).

The article concludes with a brief comparison between high-velocity tubes and low-velocity tubes, and in particular between the new image iconoscope described and the image orthicon.

NEW DEVELOPMENTS IN THE GRAMOPHONE WORLD

by L. ALONS.

681.844.081.48

Developments of the last few years have been of great importance for gramophone enthusiasts, in that it has been made possible to play the whole of a classical work without interruption or at most with only one interruption at a suitable place in the composition. Now this greater enjoyment can be had from music at no additional cost. Drastic changes have, of course, had to be made both in the records and in the record players.

Introduction

As is known, gramophone records as are normally obtainable in the trade have to answer certain standards as regards dimensions, speed of rotation, etc. These standards were fixed at a time when almost exclusively mechanical sound boxes were used, but since these have now for the greater part been replaced by electrical pick-ups it has become possible to revise those old standards. An incentive in this direction was given by the appearance of the magnetophone, a magnetic-tape recorder and player with a long playing time, which threatened to become a competitor of the gramophone. This led to the development of a new kind of gramophone record — the long-playing record — which, in the 12″ size, has a playing time of $22^1/_2$ minutes on each side; compared with the $4^1/_2$ minutes of a normal record of the same diameter this is indeed a great advance, especially when it is considered that, in addition, the quality of the reproduction is vastly improved in many respects.

For dance music and for short musical pieces in general a long playing time is of less importance, but also for this category some producers have taken advantage of the possibility to depart from the old standards. The records now being made for such short pieces of music have a diameter of only 7″ for a playing time of $4^1/_2$ minutes. A striking feature about these "easy-storage" records, as they have come to be called, and also the long-playing ones is the large number of extremely fine grooves cut in them, and that is why both these new kinds of records are often given the name of microgroove records.

In this article only the long-playing records will be discussed, but similar considerations apply also for the other types of microgroove records. First the factors which set a limit to the playing time of a gramophone record will be considered, from which it will be seen that one of the conditions for extending the playing time was that the pick-up has to be made as light as possible. This article will conclude with a description of the new record players developed by Philips. These are fitted with an extremely light pick-up and reproduce with a high degree of fidelity the sound recorded both on long-playing records and on those of the old standards, whilst at the same time record wear is reduced to a minimum.

Factors limiting the playing time

The gramophone records obtainable in the trade are made by moulding a thermoplastic material in a matrix which is an electrolytically produced negative of an original recording on wax or lacquer. The original is obtained in the following way.

A metal disc covered with a layer of wax or lacquer is rotated at a constant speed under a cutter travelling at a uniform speed along a straight line from the circumference to the centre of the disc. In this way a spiral groove with a certain pitch is cut in the disc (the unmodulated groove). Now when the sound that is to be recorded is converted into radial oscillations of the cutter and these oscillations are superimposed upon the uniform movement of the cutter towards the centre, the modulations are cut in the disc and thus a modulated groove is obtained.

Obviously the playing time T depends upon the number of turns N of the groove cut and the number of revolutions per second R made by the disc. Hence:

$$T = \frac{N}{R}.$$

The number of turns can be expressed by equating the diameter of the outermost turn D_{max}, the diameter of the innermost turn D_{min} and the pitch in the formula:

$$N = \frac{D_{max} - D_{min}}{2 \cdot \text{pitch}}.$$

and thus for the playing time we arrive at the expression:

$$T = \frac{D_{max} - D_{min}}{2 \cdot \text{pitch}} \frac{1}{R} \quad \ldots\ldots \quad (1)$$

If it is desired to increase the playing time without increasing the outer diameter of the record, in principle there are the following possibilities:

(1) Reducing the pitch.

Since the pitch is the sum of the width of the groove itself and the width of the wall between the grooves (pitch = wall width + groove width) the pitch can be reduced in two ways:

a) by reducing the width of the groove, or

b) by reducing the width of the wall.

A narrower groove requires a sharper needle (the point of the needle has to have a smaller radius of curvature). The dimensions of the needle must be such that the rounded point of the needle come to rest against the sides of the groove, so as to avoid the possibility of the point gliding along the upper edges of the groove, which, owing to the roughness of the edges, would cause excessive noise (see *fig. 1*). Incidentally it should be added that the point of the needle must not be allowed to glide along the bottom of the groove, because it would then be unable to follow the oscillations of the groove without some play, which would cause both distortion of the sound and excessive noise.

The sharper needle required for a narrower groove implies — given equal weight of the pick-up — higher needle pressure on the disc. Naturally the needle pressure must be kept below the critical pressure at which non-elastic deformation of the disc material takes place.

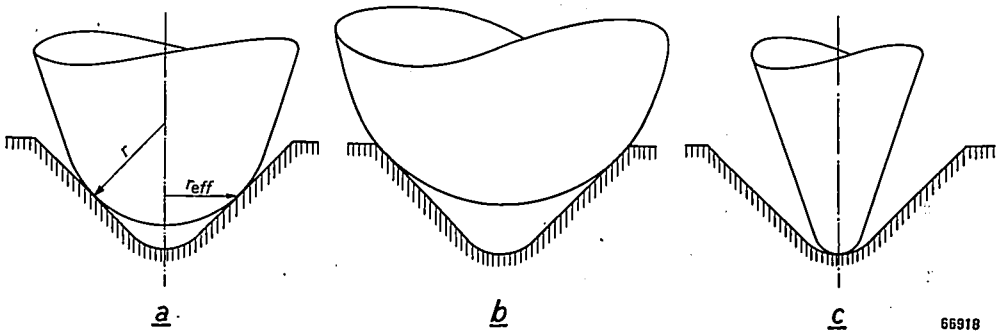


Fig. 1. Position of the needle in the groove. (*a*) How the spherically rounded tip of the needle should rest against the walls of the groove; r is the radius of curvature of the needle tip, r_{eff} the effective needle radius. (*b*) and (*c*) How it should not be. In the situation (*b*) the needle rests on the sides of the groove, which are subject to damage both in the cutting of the groove and in the further handling of the record; consequently with the needle in such a position there would be excessive noise. In the situation (*c*) the needle rides on the bed of the groove, and the forces required for "controlling" the needle can be produced only after it has run up against a side wall; this gives rise to play of the needle and thus distortion, in addition to excessive noise.

(2) and (3) Reducing either the speed of rotation or D_{min}. Anticipating the discussion of these possibilities, it is to be pointed out here that a relationship exists between R, D_{min} and the minimum groove velocity V_m, i.e. the linear velocity at which the innermost groove passes under the cutter or the playing needle, this being expressed as:

$$V_m = \pi D_{min} R \quad \ldots\ldots \quad (2)$$

It will appear that V_m, and not D_{min} or R, is the factor which counts.

Each of these possible measures will now be considered in turn.

The groove width

The lower limit of the groove width depends upon the properties of the material from which the ultimate gramophone record is made, as will be understood from what follows.

The groove width of long-playing records has been reduced, but this was possible only on account of the weight of the pick-up having been considerably reduced.

The width of the wall

The width of the wall is determined by the maximum amplitude, B_{max}, of the modulation made by the cutter. To avoid two adjacent modulated grooves from running into each other this width must be at least $2B_{max}$. Consequently the width of the wall can be reduced only by reducing B_{max}, and the question is what factors set a limit to this.

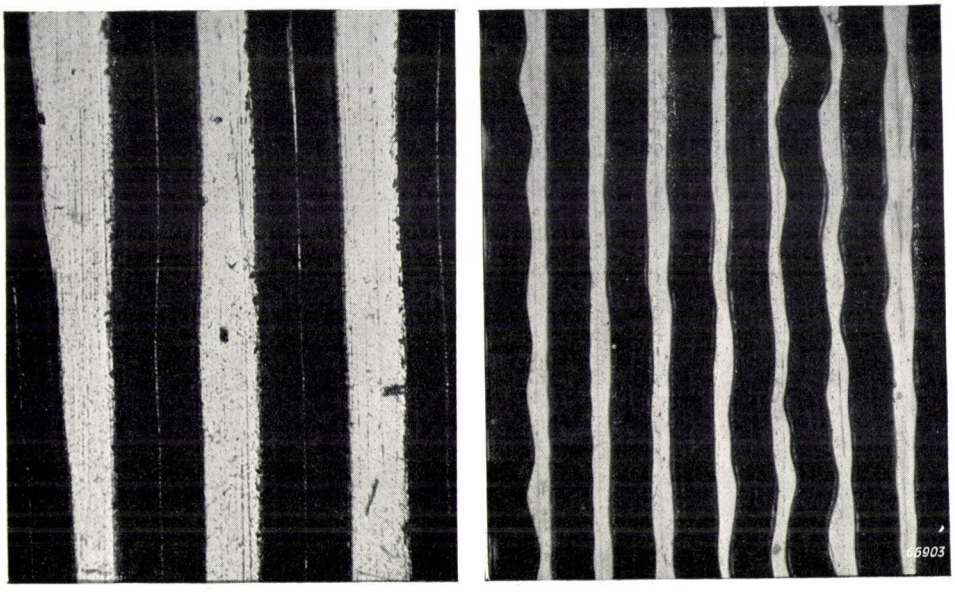
A very rough reasoning — which, however, immediately brings us to the root of the matter — is this:

The material from which the gramophone record is made is not perfectly homogeneous but has a certain granular structure, so that there are irregularities in the groove, and these cause background noise. The sounds that are still to be heard clearly during reproduction must have an amplitude greater than the irregularities of the groove — thus a minimum "signal-to-noise ratio" is required. The amplitude B_{min} corresponding to the weakest sounds to be recorded is therefore directly dependent upon the degree of homogeneity of the disc material. The amplitude with which the sounds with the highest intensity are recorded, B_{max}, must be sufficiently greater for proper expression to be given to the differences actually occurring between forte and piano passages (dynamic range). Bearing in mind that the thought from which we started was to make B_{max} as small as possible so as to obtain a longer playing time, it is evident that a compromise has to be made so as to satisfy both the artistic requirements and those for a not too short playing time.

Once an artistically acceptable relation has been found between B_{max} and B_{min}, it is clear that the width of the wall can be reduced — and thus the playing time extended — provided a finer grained material is available for the records (see *fig. 2*).

Actually the matter is more complicated than this, since account has to be taken of the differences in frequency of the sounds to be recorded (from, for instance, $\omega/2\pi = 50$ c/s up to 10,000 c/s). There is no hard and fast rule for the manner in which sounds of equal intensity but differing in frequency are recorded on the disc — this is called the cutting characteristic — and great problems are still involved in the choice of a cutting characteristic satisfying all requirements. Considered in the light of what has been set forth above, the obvious thing to do would be to give these sounds one and the same amplitude. When reproduced with so-called amplitude pick-ups (where the voltage induced depends entirely upon the deflection of the needle) the sounds so recorded induce equal voltages which are converted in the loudspeaker into sounds of equal intensity. The consequence of such a cutting method "with constant amplitude" — assuming, for the sake of simplicity, that in the sounds to be recorded all frequencies have the same maximum intensity — would be that all frequencies on the disc would have the same recorded amplitude B_{max}. This means that the accelerations to be given to the cutter for the highest frequencies ($\sim B_{max}\,\omega^2$) would be so great as to be unattainable. (Such great accelerations would give rise to difficulties also in the playing back of the records, though these do not weigh quite so heavily as the difficulties encountered in the cutting, which are of a fundamental nature.)

Another obvious solution would be to adopt a cutting method for recording with "constant



a *b*

Fig. 2. Two pieces of gramophone records: *a*) with normal groove, *b*) with microgroove (enlarged 80 times). Note the smoothness of the surface in the second case.

velocity": sounds of equal intensity are then recorded with the same amplitude of velocity $B\omega$, thus with an amplitude inversely proportional to ω. In the case of reproduction with a so-called velocity pick-up (equal velocity amplitude produces equal voltage) a faithful rendering of the sound recorded would then be obtained. But with such a cutting characteristic we are farther off than ever from attaining our object of reducing the width of the wall; because if for the highest frequency B_{min} is chosen as large as is necessary for an adequate signal-to-noise ratio and B_{max} is given a value such that the forces of acceleration then occurring are still permissible, the amplitudes for the low notes would assume impossible proportions. Moreover, for the highest notes the ratio $B_{max}:B_{min}$ would anyhow be smaller than that required for a good dynamic range.

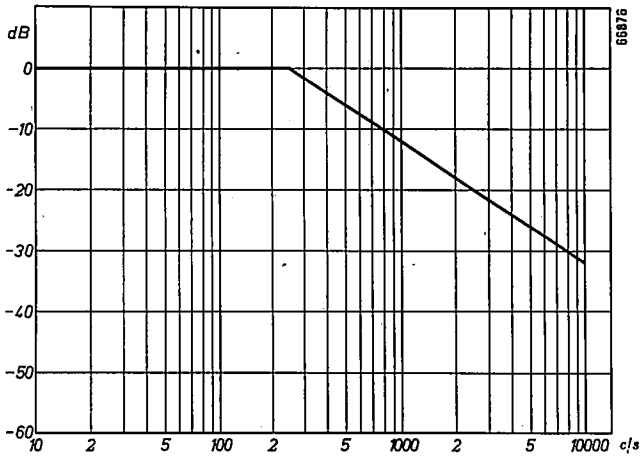


Fig. 3. Cutting characteristic of normal gramophone records. The line represents the amplitude of the modulation for tones of equal strength but with different frequencies, the amplitude being expressed in decibels. Apart from an additive constant the ordinates are therefore equal to 20 log B. Note the logarithmic scale on the abscissa axis. The cross-over frequency, where the knee occurs in the curve, differs slightly for records of different make.

In the cutting characteristics used in practice there is, therefore, always a compromise between the two methods mentioned; see *fig. 3*, where a cutting characteristic as hitherto commonly used is given by way of illustration. Though only amplitude pick-ups or velocity pick-ups may be available, this compromise does not in principle constitute any objection if the reproduction is by electrical means, since in that case the (linear) distortion which would occur in the reproduction can be avoided by using a suitable filter. Bearing in mind that a mechanical player with a gramophone horn is in principle a velocity pick-up, and that mechanical filters are not so easy to make, it can be understood why in the past the method of cutting with a constant velocity was adhered to as closely as possible.

This more precise approach to the problem of the maximum amplitude B_{max} that is to be chosen does not, however, detract from the truth of the conclusion derived from the rough reasoning first given. B_{min} is determined by the granulation of the disc material and B_{max} by the requirement that the ratio $B_{max}:B_{min}$ must not be too small, for the sake of reasonable dynamic range. The precise reasoning however shows that, after B_{max} has been fixed and thus also the playing time, the difference between forte and piano will not be given such good expression in the high frequencies as in the low ones.

The speed and the diameter of the innermost turn

The speed R and the diameter of the innermost turn D_{min} are limited by the fact that the minimum groove velocity V_m, to which they are both proportional (eq. 2), must not be below a certain limit, as otherwise the sound will be badly distorted and the record will suffer from excessive wear. This is related to two effects which are known as "tracing distortion" and "pinch effect". These two effects are in turn a direct consequence of the fact that the cutting and the playing back of the gramophone record are done with two instruments (the cutter and the needle) which have a different profile. The cutter must have a flat profile with cutting edges (see *fig. 4*), the groove being cut in such a way that the walls of a sinusoidal groove enclose a smaller angle (α) in the sides than in the peaks of the sine curve. The needle used for playing the record must

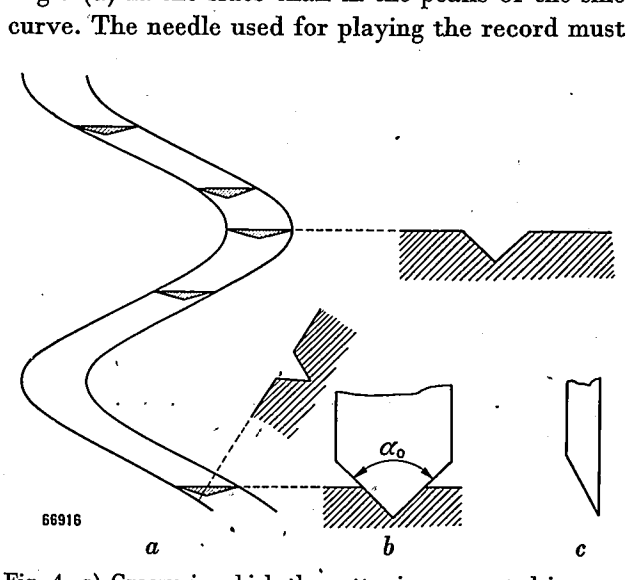


Fig. 4. *a*) Groove in which the cutter is represented in cross section at different places. The groove becomes narrower the more it deviates from the direction of the unmodulated groove. *b*) Front view of the cutter. *c*) Side view of the cutter. The angle α_0 is the angle given by the cutter to an unmodulated groove and to a modulated groove in the peaks of the sine. In this illustration and in fig. 6 the ratio of the modulation amplitude to the width of the groove is exaggerated (see fig. 2).

not, of course, have the same profile as that of the cutter, because otherwise each time the record is played it would, as it were, cut the groove again, which would result in prohibitive wear of the record. The needle should preferably have a spherically rounded tip and rest in the groove in the manner already described (see fig. 1).

It will now be shown what factors influence these two distortion effects.

Tracing distortion

Let us consider a short length of the unmodulated groove as being straight, with its centre line as x-axis (*fig. 5*). Denoting the direction perpendicular

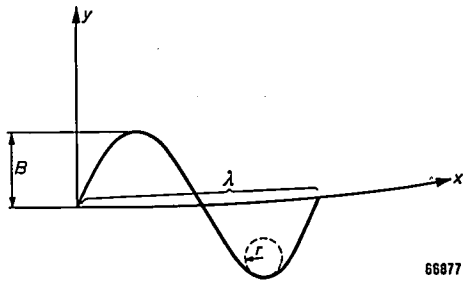


Fig. 5. The elements of the modulated groove. λ wavelength of the modulation, B amplitude of the modulation, r minimum radius of curvature of the modulation. The x-axis is the unmodulated groove.

to the unmodulated groove in the plane of the disc as the y-direction, then for the centre line of a groove modulated with a pure sinusoidal tone we have the equation:

$$y = B \sin \frac{\omega x}{V} \quad \ldots\ldots \quad (3)$$

where ω is the angular frequency of the vibration recorded and V the groove speed of the turn in question. (For the innermost turn the groove speed is therefore $V = V_{\min}$.) The sine curve given by (3) is most strongly curved in the peaks, the radius of curvature there being approximately

$$r = \frac{V^2}{B\omega^2} \quad \ldots\ldots\ldots \quad (4)$$

Thus it is seen that the minimum radius of curvature decreases with decreasing groove speed. Now, as *fig. 6* shows, if the radius of curvature of the groove is comparable to the "effective" radius r_{eff} of the needle (i.e. the radius of that part of the needle point which rests against the walls of the groove; see fig.1) then the needle can no longer follow faithfully the modulation of the groove. The resultant distortion is called the tracing distortion. If this distortion is to be kept within a certain limit it necessarily implies that a limit is set for the groove speed of the innermost turn, viz. roughly $V_m^2 > B_{\max}\omega^2 r_{\text{eff}}$. If $B_{\max}\omega^2$, i.e. the acceleration at the highest frequencies, is chosen as large as is permissible on the foregoing grounds — this has been done for the long-playing records, as we shall presently see — then V_m can be reduced only if at the same time r_{eff} is also reduced, or, in other words, if a sharper needle is used.

Mention has already been made in this article of the favourable effect that a sharp needle has on the playing time, on account of this being related to a small groove width. It has also been pointed out that a sharper needle makes it essential to have a lighter pick-up so as to avoid non-elastic deformation of the record material resulting from too great a static needle pressure. There is, however, yet another factor, already mentioned in passing, namely the dynamic needle pressure, which results from the radial accelerations of the needle in the groove. The maximum value of these acclerations is $B_{\max}\omega^2$. It has already been stated that for the long-playing records this quantity has been chosen of the same order as that for normal records, and when it is pointed out that the dynamic pressures occurring in the playing of normal records with a normal pick-up certainly lie round about the limit of elasticity of the disc material, it is obvious that a limit is set for the sharpness of the needle. A sharper needle can be chosen only if the mass of

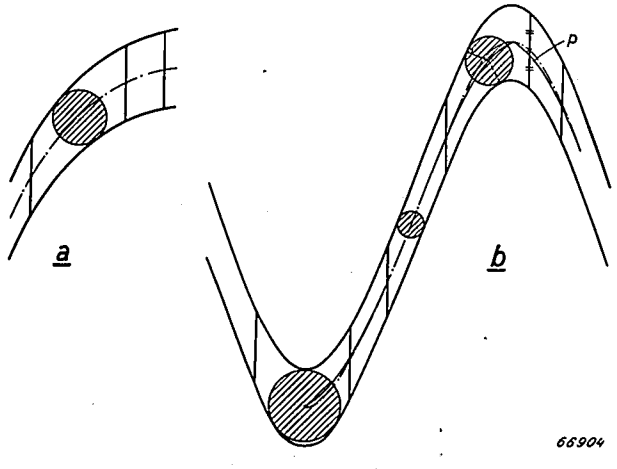


Fig. 6. How tracing distortion arises. *a*) The centre of the needle follows the centre line of the groove if the radius of curvature of the needle point is smaller than that of the groove. In *b*) the two radii of curvature in the peaks of the sine are equal. There the needle can no longer exactly follow the groove, as may be seen in the part of the curve indicated by p. The narrowing of the groove (reduced groove angle) along the side of the sine curve gives rise to pinch effect (see fig. 4). The cutter is represented schematically at different points.

the pick-up, in so far as it is rigidly coupled with the needle in respect to its horizontal movements, is made sufficiently small [1]).

Pinch effect

As already stated, the needle used for playing a gramophone record should bear against the sides of the groove. The height at which the point of the needle "rides" above the bottom of the groove depends upon the aforementioned groove angle α. Simultaneously with the periodical variations of α the needle together with the mass rigidly coupled to it is forced slightly upward (see figs 4 and 6). This periodical vertical movement of the needle is called the pinch effect. The variation in the height of the needle, and therefore the amplitude of the pinch effect, increases in direct proportion to the variation of the groove angle α. Now the steeper the slope of the sine curve the smaller is the groove angle α in the sides. As follows directly from equation (3), the steepest slope of the groove is given by:

$$\left(\frac{dy}{dx}\right)_0 = \frac{\omega B}{V} \quad \ldots\ldots\ldots \quad (5)$$

Therefore the periodical variation of the groove angle α and consequently the amplitude of the pinch effect will increase with decreasing groove speed.

The frequency at which the pinch effect takes place — the "pinch frequency" — is twice the modulation frequency, since each sine curve has two peaks and two sides per cycle.

This pinch effect has two adverse consequences: (1) It leads to wear of the record, owing to the forces exerted by the needle on the record by the periodical vertical movement of the needle. It has already been shown that the periodical variation of the groove angle (to which the pinch effect is proportional) increases with decreasing minimum groove speed. The reduction of the latter quantity, which is what is aimed at, will therefore result in an increase of the pinch effect. Furthermore this effect will be greater still on account of the fact that a smaller V_m makes it necessary to use a finer needle (smaller contact surface); given equal force, this results in increased dynamic needle pressure. The only possible way to avoid greater wear of records having a smaller V_m is to keep as small as possible the mass m_1 which, in respect to the vertical movements, is rigidly coupled with the needle point. This is the course that has been successfully followed in the new Philips' pick-up which will be described later.

(2) The second consequence of the pinch effect is a certain distortion of the reproduced sound, due to the fact that in general the vertical movements of the needle also induce voltages in the pick-up. It will be seen that with the solution of the problem sub (1) this distortion can also be limited to a negligible minimum.

Long-playing records

It has already been said in what respects long-playing records differ from the old type, and we shall now go into these points in more detail.

Long-playing records are made from a new material named "Vinylite", which belongs to the family of plastics. It combines the advantages of a much smaller size of grain (see fig. 2) with unbreakability. It has already been explained how a smaller grain structure allows of a smaller width of wall and thus a smaller pitch of the groove.

In fact the groove pitch has been reduced from 0.26 mm in the old standard records to 0.1 mm, whilst the groove width has been narrowed down from 0.13 mm to 0.07 mm. At the same time the speed of rotation has been reduced from 78 r.p.m. to $33^1/_3$ r.p.m., and the diameter of the innermost turn has been increased from 4" to 4.75". Owing to these changes a playing time of $22^1/_2$ minutes has been reached. Compared with the old standard records the minimum groove velocity has been reduced by a factor of $\frac{33^1/_3}{78} \times \frac{12}{10} = 0.51$.

It might at first sight seem strange that the inner diameter should have been increased. It can be proved, however, that with a given minimum groove velocity there is an optimum value of the inner diameter and of the speed of rotation at which the playing time is the maximum. In other words, if a smaller inner diameter were chosen then, to keep the minimum groove velocity above the given limit, the speed of rotation would have to be increased to such an extent that in the ultimate the playing time would be shorter.

The playing time mentioned applies for the largest microgroove records having an outer diameter of 12". In addition to these there are also smaller records of 10" and the so-called "easy-storage" records, of only 7" diameter and a playing time of $4^1/_2$ minutes. Not all microgroove records are played at a speed of $33^1/_3$ r.p.m., some records of 7" having been made for a speed of 45 r.p.m.

[1]) When speaking here of certain masses being rigidly coupled to the needle in respect to movements in certain directions we mean certain effective masses which may be imagined as being rigidly connected to the needle. This occurrence of different effective masses for different directions of movement is related to the fact that various parts of the pick-up are elastically coupled to each other.

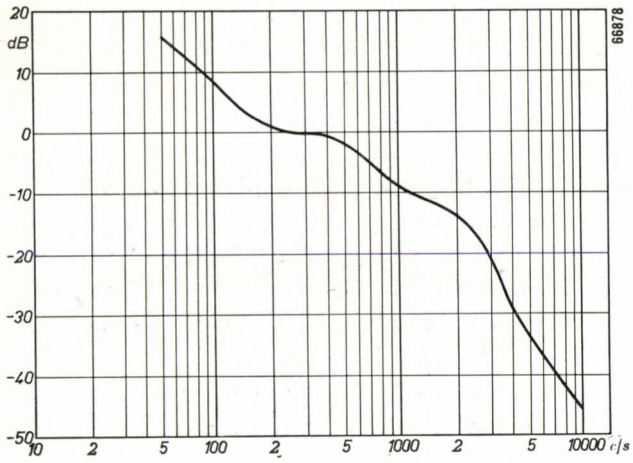


Fig. 7. Cutting characteristic of a long-playing record.

The cutting characteristic of a long-playing record is given in *fig. 7*, from which it is seen that over the whole range the amplitude increases with decreasing frequency. Thus the compromise between cutting at a constant velocity and cutting at a constant amplitude has been reached in a different way from that indicated in fig. 3.

It should be pointed out that, by choosing for B_{max} for the high tones a value of the same order as that of normal records, any possible gain in playing time that might be reached by further reducing V_m has been sacrified for the sake of better quality in the reproduction of the high frequencies.

Description of the pick-up used in the new Philips record player

All kinds of gramophone records, both the normal and the microgroove ones, can be played on the Philips type 2508 record player. This has one motor for three speeds, 78, 45 and $33^1/_3$ r.p.m., and further a record changer for automatically changing 12″, 10″ and 7″ records.

A simpler design is the type 2978 record player (see *fig. 8*), which is without a record changer, while there is a choice of two speeds, 78 and $33^1/_3$ r.p.m.; the lack of the third speed however is not of any great consequence for all countries outside the U.S.A. since there are relatively few 45 r.p.m. microgroove records in these countries.

Both these types of record players are equipped with a new design of pick-up fitted with a specially made sapphire needle.

The needle

It appears that even to the trained ear a needle rounded off at the tip with a radius of curvature of 15 μ gives no perceptible tracing distortion, even on the innermost groove.

With a groove angle of 90° such a radius of curvature for the rounding off of the tip gives an effective needle radius (see fig. 1) of:

$$15 \cos \frac{\alpha}{2} = 15 \cos 45° = 11 \mu.$$

Fig. 8. The Philips type 2978 record player.

Such a fine needle point, 1 mm in length, can indeed be made, but then it could be used only for playing microgroove records, since on a normal record this fine point would ride on the bottom of the groove (see fig. 1*c*), which is not permissible. To be suitable for playing both microgroove and normal records the pick-up would therefore have to be fitted with two different needles, and as a matter of fact this is indeed done upon request for connoisseurs who want the reproduction to answer the very highest demands. For the normal type, however, it is desired to avoid this complication because if the gramophone were to be used thoughtlessly it might lead to trouble. A compromise has been found by making a needle which can be used for records with a normal groove as well as for those with a microgroove. The needle point is made of sapphire and has the shape of a truncated cone with the edges rounded off (see *fig 9*). The effective radius is 26μ, the total length 1mm, the thickness 0.4 mm and the weight 0.2 mg, i.e. approximately one-eighth of the weight of the sapphire needles hitherto used. Thus a needle

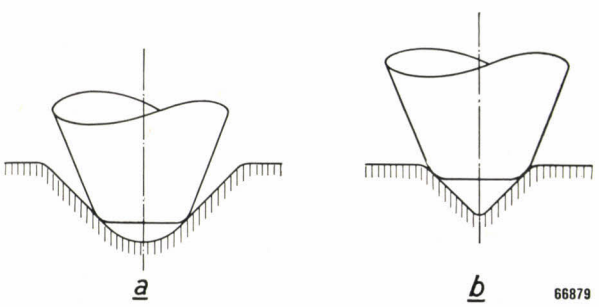


Fig. 9. The truncated sapphire needle in a normal groove (*a*) and in a microgroove (*b*). It is seen that the needle has been given such a shape and dimensions that while it does not ride on the bed of a normal groove it is nevertheless small enough to fit in a microgroove.

has been produced which by its simplicity in use amply compensates the small loss in quality of reproduction of the high notes due to tracing distortion.

The piezo-electric crystal and the armature

The pick-up developed for playing microgroove records is of the piezo-electric type. This appeared to be the most suitable for a simple, light construction which will generate sufficiently high voltages without any additional amplification being required over and above that available in the A.F. part of a radio receiver.

This choice was made in spite of a certain prejudice felt against piezo-electric instruments on account of their susceptibility to atmospheric influences (e. g. very humid or extremely dry climates). Special measures have been taken against this in the packing of the pick-ups for the trade and care has also been taken to provide for adequate protection of the instrument while in use; these measures will be described later.

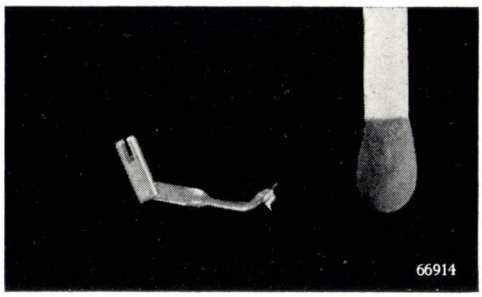


Fig. 10. The drag spring with needle. On the right the head of a match to give an idea of the dimensions.

The sapphire needle point is fixed to a drag spring (see *fig. 10*) connected to the piezo-electric crystal via the so-called armature (see *fig. 11*). Between the fork of the armature and the crystal is a piece of elastic, rubber-like material. The armature is rotatable about its longitudinal axis, with rubber rings serving as bearings. In this way horizontal deflections of the needle are translated into torsional movements of the crystal, which is so sawn out of a larger crystal as to react mainly to those movements. The part of the pick-up in which the crystal is contained is filled with a gel. The whole pick-up weighs no more than 15 grammes. The force exercised by the needle on the record is 7 grammes.

It has been pointed out already that not only the static needle pressure but also the dynamic pressures have to be kept sufficiently low. This means that the mass m_1, which plays a part in the pinch effect, has to be small because the frequency at which m_1 vibrates is twice the modulation frequency. Provision has therefore been made for this by connecting the needle to the armature via the drag spring, thereby coupling the needle to the crystal so loosely, as far as the vertical movements are concerned, that the vibrations of the needle due to the pinch effect are not conveyed to the crystal. The effective mass m_1 has thereby been reduced to a few milligrammes, and as a consequence no trouble is experienced from the above-mentioned distortion accompanying these vertical movements.

Owing to the accelerations taking place in the horizontal direction, the mass m rigidly connected with the needle in this direction has also to be kept small. This means primarily that the armature, which as a matter of fact also yields a small contri-

bution towards m_1, has to be made as light as possible. How drastically the mass of the armature has been reduced may be judged from the fact that the new armature weighs only 30 mg, as compared with the 2500 mg of the armature formerly used, which makes an enormous difference (see *fig. 12*).

The function of the elastic layer between armature and crystal may be explained as follows:

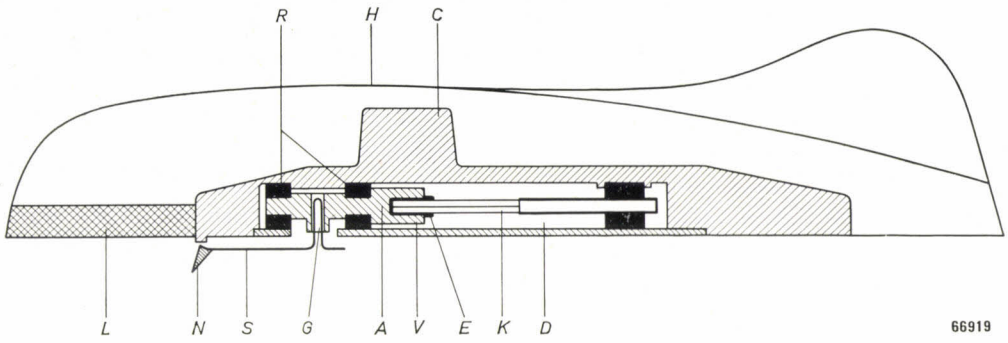


Fig. 11. The pick-up head with its most important parts. All that can be seen of it from the outside is the streamlined cover *H*. In the front of the head is a piece of lead *L* serving to increase the moment of inertia about the vertical axis. The moving parts of the pick-up head are mounted in the chassis *C*. The small rubber rings *R* serve as bearings for the armature *A*. Fitted in the aperture *G* of the armature is the drag spring *S* with sapphire needle *N*. The movements of the armature are transmitted via the fork *V* to the piezo-electric crystal *K*. Between the crystal and the fork is an elastic layer *E* which, owing to its frequency-dependent rigidity, ensures distortion-free sound reproduction. The space *D* around the crystal is filled with a gel protecting the crystal and damping parasitic vibrations.

The rigidity of this material depends greatly upon the frequency at which it is deformed, increasing with increasing frequency, so that vibrations of higher frequencies are conveyed to the crystal with a relatively larger amplitude than those of low frequencies. This linear distortion, together with that originating from the cutting characteristic, is of such a nature that in the reproduction it can quite easily be neutralized by means of a simple *R-C* filter.

Thanks to the presence of the elastic layer the torsional angles made by the crystal at low frequencies are relatively small, but owing to the great sensitivity of the crystal they are still large enough to induce the voltages necessary for good reproduction.

The gel with which the space around the crystal is filled has a two-fold function. In the first place it serves for damping parasitic vibrations likely to arise from resonance effects of various parts. In the second place it affords protection of the crystal against harmful atmospheric influences.

This need of protection has been further provided for in the packing of the pick-up, which is such as to ensure that the crystal will not suffer from long storage in any climate. The packing consists of an air-tight sealed metal tube containing a chemical which maintains the correct moisture content.

It has already been seen how the relatively large accelerations inevitably resulting from the movements of the armature essential for the reproduction have been kept within bounds. When the record is being played however, in addition to these accelerations and the static needle pressure already mentioned, there are also vertical forces of inertia as a result of the fact that neither gramophone records nor turntables are absolutely flat. Notwithstanding the much lower frequencies of the resulting movements, the forces that are brought into play may be considerable, on account of the attendant movement of the mass of the whole pick-up

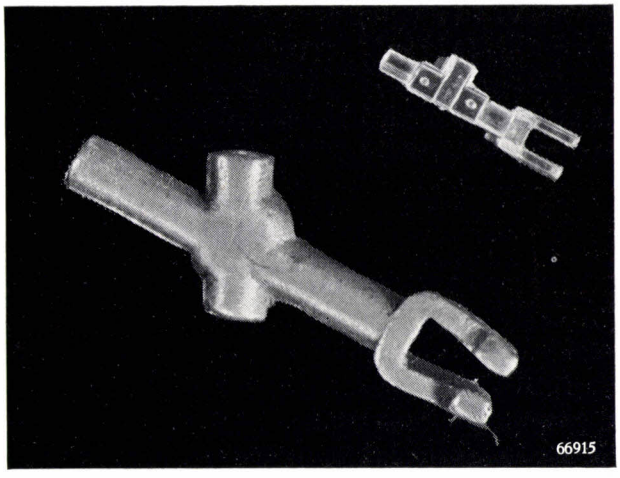


Fig. 12. A new armature (right) and an old one (left). The enlargement of the photograph is the same as that of fig. 10.

head plus the arm. Naturally the problems presented by these forces also become more serious when a sharper needle is used. It will now be explained how these problems have been solved in construction of the pick-up arm.

The pick-up arm

The pick-up arm is made of a very light thermoplastic material. The arm proper, which is remarkably slender, is set at a certain angle to the streamlined head carrying the crystal (see fig. 8). In the front of the head is a small piece of lead. At the other end of the arm are the bearings for the horizontal and vertical movements. A striking feature is the absence of a ball bearing for mounting the arm and the absence of any counterweight for reducing the static needle pressure.

Any such counterweight is superfluous here because the weight of the whole is so small that the correct needle pressure is obtained automatically. In fact the use of a counterweight is a most unsatisfactory method of reducing needle pressure, because then the forces of inertia arising from the periodical accelerations in the vertical direction are increased. Denoting the frequency at which these accelerations take place by $\Omega/2\pi$, the amplitude of the torque that the pick-up arm brings to bear upon the record may be expressed as $K \sim I_{\text{hor}}\Omega^2$, where I_{hor} represents the moment of inertia about the horizontal axis of rotation. Thus the force exercised upon the record is proportional to I_{hor}. From this it is seen that although a counterweight would reduce the static needle pressure, owing to the increase of I_{hor} it would amplify the variations of the dynamic pressure due to the wobbling of the disc. At the moments when this dynamic pressure is increased the record suffers more wear, and half a cycle later the needle pressure may be so small as to cause the needle to be forced upward along the wall of the groove by the horizontal accelerations, thus giving rise to distortion and background noise. In the worst case, if I_{hor} becomes too great, it may even happen that the needle jumps out of the groove.

After this exposition, from which it follows that the moment of inertia about the horizontal axis should be kept as small as possible, it may seem surprising that a weight has been introduced in the head of the pick-up, thereby increasing the moment of inertia I_{hor}. To understand properly the function of this weight it is necessary to consider the moment of inertia about the vertical axis.

The needle, together with the parts rigidly coupled with it in the horizontal direction (total mass m), follows precisely the modulations of the groove. The mass m is thereby displaced with respect to a mass M, the centre of gravity of which projected onto the plane of the gramophone record is, in the ideal case, always at a certain distance from the centre line of the unmodulated groove (see *fig. 13*).

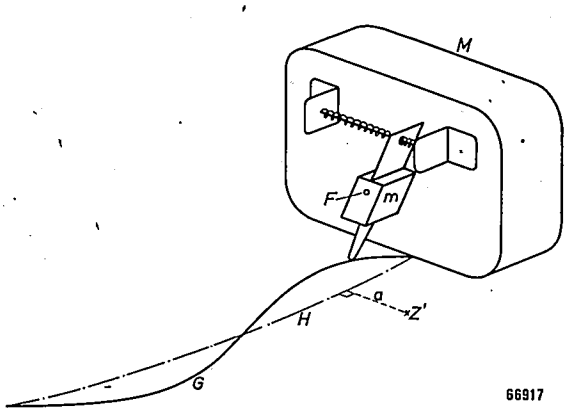


Fig. 13. Schematic representation of a pick-up. A small mass m is connected to a large mass M via a rotational movement about an axis F. Provided it is sufficiently small, the mass m, being rigidly coupled to the needle, follows precisely the modulations of the groove G, thereby describing a movement with respect to M, the centre of gravity of which when projected onto the plane of the record (Z') is at a constant distance a from the centre line H of the unmodulated groove.

In order to approximate this ideal case as closely as possible, the pick-up arm to which M is fixed must have the largest possible moment of inertia about the vertical axis, I_{vert}, for it is only in that case that the pick-up arm will be prevented from following the rapid deflections of the needle in the horizontal direction.

The object of the lead weight, therefore, is to give the pick-up arm a sufficiently large moment of inertia about the vertical axis. Furthermore, being placed in front of the needle, it causes the bearing pressure of the arm to be so small that a complicated and thus expensive ball bearing can be dispensed with. It may seem strange that what is done here is just the reverse of what was to be expected from the arguments previously expounded: the bearing pressure is reduced at the cost of increased needle pressure. But the fact that the weight is placed in the front end of the head results in a minimum increase in needle pressure. The explanation of this is illustrated in *fig. 14*.

The two requirements of small I_{hor} and large I_{vert} are naturally opposed to each other, but the construction indicated proves to be a compromise which in practice satisfies the most stringent demands.

The extent to which, for instance, it has been possible to reduce I_{hor} may be judged from the

following simple but instructive test: when a cigarette is placed between the turntable of the gramophone and a record thus causing an abnormal wobbling of the record, it will be seen that the quality of the reproduction is scarcely affected at all.

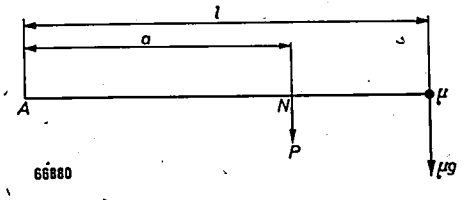


Fig. 14. Increase of the moment of inertia about the vertical axis A. N = needle, μ = the mass applied in order to increase I_{vert}. Assuming that for good sound reproduction a certain increase I_v of the moment of inertia about the vertical axis is needed, then $I_v = \mu l^2$. The increase of the needle pressure due to the mass μ is $P = \mu l/a = I_v/al$. Thus, with a given I_v, P decreases with increasing l.

Another point clearly but simply demonstrating the advantages of the new type of pick-up is connected with the needle talk, which is mainly a result of the vertical movements of the needle and the mass rigidly coupled to it in those directions. Owing to the great reduction of the mass this needle talk has been so minimized that there is no need to enclose the turnable and the pick-up in a cabinet, as is the case with other types of pick-up (see fig. 8).

Finally it is to be mentioned that a record of reasonable quality suffers so little wear that even after it has been played a hundred times there is hardly any noticeable difference in the reproduction. Thus it may be claimed that owing to the extreme care taken in the construction of the pick-up gramophone records have been given an almost unlimited life.

Summarizing, it is seen that thanks to the measures described it has been possible to combine three important features, viz: a longer playing time, better reproduction and very little record wear.

Summary. The playing time of a gramophone record is limited by the width of the groove, the width of the wall between adjacent grooves and the minimum groove speed (linear velocity of the innermost turn). The groove width can be reduced only if a needle is used which has a smaller rounding-off radius of the tip, and this is impossible unless the weight of the pick-up is reduced. The width of wall is determined by the desired amplitude in recording and therefore, due to the signal-to-noise ratio, by the granulation of the record material. The width of the wall can therefore be reduced if a material of a finer structure is used. Owing to tracing distortion and pinch effect the minimum groove speed can be reduced only if all parts of the pick-up are simultaneously so changed that none of the pressures arising during the playing of the record exceeds the limit of elasticity of the record material. The pick-up of the new Philips record players satisfies all these requirements, so that long-playing records ($22^1/_2$ minutes for a 12″ cm record) can be played with an extremely small amount of distortion. The new record players also offer the possibility of playing records at different speeds, so that both ordinary and long-playing (microgroove) records can be played on the same instrument.

VOL. 13 No. 6, pp. 145-180. DECEMBER 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE WIRE CAPACITOR AND OTHER COMPOSITE DRAWN PRODUCTS

by J. L. H. JONKER and P. W. HAAIJMAN. 621.319.4:621.396.6:621.771.31

The attempts that are being made to reduce more and more the dimensions of electrical components have led to the development of a new product, the wire capacitor, which not only has very small dimensions but possesses excellent electrical properties. This product is made by means of a special drawing process. Together with the magnetic material Ferroxcube this drawn capacitor makes it possible, for instance, to construct I.F. transformers which are considerably smaller than those hitherto employed.

The demand for small capacitors

The requirements to be met by electronic apparatus are increasing in number and becoming more stringent day by day. As a result a large number of components have to be incorporated in almost any apparatus, and yet at the same time the apparatus is required to be of limited dimensions. As examples we would mention the electronic computing machines, with their thousands of valves, relays, resistors, capacitors and inductors and the modern military electronic equipment, in which it is imperative that the dimensions and weight be minimized.

As a simpler case, let us consider a normal radio receiver. Generally speaking this contains 5 radio valves (some of which perform more than one function) and about 40 resistors, 20 inductors and 40 capacitors. For the dimensions of such an instrument to be kept within reasonable limits it is necessary that all these components should be made as small as possible. It is also of importance that various components can be combined into one small "unit", as for instance the I.F. band-pass filters, consisting of two inductors and two capacitors incorporated in a small can. Hence the components must sometimes match each other.

Much progress has already been made towards reducing the size of components. The resistors now being used are almost invariably small carbon resistors. The magnetic material Ferroxcube permits of smaller inductors being made. As far as capacitors are concerned, these are of various types. Paper capacitors, which are of rather large dimensions, are used in places where no particularly high requirements are to be met as regards constancy of the capacitance and suchlike. In cases where more stringent demands are made, the small but much more expensive mica capacitors are employed, which are of a better quality. In certain circuits (for higher frequencies) mica capacitors are often replaced by ceramic capacitors, which are much smaller than the other types, at least for the lower capacitances, but when such a ceramic capacitor is required to have a high capacitance and still be of very small dimensions this usually entails a larger temperature coefficient.

In addition to being small, it should also be possible for the components to be easily made within the tolerances required. Furthermore, they have often to answer stringent requirements of constancy in their properties during their life and must not be susceptible to the fluctuations in temperature taking place in any electrical apparatus.

The wire capacitor

The Philips' Works at Eindhoven have developed a capacitor of low capacitance (about 100 pF) which is smaller than any type hitherto

known. This is the wire capacitor. In its most common form it consists of a small metal tube slightly less than 1 mm in diameter and about 5 cm long, in which is contained a metal core about half a millimetre thick, while the space between the core and the jacket is filled with a compressed insulating material with high dielectric constant. The total volume of such a capacitor is no more than 30 mm^3. This is to be compared with the 100 mm^3 of a ceramic capacitor and 1200 mm^3 of a mica capacitor of the same value and suitable for the same voltage (300 V). As regards dielectric losses, temperature coefficient, etc., the wire capacitor is comparable to the mica capacitor. Such drawn capacitor are shown in *fig. 1*.

The narrow space between the core and the jacket is filled with the insulating material in such a way as to make it highly uniform throughout. It would be impracticable to fill the capacitor as such. The basic material for the production of drawn capacitors is a metal tube 20 cm long, with outer diameter about 20 mm and a wall thickness of 2 mm. A wire core about 8 mm thick is inserted in the tube and centred with the aid of two rings acting as jigs. The annular space, which is quite wide enough for this to be done at this stage, is filled with an insulating material in powder form, which is firmly stamped in. The whole is then hammered and afterwards drawn out to a "wire" of about 40 metres in length and with the desired diameter of slightly less than 1 mm. Notwithstanding the great deformations to which the metal and the insulating material are subjected during this process, after the wire has been drawn the core and the insulation are still in place and a firm product is obtained.

The drawn wire is divided into pieces of the desired length, and at one end of each piece the jacket and insulation are removed, thus leaving the core bare. Connecting wires are then soldered onto the bared end of the core and onto the jacket. *Fig. 2* shows the product in the various stages of manufacture.

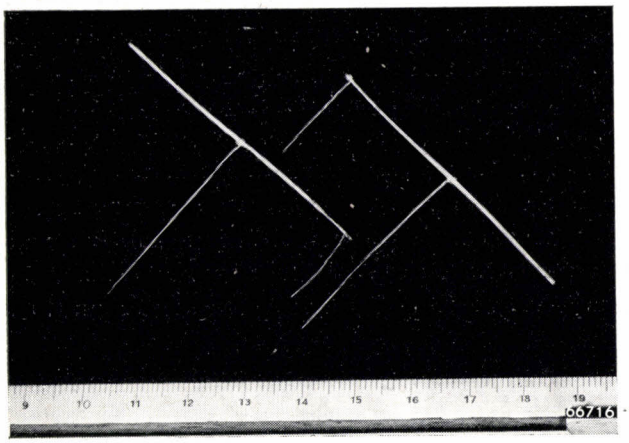


Fig. 1. Two wire capacitors.

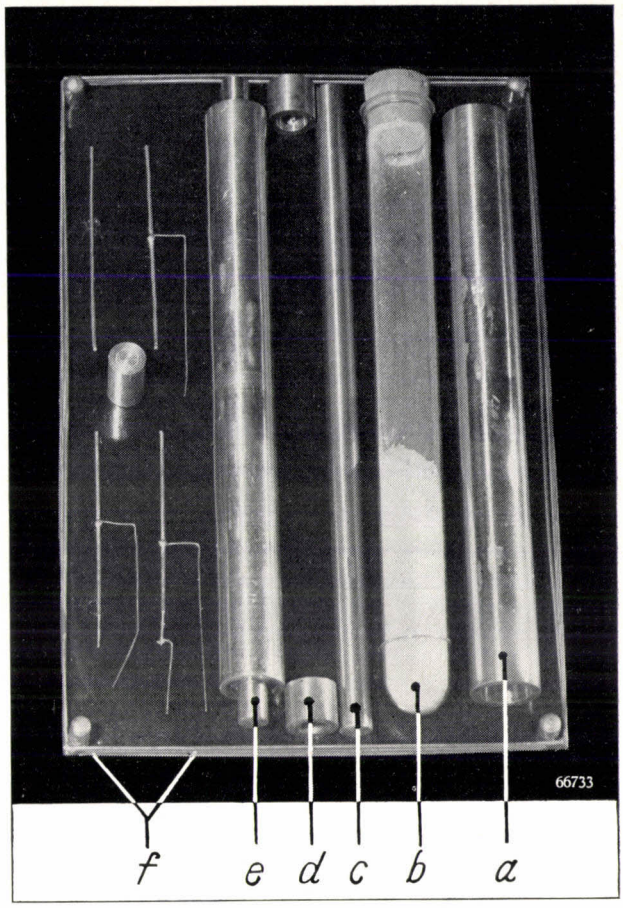


Fig. 2. Composition of the wire capacitor:
a) the jacket before drawing;
b) the dielectric with which the capacitor is filled;
c) the core prior to drawing;
d) jig rings for centering the core in the jacket prior to drawing (and after filling);
e) the composite system ready for drawing;
f) the drawn tube broken at the required length (upper left), with one connecting wire soldered on (upper right), the outer jacket partly removed by etching (lower left), with the second connecting wire soldered onto the core (lower right). Part of the 40-metres long, drawn "wire" is stretched round the whole.

In the development of this process several problems were encountered, which will be dealt with in this article. In conclusion some applications of the wire capacitor will be discussed.

In addition to capacitors, other products can also be made in the manner described here. We have in mind, for example, indirectly heated cathodes for electron valves, which likewise consist of two conductors (the filament and the cathode surface) separated by an insulating layer. Drawn cathodes can be used for various applications where normal cathodes fail, as will be shown. There are also other products of a similar construction which can be made by the drawing process.

Composition and manufacture

Materials

The materials from which the wire capacitor is built up have to satisfy certain demands. For instance, the metal of the outer wall and the inner conductor must lend itself readily to drawing, and at the same time it should show little tendency to oxidize, because oxidation would reduce the quality of the capacitor. As regards the dielectric the following conditions, among others, have to be satisfied:

1) It must have great permittivity (ε).
2) The dielectric losses must be low, tan δ (δ is the loss angle) should be less than 10^{-3}.
3) The temperature coefficients of ε and tan δ must be small; when used in electronic apparatus the capacitor is subject to temperature fluctuations, and these must not greatly affect the capacitance and the loss angle. This condition is already partly satisfied if the second requirement is fulfilled.
4) Only small differences are allowed between the thermal expansion coefficients of the dielectric and the metal, a point which will be reverted to later.
5) The insulating material must be easily pulverized; the grains may not be too hard, nor too large, as otherwise, during the process of drawing, they would scratch the surface of the metal wall and the inner conductor, thereby increasing the area liable to oxidation.

Extensive investigations have led to a combination of materials being found with which a capacitor can be made that has small losses (tan $\delta < 6 \cdot 10^{-4}$) and a very low temperature coefficient.

Details of manufacture

The 40-metre long "wire" obtained by drawing is stretched and divided into pieces of the length required. When dividing the wire there must not be any short-circuiting between the inner and outer conductors. There is no danger of this, however, since the jacket always bends somewhat outward when being broken. The insulating material is so firmly packed between the core and the jacket that there is no possibility of any of it falling out of the capacitor along the plane of rupture.

To bring about electrical contact with the inner conductor, the jacket and the insulation have to be removed over part of the length of the capacitor. This also affords an opportunity to give the capacitor exactly the desired value, by removing just as much of the jacket as is necessary to reach that capacitance. (The capacitors must therefore have a somewhat higher capacitance before part of the jacket is removed.) In practice the jacket is removed by electrolytic etching, with the jacket serving as anode. A hundred of the capacitors at a time are immersed in the etching bath to a depth corresponding to the length of jacket to be removed, and a voltage is then applied between all the jackets and an electrode placed in the bath. The current passing through the bath is continuously controlled; by the time that the metal of the jacket has dissolved the current has dropped to zero (or practically zero); the inner conductor of the capacitor cannot act as an electrode because it is insulated from the outer conductor. The insulating material is so firmly compressed during the drawing process that after the outer jacket has been etched off it does not of itself break away from the inner conductor, but has to be removed separately. After the etching the capacitors are washed in water in order to remove any traces of the etching medium, which might cause conductance between the inner and outer conductors. This treatment does not in any way impair the quality of the capacitors.

An entirely different method of removing the jacket is to shave it off with a lathe. For this purpose a special lathe has been designed, which is characterized by the fact that the cutter rotates, while the capacitor is clamped tight close up against the cutter; this is necessary on account of the natural lack of rigidity of the thin capacitor. The capacitor

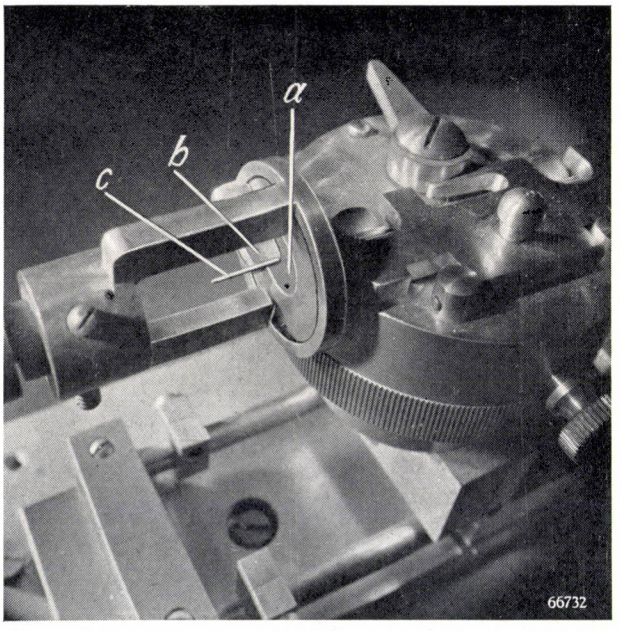


Fig. 3. Part of the lathe on which the jacket of the capacitor is removed over a length of about 1 cm. The capacitor is moved from right to left by the guiding die *a*. The rotating cutter (of diamond) *b* can just be seen behind the (turned) capacitor *c*. The automatic regulating device is not shown in this photograph.

is centred in the lathe by means of a diamond guiding die, while a diamond cutter shaves off the jacket over the length necessary to give the capacitor the value desired. The capacitance is measured while the jacket is being shaved off and as soon as it reaches a value within the required tolerances the rotating cutter is automatically stopped. By this shaving process it is possible to reach a tolerance of 0.2% in capacitance. By the etching method the degree of accuracy that can be reached is not so high, so that after this process the capacitors have to be sorted. *Fig. 3* shows a part of the lathe in question.

Physical properties of the wire capacitor

This is not the place to discuss fully the electrical properties of the wire capacitor, but we shall consider briefly two very important properties which determine the quality of this component and are largely influenced by the remarkable process of manufacture. These properties are the temperature coefficient of the capacitance, to be denoted here by α, and the losses in the capacitor, which are proportional to tan δ.

Temperature coefficient

Although every possible precaution is taken to ensure that the capacitor is filled as compactly as possible, air cavities between the grains of the powder are unavoidable. When the temperature of the capacitor is raised, both the jacket and the core as well as the grains expand. Since the expansion coefficients of the materials are not exactly equal, the "filling factor" of the capacitor changes with its temperature. The permittivity ε of the powder likewise changes with the temperature. Consequently the capacitance of the capacitor varies.

The materials, therefore, have to be so chosen that the temperature coefficient due to the difference in expansion coefficients is compensated as completely as possible by the temperature coefficient of the powder. Since, however, the materials have to satisfy many other requirements, complete compensation is very difficult to obtain. The temperature coefficient of the permittivity of the powder used is slightly negative. It appears that the powder expands somewhat more than the metal of the jacket and the core, so that upon the temperature being raised the filling factor is increased, which in itself leads to a positive temperature coefficient. For the resultant relative temperature coefficient of the capacitance, $\alpha = 1/C \cdot dC/dT$, we find (at temperatures up to about 60 °C) roughly $+1\times10^{-4}$ per degree centigrade for a capacitor having a copper jacket and a copper core. A capacitor with a copper jacket and an aluminium core shows a higher value of α, viz. about $+5\times10^{-4}$ (see *fig. 4*). As fig. 4 shows, the curve representing $\Delta C/C$ as a function of T describes a loop: the variation in capacitance while heating up differs from that during the cooling of the capacitor. The curves also show maxima, especially those for the capacitor with copper core. No exact explanation has yet been found for these two phenomena; they no longer occur when the capacitor has been aged, in a manner presently to be discussed.

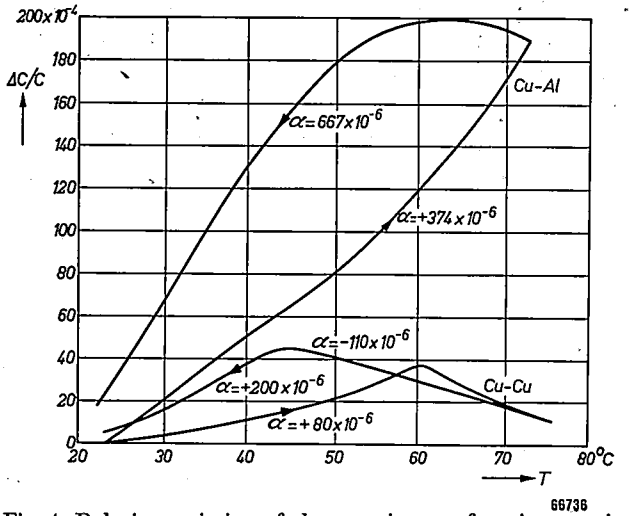


Fig. 4. Relative variation of the capacitance of a wire capacitor as a function of the temperature. No exact explanation has been found for the fact that the curve for increasing temperature differs from that for decreasing temperature. After the capacitor has been aged this "hysteresis loop" disappears. The two loops are respectively that of a capacitor with copper jacket and copper core and that of a capacitor with copper jacket and aluminium core. The latter shows a much greater temperature coefficient (slope of the curve). The values of the temperature coefficients are given in the drawing.

The temperature coefficient of the capacitance can be calculated if the expansion coefficients of the materials and the temperature coefficient of the permittivity are known. The results of calculations made agree well with the temperature coefficients as actually observed. Partly with the aid of these calculations it has been found possible to control the temperature coefficient within certain limits by varying the ratio of the dimensions of the jacket and the core.

Dielectric losses; ageing of the capacitor

The hammering and drawing processes cause the temperature of the capacitor tube to rise. Particularly during the hammering, when the temperature reaches about 80 °C the metal (copper) of the jacket and the inner conductor tends to oxidize, forming, for instance, Cu_2O, which is a semiconductor and thus causes losses. This oxidation is, of course,

limited by the amount of air available inside the capacitor, which amounts to about 0.6 vol. %. The fact that oxidation is indeed the main cause of the losses is proved when a non-oxidizing metal (e.g. silver) is used for the jacket and the core; tan δ is then much smaller than the value of 6×10^{-4} mentioned in this article.

During the fluctuations in temperature, which regularly take place in practice, the capacitor alternately ejects and draws in air, so that the air inside the capacitor is partly renewed each time. Consequently it is to be expected that oxidation is started again time after time. And, indeed, tan δ appeared to increase, at least for some time after the capacitor had been made. Later, however, no further change in tan δ was to be observed, probably due to the whole of the surface being covered with a very thin layer of oxide. In order, therefore, to avoid trouble from an increase in tan δ during use, the capacitors are aged before delivery by keeping them for a few hours in an oven in which the temperature is alternately raised and lowered. The capacitors aged in this way have a constant tan δ also immediately after manufacture.

Applications of the wire capacitor

The wire capacitor has been specially developed for use in I.F. transformers. The principle of such a transformer is represented in *fig. 5*. It consist of two mutually magnetically coupled L-C circuits.

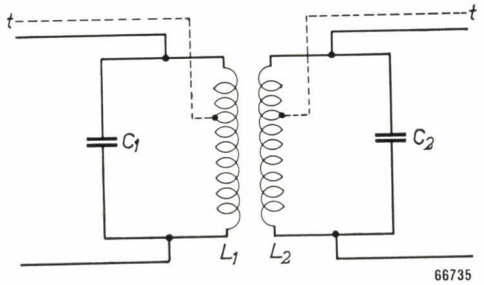


Fig. 5. Principle of an I. F. transformer consisting of two magnetically coupled L-C circuits. The tappings denoted by t can be used when it is desired to reduce the influence of resistances and capacitances outside the filter.

The two coils L_1 and L_2 are each wound directly on a core of Ferroxcube, a ceramic magnetic material with high permeability [1]). The coupling is brought about by a rod of ferroxcube. The exceptional properties of ferroxcube, compared with the powder cores formerly used, allow of the coils being made so small that when mica capacitors are used the

[1]) J. L. Snoek, Non-metallic magnetic material for high frequencies, Philips Techn. Rev. **8**, 353-360, 1946.

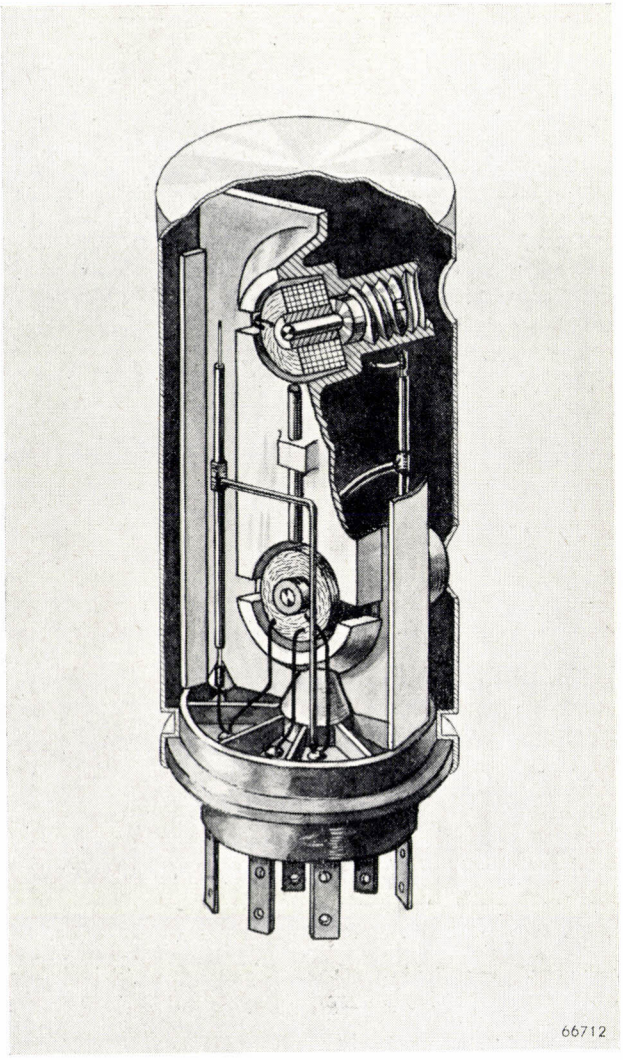


Fig. 6. Drawing of the I. F. transformer, type 5730, cut open. Note the ferroxcube cores, which can be screwed farther in or out of the coils by means of screws from the outside of the transformer. The wire capacitors can be seen on the left and right sides; the vertical rod in the middle is the ferroxcube coupling rod.

dimensions of the filter are almost solely determined by those of the capacitors. Thus it was desired to replace the mica capacitors by smaller ones, for example by ceramic capacitors. Now, however, the wire capacitor proves to be a still better solution for the I.F. transformer, since it is much smaller and, moreover, for the capacitances required it has a lower temperature coefficient than the smallest possible ceramic capacitor. *Fig. 6* is a drawing of a type 5730 I.F. transformer, cut open, in which two wire capacitors are employed; *fig. 7* is a photograph showing the external appearance of these I.F. transformers. The dimensions of this cylindrical I.F. transformer are 60 mm $\times$ 27 mm, and the bandwidth for which it is suitable is 18,000 c/s at the (standardized) intermediate frequency of 452 kc/s. (I.F. transformers are also being made for

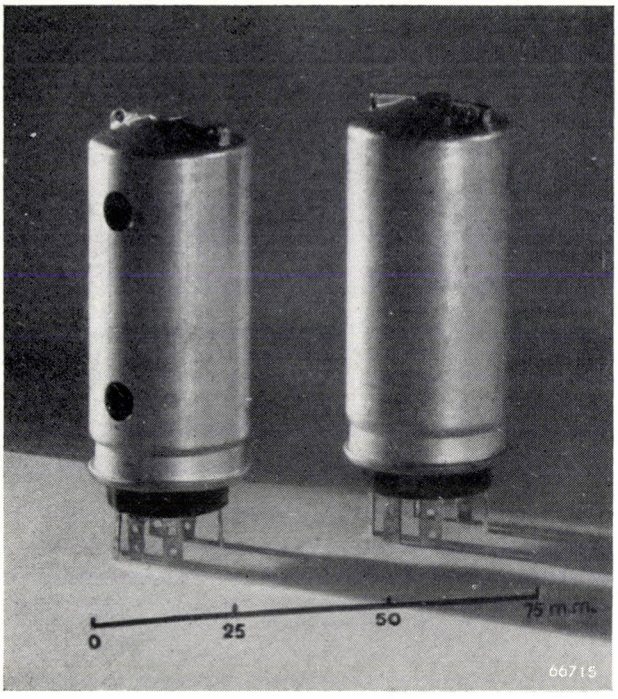


Fig. 7. The type 5730 I.F. transformer. The openings are made in the side for adjusting the screws with which the cores can be screwed in and out of the coils.

slightly different frequencies.) The circuits can be tuned by screwing the ferroxcube cores more or less deep into the coils, it being possible to do this from the outside. The temperature coefficient of the I.F. transformer is very small; for 1 °C variation in temperature the frequency drift is no more than 5 c/s.

In recent years Philips have brought onto the market a still smaller I.F. transformer of very similar properties, the dimensions of which are 36 mm × 25 mm × 10 mm. This has been made possible by employing somewhat shorter wire capacitors of the same value. Two of these "microtransformers" (type AP 1000) are illustrated in *fig. 8*.

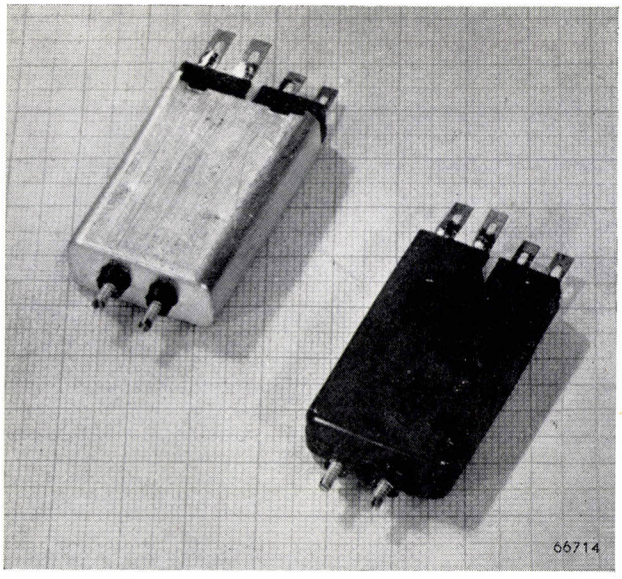


Fig. 8. The type AP 1000 microtransformer. The inductance of the coils can be adjusted with the two screws. The filter on the right has had its screening aluminium jacket removed.

The wire capacitor is likewise often employed in simple high-frequency tuning circuits. These consist of a small coil with a movable ferroxcube core and a wire capacitor. By screwing the core farther in or out the coil it is possible to vary the resonant frequency of the circuit. This device replaces the conventional tuning circuits built up from one or more coils and one or more variable capacitors. *Fig. 9* is an illustration of a tuning unit consisting of a high-frequency circuit and an oscillator circuit.

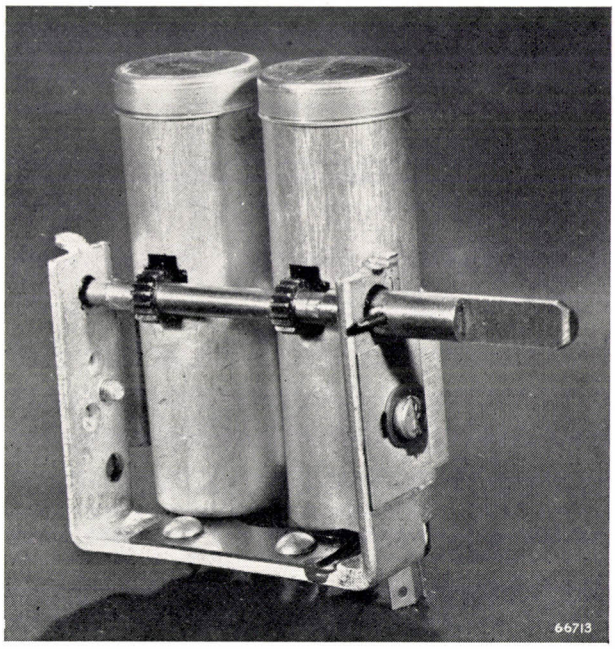


Fig. 9. High-frequency tuning unit, consisting of an aerial circuit and an oscillator circuit. Tuning is done with the aid of the spindle carrying gear wheels, each engaging a toothed rail affixed to the cores of the coils.

Other composite drawn products

As already stated, the drawing process can be employed also for making indirectly heated cathodes. Such cathodes consist of a filament separated by an insulating layer from a metal envelope which in turn is coated with an electron-emitting substance and forms the cathode proper. Thus in principle the construction of the drawn cathode does not differ from that of the wire capacitor, the only difference lying in the materials used and the dimensions.

Owing to the solid construction of the drawn cathode there is very good thermal contact between

the filament and the cathode surface. As a consequence the difference in temperature between filament and surface is no more than some tens of degrees, whereas in a normal indirectly heated cathode the temperature of the filament is usually some hundreds of degrees higher than that of the emitting surface. It is therefore possible to use nickel for the filament in a drawn cathode (as also for the outer jacket), so that one is no longer bound to use high-melting metals such as tungsten.

The filament is made as thin as possible so as to minimize the filament current required. The diameter of the outer jacket depends upon the kind of electron valve in which the cathode is to be used; it varies from a few mm to $^1/_4$ mm. Thus it is possible to make a drawn, indirectly heated cathode much smaller than a conventional indirectly heated cathode.

The drawn cathode can be rolled or laminated, forming a ribbon-shaped cathode, without the filament making contact with the jacket. The cathode can also be bent and wound into various shapes. For instance, the cathode can be rolled up into a cylinder, closely resembling a cylindrical, hollow, cathode. It can also be wound into a grid-shaped helix, with the turns spaced apart, which is a great advantage for special applications, or into a flat spiral like a watch spring, closely approximating a flat cathode (see *fig. 10*). Thus the drawn cathode makes it possible for the desired cathode shape to be obtained in a simple manner for all sorts of electron valves,

As other applications of the composite drawing process we would mention here the manufacture of heating elements, of thermo-elements, and the making of vacuum-tight lead-in ports serving, for example, for reaching the inside of an engine cylinder.

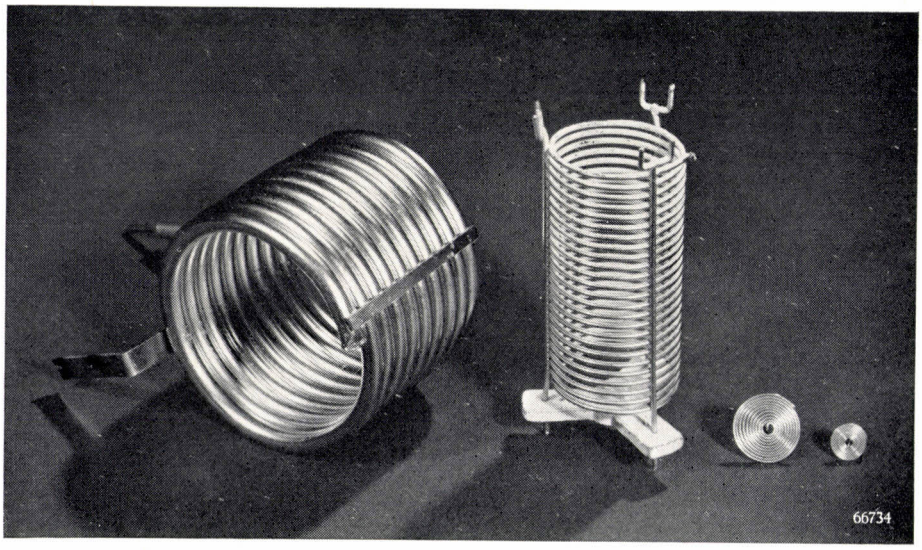


Fig. 10. Various forms into which a drawn cathode can be shaped. From left to right: cylindrical cathode, helical ("grid") cathode, two flat cathodes in the form of a watch spring.

Summary. By means of a special drawing process — the drawing of a tube containing a wire core, with the space in between filled with an insulating material — it has been found possible to manufacture capacitors possessing very good properties. These wire capacitors, being of the smallest dimensions hitherto known, are ideally suitable for use in assembled parts, such as I.F. transformers. The special process of manufacture sets high requirements for the materials from which the capacitor is made. Some details of the manufacture are described, such as the baring of the inner conductor, which can be done either by etching or by shaving on a lathe. A discussion of the temperature-dependency of the capacitance shows that to a certain extent this can be controlled by the choice of material and dimensions. The dielectric losses, which are small, appear to be caused by oxidation of the metal; this makes it necessary to age the capacitors. Some applications of the wire capacitor are discussed, followed by a brief reference to the drawn cathode, an indirectly heated cathode with special properties, e.g. exceptionally easy shaping.

A SIMPLE IONOSPHERE SOUNDER

by R. ASCHEN *) and P. GAILLARD *).

621.396.91

Apart from its great scientific value the investigation of the ionosphere is also of direct practical value, especially for predicting the most favourable wavelength for a radio communication between different points on the earth at specified times. The American Bureau of Standards, for instance, issues monthly predictions, based on investigations of the ionosphere at various places on the earth, from which the most favourable wavelength for a given radio link can be deduced with a fair degreee of certainty. P.T.T. services, radio and airways companies and suchlike are making good use of this information. This explains the increasing interest being shown in the investigation of the ionosphere, a fact which is to be welcomed because the more information that is collected from observations the more accurate will be the predictions.

One of the ionosphere sounders designed by the authors has purposely been kept as simple as possible so that it may be brought within the reach of serious radio amateurs, who — after having "discovered the short waves" — may yet again yield an important contribution towards the common good by regularly carrying out ionosphere soundings.

As may be kown, it is convenient to divide the atmosphere of the earth into zones, the most important of which are called the troposphere, the stratosphere and the ionosphere. The transitional layer, rich in ozone, between the last two zones is called the ozonosphere[1]). *Fig. 1* is a graphical representation of this division. In this article we are concerned only with the ionosphere.

One of the reasons why the investigation of the ionosphere is now being taken up on such a large scale is that the ionosphere has such a considerable influence upon radio traffic. This immediately became a matter of importance at the beginning of this century when Marconi first established transatlantic wireless communication and the question arose as to how the radio waves were able to follow the earth's curvature. The hypothesis — proposed by Kennelly and Heaviside independently of each other — is that in the atmosphere there is an ionized layer acting as reflector for radio waves and making it possible for them to be propagated over long distances in one or more "hops". Though this hypothesis dates from 1902, it was not until 1924 that Appleton succeeded in proving the existence of such a layer, and even of several layers, which he named the D, E and F regions, the last one being split up in the daytime into two layers (F_1 and F_2; see fig. 1).

Appleton carried out his investigations by using radiotechniques, and even at the present day the same method (in a somewhat modified form) is being mainly followed, but it is no longer the only method. Direct exploration by means of instruments sent up into the ionosphere has now been made possible by the adoption in America

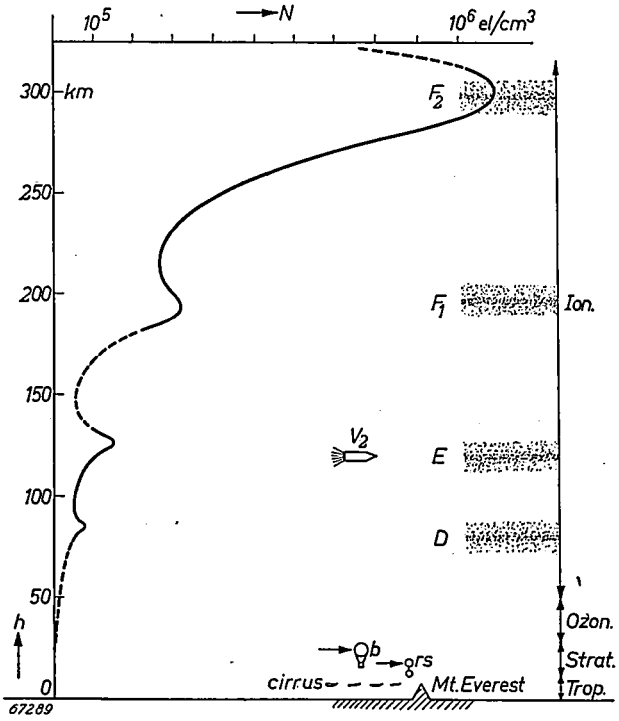


Fig. 1. Division of the atmosphere into the troposphere, the stratosphere, the ozonosphere and the ionosphere. The vertical scale gives the height h. The greatest heights reached by a normal radio sonde balloon (*rs*), a manned balloon (*b*) and — as far as is known — by a V2 rocket are indicated to scale. D, E, F_1 and F_2 are the ionized regions present in the ionosphere in daytime. At night the D region is entirely absent and the E region has disappeared for the greater part, while the F_1 and F_2 layers are combined into the F region. The height of these layers, especially that of F_1 and F_2, is changeable. The curve gives the electron density N as a function of h.

*) Laboratoire d'Applications Electroniques Philips-France, Paris.

1) Proposals have been made for a more detailed and more logical classification. See S. Chapman, Upper atmospheric nomenclature, J. geophys. Research **55**, 395-399, 1950 (No. 4).

of the technique of V2 rockets. With these rockets it has been possible to measure the concentration of ions at an altitude of 100 km [2]) and also the temperatures in or above the E layer. Unfortunately this is a very expensive technique, even when employing the small rocket ("aerobee") specially designed for physical observations. For this reason alone it is not suitable for use at many places on earth and for a number of times per day, as is necessary if a good picture is to be formed of the state of the whole of the ionosphere and of the changes to which this is continually subject.

For such routine investigations on a large scale it is, therefore, still exclusively radio methods that are to be considered. Five years ago an account of the results of this radio investigation was published in this journal [3]). Some of these results will be reviewed in the present paper, dealing rather more with the practical value and the method of the investigations, and this will be followed by a description of a simple apparatus which has been designed for investigating the ionosphere.

Principle of an ionosphere sounder

For quite a number of years the investigation of the ionosphere by radio has no longer been carried out by the method employed by Appleton (a description of which can be found in the article by Bakker quoted in footnote [3])), but by a system developed by Breit and Tuve, upon which radar is also based. Short wave trains are sent out by a transmitter (*fig. 2a*), and a receiver in the immediate vicinity picks up both the direct wave and the reflections from some object (in the case of ionosphere investigations, the reflections from one of the ionized layers in the atmosphere). While the transmitter is working, the signal received direct has to be limited in order to avoid adverse effects. Both the direct signal and the reflected signal as picked up by the receiver are displayed on the screen of a cathode-ray tube. In the simplest form of the apparatus, which is called the ionosphere sounder, these signals appear as a vertical deflection in the form of pulses, while in the horizontal direction a linear time base is traced, synchronized with the repetition frequency of the wave trains. This frequency is chosen such that when the wave train received direct produces a pulse on the extreme left of the oscillogram, the echo received from the ionosphere produces a pulse on the screen at a measurable distance to the right of the first one fig. 2*b*). The known time-base frequency gives the echo time t_e between the two pulses and from this can be found the altitude h' of the reflecting layer ($h' = \frac{1}{2}ct_e$, where c is the velocity of light). A kilometre scale drawn along the time axis gives direct readings of h'.

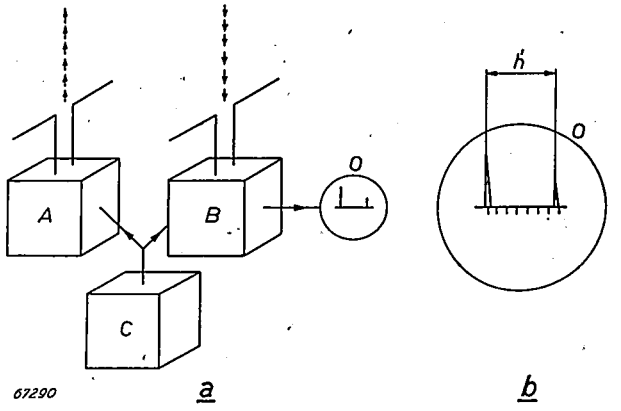


Fig. 2. Schematic representation of the principle of an ionosphere sounder according to Breit and Tuve.
a) *A* the transmitter sending out pulse-shaped wave trains, *B* the adjacent receiver connected to the cathode-ray tube *O*. The pulses for the transmitter and the time-base voltage for the oscilloscope are supplied by the control unit *C*.
b) When the wave is reflected back from the ionosphere two pulses are seen on the oscilloscope, one from the signal received direct and the other as an echo from the ionosphere. The distance between the two pulses is a measure for the (virtual) height h' of the reflecting layer in the case of perpendicular incidence.

Two characteristic phenomena of ionospheric reflection are the following:

1) the quantity h' measured in the manner described appears to be dependent upon the frequency f of the transmitter, and
2) the result varies with the time of day, the season and the phase of the year with respect to the eleven-year cycle of the sun-spots, whilst it also shows irregular fluctuations.

These two effects call for a closer consideration.

Measurement as a function of the frequency

It is not surprising that h' depends upon f when it is borne in mind that the signal is not reflected from a sharply determined surface but from a layer whose electrical properties vary continuously with altitude, this corresponding to a gradually changing refractive index. These electrical properties depend primarily upon the density of the free electrons present in the ionosphere. The influence of the ions, likewise present, upon the propagation of the radio waves can be ignored, because their relatively large mass as compared with that of the electrons makes their effect negligible.

[2]) A. Reifman and W. G. Dow, Phys. Rev. **75**, 1311-1312, 1949 and **76**, 987-988, 1949.
[3]) C. J. Bakker, Radio investigation of the ionosphere, Philips techn. Rev. **8**, 111-120, 1946. See also T. W. Bennington, Short-wave radio and the ionosphere, Iliffe & Sons Ltd., London, 1950.

In each of the layers the electron density is greatest and the refractive index smallest in the middle of the layer. An incident ray with frequency f_1, for instance, therefore follows a partly curved path as represented in *fig. 3a*. The greatest height it reaches, h_{max1}, is less than the virtual height h_1' found from the echo time. (It is only for the sake of clarity that in fig. 3 the distance between A and B has been drawn rather large. In practice the transmitter and the receiver for ionosphere sounding are placed so close together that the distance can be ignored, so that the angle of incidence α may be taken as 0.) In the curved part of the path the velocity [4]) is less than c.

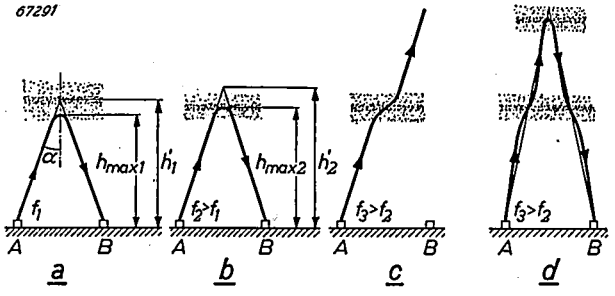


Fig. 3. The path followed by radio waves in ionosphere sounding (here for the sake of clarity the transmitter A and the receiver B are placed so far apart that the angle of incidence α is much greater than it is normally).
In the case (*a*) the frequency (f_1) is so low that the wave is already turned back to earth when it reaches the bottom of the reflecting layer. In the case (*b*) the frequency (f_2) is a little higher and the waves penetrate deeper into the layer. In the cases (*c*) and (*d*) the critical frequency is exceeded and the waves pass right through the layer. In the case (*d*) there is a higher layer reflecting the waves back to earth. h_{max} = the greatest height reached by the reflected ray, h' = the virtual height measured.

If a somewhat higher frequency is chosen (f_2) then either a greater density of electrons or a longer refractive path is required to reflect the ray completely. It therefore penetrates deeper into the layer ($h_{max2} > h_{max1}$, fig. 3*b*) and a greater virtual height h'_2 is recorded. Upon the frequency being increased still further a limit is reached where the ray penetrates into the zone of maximum electron concentration N and minimum index of refraction. When the frequency exceeds this limit the ray passes right through the layer. The frequency limit in the case of perpendicular incidence ($\alpha = 0$) is called the critical frequency (f_{crit}). From the mechanism of the interaction between the free electrons in the ionosphere and the field of the incident radio wave it is possible to derive a simple relation between the maximum electron density N_{max} and the critical frequency (see, e.g., the article by Bakker quoted in footnote [3])):

$$N_{max} = 1.24 \cdot 10^{-8} \cdot f_{crit}^2, \quad \ldots \quad (1)$$

where N_{max} is expressed in cm^{-3} and f_{crit} in c/s.

For a frequency higher than f_{crit} the electron density of the layer is not great enough to deflect the ray back to earth. There are then two possibilities: either the ray disappears into space (fig. 3*c*) or it is reflected back to earth by a higher layer with sufficient density of electrons (fig. 3*d*); examples of the latter phenomenon will be given later.

The critical frequency for a certain layer — i.e. the highest frequency at which an echo can be received from that layer at normal incidence — can in most cases be sharply determined. According to eq. (1) the value found for f_{crit} gives at once the greatest electron density in the layer.

From fig. 3 it is seen that the virtual height h' as found by measurement is always greater than the maximum height h_{max} reached by the radio wave. As long as f is much lower than f_{crit} the difference between h' and h_{max} is very small, but at frequencies near f_{crit} the velocity in the layer is much smaller than c, so that the echo time for an observed echo is thereby increased and the value of h' deduced from it is much greater than the height at which the layer has the greatest electron density.

These phenomena find expression in the graph obtained by plotting the measured values of h' as a function of the frequency, which is increased in small steps from, say, 1 to 12 Mc/s. *Fig. 4a* shows the situation as found about noon on a winter day at a temperate latitude.

Between 1 Mc/s and 1.5 Mc/s no echo at all is observed, this being due to absorption of the radio waves in the D region, which particularly develops in daytime. This region is also responsible for the well-known fact that many remotely situated broadcasting stations working on a wavelength between 600 and 200 m cannot be heard in the daytime, whereas in the evening they can be picked up quite well. The D region, having a critical frequency much lower than 1 Mc/s, does not reflect waves with frequencies higher than 1 Mc/s.

At $f = 1.5$ Mc/s an echo can be observed corresponding to a virtual height of slightly over 100 km: the E layer. This proves that for this and higher frequencies the absorption in the lower-lying D region is sufficiently reduced to make reception possible. As explained with reference to fig. 3, as the frequency is raised so h' gradually increases, until at $f = 3.2$ Mc/s the critical frequency for the

[4]) Here the group velocity is meant, and not the phase velocity.

E layer is reached: h' then suddenly increases from about 200 km to 300 km. This means that from this frequency onward the E layer suddenly becomes "transparent" and that there is a layer at a higher altitude — the F_1 layer — reflecting the radio waves

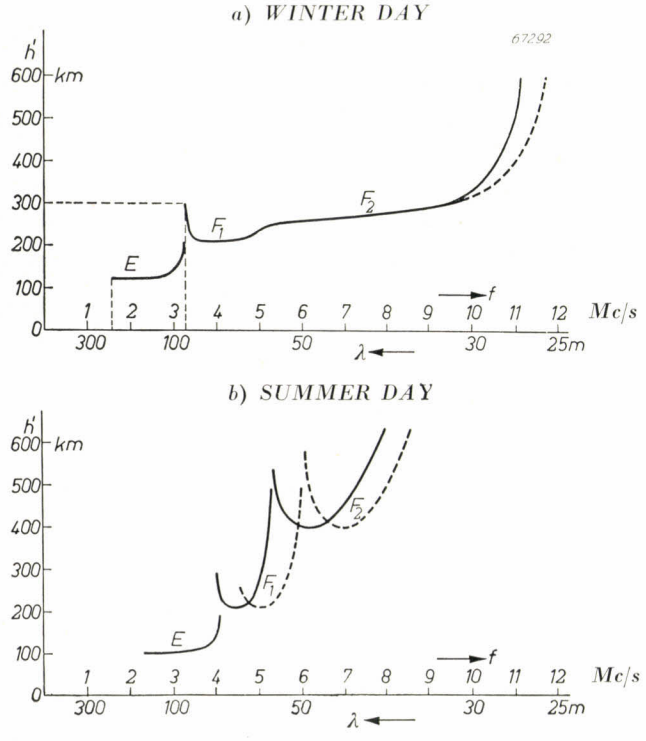


Fig. 4. Examples of sounded virtual heights h' as functions of the frequency f or of the wavelength λ, (*a*) at noon on a winter day, (*b*) at noon on a summer day, both on a temperate geographical latitude. The letters E, F_1 and F_2 denote the layers acting as reflector. The curves drawn in broken lines relate to the "extraordinary wave" formed by double refraction as a result of the earth's magnetism.

(and thus apparently having a greater electron density than the E layer). The great retardation to which signals of about the critical frequency are subjected in the E layer finds expression in the steepness of the curve either side of the discontinuity.

The fact that the greatest value of h' immediately preceding the discontinuity — about 200 km — is much greater than the actual height of the E layer has been proved by other observations, from which it has been possible to conclude that the effective thickness of this layer is no more than a few kilometres (1 to 4 km), whilst the height of its centre is about 120 km.

As the frequency is increased further, h' at first diminishes rapidly (due to reduced retardation in the E layer) but soon begins to rise again gradually. At about 5 Mc/s the critical frequency of the F_1 layer is passed, after which the higher F_2 layer takes over the function of reflector. At this transition there is often no decided discontinuity in the curve owing to the fact that in winter there is no greatly pronounced separation between these two layers. The absence of a long flat part in the curve is due to the fact that the two F layers are much less sharply defined than the E layer.

From about 9 Mc/s onward a peculiar phenomenon arises: the echo pulse in the oscillogram is seen to split up into two parts, corresponding to the two branches of the right-hand part of the curve in fig. 4*a*. This phenomenon is related to the earth's magnetic field causing double refraction, so that an incident wave is reflected back as an "ordinary" and an "extraordinary" wave [5]). It yields two critical frequencies for the F_2 layer, the difference between them (mostly about 0.7 Mc/s) depending upon the terrestrial magnetic field strength at the point of reflection; as far as the present considerations are concerned it may be ignored. The fact that this double refraction is apt to occur also at lower frequencies is proved by the oscillogram in *fig. 5*, recorded at a frequency of 5.75 Mc/s with the ionosphere sounder to be described.

As the frequency is still further increased there is no longer any echo to be observed, from which it is to be concluded that above the F_2 layer there are no other layers having a sufficient density of electrons to reflect the waves penetrating through the F_2 layer.

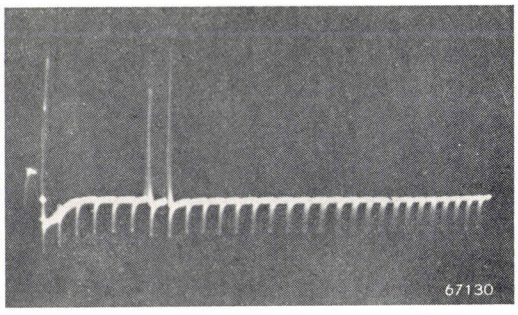


Fig. 5. An oscillogram recorded with an ionosphere sounder (at Paris, 30th March 1951 at 16 hrs) at a frequency of 5.75 Mc/s. On the extreme left the pulse received direct. To the right of that are two echo pulses caused by double refraction in the ionosphere. The interval between the pulses directed downward corresponds to a difference in height of 50 km.

Others factors affecting the results of a sounding

The values given for the critical frequencies in the preceding section are to be regarded as being typical for the situation about noon on a winter day. At other times and in other seasons greatly different values are often found. Considering that in the first place the sun is responsible for the

[5]) See, i.a., the literature cited in footnote [3]).

ionization of the atmosphere, it is not surprising that the position of the earth with respect to the sun, and more particularly the local time and the geographical latitude of the point in the atmosphere considered, will be of great influence. Fig. 4*b* gives a graph for $h' = f(f)$ for a temperate latitude at at noon on a summer day. The occurrence of a pronounced critical frequency of the F_1 layer, the great difference in altitude between F_1 and F_2, and the lowered critical frequency of F_2 are factors which contrast with the situation in winter.

Fig. 6 gives an idea of the variation in the course of 24 hours for a summer day and a winter day. The expectation that the ionization will be most

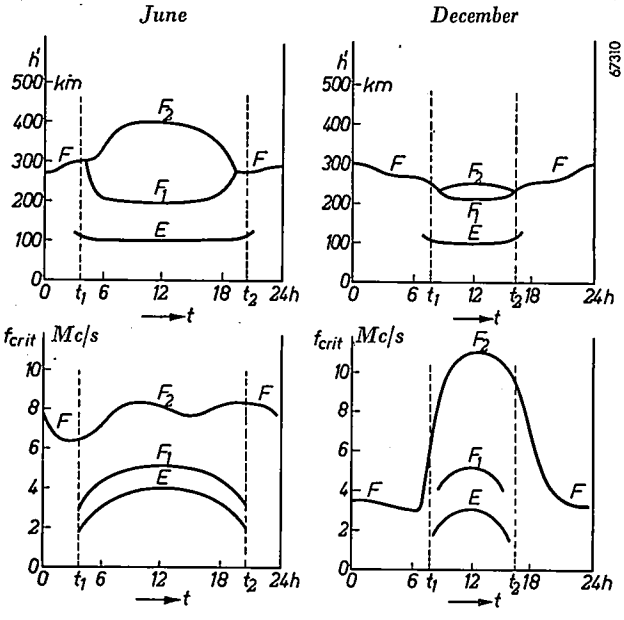


Fig. 6. Some results of periodical ionosphere soundings. Changes taking place in the virtual height h' (upper graphs) and in the critical frequencies f_{crit} (lower graphs) of the E and F layers during 24 hours. h' is the virtual height in the roughly horizontal part of the characteristic $h' = f(f)$, thus at $f < f_{crit}$. The graphs on the left were recorded in June, those on the right in December, both in a year when there was a large number of sun-spots, and at a northern temperate latitude. t_1 = the time of sunrise, t_2 = sunsset.

intense at the highest position of the sun is confirmed with respect to the E and the F_1 layers. The fact that, contrary to expectations, in the course of a summer day the critical frequency of the F_2 layer does not reach such a high maximum as on a winter day may be related to the rising (and thus expansion) of this layer on a summer day, the resultant reduction in electron density being greater than the increase caused by ionization by the sun's rays.

Conditions in the ionosphere are also closely connected with the number of sun-spots, which, as is known, follows an average cycle of 11 years. Fig. 6 applies for a period when there are many sun-spots. When there are only a few sun-spots as a rule the critical frequencies are lower (especially that of the F_2 layer).

Importance of ionosphere soundings for radio communications

The direct rays of a transmitter have only a very limited range. Radio communication over any great distance can be established only with the aid of the ionosphere [6]). Here we shall confine our considerations to what are known as one-hop transmissions, where the wave is reflected back to earth by the ionosphere only once. The maximum distance D that can be covered in one hop depends upon the height H of the reflecting layer (for the sake of simplicity the thickness of the layer will be ignored) and the radius R of the earth. From an elementary calculation it follows that $D \approx 2\sqrt{2HR} = 226\sqrt{H}$ (D and H expressed in kilometres), from which it is found that $D \approx 2500$ km for $H = 120$ km (E layer) and almost 4000 km for $H = 300$ km (F layer).

This formula represents the situation rather too favourably, since it applies for waves leaving the transmitter in an absolutely horizontal direction (angle of elevation zero). Such waves, however, are strongly absorbed by the earth. In practice, therefore, an angle of elevation of at least a few degrees has to be reckoned with, and D is then less than $226\sqrt{H}$.

What counts only is the state of the ionosphere at the point of reflection, which is above the point lying half way between the transmitter P and the receiver Q (see *fig.* 7). If we set up at that half-way point an ionosphere sounder A-B and with it determine one of the critical frequencies — for instance that of the F layer — then that f_{crit} is by no means the maximum frequency with which communication can be established from P to Q (at a particular time and via the F layer). The reason for this lies in the difference in the angle of incidence α, which with the ionosphere sounder is zero but much greater for the connection P-Q. For the maximum usable frequencies (M.U.F.) f_{max}, (again ignoring the thickness of the layer):

$$f_{max} = f_{crit}/\cos\alpha \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The angle α depends upon the distance P-Q and the height H of the layer in question. Thus it is seen that eq. (2) contains the two data that the ionosphere sounder gives for a certain layer (f_{crit}

[6]) Communications by means of centrimetric or decimetric waves via a tropospheric "radio duct" are left out of consideration here.

and h', here $= H$), these together determining the maximum usable frequency with which a given distance (provided it is less than D) can be covered in one hop. Actually, however, owing to the considerable depth to which the waves penetrate into the more or less diffuse layer, the dependency of the relation between f_{max} and f_{crit} — the "M.U.F. factor" — upon the angle α is much more complicated than that given by eq. (2). The value of this M.U.F. factor lies between 1 and about 5, so that under certain conditions it is possible to work with frequencies much higher than f_{crit}.

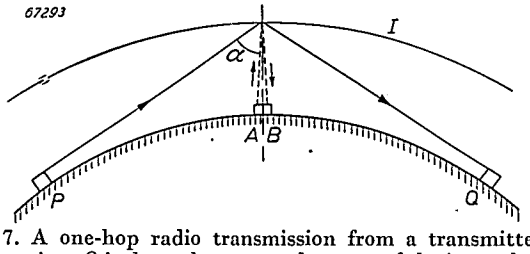


Fig. 7. A one-hop radio transmission from a transmitter P to a receiver Q is dependent upon the state of the ionosphere I halfway between P and Q and upon the angle of incidence α. A-B is an ionosphere sounder set up halfway between P and Q.

What, then, is the most effective frequency for a certain transmission over a distance of some thousands of kilometres? It has to be taken into account that over such a long distance part of it will nearly always lie in daylight, so that, owing to the absorption in the D region of waves shorter than 200 m greatly diminishing with increasing frequency, the frequency has to be chosen as high as possible. For radio-telegraphy, radio-telephony and broadcasting over great distances 6 Mc/s (wavelength 50 m) is considered to be about the lowest frequency. Preferably the frequency is chosen fairly close to the maximum usable frequency f_{max}. According to an American calculation — described below — the optimum working frequency in respect to the F_2 layer is taken to be 0.85 f_{max} and that for the more stable E layer 0.97 f_{max} (here f_{max} is a predicted value).

As we have seen, for the differently ionized layers there are different values of f_{crit}, and for communication between given points there are therefore, according to eq. (2), also different values of f_{max}. Here it is a question of the greatest f_{max}. As a rule this will be the f_{max} for the F_2 layer (at night the F region).

Sometimes, however, the E layer may be decisive. At noon on a summer day, for instance, f_{crit} of the E layer is only slightly lower than f_{crit} of the F_1 layer (see fig. 6). For a one-hop transmission between given points however the angle of incidence on the E layer is greater than that on the F_1 layer ($\alpha_E > \alpha_F$, *fig. 8*), thus also $\cos^{-1}\alpha_E \gtrless \cos^{-1}\alpha_F$, so that it is not precluded that the product of f_{crit} and $\cos^{-1}\alpha$ for the E layer is greater than that for the F_1 layer.

In the last ten years the Americans have taken the lead in working out a system for predicting f_{max} values from periodical ionosphere soundings taken at a large number of places on the earth. The Central Radio Propagation Laboratory of the National Bureau of Standards at Washington issues graphs with a large number of "predicted" curves along each of which f_{max} has a constant value; the coordinates are the geographical latitude and the local time at the point of reflection. Separate sheets with f_{max} contours are issued for a certain ionospheric layer, for a certain distance and for a particular terrestrial zone (the latter with a view to taking into account, roughly, the influence of the terrestrial magnetism). The curves are derived by extrapolation of the mean variation of values of f_{crit} obtained from numerous ionosphere soundings. They are valid for one month and are published three months in advance.

As regards the distance, the data given in these graphs are limited to the distances zero (at which $f_{max} = f_{crit}$), 2000 km (only for the E and E_s layers [7])) and 4000 km (only for the F region and F_2 layer). With the aid of (rather complicated) directions for use, a world map and nomograms, the optimum working frequency for any transmission can be predicted with a fair degree of certainty.

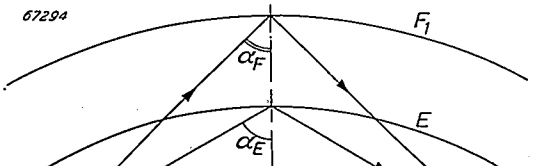


Fig. 8. In the case of a one-hop radio transmission from P to Q via the E layer the angle of incidence is greater than that for a similar transmission via the F region ($\alpha_E > \alpha_F$). Consequently under certain conditions a higher frequency can be used for the transmission via the E layer than for that via the F region.

The foregoing should be sufficient to show the scientific and practical values of ionosphere soundings. We shall now proceed to describe one of the sounders that we have developed. This has been designed mainly with a view to making it as simple as possible, so that it may be within the reach of advanced amateurs, who may furnish a welcome contribution to the information being collected from observations. It is not the intention to go deeply into technical details here; any

[7]) E_s denotes the sporadic E layer, which occurs more or less locally at irregular times. Its height is roughly equal to that of the normal E layer but it has a much greater electron density (thus also a greater f_{crit}).

readers interested in such details are referred to an article published elsewhere [8]), whilst data can be had by applying to the laboratory mentioned in the footnote *) on p. 152.

Description of a simple ionosphere sounder

Fig. 9 is a photograph showing this simple ionosphere sounder. It was with this apparatus that the oscillogram in fig. 5 was obtained. The frequency is variable from 5 to 10 Mc/s. This band is quite sufficient for studying the main properties of the F layers (cf. fig. 6) and it enables us to observe echoes from the sporadic E_s layer (cf. footnote [7])) when this layer is present. For the repetition frequency of the wave trains the mains frequency has been chosen (50 c/s); the ionospheric conditions do not change so quickly as to justify any higher frequency.

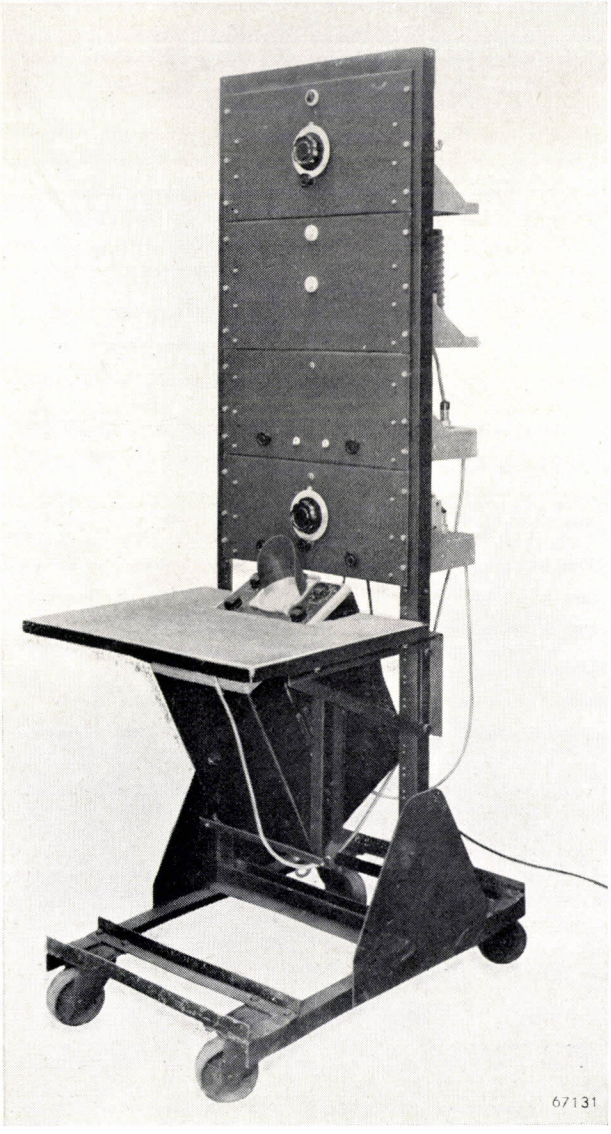


Fig. 9. An ionosphere sounder of simple construction. From top to bottom; the transmitter, the modulator, the control unit, the receiver and type MS 476 oscilloscope.

[8]) P. Maguer, Le sondage ionosphérique, Radio Française 1949 (No. 4), pp. 7-12.

The wave trains must not be so short as to make it necessary for the circuits to be given an exceptionally large band width. On the other hand the wave train must be shorter than the echo time t_e that the signal takes to travel the distance from the earth to the lowermost layer to be sounded and back to earth again; if this condition were not satisfied then the echo pulse on the screen of the oscilloscope would interfere with the pulse received direct. Since the lowermost layer with which we are concerned is the E_s layer, at an altitude of about 100 km, the time t_e is at least about 700 μsec. For the duration of the wave trains a much shorter time has been chosen, in the order of 100 μsec, thus ensuring a clear separation of the two pulses without it being necessary for the transmitted band to be exceptionally wide.

The principle of the apparatus is entirely in accordance with the explanation given with reference to fig. 2, the oscilloscope showing a pulse received direct and an echo pulse. Consequently the main components to be distinguished are the control unit, the transmitting unit, the receiving unit and the aerials.

Control unit

The control unit has to supply the signals for coordinating the working of the transmitter with that of the receiver. Every 1/50th second it produces a pulse (see *fig. 10*) which brings the transmitter into action, a linearly increasing and decreasing voltage for the time base of the cathode-ray tube, a voltage for blanking out its beam during the flyback, and a series of pulses forming a distance scale along the horizontal axis.

The wave forms and phase relations of the various signals are represented in fig. 10. All these signals are derived from one single square-wave voltage, called the control voltage, so that they are all perfectly constant in phase (the phase of the pulses controlling the transmitter is slightly variable; this will be dealt with presently).

The following circuits are to be distinguished: a) Control circuit. The square-wave voltage controlling the whole apparatus is derived from a sinusoidal alternating voltage of 50 c/s. This is taken, via a resonant circuit tuned to 50 c/s (see *fig. 11*), from a winding of the supply transformer and applied, via a resistor of 0.5 MΩ, to the grid of the first triode section of an ECC 40 valve connected in cascade with the second section. In this

circuit a large part of the two sine peaks is clipped. At the anode of the second triode section a practically rectangular voltage is obtained, the wave form of which shows steep flanks; this voltage is suitable for use as control voltage.

tains an EF 42 pentode, the control grid of which is fed via a resistor R_2 of 2 megohms and is connected to the anode via a capacitor C_2 of 50,000 pF. The time constant R_2C_2 is 0.1 sec, thus greatly exceeding the duration of a cycle of the control voltage to

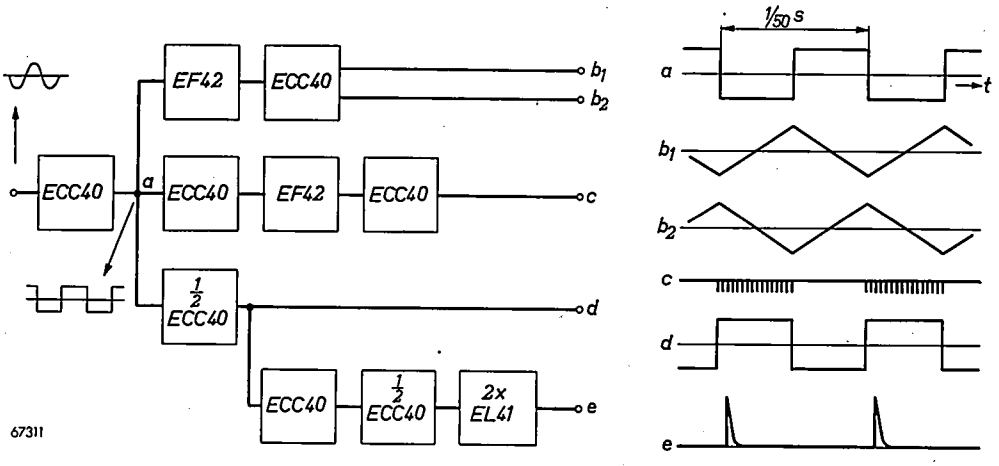


Fig. 10. Block diagram of the control unit. The ECC 40 valve (on the extreme left), to which is applied a sinusoidal voltage of 50 c/s, produces a square-wave voltage (control voltage) as represented at (*a*). From this control voltage are derived: the push-pull voltage for the horizontal deflection (b_1, b_2), the voltage for the electronic altitude scale (*c*), the voltage for periodically blanking out the cathode ray (*d*), and the pulses for periodically starting the transmitter (*e*).

The accuracy of the synchronizations brought about by the control voltage is proportional to the steepness of the flanks. The above-mentioned resonant circuit removes any harmonics, which might distort these flanks.

b) Circuit for generating a triangular deflecting voltage. This circuit (*fig. 12*) supplies a voltage for the horizontal deflection in the cathode-ray tube. It comprises, i.a., an integrator fed with the square-wave control voltage, so that the output voltage assumes the shape of an isosceles triangle (fig. 10, b_1).

The integrator is of the Miller type [9]). It con-

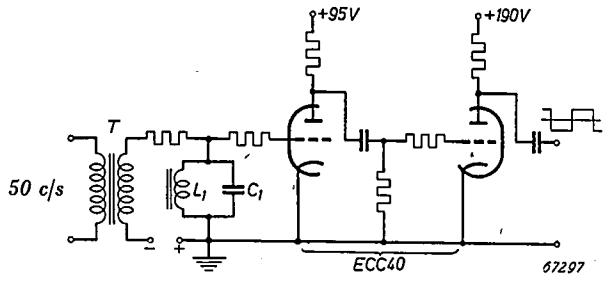


Fig. 11. Circuit diagram of the control unit. T = mains transformer. L_1-C_1 = circuit tuned to 50 c/s. ECC 40 a double triode in which both peaks of the sinusoidal voltage are clipped off, producing a square-wave voltage (the control voltage).

[9]) The Miller integrator has been discussed, i.a., in Philips techn. Rev. **12**, 328-330, 1951 (No. 11), fig. 15 and footnote [7]).

be integrated. With this value of R_2C_2 the form of the output voltage deviates less than 1% from linearity.

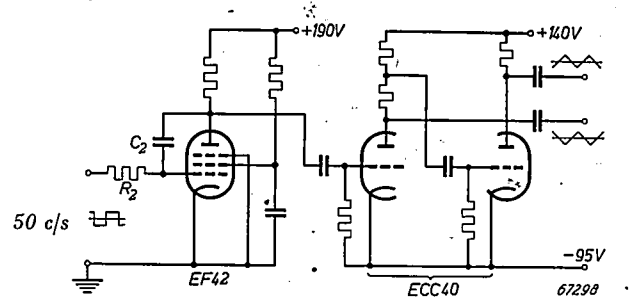


Fig. 12. Circuit for generating the deflection voltage. The pentode EF 42, the resistor R_2 and the capacitor C_2 form a Miller integrator. With a square-wave voltage applied to the input this supplies a voltage having the shape of an isosceles triangle. This voltage is amplified by an ECC 40 valve with balanced output.

The integrator is followed by an amplifier, fitted with an ECC 40 valve. One triode section of this valve amplifies a proportion of the output voltage from the other section, in such a way that the two output voltages are symmetrical with respect to earth (see fig. 10, b_1 and b_2).

c) Electronic altitude scale. The electronic graduation is formed by pulses repeated at a frequency of 3000 c/s, the distance between two pulses corresponding to a difference in height of 50 km.

These pulses are obtained with the aid of an

oscillator with "phase restoration". This consists mainly of a circuit L_2-C_3 tuned to 3000 c/s in connection with the two triode sections of an ECC 40 valve (*fig. 13*). Together with this tuned circuit the right-hand triode section (*I* in fig. 13) forms a Hartley oscillator. A variable resistor R_3, connected between the cathode and the tapping on L_2, is so adjusted that the oscillations are just maintained.

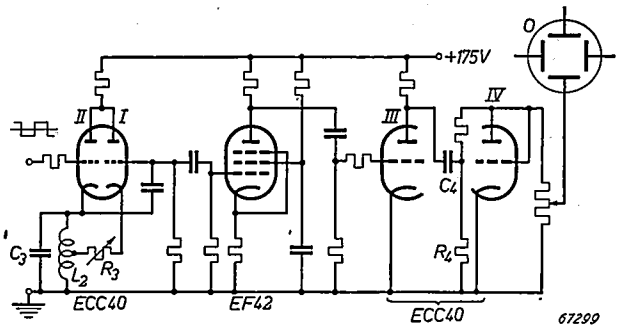


Fig. 13. Circuit for the electronic altitude scale. The circuit L_2-C_3, tuned to 3000 c/s or 1500 c/s, and the triode section *I* of an ECC 40 valve form a Hartley oscillator. Its phase is restored 50 times per second by the section *II* of the same valve, to the input of which the control voltage is applied. EF 42 is an amplifying valve, *III* a triode clipping the peaks. C_4-R_4 a differentiator, *IV* a diode suppressing the pulses of undesired polarity. *O* cathode-ray tube.

Generally the frequency of these oscillations is not exactly a multiple of 50 c/s, so that some device is required to ensure that the oscillator always commences to oscillate with the same phase, with respect to the flanks of the control voltage, each time the horizontal axis begins to be traced. This restoration of the correct phase is brought about by the other triode section (*II*) of the ECC 40, to the grid of which the control voltage is applied, in such a way that the oscillator is started and stopped 50 times per second. At the moment that the control voltage changes from positive to negative, the anode current of *II*, flowing through the coil L_2, is suddenly interrupted and the circuit L_2-C_3 begins to oscillate. With the aid of R_3 the damping is reduced almost to zero, so that the amplitude of the oscillations is practically constant (the transient is of very short duration) up to the end of the half-cycle in which the control voltage is negative. Then the triode *II* is again unblocked, and since its small internal resistance is shunted across the oscillatory circuit the oscillations are quickly damped. The oscillations begin anew, in the resonant frequency of the circuit L_2-C_3, as soon as the control voltage again changes from positive to negative.

The sinusoidal voltage of 3000 c/s is taken from the grid of the triode section *I*. It is amplified by an EF 42 pentode and the peaks of the sine are clipped in the usual way by the triode ECC 40 (*III*), the output from which is thus a square-wave voltage of 3000 c/s, This is applied to a circuit formed by a capacitor C_4 and a resistor R_4 connected in series. The time constant of this circuit is small compared with $^1/_{3000}$ sec. Thus the circuit functions as a differentiator, so that the voltage across the resistor has the form of the derivative of the square-wave voltage applied, viz. a series of pulses of alternating polarity. The second triode section of the ECC 40, connected as a diode, cuts off the negative pulses, so that only positive pulses remain, at intervals of $^1/_{3000}$ sec. When these pulses are applied to the lowermost deflection plate of the cathode-ray tube they produce along the horizontal axis graduation marks, directed downward, corresponding to differences in height of 50 km. (The reason for cutting off the negative pulses is that these would produce graduation marks directed upward, which might easily be confused with the likewise upward-directed echo pulses).

The oscillator can be made to work also at a frequency of 1500 c/s, in which case a scale is obtained with graduations of 100 km.

d) Blanking out the beam during the flyback. The control voltage is also applied to one half *V* of another ECC 40 valve (*fig. 14*) so that it supplies a square-wave voltage of opposite polarity. This is applied to the control grid of the cathode-ray tube and blanks out the beam during the half-cycles when it is negative (thus when the control voltage is positive). This makes it possible for the oscillogram of what takes place in the other half-cycle to be spread over the whole width of the screen.

e) Circuit for the intermittent working of the transmitter. This circuit supplies synchronization pulses to a multivibrator which modulates the high-frequency oscillator. These pulses, too, are derived from the control voltage. Unless certain measures were taken, however, owing to the differences in time constants of the various parts of the apparatus, these pulses would not exactly coincide with a pulse of the 3000 c/s oscillator. The transmitting pulse, therefore, would not exactly coincide with a division mark of the altitude scale, and this would affect the reading. To provide against this the pulse which starts the transmitter is passed through with a small variable delay after the control voltage has passed through zero.

The circuit employed for starting the transmitter consists of a flip-flop circuit, providing for the variable delay, an amplifying stage and a pulse generator.

The flip-flop circuit contains an ECC 40 valve the two halves of which (*VII* and *VIII*) form a blocked multivibrator. In the state of rest *VII* is blocked while *VIII* is conducting. From the square-wave voltage at the point *b* — the same voltage that controls the cathode-ray tube — the differentiating circuit C_5-R_5 forms a series of pulses, of which the positive ones cause the triode *VII* to conduct, while the triode *VIII* is blocked. After a certain interval of time the system returns to the state of rest, and earthed. This yields a positive voltage of about 190 V and a negative one of 95 V serving as grid bias. The supplies for the various parts are by passed to earth.

Transmitting unit

The transmitting unit (*fig. 15*) comprises the modulator, the supply unit and the transmitter proper. The latter consists of a self-oscillator fed with voltage pulses.

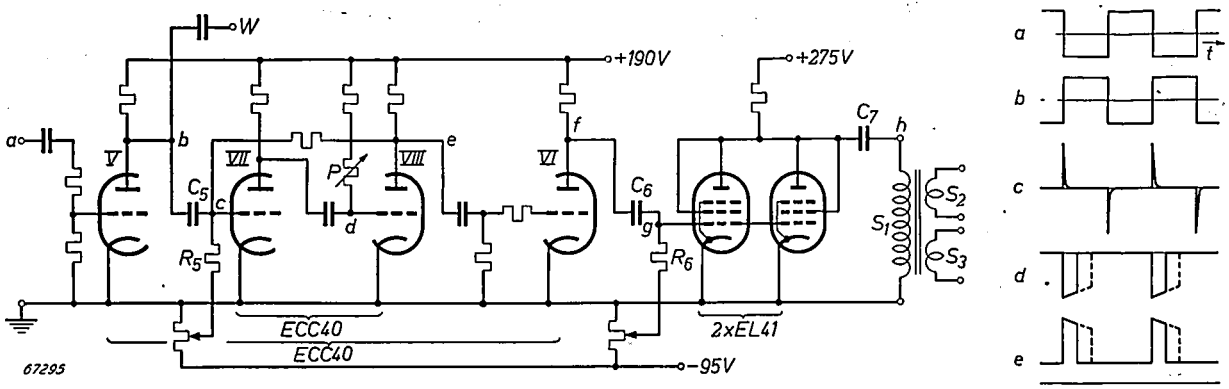


Fig. 14. Circuit for periodical starting the transmitter. The voltage at the points *a* ... *g* is represented, as a function of *t*, in the diagrams at the side denoted by the corresponding letters. *V* is a triode supplying a square-wave voltage of a polarity opposite to that of the control voltage applied to the input. The terminal *W* is the point from which the voltage is taken for periodically blanking out the beam. C_5-R_5 differentiating circuit. *VII* and *VIII* triodes of a flip-flop circuit. *VI* amplifying triode. C_6-R_6 differentiator. $2 \times$ EL 41 output valves discharging the capacitor $C_7 = 0.5$ μF across the coils S_1 every time a positive voltage peak occurs at *g*. *VII* conducts and *VIII* is blocked each time a positive peak occurs at *c*. The original position (*VIII* conducting, *VII* blocked) is restored a certain time later, this time being adjustable by means of the variable resistor *P*. Thus the pulses at *d*, *e*, and *f* are of variable duration, and the positive peaks at *g* have a variable phase.

his interval can be varied by means of the resistor *P* (fig. 14), across which the capacitor, which couples *VII* to *VIII*, is discharged.

The output from *VIII* is amplified by a triode *VI* (forming part of the same ECC 40 valve of which the part *V* has already been mentioned) and differentiated by a circuit C_6-R_6. At the instants that the flip-flop circuit returns to the state of rest the voltage across C_6-R_6 has a steep front and by differentiation this yields a positive pulse at the point *g*. This is applied to the control grids of two EL 41 valves connected in parallel. Normally these valves are blocked, but when they are triggered by the pulses they cause a capacitor C_7 of 0.5 μF to discharge through a winding of a transformer forming part of the modulator (see below).

f) Supply. The D.C. supply for the control part is provided by a rectifier equipped with an 1882 rectifying valve. The direct voltage produced is smoothed and then stabilized by means of three 4687 neon tubes connected in series. The point between the first and the second stabilizers is

The anode of the modulator valve (TB 2/200 [10])) is fed, via the coil S_2 of a transformer with three windings, with 1100 V direct voltage supplied by a rectifier equipped with a DCG 4/1000 mercury-vapour rectifying valve. Current pulses from the control unit are fed to the coil S_1 (fig. 14). In the grid circuit the coil S_3 and a negative bias source are included.

Each pulse produces an anode current pulse in the otherwise blocked modulator valve [11]). The coupling between S_2 and S_3 is such that each time

[10]) The TB 2/200 valve of the modulator can to advantage be replaced by a QB 3/300 tetrode and the two TB 1/60 valves of the oscillator by a QQE 06/40 double tetrode (for a description of the latter see: E. G. Dorgelo and P. Zijlstra, Two transmitting valves for use in mobile installations, Philips techn. Rev. **12**, 157-165, 1950 (No. 6)). These more modern types were not available at the time that this ionosphere sounder was designed.

In the apparatus described here the maximum peak values of the anode current and anode voltage permissible for continuous operation are greatly exceeded, so that the normal guarantee for the valves does not hold in this case.

[11]) For a more detailed description of this circuit see, e.g. L. Liot, Le générateur d'impulsions à auto-excitation, Electronique, April 1951 (No. 53), pp 8-11.

this valve is triggered an oscillation is produced which lasts only one cycle. The sudden interruption of the anode current produces on the anode of the valve a large voltage pulse of about two thousand volts which causes the transmitting oscillator to work momentarily; during the rest of the time this oscillator remains inoperative.

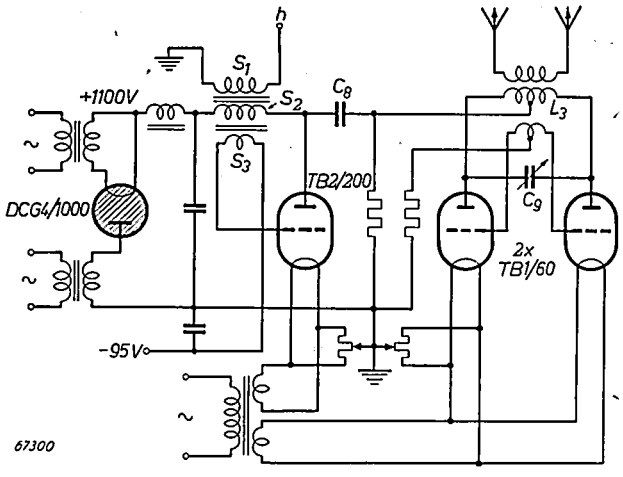


Fig. 15. Circuit diagram of the transmitter. TB 2/200 modulating valve discharging the capacitor $C_8 = 0.25$ μF via the Mesny oscillator (valves TB 1/60 and oscillator circuit L_3-C_9). each time a current pulse is sent through the coil S_1 (cf fig. 14). By means of C_9 the frequency can be adjusted between 5 and 10 Mc/s.

Receiving unit

The receiver is a sensitive superheterodyne receiver (see *fig. 16*) containing an H.F. amplifying stage with a symmetrical input and equipped with an EF 42 valve. As mixer a triode-hexode ECH 41 is employed. The I.F. amplifier is adjusted to 472 kc/s and comprises two stages each with an EF 42 valve. The bandwidth of the I.F. transformers is larger than that of an ordinary receiver, in order to obtain a better pulse response. The band covers 18 kc/s between the frequencies at which the gain drops 6 db.

Detection takes place in a diode EB 40 in combination with an adjustable threshold voltage. The latter enables interferences to be eliminated when the field strength of the ionospheric echo is sufficiently strong, The A.F. amplifier has an EF 41 valve loaded with a resistor. The output voltage provides for the vertical deflection of a type MS 476 cathode-ray oscilloscope.

In this apparatus there was no need to take steps for purposely blocking the receiver for the direct rays while the transmitter is working. With the low anode voltages in the receiver (60 V) the signal received direct is already sufficiently limited to avoid any adverse effects.

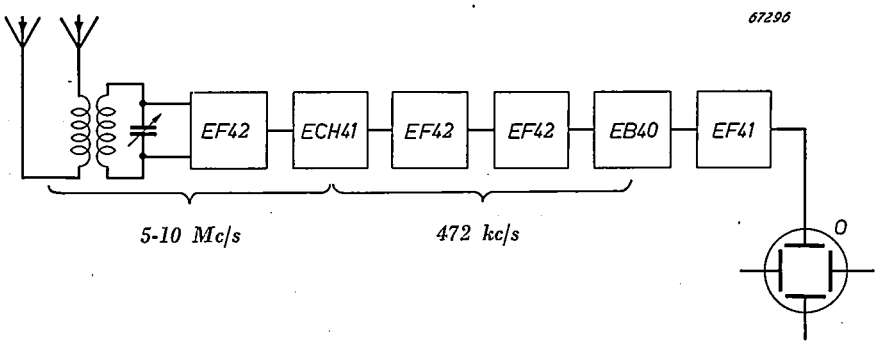


Fig. 16. Block diagram of the receiver, consisting of an H.F. amplifying stage (EF 42), a mixing stage (ECH 41), two I.F. amplifying stages (EF 42), a detector (EB 40) and an A.F. amplifying stage (EF 41) connected to the oscilloscope *O*.

The transmitting oscillator works with two TB 1/60 [10]) triodes in a symmetrical circuit with inductive feedback (Mesny circuit). The peak output is about 2 kW. As already mentioned, the transmitting frequency is variable from 5 to 10 Mc/s. The tuned circuit is formed by a coil wound on a ceramic former (likewise carrying the grid coil) and a variable capacitor of 130 pF insulated for 5000 V (the H.F. voltage reaches a peak value of 4000 V).

Aerials

The transmitting and receiving aerials, which are exactly alike, have been so designed as to give a vertical directional effect and to require only two points of suspension.

Each aerial (see *fig. 17*) has the shape of a rectangle 1 m high and 40 m long in the horizontal direction (i.e. $^2/_3$ to $^4/_3$ times the wavelength at which the apparatus works). The two aerials are spaced 1 m apart. In the middle of the wire along

the bottom of each rectangle is an insulator, on either side of which a feeder 10 m long is connected. At the other end these feeders are connected to two coils, one coupled to the transmitter and the other

porated, so that, for instance, the diagram $h' = f(f)$ is produced direct on the screen of the oscilloscope. We hope to be able to describe this type of sounder on another occasion.

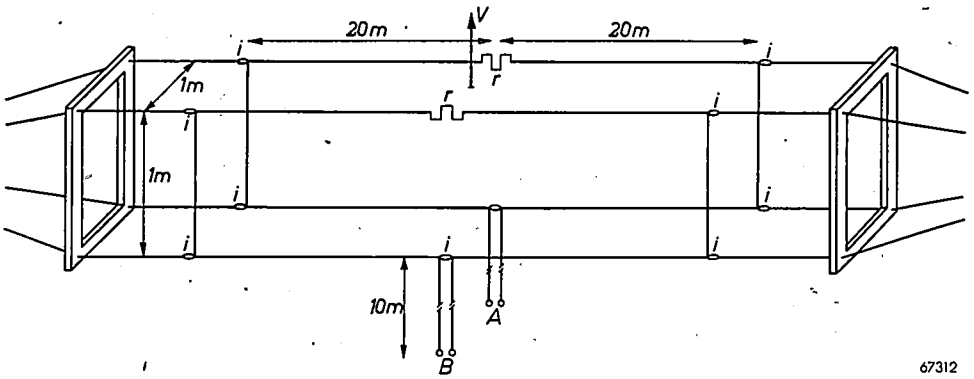


Fig. 17. Transmitting and receiving aerials, each 40 m long and connected via a 10-m lead respectively to the transmitter A and the receiver B · i = insulators, r = resistors of 800 ohms (2 W). The arrow V denotes the direction of the vertical.

to the receiver. The impedance of the feeders is about 1000 ohms.

In the middle of the upper wires of the rectangles is a resistor of 800 ohms. These resistors, in which a power of a few watts is dissipated, suppress stationary waves. In this way a current distribution is obtained which results in a maximum radiation to (and reception from) the zenith.

The form of aerial described is a special adaptation of the more common rhombic aerial *(fig. 18.)*. Calculations show that the maximum directional effect with respect to the zenith is obtained when the top half-angle β is 90°. Thus we arrived at the shape of aerial illustrated in fig. 17.

In addition to the apparatus described here, another has been developed in which various improvements and refinements have been incor-

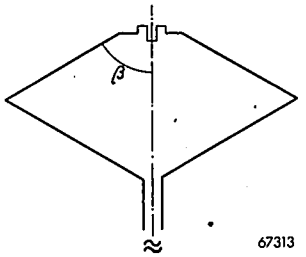


Fig. 18. Rhombic aerial. By making the top half-angle $\beta = 90°$ an aerial is obtained of the shape illustrated in fig. 17.

Summary. Investigations of the ionosphere are mainly carried out by means of the ionosphere sounder according to Breit and Tuve. By this method the echo time elapsing between the transmission of a wave train and the reception of the signal reflected back by the ionosphere in a receiver set up very close to the transmitter is measured by means of an oscilloscope. The measured echo time gives the virtual height of the reflecting layer, and from the highest frequency at which the layer in question still reflects the signal it is possible to find the maximum electron density in that layer. These two quantities as functions of time show various cyclic changes and also irregular fluctuations, whilst they are also dependent upon the geographical latitude of the point of observation. Once the curves for these quantities at a number of places on the earth are known it is possible to predict, some months in advance and with a fair degree of accuracy, the optimum working frequency for any radio transmission. The more data that can be collected from observations of ionospherical conditions, the more reliable will such predictions become. With this end in view the authors of the present article have developed an ionosphere sounder of simple construction, intended for the investigation of the F layers and the sporadic E layer.

The transmitter works with two TB 1/60 valves in a Mesny circuit at a frequency variable between 5 and 10 Mc/s and produces 50 wave trains per second with a duration of about 100 μs and a peak power output of about 2 kW. The receiver is of the superheterodyne type and has a bandwidth of 18 kc/s. The signal received direct and the echo signal are displayed on the screen of an oscilloscope as two pulses. The distance between the two pulses is a measure for the virtual height of the reflecting layer, and this height can be read direct from an "electronic altitude scale". The working of the transmitter is synchronized with that of the receiver by means of control signals all derived from the mains voltage. The aerials give maximum radiation to (and maximum reception from) the zenith.

FROM THE MANUFACTURE OF CATHODE-RAY TUBES FOR TELEVISION RECEIVERS

The glass bulb of cathode-ray tubes has to be coated on the inside with a conducting layer for collecting the secondary electrons emitted by the fluorescent screen; this layer is connected to the anode of the electron gun.

The photo shows a picture tube being coated with such a layer. Aquadag — a suspension of colloidal graphite in a solution of certain organic substances in water — is fed through a rubber tube to a small brush affixed to a bent leaf-spring and passed through the neck of the bulb. The bulb is centred by three guides and rotated about its axis by the action of the atmosphere pressing it against a rotating rubber ring, the space between the ring and the bulb being kept below atmospheric pressure. Starting from the top, the brush is gradually withdrawn, so that first the conical part of the bulb and then the neck is evenly coated with aquadag. The operator regulates the supply of aquadag with his left hand.

AGEING PHENOMENA IN IRON AND STEEL AFTER RAPID COOLING

by J. D. FAST. 620.193.918:669.14-156:621.785.618:621.791.75

After mechanical working or thermal treatment commercial steels often show a peculiar behaviour. For instance, after rapid cooling (quenching), in course of time a steel often shows phenomena of ageing, i.e. changes in its mechanical properties such as its hardness. These phenomena are apt to occur, inter alia, after arc welding, and they appear to depend upon constituents, such as carbon and nitrogen, present in the metal. The atomic processes responsible for this ageing will be discussed in two articles. The first of these — the present one — deals with quench ageing as determined by hardness measurements. In the second article, following upon the first, the processes causing this phenomenon are discussed from an entirely different point of view. A third article in this series, to be published later, will deal with the ageing taking place after mechanical deformations (strain ageing).

Introduction

The normal commercial steels, such as are used on a large scale for constructional purposes, show ageing phenomena both after rapid cooling from high temperatures (quenching) and after mechanical deformation at room temperature. Among these phenomena there is a gradual but decidedly adverse changing of the mechanical properties of the steel, as for instance its hardness. When such changes take place after rapid cooling one speaks of "quench ageing", and in the other case of "strain ageing".

It will be made clear that the cause of this ageing in both cases lies in the presence of certain constituents in the steel. Of these, for our experiments and investigations, manganese, carbon, nitrogen and oxygen are of particular importance. These elements frequently occur in the steels in question in roughly the following weight percentages: 0.5 % Mn, 0.1% C, 0.01% N, 0.01% O. The statements to be found in the literature about the effect of each of these elements separately upon ageing contradict each other. This is due to the fact that the conclusions drawn are usually based upon experiments with the commercial steels themselves, that is to say with iron in which the said impurities and some others (particularly sulphur and phosphorus) are present simultaneously.

By employing the apparatus already described in this journal [1]), by means of which it is possible to prepare iron with an exactly known content of one or more impurities, we have been able to study separately the individual effects of carbon, nitrogen and oxygen in otherwise pure iron and in iron containing a little manganese.

In the course of the investigations dealt with in this article quench ageing was studied with the aid of hardness tests after various thermal treatments. New results, of importance for the theory of ageing and the technology of metals, were obtained by studying alloys containing nitrogen and others containing both manganese and nitrogen [2]).

The internal structure of iron and iron alloys

Before proceeding to discuss the experiments it is necessary to say something about the internal structure of the alloys, the degrees of solubility of the constituents mentioned, and the atomic processes in the metal which are to be held responsible for quench ageing. The subject matter of the following sections also forms the basis for the discussions in the second article.

Iron occurs in two modifications distinguished one from the other by the arrangement of their atoms. Below 910 °C and from 1400 °C up to its melting point (1540°C) iron crystallizes into the body-centered cubic structure, a structure the smallest unit of which (the elementary cell from which an iron crystal of any size can be imagined to be built up by a continuous repetition in three mutually perpendicular directions) is illustrated in *fig. 1*. The cell contains only two atoms, since each of the eight atoms at the corners is common to eight cells. Below 910 °C this structure is denoted as α-iron and above 1400 °C as δ-iron. Iron in which the atoms are arranged in this manner is also sometimes

[1]) J. D. Fast, Philips techn. Rev. **11**, 241-244, 1949.

[2]) These results have been briefly reported in: J. D. Fast, Revue de Métallurgie **47**, 779-786, 1950 (No. 10).

called ferrite or ferritic[3]) iron, whilst iron alloys having this crystal structure are called ferritic alloys.

Between 910 °C and 1400 °C iron crystallizes into a face-centered cubic structure, a structure whose elementary cell is depicted in *fig. 2*. This cell contains four atoms, since the atoms at the corners lie only for one-eighth part in the cube and those in the middle of the sides for only one half. In this case one speaks of γ-iron, or austenite, and of austenitic alloys.

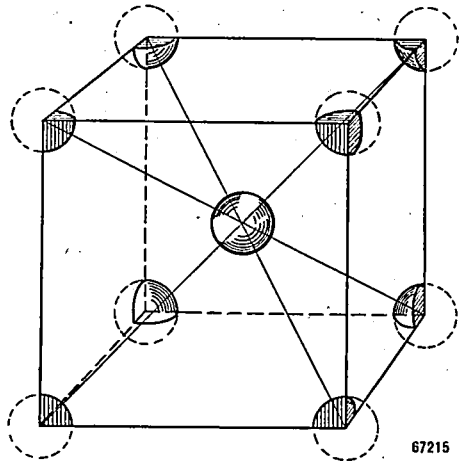


Fig. 1. Elementary cell of a body-centered cubic crystal, i.e. the structure of α-iron. In order to demonstrate that the elementary cell contains only two atoms, only those parts of the atoms at the corners are shown which lie within the cube.

In the case of a solution of manganese in solid iron (α- or γ-iron) one has to imagine iron atoms arbitrarily distributed over lattice sites as being replaced by manganese atoms. In a homogeneous austenitic alloy of iron and manganese with 25 atomic % Mn, each elementary cell (fig. 2) will therefore on an average contain one atom of Mn and three atoms of Fe. In such cases one speaks of substitutional solid solutions or "substitution mixed crystals".

When, however, elements such as carbon, nitrogen and oxygen, with small and comparatively readily deformable atoms, are present in solid iron as a solute, these atoms are located in the spaces between the iron atoms. Each elementary cell retains its normal number of iron atoms, but here and there one of them contains in addition an atom of the foreign element. In that case one speaks of interstitial solid solutions.

A closer investigation into the dimensions of the interstices available shows that both in α-iron and γ-iron there are two kinds of cavities to be distinguished. Supposing that the iron atoms in the crystal are touching spheres with radius R, then the radii of the largest spheres fitting in the cavities between the "iron spheres" have the following values:

in α-iron 0.29 R and 0.155 R,
in γ-iron 0.41 R and 0.23 R.

Thus the cavities in a body-centered crystal are much smaller than those in a face-centered crystal. However, the number of cavities in the α-structure is so much larger that the total space available in γ-iron is less than that in α-iron[4]). This is in agreement with the fact that the atoms in γ-iron form a close-packed structure, while the packing in α-iron is less dense.

The radii of the carbon and nitrogen atoms are much larger than those of the largest cavities in α-iron and γ-iron, amounting respectively to $r_C = 0.62\ R$ and $r_N = 0.56\ R$. The dissolution of carbon or nitrogen must, therefore, result in a considerable deformation of the elementary cell. This also holds when it is assumed that the carbon atoms are present in the lattice as monovalent or divalent positive ions and the nitrogen atoms as monovalent or divalent negative ions[5]).

When carbon or nitrogen is dissolved in γ-iron the deformation is equivalent with respect to the cube axes regardless of the kind of cavities occupied (there are indications that the carbon and nitrogen atoms occupy only one of the two kinds of cavities).

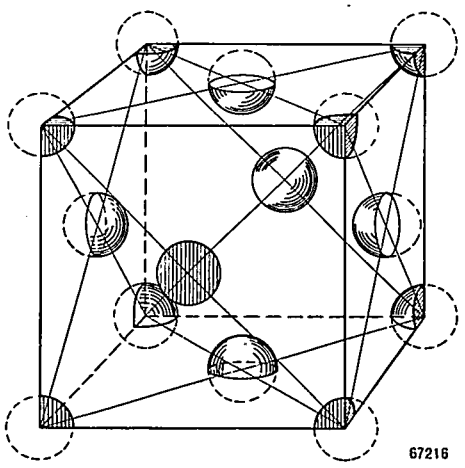


Fig. 2. Elementary cell of a face-centered cubic crystal. i.e. the structure of γ-iron. In this diagram and in fig. 1 the atoms have been drawn on too small a scale for the sake of clarity. To show that the elementary cell contains four atoms, only those parts of the atoms have been drawn which lie within the cell.

[3]) Not to be confused with compounds of the type $MeO.Fe_2O_3$ (in which Me represents a divalent metal), which have become known as ferromagnetic materials and are likewise called ferrites.

[4]) The α-structure contains six large (0.29 R) and three small (0.155 R) cavities per iron atom, the γ-structure only one large (0.41 R) and two small (o.23 R) cavities per iron atom.

[5]) Cf. W. Seith, Diffusion in Metallen, Springer, Berlin, 1939, pp 134 et seq.

In α-iron however the deformation is not symmetrical. This is illustrated in *fig. 3*. An atom situated at the point *1* (one of the smallest cavities available in the α-iron crystal) will bring about a tetragonal deformation of the elementary cell in the direction indicated in the diagram by y. By tetragonal distortion is to be understood such a distortion of the cube

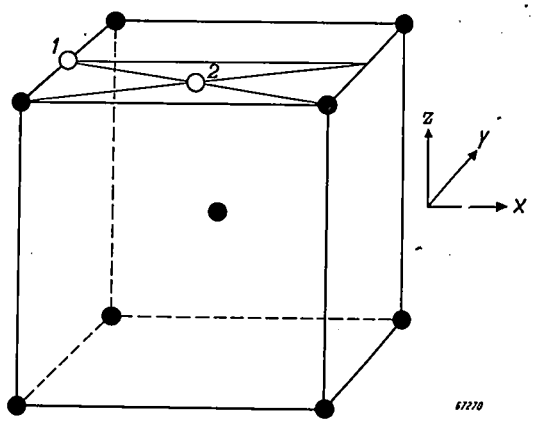


Fig. 3. The numbered positions are two so-called y-cavities in the lattice of α-iron. The position *1* corresponds to the small cavities between the iron atoms, the position *2* corresponding to the larger ones. Foreign atoms at either of these positions cause a tetragonal deformation of the elementary cell in the y-direction.

that one of the three mutually perpendicular edges is made longer or shorter than the others, as the case may be. An atom in the position *2* (one of the large cavities in the α-iron crystal) also causes a tetragonal but smaller deformation in the y direction. Atoms in the other available cavities will likewise cause tetragonal deformations but not necessarily always in the y direction; according to their position this deformation may take place in the x, y or z directions. It is therefore useful to distinguish between the so-called x, y and z positions, both for the large cavities and for the small ones; atoms in these positions will cause deformations in the x, y and z directions respectively.

In an equilibrium solution of carbon or of nitrogen in α-iron, not subject to any external forces, the solute atoms will be evenly distributed over the x, y and z positions. It must not be imagined that the C or N atoms are always in the same place; on the contrary they migrate from place to place through the crystal. There is, however, a random distribution at any moment, the lattice then being equally deformed in all directions and remaining cubic. A solid solution of carbon or nitrogen in α-iron having a tetragonal structure can, however, be obtained by quenching this C- or N-containing iron from the γ-phase. In the α-iron then obtained, the C or N atoms are found to be present mainly in interstices of only one of the three classes (x, y or z). Such tetragonal solutions are named martensite; they may be formed at the technical hardening of iron and steel. From now on we shall consider only quenching from the α-phase itself; the phenomena connected with technical hardening will not be discussed here.

Solubility of Mn, C, N and O in α-iron

The foregoing might create the impression that a metal like manganese, which with iron forms substitutional solid solutions, could do so when mixed in any proportions: in the picture given one could continue replacing Fe atoms by Mn atoms until pure manganese is obtained. Similarly one might gain the impression that the introduction of C or N atoms in, say, α-iron could be continued until all the available places are occupied. As the small cavities are filled this would lead to the formula FeC_3 or FeN_3 respectively, and as the large cavities are filled it would lead to FeC_6 and FeN_6 respectively [4]). This, however, is contrary to what is actually found in practice, which shows that Mn, C or N can be taken up in solid iron only in limited quantities. Predictions as to limits of solubility cannot be based upon the primitive geometrical considerations of the preceding section. They would have to be based upon atomic and thermodynamic calculations, but the theory of the solid state is not yet sufficiently developed to make such calculations possible [6]).

Thus one has to be guided almost entirely by experience, and this teaches that in the cases under consideration here, upon a certain quantity of the added element being exceeded, the temperature remaining the same, a second phase arises in the alloy with a composition differing from the first one. Crystals of this new composition are then formed at different places at the surface of the metal. The more of the foreign element is added, the larger these crystals become, until at last the whole metal is transformed into the new phase. The compositions of the phases in equilibrium with each other are, as a rule, greatly dependent upon the temperature. These relations can be read from diagrams called equilibrium or phase diagrams.

As an example *fig. 4* shows the part of the phase diagram for iron-manganese which is of interest to us here. The lines in this diagram indicate the limits within which different phases are stable. Along

[6]) The same applies, in fact, for answering the obvious question as to why below 910 °C and above 1400 °C iron crystallizes in the body-centered cubic structure, whereas between 910 °C and 1400 °C the structure is face-centered cubic.

the temperature axis the composition is that of pure iron, with as characteristic point the transition, already mentioned, at 910 °C. As is seen, when manganese is added the transition α-γ is shifted to lower temperatures. It is not our intention to go into all details of the diagram, but it should be pointed out that — as is always the case in such diagrams — the domains within which only one phase occurs are separated by a region in which the two separate phases are co-existent. As the diagram shows, the solubility of manganese in α-iron at room temperature is at most about 3%; when more manganese is added the γ-phase and the α-phase exist at the same time.

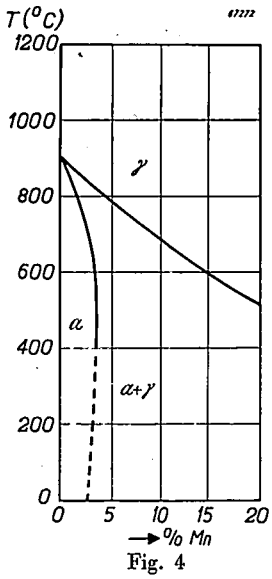


Fig. 4

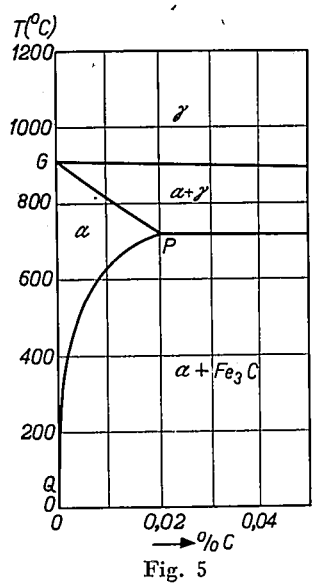


Fig. 5

Fig. 4. Part of the phase diagram for iron-manganese. The zone above 1200 °C (where, i.a., the γ-δ transition takes place and the liquidus line is situated) has not been drawn. (Taken from A. R. Troiano and F. T. McGuire, Trans. Am. Soc. Met. **31**, 340-364, 1943.)

Fig. 5. Part of the phase diagram for iron-carbon. The zone above 1200 °C has not been drawn. The line PQ represents the equilibrium between α-iron with carbon in solid solution and iron carbide. This diagram clearly shows that by rapid cooling from the α-phase a supersaturated solution of carbon in α-iron can be obtained. From this solution there will be a gradual precipitation of iron carbide.

The solubilities of carbon, nitrogen and oxygen in α-iron are much less than that of manganese. For carbon this is demonstrated in *fig. 5*, representing part of the equilibrium diagram Fe-C. Just as in the case of managnese, the solubility in the γ-phase is much greater than that in the α-phase.

If the α-iron contains more carbon than corresponds to the solubility line PQ then, in the state of equilibrium, in addition to the saturated solution there is a quantity of iron carbide (Fe_3C; cementite) present [7]).

[7]) Strictly speaking, in the state of equilibrium graphite is present as second phase. Due, however, to nucleation difficulties this completely stable state of equilibrium is not reached, but rather the metastable state mentioned.

The curve GP, covering the temperature range between 910 °C and 720 °C, is the solubility curve of carbon in α-iron in equilibrium with carbon in γ-iron. The maximum quantity of carbon that can be dissolved in α-iron is reached at the so-called eutectoid temperature of 720 °C, where the two solubility curves coincide.

In the region of small nitrogen contents the iron-nitrogen diagram has an appearance similar to the iron-carbon diagram, but the eutectoid temperature lies lower, viz. at 580 °C. Below this temperature, upon the solubility limit being exceeded, the nitride Fe_4N occurs as a second phase [8]).

Only recently the position of the line PQ became known with any accuracy [9]), the solubilities of C and N, expressed in weight %, being represented by the formulae:

$$\log (\%C) = -\frac{2200}{T} + 0.50 \quad \text{and}$$

$$\log (\%N) = -\frac{1550}{T} + 0.70,$$

where T is the absolute temperature. These formulae are based upon measurements between 200 °C and the eutectoid temperatures. They show that the solubility of nitrogen is much greater than that of carbon.

For the solubility of oxygen in α-iron the literature on the subject gives greatly conflicting values. From the fact, however, that in iron containing oxygen the same transitional points are found as in pure iron, and from our own experiments on the ageing of oxygen-containing iron it can be concluded that the solubility of oxygen is almost zero and in any case much less than that of carbon and nitrogen. Even at high temperatures, nearly all the oxygen in oxygen-containing iron is thus present in the form of oxide in solid iron.

Nucleation, diffusion and precipitation

In the foregoing it has been seen that below the eutectoid temperature the solubility of carbon and nitrogen in α-iron decreases with the temperature (see the line PQ in fig. 5 and the formulae given above). When, therefore, carbon- or nitrogen-containing α-iron is cooled from a high temperature, according to the phase diagram a precipitation of carbide, or nitride, respectively, may be expected. For such a precipitation to take place, first nuclei of the second phase have to be formed in the homogeneous solution. These nuclei will then grow, because the carbon or nitrogen atoms diffuse through the solution towards them.

[8]) Here, too, strictly speaking, the state of equilibrium is metastable, since at the temperatures in question and under a pressure of 1 atm Fe_4N should spontaneously dissociate into iron and nitrogen, which, however, does not take place at a perceptible rate.

[9]) L. J. Dijkstra, Trans, A.I.M.E. **185**, 252-260, 1949.

Now the processes of nucleation and of diffusion both take some time. If the cooling, say from the eutectoid temperature to room temperature, takes place quickly, then supersaturated solutions will be obtained. As a rule, it is difficult for the nuclei to form when cooling is stopped at comparatively high temperatures, whilst diffusion takes place very slowly at comparatively low temperatures. Thus the rate at which the total process of precipitation takes place may be determined by the rate of nucleation at relatively high temperatures and by the rate of diffusion at relatively low temperatures. In the case now under consideration (C or N in α-iron), at room temperature and at temperatures not much higher it is in all probability diffusion that determines the rate of the total process of precipitation.

From the foregoing considerations, quench ageing — the change in the hardness of the metal which is apt to take place in commercial steel rapidly cooled from the α-phase — can be correlated with a phenomenon of precipitation, namely the precipitation of carbide or nitride, from a supersaturated solution of carbon or nitrogen in α-iron [10]). It should be borne in mind that the plastic deformation of a metal is brought about owing to parts of the metal crystals slipping over each other along certain crystallographic planes. The resistance of the metal to slip increases by the presence of the interstitial solute atoms. There is, however, a much greater increase in slip resistance and thus a much greater increase in hardness if a precipitate is present in the form of a large number of minute carbide or nitride particles in the metal crystals.

If the precipitation is made to take place at a constant temperature (say 100 °C), first the hardness is seen to increase gradually and then to decrease again very gradually. Such a change occurs also in the case of the frequently investigated substitutional solid solutions of copper in aluminium. In an attempt to explain this, in 1919 Merica, Waltenberg and Scott assumed that first a very fine precipitate of a Cu-Al compound (Al_2Cu) is formed, which gradually increases in size. The maximum hardness would be reached at a "critical degree of dispersion" of the Al_2Cu lying below microscopic resolution. Later it appeared that the process of precipitation in copper-containing aluminium is much more complicated: there is not a decidedly heterogeneous (two-phase) system right from the beginning of the precipitation, as in the old picture, but rather a system in which the degree of heterogeneity increases gradually, with intermediate stages where there is neither a distinct homogeneous nor a distinct heterogeneous system [11]).

There are no reasons to assume, however, that similar intermediate stages occur in the process of precipitation taking place in the simple interstitial alloys Fe-C and Fe-N (at least if these alloys have not previously been plastically deformed). With these alloys the precipitation seems to follow a simpler course, such that a second phase is at once formed in the shape of minute carbide or nitride crystals, but with this complication (see [9])) that within a certain temperature range in the Fe-N system a nitride phase is first formed (probably $Fe_{16}N_2$ [12])) which is less stable than Fe_4N. Nuclei of $Fe_{16}N_2$ are formed more readily and thus in larger numbers than those of Fe_4N. Most of the nitrogen, therefore, begins to precipitate in the form of $Fe_{16}N_2$, which, however, gradually dissolves as the more stable Fe_4N nuclei increase in number.

In the Fe-C system it seems that in precipitation Fe_3C alone occurs as second phase. In this case, too, however, an equilibrium state is not yet reached when the supersaturation of the solid solution has ceased by the formation of Fe_3C; some Fe_3C crystals will gradually grow at the expense of others, which again dissolve because the total interfacial energy is thereby decreased. Whereas the precipitation in its first stages causes an increase in hardness and a reduced malleability, as the precipitate becomes coarser — the process being called "coagulation" (occurring also in the case of the nitride) — hardness again decreases and malleability increases. It is difficult, however, in this connection to ascribe any sharply defined significance to the "critical degree of dispersion" of Merica c.s., since at any moment there will always be a range of particle dimensions owing to the simultaneous growth of some carbide particles and the dissolution of others (which in fact already takes place before the supersaturation is entirely overcome).

Experiments on ageing

In order to study the influence of each alloy element separately on quench ageing, alloys of exactly known compositions were prepared in the apparatus already referred to [1]). From these alloys flat test bars about 4 mm thick were forged, these then being heated to 940 °C and slowly cooled

[10]) This phenomenon was first pointed out by Köster in 1929; see W. Köster, Arch. Eisenhüttenwesen **2**, 503-522, 1928-'29; **3**, 553-558, and 637-658, 1929-'30.

[11]) Cf. e.g. A. Guinier, Physica **15**, 148-160, 1949 and G. D. Preston, Phil. Mag. **26**, 855-871, 1938.

[12]) K. H. Jack, Acta Cryst. **3**, 392-394, 1950 (No. 5).

before the tests were begun. This took place in a protective atmosphere to prevent change in composition of the bars.

The bars were then heated for some time in vacuo at the quenching temperature, this temperature being chosen for each alloy to be slightly below its eutectoid temperature [13]). After this heating the bars were rapidly cooled by immersion in a large quantity of cold water. Immediately after quenching the Vickers hardness was measured at a temperature of 0 °C. This hardness value is measured by pressing with a certain load a quadrilateral pyramid of diamond into a smooth-ground surface of the metal. The impression left in the surface of the metal is square-shaped, and the diagonals of the square are measured. By the term hardness H is understood the force in kilogrammes divided by the pyramidal area (the sum of the areas of the four sides) of the indentation in mm^2.

After this measurement had been taken at 0 °C the hardness was measured again a number of times at 20 °C, namely after two hours' heating at 50 °C (followed by slow cooling), then after two hours' heating at 100 °C, then again after two hours' heating at 200 °C, and so on. In this way a picture of the phenomenon of precipitation can be obtained much more quickly than by measuring the hardness as a function of time at each of these temperatures, though in principle this is the most elegant method.

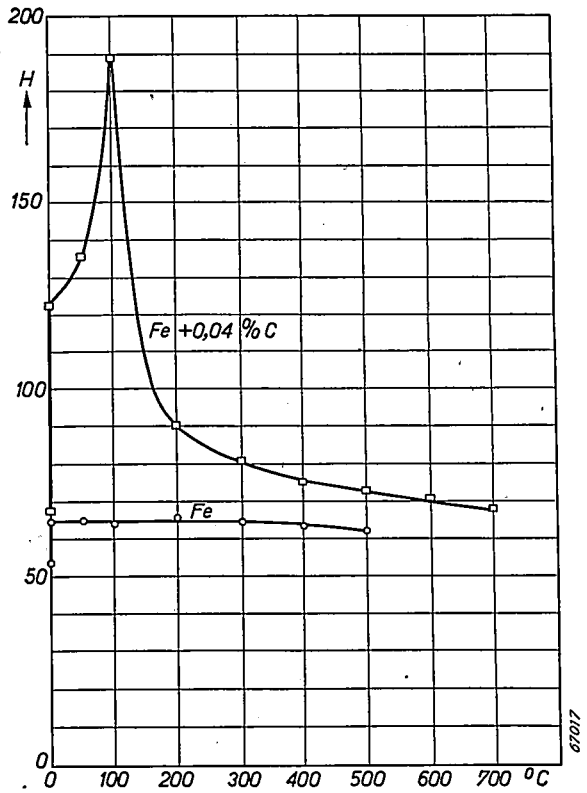


Fig. 6. Changes taking place in the Vickers hardness value H of iron with 0.04% carbon (squares) and of pure iron (circles) after the heat treatments described in the text.

[13]) The iron-oxygen system does not show any eutectoid point, or at least not one lying noticeably below the transition point of pure iron. In this case quenching was done from 750 °C.

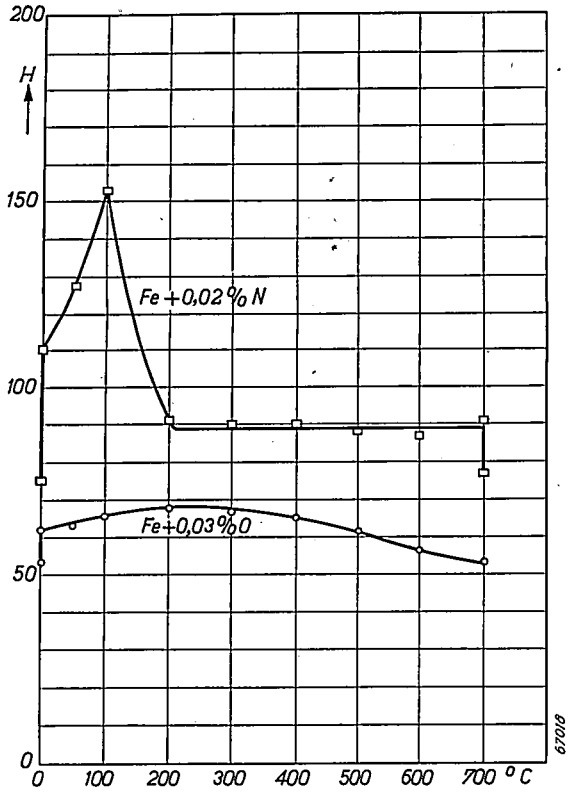


Fig. 7. As in fig. 6 but for iron with 0.02% nitrogen (squares) and for iron with 0.03% oxygen (circles). The lowest hardness value on the ordinate for 700 °C of nitrogen-containing iron was reached after re-heating to 700 °C and then very slowly cooling (during some hours) from that temperature. This further decrease in hardness occurs only with nitrogen-containing iron.

As already stated, the general trend, at a constant temperature, in the change of the hardness of a material that shows quench ageing is characterized by first a rather rapid increase followed by a slow decrease, until approximately the same hardness is reached which the metal possessed prior to the quenching. To be able to observe the slow change within a limited space of time, the bars have to be heated to a relatively high temperature, because the higher the temperature the more quickly the process takes place, as will be evident from what has been said above. The method of working described is based on this fact alone and has no further physical significance.

The materials used for these experiments were pure iron, iron with 0.03% oxygen, iron with 0.02% nitrogen, iron with 0.04% carbon and, further, the last two alloys with an extra admixture of 0.5% Mn. The results are shown graphically in *figures 6, 7* and *8*.

N. V. PHILIPS' GLOEILAMPENFABRIEKEN

In addition to the hardness values determined after the various heat treatments described above, the graphs also show the hardness values of the alloys prior to quenching, which are lower.

As was expected, pure iron (fig. 6) shows no quench ageing. The quenching does, it is true, set up internal stresses in the metal which cause a slight increase in hardness, but the heat treatments at 50 °C, 100 °C and higher cause practically no further increase. The picture is quite different with the iron containing only carbon and that containing only nitrogen (figs 6 and 7). Owing to the dissolution of the carbon and nitrogen, the hardness after heating to the eutectoid temperature and quenching is greater than that prior to quenching.

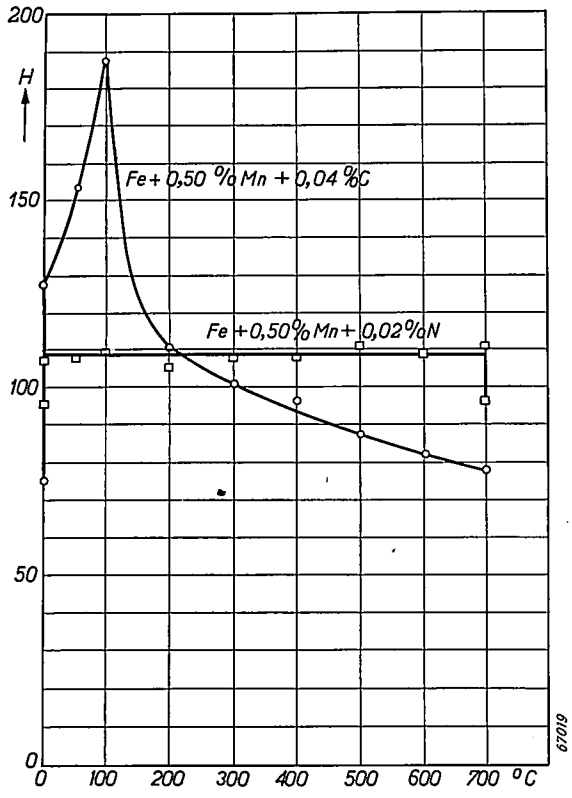


Fig. 8. As in fig. 6 but for iron with 0.50% manganese and for iron with 0.04% carbon and 0.02% nitrogen. The quench ageing of nitrogen-containing iron is suppressed by the manganese. Here again there is a further decrease in hardness of nitrogen-containing iron after slow cooling from 700 °C.

Under the heat treatments following upon quenching there is, however, a further considerable increase in hardness, the maximum being reached after heating to 100 °C. This temperature has no physical significance, on account of the arbitrarily chosen treatment to which the material was subjected. Oxygen-containing iron (fig. 7) shows only very little quench ageing.

The experiments show that the presence both of carbon and of nitrogen may lead to phenomena of precipitation in iron. The fact that oxygen has so little effect has to be ascribed to the extremely small degree of solubility of this element in α-iron, even at such a high temperature as 750 °C.

A very important result of the experiments is that manganese appears to have but little influence upon the precipitation behaviour of carbon-containing iron (cf. figs 6 and 8), but that this element tends to suppress the precipitation in nitrogen-containing iron (cf. figs 7 and 8).

The fact that in the presence of 0.5% manganese no perceptible quench ageing occurs in iron containing nitrogen may be accounted for in the following way. It is presumed that the nitrogen atoms are preferably present in such interstitial sites that at least one of their immediate neighbours is a manganese atom, for the reason that in such places a nitrogen atom will possess less energy than when it has exclusively iron atoms for its immediate neighbours. This is quite feasible considering that manganese and nitrogen show a greater chemical affinity than iron and nitrogen. Consequently for the nitrogen atoms to break away from the manganese atoms and then to be able to precipitate, the temperature has to be so high that coagulation also proceeds at a fast rate. Thus there would not be any perceptible temporary increase of the hardness.

It is possible to put the foregoing supposition to the test by means of damping experiments. These will be dealt with in the following article.

Summary. After quenching (rapid cooling) of iron or steel a slow change often takes place in the mechanical properties of the material. This quench ageing is to be ascribed to the presence of certain constituents in the metal. The hardness of α-iron containing small quantities of carbon, nitrogen or oxygen separately has been determined after various heat treatments. Also the influence of manganese on the ageing of α-iron with carbon or nitrogen has been studied. This investigation has been confined to a study of the changes in hardness taking place after quenching from the α-phase, thus from a temperature lying below the eutectoid temperatures of the materials; the technically still more important case of quenching from the γ-phase has been left out of consideration. It appears that oxygen causes virtually no quench ageing; this element is, in fact, almost insoluble in α-iron. Carbon and nitrogen are to some extent soluble in α-iron, and both cause considerable quench ageing. This is to be ascribed to precipitation of carbon and nitrogen in the form of carbide and nitride, from their supersaturated solution in the iron formed by rapid cooling. Manganese appears to be capable of suppressing the quench ageing of iron containing nitrogen but not that of iron containing carbon. Considering the comparatively great affinity between manganese and nitrogen, an explanation is sought by assuming that in the crystal lattice the nitrogen atoms will preferably occur in the vicinity of manganese atoms, where they are bound much more strongly than when surrounded exclusively by iron atoms, so that the formation of iron nitride, thus the precipitation, takes place at a perceptible rate only at high temperatures.

INTERNAL FRICTION IN IRON AND STEEL

by J. D. FAST and L. J. DIJKSTRA.

539.312 : 534.372 : 620.193.918 : 669.14-156

The internal friction of torsional oscillations of small amplitude of iron and steel wires is an after-effect phenomenon which under certain conditions arises from the diffusion of foreign atoms in the iron lattice. An application of this phenomenon is to give an insight into the precipitation of carbon or of nitrogen in α-iron, which precipitation — as shown in the preceding article — is responsible for quench ageing. The damping experiments which have been carried out throw more light upon the part played in this connection by manganese in nitrogen-containing iron.

Introduction

In the preceding article[1]) some phenomena were discussed which are liable to take place in iron and manganese-containing iron when carbon or nitrogen are also present. Following upon those discussions, an account will be given of some experiments which have been carried out with a view to obtaining a better insight into the atomic background of these phenomena. The experiments are based upon the fact, now known for about ten years, that the presence of carbon or nitrogen in iron may lead to a strong "internal friction", i.e. a rapid decay of mechanical vibrations set up in the metal.

Before discussing the experiments and the interpretation of their results, it is necessary to consider briefly the more widely known phenomenon of "external friction" and then to explain how it is possible for the presence of carbon or nitrogen *in* the metal to have the same effect upon mechanical vibrations as that of friction with a medium outside the metal. It will then appear, inter alia, that carbon and nitrogen can cause friction only if they are present in the iron in the solute state, and not if they are present in the form of carbide and nitride.

Torsional oscillations and external friction

In our damping experiments use was made of wires of the alloys to be investigated. Such a wire formed the elastic element of a torsional pendulum; its upper end was tightly clamped, while a body with a known moment of inertia was suspended from the other end. When the free end is twisted over a certain angle and then released the wire starts to make torsional oscillations. The potential energy of the twisted wire is converted into kinetic energy of the inert mass of the body and vice versa. The deformation of the wire is determined by an elastic couple K directed opposite to the deflection, which couple in the case of a small deflection is proportional to the angle of deflection φ:

$$K = -a\varphi,$$

where a is called the rigidity with respect to torsion or, briefly, the torsional rigidity of the wire. According to Newton's equations of motion this couple should always equal the product of the moment of inertia I of the pendulum and the angular acceleration $\mathrm{d}^2\varphi/\mathrm{d}t^2$. Thus we have the relation:

$$I\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} = -a\varphi \quad \ldots\ldots\ldots \quad (1)$$

In our case the solution of this differential equation is:

$$\varphi = \varphi_0 \cos \omega t, \quad \ldots\ldots\ldots \quad (2)$$

and hence:

$$K = -a\varphi_0 \cos \omega t = a\varphi_0 \cos (\omega t + \pi),$$

provided the zero point of the time is fixed at a moment when the angle φ has the maximum value φ_0. The angular frequency, ω, is given by:

$$\omega = 2\pi\nu = \frac{2\pi}{T} = \sqrt{\frac{a}{I}}, \quad \ldots\ldots\ldots \quad (3)$$

where ν is the frequency and T the periodic time of the oscillations.

According to eq. (2) the torsional pendulum would continue to oscillate for an infinitely long time. In reality such oscillations always decrease in amplitude, i.e. free vibrations are always damped. The best known cause of this is external damping, due to friction between the oscillating body and the surrounding medium (e.g. air). The frictional couple is always opposed to the angular velocity and for not too high velocities is proportional to it; it can therefore be expressed by the formula:

$$W = -c\frac{\mathrm{d}\varphi}{\mathrm{d}t}. \quad \ldots\ldots\ldots \quad (4)$$

[1]) J. D. Fast, Ageing phenomena in iron and steel after rapid cooling, Philips techn. Rev. **13**, 165-171, 1951 (No. 6), hereinafter referred to as article I.

Provided the friction is only small, the decrease in amplitude due to such a frictional couple follows an exponential law; in this case the relation between the angle of deflection and the time is represented to a good approximation by:

$$\varphi = \varphi_0 \, e^{-\frac{\gamma}{2}t} \cos \omega t, \quad \ldots\ldots\ldots \quad (5)$$

in which γ is given by:

$$\gamma = \frac{c}{I}. \quad \ldots\ldots\ldots\ldots \quad (6)$$

The zero point of the time lies at the instant when, after an angular twist $\varphi = \varphi_0$ is given to the system, it is left to itself.

The extent to which the amplitudes of vibration decrease with time can be expressed in various ways. A quantity frequently employed is the logarithmic decrement δ, this being the natural logarithm of the ratio of two amplitudes φ_1 and φ_2 lying one periodic time T apart. From eq. (5) we find for δ in our case:

$$\delta = \log_e \frac{\varphi_1}{\varphi_2} = \log_e \frac{e^{-\frac{\gamma}{2}t} \cos \omega t}{e^{-\frac{\gamma}{2}(t+T)} \cos \omega (t+T)} = \gamma \frac{T}{2} = \pi \frac{\gamma}{\omega}. \quad (7)$$

Another method of expressing that energy is dissipated, thus that damping occurs, is to state that the angle of deflection φ and the total counteracting couple V — now the sum of the elastic couple K and the frictional couple W — are no longer exactly in anti-phase, or, in other words, that they no longer show a phase difference π.

The deviation from π equals an angle β, the loss angle, which, in the case of not too great a friction, is given by:

$$\beta = \frac{\gamma}{\omega} = \frac{\delta}{\pi} \quad \ldots\ldots\ldots \quad (8)$$

(all angles are expressed in radians).

The above results follow from the differential equation for the damped free vibration, which reads:

$$I \frac{d^2\varphi}{dt^2} = -a\varphi - c\frac{d\varphi}{dt}, \quad \ldots\ldots \quad (9)$$

as becomes obvious from the text (formulae (1) and (4)).

In the case that $\gamma = c/I \ll \omega$ (the angular frequency of the undamped vibration), to a very good approximation the solution of this equation is the expression given in (5).

The couple V opposing the motion is given, according to (9), by:

$$V = -a\varphi - c\frac{d\varphi}{dt}.$$

After some calculation, employing the relations (3) and (6) and the condition $\gamma \ll \omega$, substitution of (5) in the formula for V gives:

$$V = -a\varphi_0 \, e^{-\frac{\gamma}{2}t} \left(\cos \omega t - \frac{\gamma}{\omega} \sin \omega t\right),$$

which expression, again because $\gamma \ll \omega$, can to a good approximation be written as:

$$V = a\varphi_0 \, e^{-\frac{\gamma}{2}t} \cos (\omega t + \pi + \beta).$$

Here the loss angle β is given by:

$$\sin \beta \approx \beta = \frac{\gamma}{\omega},$$

which, according to (7), is equal to δ/π.

Thus the couple V differs an angle $\pi + \beta$ in phase from the angle of deflection φ, thereby proving the statement set forth in the foregoing.

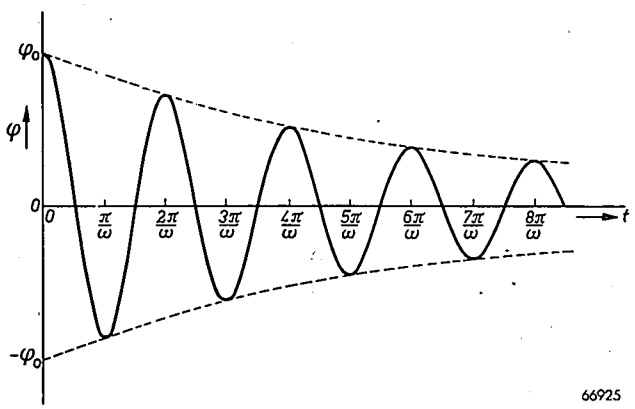


Fig. 1. Torsional oscillations with angular frequency ω decay exponentially when they are externally damped. The enveloping curves $\varphi = \varphi_0 \, e^{-\frac{\gamma}{2}t}$ and $\varphi = -\varphi_0 \, e^{-\frac{\gamma}{2}t}$ are given in broken lines. The points where these enveloping curves coincide with the fullydrawn curve correspond almost to the maximum and minimum values of φ. It will be shown that a similar relation exists between φ and t for the internal friction. This diagram applies for a value of $\gamma/\omega = 0.1$. In practice the damping is mostly much less than this.

In *fig. 1* it is illustrated how the amplitude of the damped oscillations decreases with time. It demonstrates the significance of the condition $\gamma/\omega \ll 1$ (viz. that the relative decrease of the amplitude during one periodic time is small).

After-effect, relaxation and phase shift

To understand how a vibration can be damped also as a result of the presence of dissolved carbon or nitrogen (thus, how vibrational energy can be converted into heat by internal causes), we have to start from the fact, discussed in article I, that the deformation caused by interstitially dissolved atoms of these elements in α-iron is not uniform with respect to the axes of the cube. A distinction was made between x, y and z positions. Carbon or nitrogen atoms in x positions cause a tetragonal deformation in the x direction, those in y positions a similar deformation in the y direction, and so on.

Let us now imagine a single crystal of α-iron in the shape of a cube in which a quantity of carbon is present in homogeneous solution *(fig. 2)*. So long as the cube is not subjectief to external forces, the carbon atoms are uniformly distributed among the x, y and z positions. If, however, the cube is elongated elastically in the z direction (which is assumed to coincide with the z directions of the elementary cubes in the crystal) then in the z positions more space becomes available for the carbon atoms, whereas, owing to the Poisson

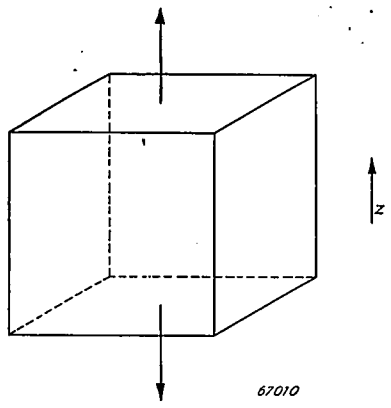


Fig. 2. A single crystal of α-iron in the form of a cube stretched in the z direction; as a result there will be some contraction in the x and y directions.

contraction, in the x and y positions there will be less space. Owing to this deformation the positions are no longer energetically equivalent: the carbon atoms will show a preference for the z positions. Whereas prior to the elastic deformation the time that the carbon atoms stay in any one interstice was on an average the same for all interstices, after the deformation this mean time of stay will be greater for the z positions than for the others. In other words, under the influence of the elastic deformation the diffusion equilibrium is shifted in such a way that more z positions and less x and y positions are occupied.

It is clear that this increased occupation of the z positions will lead to relaxation of the elastic force in the case of a constant deformation. When a constant load is applied the body will at once respond with a certain deformation, but under the influence of the displacement of the carbon atoms this deformation will gradually increase. Upon the load being removed the reverse takes place: part of the deformation disappears immediately, but the remaining deformation requires some time to do so, because it takes some time for the diffusion equilibrium to be restored. In the case under review one speaks of elastic after-effect. The fact that carbon and nitrogen may, in the manner described, lead to after-effect phenomena in α-iron was first recognized by Snoek[2]). (Other possible causes of after-effect, such as the thermo-elastic effect, will not be considered here.)

The foregoing considerations were related to a single crystal, whereas our experiments were carried out with polycrystalline wires. In a single crystal elongation in a direction parallel to the axes of the cube results in relaxation, but elongation in the direction of a solid diagonal of the cubic elementary cell has, of course, no such result. Distortion of a polycrystalline material in which the crystals have all possible orientations will therefore result in a relaxation lying somewhere in between the maximum value and zero.

It will now be investigated in what manner this relaxation affects torsional oscillations. As is known from the theory of elasticity, the torsion of an iron wire results in a compression of the volume elements of the iron in one direction (at an angle of 45° to the longitudinal axis) and an equally large elongation in a direction at right angles to the first. Thus also due to the torsion the diffusion equilibrium is shifted. Upon the direction of torsion being reversed the compressions change into elongations and the elongations into compressions. If the wire is twisted suddenly over a certain angle and that angle is thereafter kept contant, then the resultant elastic couple will gradually decrease owing to displacement of the carbon and nitrogen atoms. This is illustrated by *fig. 3*, where the rectangular line drawn in *a*

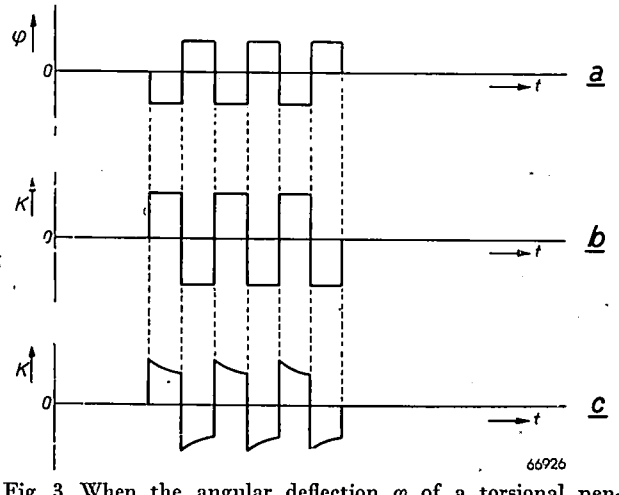


Fig. 3. When the angular deflection φ of a torsional pendulum varies discontinuously in the manner indicated (*a*) a medium without after-effect will react to this with an elastic counteracting couple K in the form of the curve (*b*). In a medium with after-effect the couple will vary as a function of time according to curve (*c*). During each period of a constant angular deflection the absolute value of the couple decreases. This decrease takes place practically exponentially.

[2]) J. L. Snoek, Physica **9**, 711-733, 1941; **8**, 862-864, 1942.

represents an angular deflection kept constant for a very definite time and then changed in sign in an immeasurably short time. A material without after-effect responds to this with a counteracting couple of the form represented in *b*, whilst a medium having after-effect reacts according to *c*.

For a relaxation based on a simple diffusion phenomenon as that described here it may be assumed that the couple K required for a constant angular deflection decreases with time such that dK/dt is always proportional to the instantaneous value of the variable part of K. Thus it may be assumed that

$$-\tau \frac{dK}{dt} = K - K_e, \quad \ldots\ldots \quad (10)$$

where τ is a proportionality constant with a time dimension, which is called the relaxation time of the elastic after-effect. If $K = K_b$ at the instant $t = 0$ then integration of (10) yields the relation:

$$K = K_e + (K_b - K_e)\, e^{-\frac{t}{\tau}}. \quad \ldots \quad (11)$$

K therefore decreases exponentially with time to a final value K_e which would be reached after an "infinitely long time". The relaxation time τ denotes how quickly this decrease takes place: after τ sec the difference between K and its final value has been reduced to the fraction $1/e = 37\%$, after 2τ sec it has been reduced to $1/e^2 = 13\%$, and so on.

If the angular deflection is made to reverse in sign after intervals of time which are very short compared with τ then the relaxation will have only very little effect; during the short periods of constant angular deflection the couple retains practically the value K_b (*fig. 4a*). In this case, therefore, one has to consider exclusively the relatively high initial value of the torsional rigidity, the "unrelaxed torsional rigidity". When, on the other hand, the angular deflection is made to change in sign after intervals of time which are long compared with τ, then during the long periods of constant angular deflection the couple has an average value very close to K_e (figure 4*b*). In that case one has to do with a smaller value of the torsional rigidity, the "relaxed torsional rigidity", corresponding to K_e.

Similar considerations hold when, instead of discontinuously as represented in figs 3 and 4, the value of φ changes gradually, for instance periodically according to a sine curve with angular frequency ω, as is the case with torsional oscillations. If the periodic time of the oscillations is short compared with the relaxation time, in other words if $\omega\tau \gg 1$, then the deformation is continuously changing sign so quickly that there is no perceptible displacement of the diffusion equilibrium and thus no perceptible relaxation; one has to consider the unrelaxed torsional rigidity. If, on the other hand, $\omega\tau \ll 1$ the diffusion equilibrium has every opportunity to adapt itself to the changing conditions and the wire behaves as a less rigid medium than in the first case: the relaxed torsional rigidity is then measured. In both cases the counteracting couple is continuously in anti-phase with the deformation.

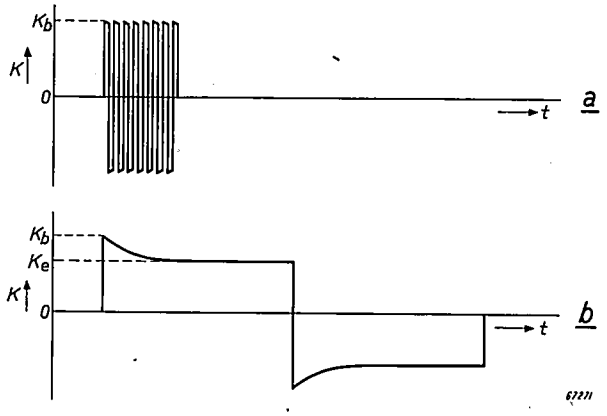


Fig. 4. *a*) When the angular deflection periodically alternates in sign at a frequency largely exceeding $1/\tau$, then during the short periods of constant angular deflection the counteracting couple K has a virtually constant value determined by the unrelaxed torsional rigidity. *b*) If the frequency is low with respect to $1/\tau$ then the mean absolute value of K during the moments of constant angular deflection is much smaller; one then has to do with the relaxed torsional rigidity.

In the frequency range for which $\omega\tau \approx 1$, that is to say in the range where the periodic time of the oscillations is of the same order as the time (a few times τ) in which the diffusion equilibrium can be almost entirely established, the influence of the interstitial solute atoms is quite different. To investigate this influence let it be assumed that the cube in fig. 2 is periodically compressed and elongated in the z direction. Within the range under consideration ($\omega\tau \approx 1$) the frequency of this compressing and elongating is not high enough to prevent a periodical variation in the occupation of the z cavities, but on the other hand it is not low enough to prevent a phase difference arising between the degree of occupation of the interstices and the deformation. The deviation from the average occupation becomes zero later and thus the counteracting force becomes zero earlier than the deformation.

Applied to our torsional oscillations this means that, apart from the difference in sign, the curve representing the variation of the couple is displaced over a certain distance β/ω, depending upon $\omega\tau$, with respect to the sine curve denoting the variation of the angular deflection. This is illustrated in *fig. 5*, where φ indicates the periodical angular

deflection and K the corresponding counter-acting couple for a medium with relaxation.

If the angle of deflection is given by $\varphi = \varphi_0 \cos \omega t$, then for K the equation reads:

$$K = K_0 \cos (\omega\tau + \pi + \beta).$$

It has already been seen that the loss angle β occurring in this expression is small both for $\omega\tau \gg 1$ and for $\omega\tau \ll 1$. The variation of β in the intermediate range is represented in *fig. 6:* a maximum

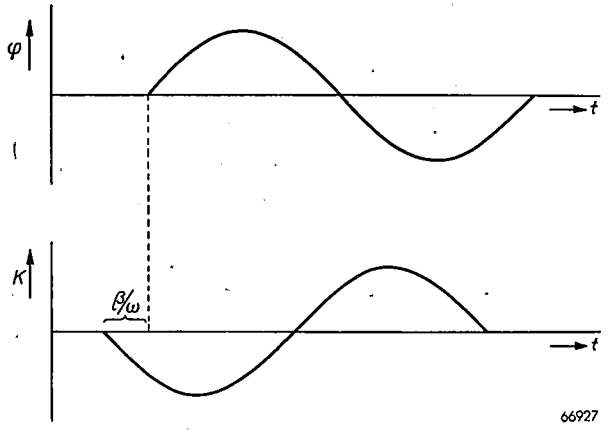


Fig. 5. When the relaxation time τ of the medium and the angular frequency ω of the pendulum satisfy the condition $\omega\tau \approx 1$ there is a phase difference $\neq \pi$ between the opposing couple K and the angle of deflection φ. The couple is a time $(\pi+\beta)/\omega$ in advance of this angle.

is reached where $\omega\tau = 1$, the level of this maximum being determined by the magnitude of the after-effect. The same diagram shows the variation of K_0/φ_0, which factor gradually changes with rising $\omega\tau$ from the relaxed to the unrelaxed torsional rigidity.

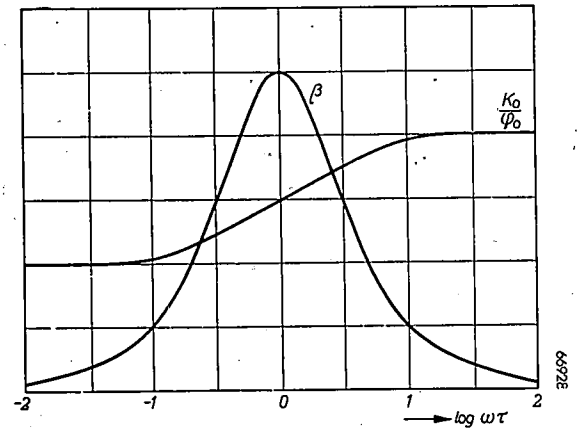


Fig. 6. β and K_0/φ_0 (the relation between the greatest values of the opposing couple and the angle of deflection) as functions of log $\omega\tau$.

It may sometimes happen that there is more than one relaxation time, for instance when different processes of diffusion are playing a part simultaneously. In that case the curve for β is a sum of two or more curves, each with its maximum at the value of ω corresponding to the respective relaxation time. If there are not too great differences between the relaxation times this superposition means a widening of the maximum in the curve for β.

Internal friction. Influence of temperature

From the discussion of external friction it has been seen that the occurrence of a phase shift β is accompanied by a dissipation of energy, thus, in the case of free oscillations, by a decay of the oscillations. This is the general rule, no matter what the cause of the phase shift may be. Therefore the loss angle β caused by the jumping of the carbon or nitrogen atoms from one interstitial position to another also leads to damping, which is denoted as "internal friction" as distinguished from the external friction dealt with in the beginning of this article. For $\omega\tau \gg 1$ and for $\omega\tau \ll 1$ the value of β was very small, which means that the jumping of the carbon or nitrogen atoms in these frequency ranges does not cause any damping. In the range where $\omega\tau \approx 1$, however, the internal friction in the case of our torsional pendulum is far from negligible, being even much greater than the external friction. In what follows, the latter will therefore be left out of consideration.

In absolute value the loss angle β determining the internal friction is also small for $\omega\tau = 1$, being at most a few degrees (the phase shift which determines the external friction is even smaller). It is therefore permissible to apply the arguments given in the beginning of this article; we then find that the amplitude of the free oscillations decreases exponentially. The order of the decrease is determined by the logarithmic decrement δ, which, according to (8), is related to β as

$$\delta = \pi\beta.$$

In the experiments to be described below the value of δ was determined by measuring the amplitude φ some time after the free oscillations began and subsequently the amplitude φ' after the wire had made a further number of n oscillations. Thus one obtains:

$$\frac{\varphi}{\varphi'} = e^{n\delta}$$

and hence

$$\delta = \frac{1}{n} \log_e \frac{\varphi}{\varphi'}.$$

The number of oscillations $n_{0.5}$, after which the

amplitude has decreased to its half-value, is given by

$$n_{0,5} = \frac{1}{\delta} \log_e 2 \approx \frac{0.7}{\delta}.$$

In our experiments δ was always between the limits 0.1 and 0.001; thus $n_{0,5}$ varied between 7 and 700 oscillations.

As already stated, β, and thus also δ, reaches its maximum value when the angular frequency ω of the oscillation and the relaxation time τ obey the relation:

$$\omega\tau = 1.$$

The level of the maximum in the curve representing the variation of δ with ω is a measure of the magnitude of the after-effect and thus a measure of the number of foreign atoms capable of jumping. Thus it is possible to learn something about this number by investigating the damping curve in the critical range.

To this end, instead of varying the frequency, the temperature of the wire may also be varied and the frequency kept constant, as will now be shown.

In order of magnitude the relaxation time τ does not differ much from the mean time that a carbon or nitrogen atom stays in a certain interstitial place, and it may therefore be represented by the formula:

$$\tau \approx \frac{a^2}{D}, \quad \ldots\ldots\ldots \quad (12)$$

where a represents the average jump length of a carbon or nitrogen atom from one position to the other (thus in our case a is in the order of the lattice constant of α-iron), whilst D is the diffusion coefficient of carbon or nitrogen respectively in α-iron. Eq. (12) may be applied because the jumping of the C(N) atoms is a diffusion phenomenon, and for these phenomena eq. (12) is of general application.

Now D is strongly dependent upon the temperature, according to the relation

$$D = D_0 e^{-\frac{E}{RT}},$$

in which R is the gas constant per mole and E the activation energy for an atom to pass from one interstice to another, this energy being of fairly high value, viz. about 80,000 joules (about 20,000 cal.). As a consequence τ is likewise strongly dependent upon the temperature, the relation being:

$$\tau = \tau_0 e^{\frac{E}{RT}}.$$

To find the maximum damping it is therefore preferable to vary τ instead of ω, by changing T. When δ is plotted against $1/T$ for a constant frequency a similar curve is obtained as when δ is plotted against $\log \omega\tau$ for a constant temperature. The frequency is automatically kept constant by employing the free oscillations of a torsional pendulum. If β is small, and this condition is always satisfied, the frequency of the damped oscillations is virtually the same as that of the undamped ones, and is not dependent upon time either (this had already been tacitly assumed in formula (5)).

Damping experiments

The fact that the internal friction can be used to get a better insight into the phenomena of quench ageing in iron and steel is due to the difference in behaviour of carbon or nitrogen in respect to damping, according to whether they occur in solid solution or as carbide and nitride precipitates. The reason for this is that the structure of iron carbide or iron nitride is such that there cannot be jumps of the C or N atoms in it, leading to a non-symmetrical deformation of the lattice. Any precipitation of a part of the dissolved carbon or nitrogen must, therefore, immediately find expression in a lowering of the maximum damping.

The measuring of the damping is a more sensitive and more absolute method of studying precipitation phenomena than the measuring of the hardness, since the latter depends also greatly upon the dimensions of the particles precipitated. For one and the same degree of hardness there may, therefore, be greatly different quantities of dissolved C or N. On the other hand the height of the damping peak gives at once the quantity of dissolved carbon or nitrogen, it having been found [3]) that there is a simple relationship between the two. It happens to be that the amount in solution, in weight %, corresponds fairly well with the value of β.

Before proceeding to discuss the experiments carried out to check the results of the hardness tests, and in order to gain a deeper insight into the underlying atomic processes, it should be recalled that the most important fact revealed by the hardness measurements (see I) was that the presence of 0.5% manganese in iron appreciably delays the precipitation of nitrogen but has no noticeable effect upon the precipitation of carbon.

The latter fact was immediately confirmed by damping experiments with wires made of the alloys Fe+0.04% C and Fe+0.5% Mn+0.04% C. The diameter and length of the wires and the moment of inertia of the body suspended from them were

[3]) L. J. Dijkstra, Philips Res. Rep. 2, 357-381, 1947.

such that the periodic time of the oscillations amounted to approximately 1 second. As shown in *fig. 7* [4]), the measurements of the damping as a function of temperature teach that immediately after quenching from 720 °C (the curves *a*) a sharp maximum in the damping occurs at

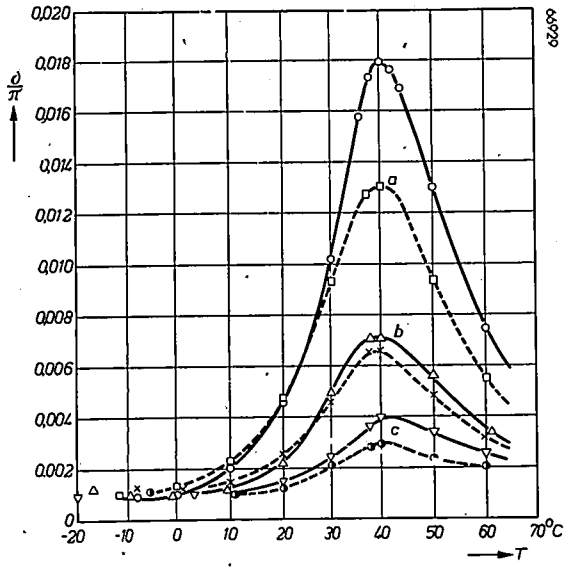


Fig. 7. The internal damping of α-iron with 0.04% carbon (fully drawn curves) and of iron with 0.5% manganese and 0.04% carbon (dotted curves) as functions of temperature. Plotted along the ordinate is the logarithmic decrement δ of the oscillations divided by π, thus the loss angle β. *a*) Immediately after quenching from 720 °C; *b*) after heating for 1 hour at 100 °C subsequent to quenching; *c*) after heating for 3 hours at 100 °C. There is considerable precipitation, which is very little affected by manganese.

about 39 °C. For Fe + 0.5% Mn + 0.04% C (dotted curve *a*) the maximum lies at the same temperature as that for Fe + 0.04% C (fully-drawn curve *a*). Also the width of the damping peak is the same in both cases. From this it is established that the rate of diffusion of carbon in iron is not influenced by the presence of manganese. Precipitation was investigated in the case of the two alloys mentioned by heating the wires for 1 hour at 100 °C after the damping measurements just described had been taken. The curves *b* show the values of δ/π after this heating. Finally the curves *c* show the damping after the wires had been heated at 100 °C for 3 hours in all. The results show that 0.5% Mn does not appreciably affect the rate of precipitation of carbon.

Whereas 0.5% Mn has no influence upon the behaviour of carbon in iron, it appears to have a very great effect upon the behaviour of nitrogen (again in conformity with the hardness measurements). This is demonstrated in *fig. 8*, the left half of which relates to the alloy Fe + 0.02% N and the right half to the alloy Fe + 0.5% Mn + 0.02% N. Heating for three hours at 100 °C leads, according to the left-hand curves, to precipitation of more than half of the dissolved nitrogen in the alloy without manganese, but according to the right-hand curves there is hardly any precipitation in the alloy containing manganese. In addition, fig. 8 reveals some other important details as regards the position and shape of the damping curve. The presence of 0.5% Mn appears to shift its maximum from 25 °C to 32 °C, whilst the half-value width of the peak appears to have increased from 24 °C to 36 °C (ignoring the small secondary maximum at 10 °C).

The fact that after the addition of manganese the damping curve still has a peak indicates that the nitrogen atoms are still able to move to and fro between the x, y and z positions. The widening of the peak, however, signifies that there is now more than one relaxation time playing a part in this diffusion process. The shifting of the maximum from 25 °C to 32 °C further shows that after the addition of manganese the value of $\omega\tau = 1$ is reached only at a temperature which is 7 °C higher than that before the addition of Mn. For the same temperature, therefore, the mean relaxation time and thus also the mean time that the atoms stay in an x, y or z position are increased. All these facts lead to the following atomic picture.

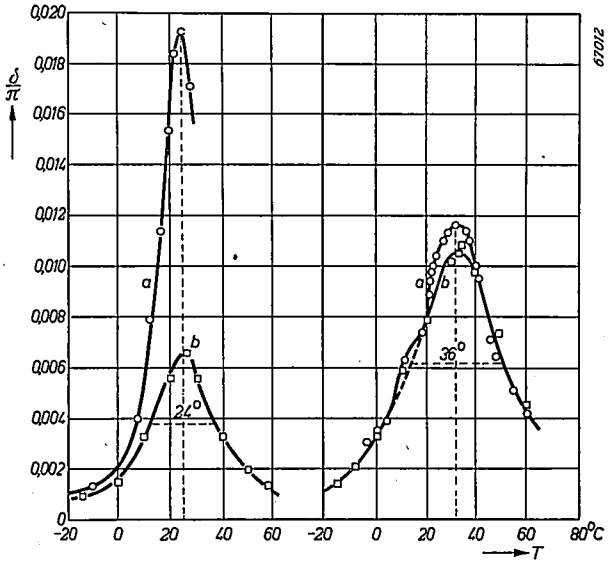


Fig. 8. The internal damping of iron with 0.02% nitrogen (left) and of iron with 0.5% manganese and 0.02% nitrogen (right) as fuctions of temperature. Plotted along the ordinate is $\delta/\pi = \beta$. *a*) Immediately after quenching from 580 °C; *b*) after 3 hours' heating at 100 °C subsequent to quenching. The alloy containing manganese shows practically no precipitation and the peak of the damping curve is displaced and broadened. The positions of the maxima are indicated by the vertical dotted lines.

[4]) For the sake of simplicity, in this and the two following diagrams the damping has been plotted direct as a function of the temperature in °C, though, in accordance with the foregoing, it is perhaps more correct to take as abscissa the reciprocal value of the absolute temperature. As ordinate the quantity $\delta/\pi = \gamma/\omega = \beta$ has been plotted.

For the nitrogen atoms it is more favourable, as regards energy, to occupy an interstice where they have a manganese atom as immediate neighbour ("Mn site") than an interstice which is surrounded exclusively by iron atoms ("Fe site"). The fact that the interstices are no longer energetically equivalent naturally leads to the occurrence of different relaxation times. According to the results of the experiments the nitrogen atoms are able to make jumps around the manganese atoms but they have difficulties in leaving those atoms. The precipitation of nitride does not take place at any appreciable rate, because each nitrogen atom that manages to liberate itself from a manganese atom is very soon "captured" again by the same or some other manganese atom. At any temperature there will be a certain equilibrium between occupied Fe sites and occupied Mn sites. The possible explanation of this behaviour of manganese in nitrogen-containing iron has already been dealt with in article I, so that it suffices here to recall that the cause is to be sought in the comparatively great affinity between manganese and nitrogen.

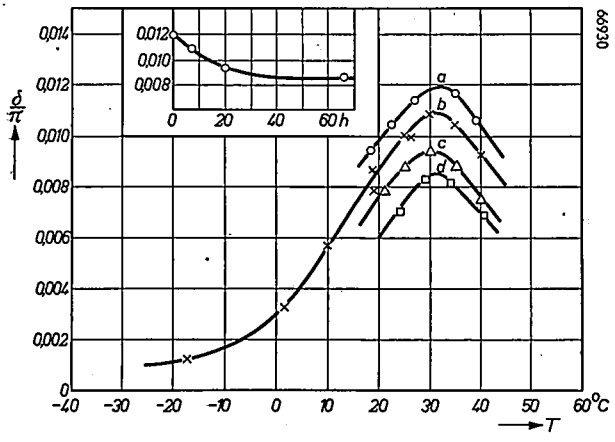


Fig. 9. Iron with 0.5% manganese and 0.02% nitrogen shows precipitation of the nitrogen (in the form of a metastable nitride) when, after quenching, the metal is heated for a long time. *a*) Damping curve immediately after quenching; *b*) after 7 hours' heating at 200 °C; *c*) after 20 hours' heating at 200 °C; *d*) after 66 hours' heating at 200 °C. In the inset, the lowering of the maximum of the damping curve due to heat treatment.

Finally, *fig. 9* shows that protracted heating at 200 °C eventually leads to the precipitation of a large part (about one-third) of the nitrogen. Even after 66 hours, however, the only precipitation is in the form of a metastable nitride, probably the phase $Fe_{16}N_2$ (also known from K. H. Jack's work; cf. footnote [12]) in I). This nitride and the more stable nitride Fe_4N are formed independently of each other and with very different rates of nucleation. The difficulties in nucleation for Fe_4N are so great that after 66 hours' heating at 200 °C this phase is still not formed in any perceptible quantity.

The results of the damping experiments form on the one hand a good confirmation of the conclusions drawn from the hardness tests, as set forth in I, and on the other hand lead to a deeper insight into the atomic processes underlying the phenomena observed. They also furnish an explanation of the technically very important fact that the precipitation of nitrogen in pure iron takes place at a much higher rate than in the commercial kinds of iron and steel (which always contain manganese!). Phenomena due to quench ageing may therefore arise in these types of steel even after comparatively slow cooling, since this "slow" cooling does after all take place quickly compared with the rate of precipitation.

Among these phenomena there is not only the gradual increase of the mechanical hardness but also that of the magnetic hardness of commercial mild steel. From the work of various investigators, particularly that of Köster [5]), it was already known that mild steel (also slowly cooled mild steel) need contain only 0.005% nitrogen for its coercive force to be doubled when heated to 100 °C for several hundred hours. This increase in coercive force is due to the precipitation of the nitrogen in the form of nitride. From the experiments described in the foregoing it may be concluded that the phenomenon of magnetic ageing is to be ascribed not to the nitrogen alone but to the simultaneous occurrence of manganese and nitrogen in the steel.

Summary. Damping experiments, at different temperatures, with a torsional wire pendulum made of iron or steel containing a little carbon or nitrogen in solution (with or without the admixture of manganese) throw some light upon the extent to which the dissolved atoms precipitate after quenching of the metal. It is shown how the internal friction, resulting from the solute atoms jumping from one interstice in the crystal lattice to another, causes the amplitude of the torsional oscillations to decay exponentially. At a certain frequency the degree of damping depends upon the rate of diffusion of the atoms and thus upon the temperature, the logarithmic decrement as a function of the temperature showing a maximum at a very definite temperature. The height of this maximum is directly proportional to the amount of dissolved atoms and thus decreases with increasing precipitation. In conformity with the conclusions drawn from hardness measurements described in the preceding article, manganese appears to have no influence upon the precipitation of carbon in carbon-containing iron after quenching, whilst the precipitation of nitrogen in nitrogen-containing iron is greatly retarded by manganese, so much so that even after 66 hours' heating at 200 °C (after quenching) only about one-third of the nitrogen is precipitated in the form of a (metastable) nitride. The results of the damping experiments throw light upon the atomic processes responsible for this behaviour and make it possible to give a new explanation of the mysterious phenomenon of "magnetic ageing".

[5]) W. Köster, Z. anorg. allg. Chem. **179**, 297-308, 1929.

BOOK REVIEW

Application of the electronic valve in radio receivers and amplifiers. A.F. amplification, output amplification and power supply, by B. G. Dammers, J. Haantjes, J. Otte and H. van Suchtelen, 452 pages, 343 illustrations. — Edited by Philips Technical Library (Book 5 of the series on electronic valves) publishers N.V. Meulenhoff & Co., Amsterdam, 1951.

The first part (book 4) of the above-mentioned work [1]) dealt with R.F. and I.F. amplification, mixing and detection. The second part now published deals with A.F. voltage amplification, output stages and supply. In a third part now in course of preparation other important subjects related to the application of the electronic valve in receivers and amplifiers will be discussed.

The second part is a worthy continuation of the first, the authors placing their vast experience at the disposal of everyone having anything at all to do with receivers and amplifiers. Practically every aspect of the application of the electronic valve is thoroughly dealt with in all its details as far as it is related to the subjects treated. Numerous calculations are given by way of illustrating the subject matter.

The book is divided into three chapters: A.F. Amplification, the Output Stage and the Power Supply. The first chapter deals in succession with A.F. amplifying circuits, phase-inverting stages, the response curve, calculation of the A.F. transformer and non-linear distortion. Under the Output Stage are discussed: the various adjustments (class A, class B, class AB, comparison of these adjustments and double-tone testing), distortion, behaviour under complex load, deviation from normal valve adjustments and overload phenomena. The third chapter deals with the filament-current supply, rectifying circuits, calculation of H.T. rectifiers and circuits for stabilizing supply voltages. For the benefit of readers desiring to go more deeply into certain subjects a bibliography is appended at the end of each chapter.

This book will prove to be an excellent vademecum for many. The uninitiated would never suspect that so much could be said about the subjects dealt with. Particularly the chapter on the output stage goes to great length, covering more than half the book and dealing with the subject in a manner unequalled elsewhere. Special attention is drawn to what is written about the double-tone method.

The clear and thorough way of treatment makes the book pleasant to read. This thoroughness might have its shadow side for those who are not yet familiar with the subject, in that they might not be able to see the wood for the trees, but the logical manner in which the subject matter is set forth ensures that the main thread need not be lost.

Notwithstanding this extensive treatment the subject of inverse feedback, which is of such great importance with A.F. amplifiers, has not been discussed in this book but kept for the third part (book 6). It might be thought that it would have been better to keep together all the problems arising in A.F. amplification, so as to facilitate reference. But with such thorough treatment as is now presented, this book would have become too voluminous if the subject of inverse feedback had been also included. It is therefore with interest that we look forward to the appearance of the third part, as an essential supplement to part 2.

TH. J. WEIJERS

[1]) Discussed in vol. **12** of this journal, page 184, 1950.

VOL. 13 No. 7, pp. 181-208 JANUARY 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE MAGNETIC AND ELECTRICAL PROPERTIES OF FERROXCUBE MATERIALS

by J. J. WENT *) and E. W. GORTER.

538.246:621.315.592.4

One of the most important results achieved from the scientific investigation of solids in the last decade is undoubtedly the development of ferromagnetic materials for use at high frequencies. Readers of this journal will already have gathered much about the application of these materials. An article devoted to this subject some years ago had of necessity to be relatively brief because the materials in question were then still in course of development. Now that this development work has been almost completed, it is possible to discuss these new ferromagnetic materials at greater length.

Introduction

The ferromagnetic materials to be dealt with in this article, and now known under the trade name of "Ferroxcube", arose from the need to avoid the losses occurring at high frequencies in the ferromagnetic metals hitherto commonly used. There are various physical causes of these losses, viz:

1) the occurrence of eddy currents,
2) the hysteresis phenomenon,
3) the so-called ferromagnetic resonance absorption [1]) and possibly other mechanisms.

The losses to be attributed to these causes are called the eddy-current losses, the hysteresis losses and the residual losses respectively.

The magnitude of the losses arising in the ferromagnetic core of a coil when a (weak) alternating current flows through it, is represented in practice by the so-called loss factor [2])

$$\frac{\Delta R}{\mu_r f L},$$

where ΔR is the equivalent resistance which (connected in series with the coil) would cause a loss equal to the losses arising from the ferromagnetic material, μ_r is the relative magnetic permeability of the ferromagnetic material ($\mu = \mu_r \mu_0$ [3])), f the frequency of the alternating current and L the inductance of the coil with core. In the case of a ring coil $L = \mu_r L_0$, where L_0 is the inductance of the coil without core. The significance of the loss factor becomes clear when it is written in the form:

$$\frac{\Delta R}{\mu_r f L} = \frac{2\pi}{\mu_r} \cdot \frac{\Delta R}{\omega L} = \frac{2\pi}{\mu_r} \frac{1}{Q},$$

where Q represents the quality of a circuit with resistance ΔR and inductance L. The factor $1/\mu_r$ has been added because the loss factor so defined does not change when an air gap is introduced in the magnetic circuit.

For the known metallic ferromagnetics it is customary to write this loss factor in the following form:

$$\frac{\Delta R}{\mu_r f L} = C_e f + C_h B_{max} + C_r, \quad \ldots \quad (1)$$

where C_e and C_h are constants, namely the eddy-current constant and the hysteresis constant. The term C_r represents the residual losses, which could be attributed to the causes mentioned under (3)

*) At present of N.V. Kema, Arnhem (Netherlands).

[1]) See H. G. Beljers and J. L. Snoek, Philips Techn. Rev. 11, 313, 1949.

[2]) See J. L. Snoek, Philips Techn. Rev. 8, 353, 1946.

[3]) In the system of rationalized Giorgi units used here the permeability of vacuum is $\mu_0 = 4\pi/10^7$ H/m and the dielectric constant of vacuum $\varepsilon_0 = 10^7/4\pi c^2$ F/m.

above. B_{max} is the peak value of the magnetic induction. In a well-conducting ferro-magnetic material the eddy-current term appears to contribute by far the most to the loss factor at high frequencies, the more so since as a rule for use at high frequencies small values of B_{max} suffice. For a plate of thickness D which is magnetized in a direction parallel to its plane, and assumed to be homogeneous and isotropic with respect to permeability, the eddy-current constant C_e assumes the following form:

$$C_e = \frac{\pi^2 \mu_0}{3} \frac{D^2}{\varrho}, \quad \ldots\ldots (2)$$

where ϱ is the resistivity of the ferromagnetic material. It is seen that instead of using a solid core it is advantageous to build up the core from a large number (n) of insulated thin plates with thickness D/n, which has led to the use of laminated cores. In fact, for a core with total thickness D constructed in this way the eddy-current constant is a factor $1/n$ smaller than that for a solid core of the same material and the same dimensions. The higher the frequency for which the core is to be used, the thinner the laminations should be. Rolling very thin sheets, however, is an expensive process and therefore at high frequencies powder cores are used, consisting of minute, insulated particles with a diameter of about 5 microns. But the use of these powder cores has great disadvantages, since the core is not homogeneously filled with the ferromagnetic material. The magnetization is therefore not uniform throughout, and as a result there is a loss of "effective" permeability and the losses mentioned are increased [4]).

These disadvantages led to the search for a ferromagnetic core material with such a high specific resistance ϱ as to make the eddy-current losses negligible (see eq. (2)), even when the material is used in a homogeneous form, i.e. neither laminated nor in powder form. For this reason ferromagnetic oxides have been developed, and those that are of technical importance have come to be known under the name of Ferroxcube. In conformity with the requirement for their development, these materials usually have such a high resistivity that the eddy-current losses become negligible. This does not imply, however, that no losses at all occur in Ferroxcube materials, but only that the losses which occur in them are mainly to be classified as hysteresis and residual losses. The residual losses (term C_r in eq. (1)) are a function of frequency (see note [2])).

With a certain frequency and a certain maximum value of the induction it can be ascertained (for instance by measuring the hysteresis curve) what part of the total losses is due to hysteresis. It is not, however, possible — and this should be borne in mind right from the beginning — to apply this separation in such a general manner as is expressed in formula (1). If this formula were nevertheless applied to Ferroxcube materials it would be found that the coefficient C_h is also frequency dependent, whilst the term C_r depends on the induction as well as on the frequency. As will be shown farther on, however, the total loss arising from the ferromagnetic material can be expressed in a useful manner.

This article will now deal with the following points: the chemical composition and the crystallographic structure of the Ferroxcube materials and, in connection therewith, the saturation magnetization [5]), which is of fundamental importance for the ferromagnetism; the behaviour of Ferroxcube materials in weak fields, attention being directed to the magnetic losses and permeability (also as a function of temperature), which is of the utmost importance for practical applications; the behaviour of these materials in strong fields, attention again being paid to permeability and losses. Finally something will be said about the exceptional dielectric properties of Ferroxcube materials which under certain circumstances may lead to additional losses.

Composition and preparation of the ferrites

As far as their chemical composition is concerned Ferroxcube materials belong to the group of ferrites with the composition $Me^{2+}Fe_2^{3+}O_4$, where Me^{2+} is a symbol for a divalent metal. In particular all Ferroxcube materials are cubic ferrites, with the same crystal structure as Fe_3O_4, which is also the structure of the mineral spinel ($MgAl_2O_4$) and is therefore called the spinel structure [6]). Ferrites

[4]) See the article quoted in note [2]). — When the particles are very small also the "intrinsic" permeability of the particles is smaller; see L. Néel, C. R. Paris 224, 1550, 1947.

[5]) As is presumably known, a ferromagnetic material in the non-saturated state is divided into Weiss domains. Within each domain the material is spontaneously magnetized and the order of magnetization equals the saturation magnetization corresponding to the temperature in question. The fact that the magnetization of the material as a whole differs from the saturation magnetization is due to the differences in direction of the magnetization in the various domains. The term saturation magnetization, without any further definition, is to be understood here as the value of the magnetization in sufficiently strong fields at a temperature of 0 °K; it is thus equal to the spontaneous magnetization at that temperature. (The values quoted in this article are those measured at the temperature of liquid nitrogen, but the resultant error is not of great consequence.)

[6]) See also: E. J. W. Verwey, P. W. Haaijman and E. L. Heilman, Philips Techn. Rev. 9, 185, 1947.

may be imagined as being derived from Fe_3O_4 by replacing the divalent ferrous ions by one or more divalent metal ions, e.g. Mn, Co, Ni, Cu, Mg, Zn or Cd.

Ferrites are usually prepared by a sintering process such as commonly employed in the ceramic industry. The component metal oxides are mixed, ground, usually pre-sintered and ground again, finally compressed into the desired shape with the necessary binders and sintered at a high temperature. This sintering is accompanied by a certain amount of shrinkage, which has to be allowed for when determining the dimensions of the mould. In the final sintering process the atmosphere plays a large part in determining the degree of oxidation of the product, which is of great importance. It is also pointed out that this sintering process does not yield an absolutely solid material but a product containing a small volume percentage of voids.

Crystal structure of the ferrites

As already remarked, Ferroxcube materials with the chemical composition $Me^{2+}Fe_2^{3+}O_4$ crystallize with the spinel structure. This structure may be described as follows (see *fig. 1*). The large oxygen

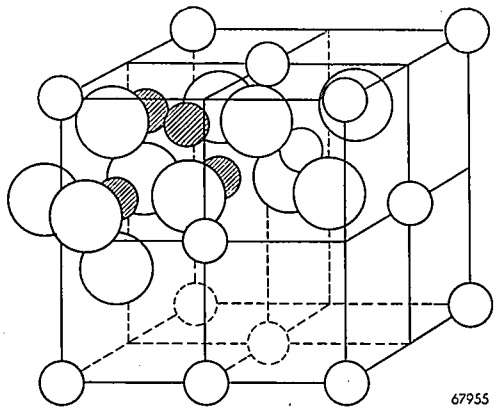


Fig. 1. The unit cell of the spinel lattice. The large spheres are oxygen ions, the small hatched spheres ions in octahedral sites, the small non-hatched spheres ions in tetrahedral sites. Only in two of the eight octants forming the unit cell have all the ions been drawn. There are twice as many ions in octahedral sites as in the tetrahedral ones.

ions form, to a very good approximation, a cubic close-packed structure of spheres. The small divalent and trivalent metal ions are found in the small interstices between the large oxygen ions. In a close packing of spheres there are two kinds of interstices, one being surrounded by four oxygen ions, the so-called tetrahedral sites, the other by six oxygen ions, the octahedral sites. In the spinel structure twice as many octahedral as tetrahedral sites are occupied. A very important point is the distribution of the various metal ions among the crystallographic sites available. The fact that there are several possibilities in regard to this distribution was first demonstrated by Barth and Posnjak [7]). This was further investigated in connection with ferrites by Verwey and Heilmann [8]), with the following result:

In the case of the so-called "normal" spinels the divalent metal ions occupy the tetrahedral sites, whilst all octahedral sites are occupied by the trivalent metal ions. This arrangement is found, with some minor deviations, for Zn and Cd ferrites. All other known simple ferrites (those containing only one kind of divalent metal ions) are said to be inverted, since the divalent metal ions are situated entirely (or mainly) in octahedral sites, while the trivalent ions occupy for one half the remaining octahedral sites and for the other half the tetrahedral sites.

Saturation magnetization

Single ferrites

The value of the spontaneous magnetization (saturation magnetization) differs for the various ferrites. It is zero for zinc and cadmium ferrites, i.e. these two ferrites (the only simple ferrites which are not inverted; see above!) are not ferromagnetic. All other ferrites are ferromagnetic.

It has now been found that the value of the saturation magnetization of Ferroxcube materials can be related to the crystal structure described, in particular to the distribution of the ions among the two crystallographic sites available in the spinel lattice.

It should first be mentioned that for the ferromagnetic metals it has hitherto not been possible to derive a theoretically calculated value for the saturation manetization agreeing with the values found. For ionic compounds such as those with which we are concerned here, one would in principle expect this to be possible, since it is known that in paramagnetic salts in which the same ions occur a certain magnetic moment can be ascribed to each ion. It is reasonable to assume that the magnetic moment of a certain ion in a ferromagnetic ionic compound, in our case a ferrite, is equal to that in a paramagnetic salt. Assuming, moreover, that all magnetic ions fully contribute towards the magnetization (i.e. that at a temperature of 0 °K the magnetic

[7]) T. F. W. Barth and E. Posnjak, Z. Krist. **82**, 325, 1932.
[8]) E. J. W. Verwey and E. L. Heilmann, J. chem. Phys. **15**, 174, 1947.

moments of all ions are parallel), it should be possible to calculate the saturation magnetization of our ferrite expressed in Bohr magnetons per "molecule" $Me^{2+}Fe_2^{3+}O_4$ by adding the moments of the ions present.

The values found experimentally, however, are much smaller than those obtained by such a calculation; see *table I*, columns 4 and 6.

Special consideration has to be given to cases such as, for instance, that of zinc ferrite, where the tetrahedral sites are exclusively occupied by non-magnetic Zn ions. Here no interaction can occur between magnetic moments on different lattice sites, such as governs the situation described above. The only magnetic interaction likely to occur in this case is that between the ferric ions in the octahedral sites themselves. Such an interaction tends to make the number of anti-parallel pairs of neighbours as large as possible. If each magnetic moment has only anti-parallel nearest neighbours then the apparent total magnetization will be zero, as is indeed found to be the case with zinc ferrite. For the complicated spinel lattice, it is true, it is not easy to imagine a condition where all the nearest neighbours of an ion have moments anti-parallel to it. Yet here, too, an arrangement is feasible whereby the total magnetization is zero. The resulting moment will of course also be zero if the interaction between the ferric ions in octahedral sites is negligibly small and thus all magnetic moments have random orientation.

Table I. The magnetic moments of various ferrites at saturation M_s, expressed in the number of Bohr magnetons per molecule [9]).

Ferrite	Ions in tetrahedral sites	Ions in octahedral sites	$M_{s_{oct.}}+M_{s_{tetr.}}$	$M_{s_{oct.}}-M_{s_{tetr.}}$	$M_{s_{exp.}}$
ferrous ferrite	Fe^{3+}	$(Fe^{2+}Fe^{3+})$	(4+5)+5=14	(4+5)—5= 4	4.2
manganese ferrite	Fe^{3+}	$(Mn^{2+}Fe^{3+})$	(5+5)+5=15	(5+5)—5= 5	5.0
nickel ferrite	Fe^{3+}	$(Ni^{2+}Fe^{3+})$	(2+5)+5=12	(2+5)—5= 2	2.3
zinc ferrite	Zn^{2+}	$(Fe^{3+}Fe^{3+})$	(5+5)+0=10	(5+5)—0=10	0
Column 1	2	3	4	5	6

The explanation for this discrepancy has been given by Néel [10]), who assumes that a magnetic dipole on a tetrahedral site has a strong tendency to align itself in anti-parallel orientation to the dipoles in adjacent octahedral sites. A similar, through less pronounced, interaction takes place also between magnetic moments in one and the same kind of lattice site.

These assumptions have, among others, the following consequences [11]):

The magnetic moments in sites of one kind have parallel orientation and the total magnetization of the ions in the octahedral sites is anti-parallel to that of the ions in the tetrahedral sites, since all "tetrahedral neighbours" of a certain "octahedral moment" will be anti-parallel to that moment. All octahedral neighbours of the tetrahedral moments will in turn be anti-parallel to the latter, i.e. they will be parallel to the dipole in the octahedral site considered first, and so on. The magnetization of the whole domain will therefore be equal to the difference of the partial magnetizations of the dipoles in the octahedral and the tetrahedral sites, this being approximately in agreement with the values found experimentally (table I, columns 5 and 6).

Mixed ferrites

The ferrites known as Ferroxcube are for the greater part mixed crystals of two or more single ferrites and as such are denoted here as "mixed ferrites". The most important of these are the MnZn ferrites (Ferroxcube III) and the NiZn ferrites (Ferroxcube IV).

Examination by X-ray diffraction shows that the Zn ions in these mixed ferrites occupy tetrahedral sites, just as is the case of zinc ferrite. The Mn and Ni ions occupy octahedral sites, as is the case for Mn and Ni ferrites. The ferric ions occupy the remaining tetrahedral and octahedral sites. The distribution may be represented symbolically by the formula:

$$(Zn_x^{2+}\,Fe_{1-x}^{3+})_{tetr.}\,(Me_{1-x}^{2+}\,Fe_{1+x}^{3+})_{oct.}\,O_4.$$

[9]) From the values given in column 6 one can find the saturation polarization in Wb/m² by multiplying those values by 7.0 times the apparent density in g/cm³ (table III, column 2) divided by the molecular weight in grammes. The saturation polarization J_s in Wb/m² equals the saturation magnetization M_s in A/m multiplied by μ_0.

[10]) L. Néel, Ann. Physique 3, 137, 1948.

[11]) For a more extensive treatise in connection with these assumptions see the article quoted in note [10]).

If the Zn content is small there are sufficient ferric ions in the tetrahedral sites to cause all the magnetic moments on the octahedral sites to be mutually parallel. As already seen, in that case the saturation magnetization is given by the difference of the partial magnetizations of the two lattice sites. Since, owing to the presence of (non-magnetic) Zn ions the partial magnetization of the ions in the tetrahedral sites will be less than that in the case of pure Ni ferrite (or Mn ferrite), the saturation magnetization of a mixed ferrite will increase with the Zn content (at least as long as the concentration of Zn ions remains small).

This has indeed been found experimentally; see *table II*.

Under certain circumstances the replacement of magnetic ions in a ferromagnetic material by non-magnetic ions thus increases the saturation magnetization, which at first sight seems remarkable.

The above-mentioned restriction that this increase occurs only for low Zn contents can be explained as follows [12]). For large concentrations of Zn ions it will happen that, at least in some small parts of the crystal, there are so few magnetic ions in tetrahedral sites that in those parts of the crystal the tetrahedral-octahedral interaction tending to give the moments on the octahedral sites a mutually parallel orientation is no longer capable of overcoming the influences, mentioned above for zinc ferrite, tending to give the magnetic moments either anti-parallel or random orientations. As a result the saturation magnetization is reduced (see *fig. 2*).

The anti-parallel orientation of the moments makes the maximum saturation magnetization attainable for a ferrite comparatively small. Metallic iron has a saturation magnetization three times as large. Consequently, for applications where a high induction is desired and high frequencies do not occur, such as for power transformers, Ferroxcube is obviously less suitable.

[12]) E. W. Gorter, Nature **165**, 798, 1950; C. R. Paris, **230**, 192, 1950.

The Curie temperature of mixed ferrites

Table II gives the Curie temperature (T_c) of NiZn ferrites with different Zn contents. It shows that the Curie temperature decreases gradually with increasing Zn content, and such appears to be the case with all mixed Zn ferrites. It is therefore possible to produce materials with a Curie temperature as low as may be desired, in the form of mixed ferrites with a suitable zinc content. It will be shown that this possibility is of great practical value.

Table II. The saturation magnetization M_s in Bohr magnetons per molecule and the Curie temperature T_c for different NiZn ferrites.

Ions in tetrahedral sites	Ions in octahedral sites	$M_{s_{theor.}} = M_{s_{oct.}} - M_{s_{tetr.}}$	$M_{s_{exp.}}$	$M_{s_{theor.}} - M_{s_{exp.}}$	T_c in °C
$Fe_{1.0}$	$Ni_{1.0}Fe_{1.0}$	(2.3+5.0)—5.0=2.3	2.3	0	585
$Zn_{0.1}Fe_{0.9}$	$Ni_{0.9}Fe_{1.1}$	(2.1+5.5)—4.5=3.1	2.9	0.2	530
$Zn_{0.3}Fe_{0.7}$	$Ni_{0.7}Fe_{1.3}$	(1.7+6.5)—3.5=4.7	4.2	0.5	435
$Zn_{0.5}Fe_{0.5}$	$Ni_{0.5}Fe_{1.5}$	(1.1+7.5)—2.5=6.1	5.0	1.1	295
$Zn_{0.7}Fe_{0.3}$	$Ni_{0.3}Fe_{1.7}$	(0.7+8.5)—1.5=7.7	3.9	3.8	85

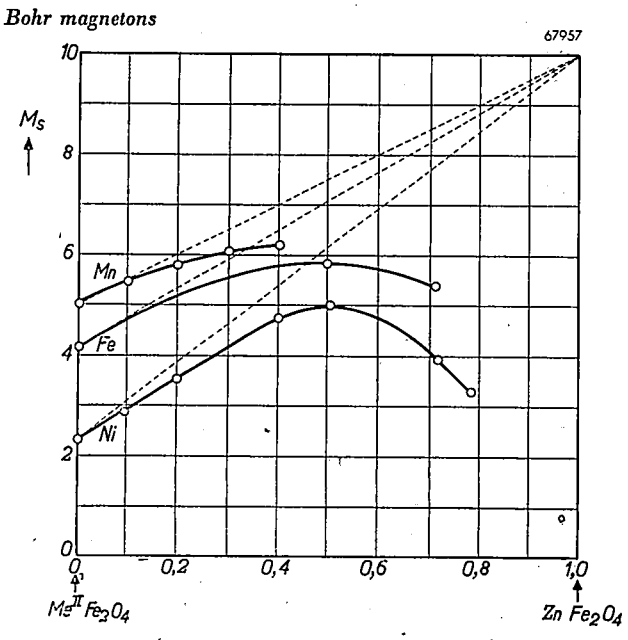


Fig. 2. The saturation magnetization of mixed ferrites, plotted in Bohr magnetons per molecule as a function of the zinc ferrite content, for different mixed ferrites. The broken lines would represent this relationship if the ions in the tetrahedral sites were always able to direct the magnetic moments of the ions in the octahedral sites. With increasing zinc content (in the tetrahedral sites) however the directing action decreases, owing to the small number of magnetic ions remaining in the tetrahedral sites. This accounts for the shape of the curves drawn, which coincide with the broken lines when the Zn content is zero but diverge from those lines as the Zn content increases.

The behaviour described may be explained qualitatively in the following way.

It is known that in a simple ferromagnetic material with only one kind of lattice sites the average energy of a magnetic moment at the Curie point due to the heat movement, kT_c, is of the same order as the average energy which would be required to invert a moment when starting from a state of complete orientation. This energy is proportional to the average number of nearest neighbouring sites occupied by magnetic ions.

This picture is easily generalized for the case of ferrites where there are magnetic moments on two kinds of lattice sites, and thus magnetic moments with two kinds of surroundings. If the ratio of the number of magnetic ions on tetrahedral sites to the number of magnetic ions on octahedral sites is t/o ($t + o = 1$), then the order of the quantity kT_c is given by $t \cdot E_t + o \cdot E_o$, where E_t and E_o are the energies which, starting from complete orientation ($T = 0$ °K), would be required to invert respectively a moment on a tetrahedral site and a moment on an octahedral site. (Here all differences between kinds of magnetic ions and the interactions between ions in equivalent sites are ignored.)

Comparing now, for instance, pure Ni ferrite $Fe(NiFe)O_4$ with NiZn ferrite $Zn_xFe_{1-x}(Ni_{1-x}Fe_{1+x})O_4$, we see that the energy required to invert a "tetrahedral moment", E_t, (to the approximation used here) will be the same in both cases, while the energy needed to invert an "octahedral moment", E_o', will be less for the NiZn ferrite than for the Ni ferrite (E_o), because in the mixed ferrite an octahedral moment has fewer magnetic "tetrahedral neighbours"; the relation will be $E_o' = (1 - x)E_o$, since the average number of magnetic neighbours of an octahedral site in NiZn ferrite will be $(1 - x)$ times the number of neighbours of an octahedral site in Ni ferrite.

Finally, therefore, for the Curie temperature T_c we have approximately in the case of $Fe(NiFe)O_4$:

$$kT_c \approx \frac{1}{3} E_t + \frac{2}{3} E_o = \frac{1}{3} (E_t + 2E_o),$$

and for $Zn_xFe_{1-x}(Ni_{1-x}Fe_{1+x})O_4$:

$$kT_c \approx \frac{1-x}{3-x} E_t + \frac{2}{3-x} E_o' =$$

$$= \frac{1-x}{3-x} (E_t + 2E_o) = \frac{1-x}{1-\frac{x}{3}} \frac{1}{3} (E_t + 2E_o).$$

The resulting Curie temperature T_c for NiZn ferrite for all values of x ($0 < x < 1$) is lower than that for pure Ni ferrite and decreases with increasing x (increasing Zn content) [13]).

Resistivity

The resistivity of Fe_3O_4 (ferrous ferrite) is about a factor 10^3 times larger than that of iron. The resistivity of the single ferrites derived from Fe_3O_4 is very much larger than even this, being a factor of about 10^{11} greater than that of iron.

The relatively low value of the resistivity of Fe_3O_4 can be understood by considering the crystallographic structure described above [14]). In this material —which can be written as $(Fe^{3+})_{tetr.}(Fe^{2+}Fe^{3+})_{oct.}O_4$ — ferrous ions of different valency are practically distributed at random among equivalent lattice sites, viz. among the octahedral sites. It requires apparently little energy to move an electron from a divalent ion to a trivalent one, because after this displacement of the electron the distribution of the divalent and trivalent ferric ions among the octahedral sites will again be arbitrary. This easy displacement of the electrons from divalent to trivalent ions means that the resistivity is low.

In mixed ferrites containing Fe^{2+} ions the influence of this mechanism, with corresponding reduction of the resistivity, can also be detected. The significance of this will be made clear below.

Behaviour in weak fields

Initial permeability

In addition to the saturation magnetization the magnetic permeability is also of great practicable importance. As a rule the highest possible permeability is desired. This permeability, μ, depends upon the strength of the field, the magnetic history (field strength at earlier moments) and the presence of a constant polarizing field. Further, μ is dependent upon the frequency — at least at high frequencies — and upon the temperature.

In high-frequency technique materials are mostly used in the demagnetized state and, moreover, with weak fields, so that particular attention must be paid to the initial permeability μ_i. The conditions for which a high value can be obtained for μ_i will now be considered.

These conditions are:

1) A sufficiently high value of the saturation magnetization; this quantity has already been discussed at some length.

2) The smallest possible amount of non-magnetic impurities and voids in the material. This point particularly needs attention in the case of ferrites because, as already mentioned, these are obtained by a sintering process, with the result that the maximum density is not reached and there is always a certain small quantity of voids.

3) The smallest possible magnetic anisotropy. In a certain sense magnetic anisotropy is the cause of a finite (not infinitely large) permeability; the term magnetic anisotropy is used to denote all phenemona which promote magnetization in a certain preferential direction (possibly differing from point to point in the material) or which, in other words, obstruct magnetization in the direction of the external magnetic field. There are three

[13]) For a more quantitative treatment see K. F. Niessen, Physica **17**, 1033, 1951.

[14]) J. H. de Boer and E. J. W. Verwey, Proc. Phys. Soc. **49**, extra part, 59, 1937.

factors contributing towards magnetic anisotropy: the crystal anisotropy, the stress anisotropy and the shape anisotropy.

a) Crystal anisotropy

From the point of view of energy it is often advantageous if the magnetization is oriented in a certain crystallographic direction. The energy E of each magnetic moment thus depends upon the angle Θ between its direction of orientation and the preferential direction in question. Thus a couple $-\mathrm{d}E/\mathrm{d}\Theta$ acts upon each elementary magnet, tending to turn it into the preferential direction. It is known that in general this crystal anisotropy, as it is called, decreases with rising temperature, and considerably below the Curie point it may often decrease to a very small value. For our considerations it is not so much the absolute temperature T that is of importance but rather the (reduced) temperature relative to the Curie point, $T^* = T/T_c$. For a given absolute temperature, the relative temperature $T^* = T/T_c$, which determines the value of the anisotropy, becomes higher as T_c is lowered. As pointed out by Snoek, it is desirable to make the Curie temperature as low as is compatible with the purpose for which the material is to be used [15]). (The temperatures which occur in the practical application must obviously always be kept below the Curie point.)

By employing the method already mentioned for reducing the Curie point (preparation of a mixed Zn ferrite) it is therefore possible to satisfy one of the conditions for the highest possible initial permeability μ_i. (See also *table III*, columns 4 and 6.) This procedure is particularly effective in the case of Ni ferrite, since the crystal anisotropy of this material has been found to be much larger than that of Mn ferrite.

b) Stress anisotropy

Stresses (internal as well as external) in a ferromagnetic material as a rule give rise to anisotropy. The magnetic moments show a preference to be oriented parallel to the direction of the stress σ or in a plane perpendicular to it, according to the sign of the magnetostriction constant λ. The magnitude of this anisotropy is always proportional to the product $\lambda\sigma$ [16]).

During the cooling of a polycrystalline material internal stresses are always set up in a solid substance, partly through non-isotropic thermal expansion and partly through other causes. The thermal expansion in cubic crystals is isotropic, and it is for this reason that all Ferroxcube materials consist exclusively of cubic ferrites. Considering, however, the small internal stresses still remaining even in a cubic material, though it may be cooled with the utmost care, it is endeavoured to keep magnetostriction small if a high initial permeability is desired.

As is the case with crystal anisotropy, magnetostriction usually decreases as the temperature approaches the Curie temperature. With respect to magnetostriction it is therefore also desirable to make the Curie temperature as low as possible [17]).

[15]) The saturation magnetization is also a function of T^*, decreasing with increasing T^* in all ferromagnetic materials so far known. From the fact that by raising T^* the value of μ_i is actually increased it can be concluded that in general the favourable effect of the decrease of the crystal anisotropy is greater than the adverse effect of decrease of the saturation magnetization.

[16]) If the material is isotropic in respect to magnetostriction, the work required per cm³ to cause the magnetization to turn over an angle φ away from the direction of the tensile stress σ is given by $E = {}^3/_2\,\lambda\sigma\sin^2\varphi$.

[17]) See also: J. L. Snoek, New developments in ferromagnetic materials, Elsevier Publishing Comp., New York — Amsterdam 1947.

Table III. Some properties of different kinds of Ferroxcube. The values apply to representative samples taken from the normal production of each material.

Material	Apparent density in g/cm³	Saturation polarization*) at 20 °C in 10^{-4} Wb/m²	Initial permeability μ_i/μ_0	$\frac{\mu_0}{\mu_i^2}\frac{\mathrm{d}\mu_i}{\mathrm{d}T}$	Curie temp. T_c in °C	μ''/μ' at 100 kc/s	Freq. at which $\tan\delta = 0.1$ in kc/s	Saturation magneto-striction $\lambda_s \times 10^6$	Resistivity ϱ for direct current in Ωm	Dielectric constant ε_r' at low frequencies
III^A	4.93	3350	1750	+ 3.7	130	0.02	300	— 1.0	1.8	appr. 10^5
III^B	4.93	4650	1245	+ 0.5	160	0.01	460	— 0.5	1.5	
III^C	4.91	4750	1280	+ 2.9	190	0.017	420	between −0,5 and 0	0.6	
III^D	4.78	5100	806	+ 4.2	235	0.010	830	— 1.1	1.0	
IV^A	4.90	3665	650	+ 9	140	0.013	1 500	— 4.2	$>4\cdot10^4$	appr. 10^3
IV^B	4.55	4170	230	+13	265	0.015	5 500	— 7.5		
IV^C	4.17	4030	90	+20	385	0.010	16 000	—15.5		
IV^D	4.07	3550	45	+27	465	0.006	29 000	—18.5		
IV^E	4.04	2460	17	+40	585	0.006	60 000	—22		
Column 1	2	3	4	5	6	7	8	9	10	11

*) As is known, the saturation polarization for metallic iron is in the order of $15000\cdot10^{-4}$ Wb/m².

As a matter of fact there is yet another way to reduce magnetostriction to a negligible value. For all ferrites except ferrous ferrite the magnetostriction is found to be negative. Fe_3O_4 has a large positive magnetostriction. Therefore by preparing a mixed ferrite with only a small quantity of Fe_3O_4 (this means the introduction of divalent iron ions), using ferrites already having a small (negative) magnetostriction, such as M_n ferrite or M_nZ_n ferrite, the magnetostriction can be reduced to a negligible value. When applying this method of increasing the initial permeability it must be borne in mind, however, that at the same time also the resistivity is decreased (and with this the eddy-current losses); this therefore limits the amount of Fe_3O_4 that can be introduced in the mixed crystal.

The ferrites denoted by the name of Ferroxcube III are all MnZn ferrites with practically zero magnetostriction. As a result, however, their resistivity is rather low, lying for the different kinds of Ferroxcube III between 0.5 and 2 Ωm (see table III, columns 9 and 10).

c) Shape anisotropy

A non-spherical body, even when crystal and stress anisotropy are absent, always has a preferential direction of magnetization, viz. the direction for which the demagnetizing field is the minimum (i.e. the direction of its greatest dimension). A non-spherical void in a ferromagnetic medium is equivalent to a non-spherical magnetic body in non-ferromagnetic surroundings, in the sense that a preferential direction exists also for the apparent magnetization of that void in a ferromagnetic medium, whose magnetization is opposed to that of the medium. This means that the magnetization of the medium in which the void is situated will preferentially be parallel to the direction of the longest dimension of the void. Therefore, in preparing the ferrite care must be taken that the inevitable voids are as far as possible spherical and as few as possible in number.

The temperature-dependence of permeability

From what has been said about the crystal anisotropy and magnetostriction it follows that in general the initial permeability μ_i increases with rising temperature and reaches a maximum in the neighbourhood of the Curie temperature. The temperature coefficient of the initial permeability is therefore as a rule positive (see *fig. 3* and table III, column 5). If, in addition to a very weak alternating field, also a constant polarizing field H_p is applied, then the permeability corresponding to the alternating field is the reversible permeability μ_{rev}. The temperature coefficient for this permeability is as a rule smaller than that for the initial permeability. For certain values of the polarizing field that coefficient may even assume negative values.

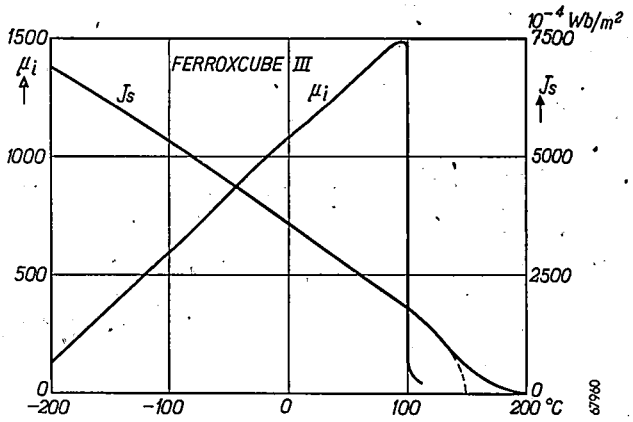


Fig. 3. The saturation polarization J_s and initial permeability μ_i of Ferroxcube III as functions of temperature.

This behaviour is easily understood. The polarization J of the material as a function of H is given for two different temperatures by the two magnetization curves in *fig. 4*. When the temperature is raised, the normal J-μ_0H curve changes into the curve given by the broken line. The general trend of this latter curve is at once clear when it is borne in mind that with increasing temperature the saturation magnetization decreases and the initial permeability increases. The steeper the slope of the magnetization curve at the point with abscisse H_p, the large the permeability μ_{rev} will be.

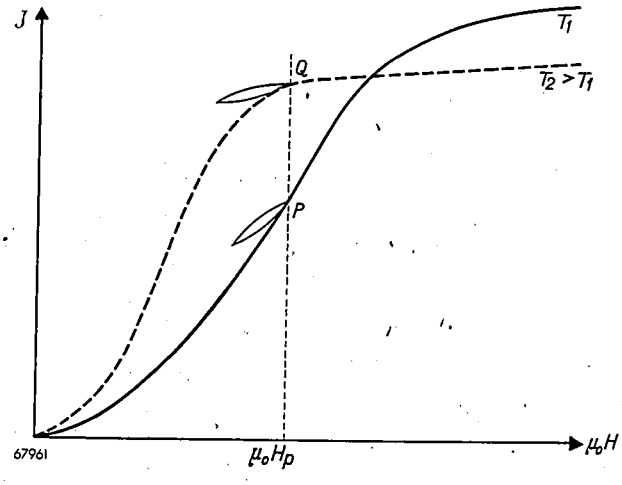


Fig. 4. The relation between the polarization J and the field μ_0H for different temperatures. The points P and Q of the magnetization curves give the values of the polarization in the field H_p at two different temperatures T_1 and $T_2 > T_1$. The reversible permeability μ_{rev} for very small alternating fields, superimposed on H_p, increases with the slope of the magnetization curve at the point with abscisse H_p. It is seen that μ_{rev} is greater at P than at Q, the temperature coefficient of μ_{rev} being negative there, while in the case of a polarizing field $H_p = 0$ it is positive.

Fig. 4 shows quite clearly that for the chosen value of H_p the quantity μ_{rev} is smaller for the high temperature than for the low one. Thus for this value of H_p the temperature coefficient of μ_{rev} is negative.

From the foregoing it can be seen that by applying a polarizing field of suitable magnitude the temperature coefficient of the permeability can be adjusted within wide limits, making it either positive or negative and even zero.

Losses at low values of the induction

In the introduction it was already pointed out that in Ferroxcube materials the eddy-current losses are generally negligible, owing to the high resistivity of these materials. Therefore only the hysteresis and residual losses have to be considered. It appears that for low values of the induction the hysteresis losses (which can be determined, for instance, from the hysteresis curve) usually form only a small proportion of the total losses measured. The residual losses, which in this case are therefore responsible for the greater part of the total losses, can be dealt with by regarding the permeability μ as a complex quantity: $\mu = \mu' - j\mu''$, thus giving expression to the fact that the induction B comprises two components, one in phase with the field applied (and determined by μ') and one 90 ° in phase behind the field (determined by μ''). The resulting induction B is therefore a certain (small) angle behind the field applied; the phase angle (loss angle) is given by $\tan \delta = \mu''/\mu'$.

The loss factor accompanying the occurrence of a complex permeability is given by:

$$\frac{\Delta R}{\mu_r' f L} = \frac{2\pi}{\mu_r'} \cdot \frac{1}{Q} = 2\pi \frac{\tan \delta}{\mu_r'}.$$

From this it follows directly that the behaviour of μ'' is just as much a matter of interest as the behaviour of μ'.

If the ferrite has been prepared with the necessary care, so that the presence of voids and of foreign phases is avoided as far as possible and the internal stresses are small, then it appears that the quantity $\tan \delta$ at low frequencies varies but little for the various kinds of Ferroxcube and is always in the order of 0.01 (see table III column 7); μ' on the other hand may assume widely different values. For use at low frequencies a ferrite with the highest possible μ' is therefore the most advantageous.

At higher frequencies complications arise owing to the fact that both μ' and μ'' are frequency-dependent. This dependence upon frequency may roughly be described as follows (see *figs 5* and *6*). Up to a certain critical frequency the real part μ' of μ is practically independent of the frequency, but above that it shows a definite decrease with increasing frequency. In the critical frequency range μ'' reaches a maximum value.

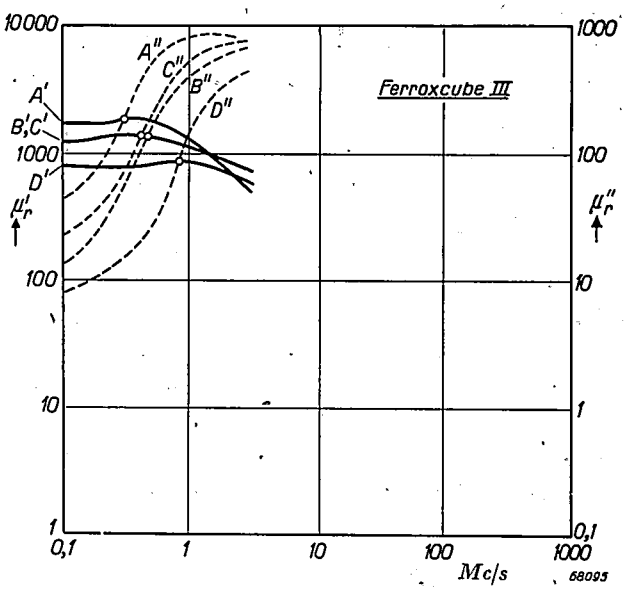


Fig. 5. The real and the imaginary parts of initial permeability, μ_r' and μ_r'', as functions of the frequency, for different kinds of Ferroxcube III.

A relationship has been found to exist between the critical frequency and the low-frequency value of μ': the smaller the value of μ', the higher is the critical frequency. This relationship, first found and interpreted by Snoek [18]), is graphically illustrated in figs 5 and 6.

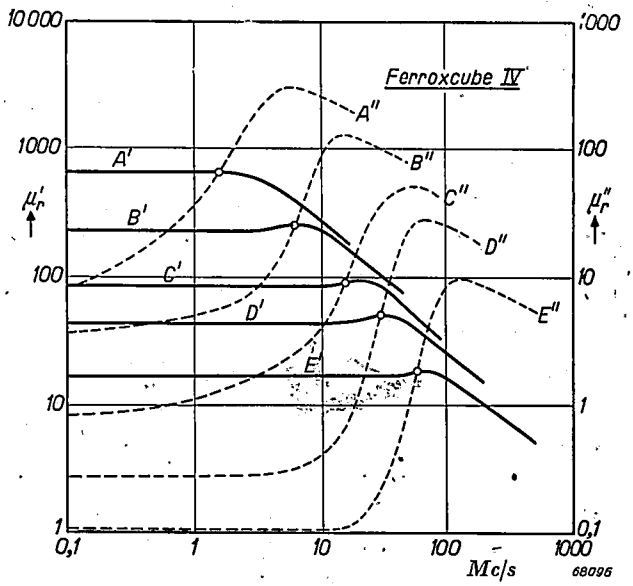


Fig. 6. The same as in fig. 5 but for different kinds of Ferroxcube IV.

[18]) J. L. Snoek, Nature **160**, 90, 1947; Physica **14**, 207, 1948.

This dependence of μ' and μ'' upon frequency in the range between 1 and 200 Mc/s is probably due to the ferromagnetic resonance absorption, a phenomenon directly related to the fact that the magnetic moments in a ferromagnetic material are, as it were, small spinning tops with a certain rotational impulse.

When such a top is subjected to a couple tending to direct it into a certain state of equilibrium it will describe a precessional movement around the equilibrium position; the corresponding frequency will be proportional to the magnitude of the acting couple. The phenomenon of ferromagnetic resonance then arises when a magnetic alternating field is applied with a component perpendicular to the equilibrium position, and when the frequency of that field equals the precession frequency of the magnetic spinning tops. When the acting couple arises from the presence of a constant external field then each top is subjected to the same couple and there is a rather sharply defined resonance frequency. The fact that the resonance frequency is not sharp is due to damping. This damping, which is caused by the mutual interaction of the magnetic moments or of their interaction with the crystal lattice, also gives rise to the losses occurring with the resonance phenomenon (the resonance absorption).

The occurrence of a directional couple in a ferrite in the demagnetized state is to be ascribed to the magnetic anisotropy, which may have widely differing values at different places in the material. As a result the spinning tops have different precession frequencies, and that is why with ferrites absorption is found to take place in a wide frequency range. Thus it is understood how a high initial permeability — which means a low magnetic anisotropy and therefore a weak directional couple — is accompanied by a low resonance frequency, as previously remarked.

Behaviour in strong fields

Permeability at high field strengths

It has been mentioned that for most applications at high frequencies only weak felds are of interest, and on that account the initial permeability was the quantity with which we were primarily concerned. However, not all applications are confined to weak fields, and moreover it is of interest also from the physical point of view to know something about the magnetization processes taking place in the presence of stronger fields, of different frequencies.

In *figs 7* and *8* the magnetization curves of Ferroxcube III^B and IV^C are represented for different frequencies. These curves reveal the remarkable fact that in the case of Ferroxcube III^B the permeability for high field strengths is frequency-dependent between 50 and 500 kc/s, whereas in

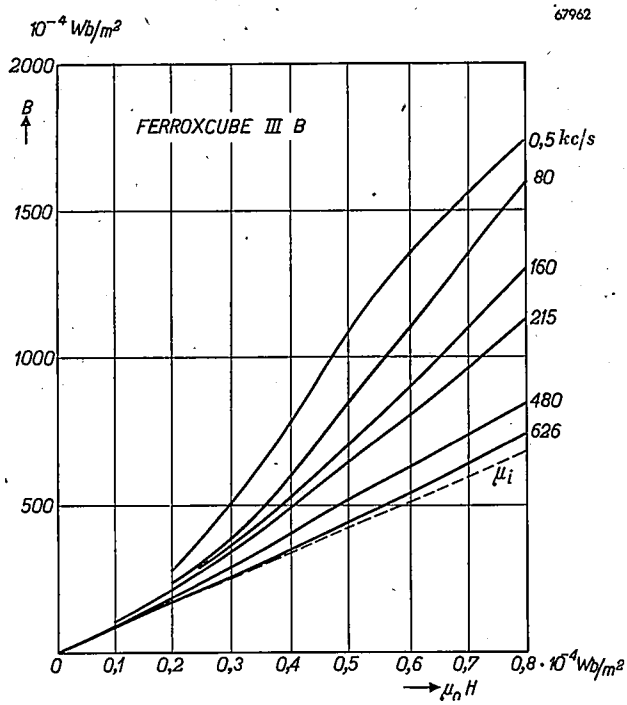


Fig. 7. Magnetization curves of Ferroxcube III^B for different frequencies.

the case of Ferroxcube IV^C it is independent of frequency up to at least 1 Mc/s. With increasing frequency the magnetization curves for Ferroxcube III^B approach more and more a straight

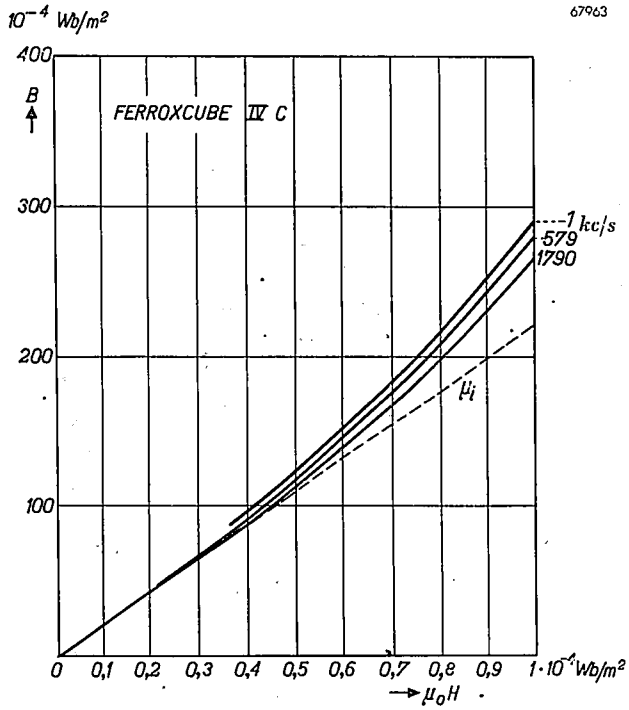


Fig. 8. Magnetization curves of Ferroxcube IV^C for different frequencies.

line, with a slope approximately equal to the initial permeability. The latter quantity, as already seen from fig. 6, is for both materials independent of frequency up to the ferromagnetic resonance frequency.

It would lead us too far afield to go into this behaviour in every detail [19]), but considering the great importance of a good understanding of the magnetization process in ferrites it is necessary to note the most important conclusions to be drawn from the curves given in figures 7 and 8.

It is known that the character of the magnetization curve (the J-$\mu_0 H$ curve) in ferromagnetic materials is directly dependent upon the fact that such a material in the demagnetized state is divided into Weiss domains. Within such a domain the magnetization has always the saturation value and it is orientated in a certain preferential direction (see note [5])). As a rule this preferential direction is not the same for different domains. The domains are mutually separated by a boundary layer (Bloch wall) in which the magnetization changes direction, gradually turning away from the preferential direction in one domain to that in another.

For the metallic ferromagnetic materials hither to commonly employed the transition from the demagnetized state to a state of total magnetization differing from zero (the magnetization process) is imagined to be as follows.

In the case of very weak fields magnetization is brought about by reversible displacements of the Bloch walls, the domains whose magnetization is orientated roughly in the direction of the field growing at the expense of the domains in which the direction of magnetization is against the field direction. In the presence of stronger fields these reversible wall displacements change into irreversible ones. Finally, in the case of even stronger fields, when the magnetization through wall displacements cannot increase any more, the magnetization in the domains themselves turns in the direction of the field.

The picture that we have of the magnetization process in ferrites differs considerably from that given above. The magnetization in very weak fields is believed to take place through rotation of the magnetization in the Weiss domains, and not through reversible wall displacements. In ferrites wall displacements occur only in the presence of stronger fields.

This picture of the magnetization process in ferrites is based upon the curves given in figures 7 and 8 together with the fact, now practically established, that the phenomenon of ferromagnetic resonance absorption originates in the rotation of the elementary magnets within a domain, and not in a wall displacement.

From the relationship found to exist between the value of the initial permeability and the value of the ferromagnetic resonance frequency it can be concluded, almost with certainty, that the initial permeability at that frequency is determined exclusively by rotations. Figures 7 and 8 show that up to the resonant frequency the initial permeability is independent of frequency. In other words, the initial permeability is exclusively governed by rotations not only at the ferromagnetic resonance frequency itself but also at the frequencies below it.

The reason why no wall displacements take place in the presence of weak fields may be understood when considering that, as already explained, in Ferroxcube materials small voids are always present, and it appears that such inhomogeneities always fix a Bloch wall to a certain extent. Apparently in Ferroxcube materials the walls are fixed by the voids at so many points that any (reversible) displacement or bulging of the walls under the influence of a weak field is so limited as to yield no perceptible contribution to the magnetization. It is only when the fields are of sufficient strength to release the walls from the voids that the magnetization process by irreversible wall displacements is superimposed on the magnetization process by reversible rotations. The introduction of a new process contributing towards the magnetization is evident in figures 7 and 8 from the fact that with stronger fields the magnetization curve shows a deviation from the straight line.

From the fact that in the case of some ferrites these deviations from a straight line disappear at frequencies above 200 to 500 kc/s — in this case the permeability above these frequencies is almost independent of the field strength — it may be concluded that obstructive mechanisms are operative which prevent any displacement of the walls above these frequencies.

Magnetic losses at high induction values

When the induction reaches high values the contribution of hysteresis losses towards the total losses is no longer negligible. As already explained, it is not possible, however, to separate the contributions of hysteresis losses and residual losses in a manner as expressed by eq. (1). One of the reasons for this lies in the frequency-dependence of the

[19]) H. P. J. Wijn and J. J. Went, Phys. Rev. **82**, 269, 1951; and in more detail in Physica **17**, 976, 1951.

magnetization curve just dealt with, as a consequence of which also the shape of the hysteresis loop and the magnitude of the hysteresis losses are dependent upon the frequency.

Fortunately, when measuring the total magnetic losses it appears that a relatively simple relationship exists between the magnitude of these losses and B_{max}. To find this relationship the total loss per cycle per m³, W_{tot}, has to be multiplied by μ_r' (taken at the frequency and induction in question) and plotted against B_{max}. When such curves are plotted for different frequencies (of course below the ferromagnetic resonance frequency) it is found that one set of curves suffices for all kinds of Ferroxcube (*fig. 9*).

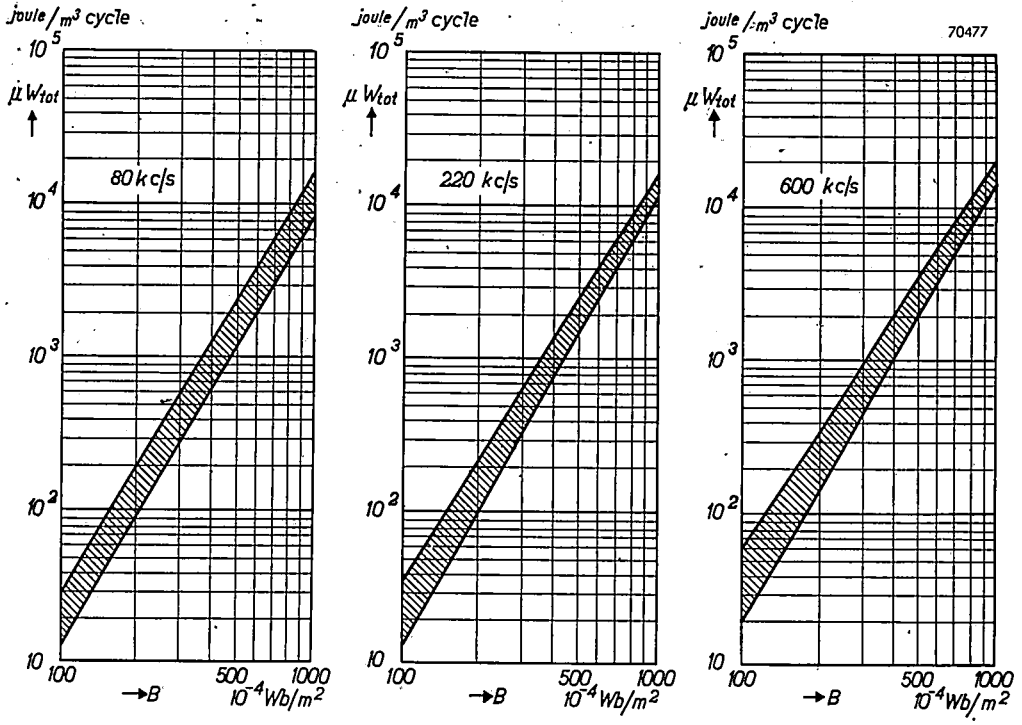


Fig. 9. Loss per cycle and per m³, W_{tot}, multiplied by μ_r', as a function of the induction B_{max}, for different frequencies. The relationships given here apply for all kinds of Ferroxcube.

Dielectric properties of the ferrites

In the introduction it was observed that the dielectric properties of the ferrites are rather exceptional. A glance at *fig. 10* will make this clear. For the expression of the dielectric behaviour one preferably uses a complex term for the dielectric constant: $\varepsilon = \varepsilon' - j\varepsilon''$, as was done for the permeability. From fig. 10, where the relative values ε_r' and ε_r'' have been plotted as functions of the frequency ($\varepsilon = \varepsilon_r\varepsilon_0$), it appears that at low frequencies the real part ε' of ε for Ferroxcube III is exceptionally large (about 10^5) and decreases appreciably only at high frequencies. The value of ε' is also large for Ferroxcube IV, but it begins to decrease at comparatively low frequencies and ultimately reaches quite normal values (10 to 20).

Further it is seen that in the same frequency range where ε' begins to decrease the imaginary part ε'', which determines the dielectric losses and at low frequencies is inversely proportional to ω, decreases at a slower rate (the product $\varepsilon''\omega$ begins to increase).

The physical causes of these interesting properties, which as a matter of fact can be observed also in other semi-conductors, will not be dealt with in great detail. It may suffice to say here that they can be understood qualitatively by assuming the materials not to be completely homogeneous electrically [20]).

However, attention is to be drawn, briefly, to one practical consequence of these properties.

When a material is placed in an electric alternating field no difference can in fact be made between ε'' and the conductivity σ. From Maxwell's equations it follows that

$$\sigma = \varepsilon_r'' \varepsilon_0 \omega.$$

At high values of the product $\varepsilon''\omega$ the resistivity ϱ is low and one would therefore expect eddy-current losses to appear.

Indeed, the high value of ε'' for Ferroxcube III does result in extra losses, but these are not to be regarded simply as eddy-current losses.

Eddy-current losses are known to become perceptible as soon as the depth of penetration of the magnetic alternating field in the ferromagnetic material is of the same order as or less than the thickness of the object (skin effect). Assuming that $\mu'' \ll \mu'$ and $\varepsilon' \ll \varepsilon''$ the following formula may be derived for the penetration depth d:

$$d = \sqrt{\frac{2}{\omega \mu_r' \mu_0 \sigma}} = \frac{\lambda}{2\pi} \sqrt{\frac{2}{\mu_r' \varepsilon_r''}}, \quad . \quad . \quad (3)$$

where λ is the wavelength of the electromagnetic

[20]) C. G. Koops, Phys. Rev. **83**, 121, 1951.

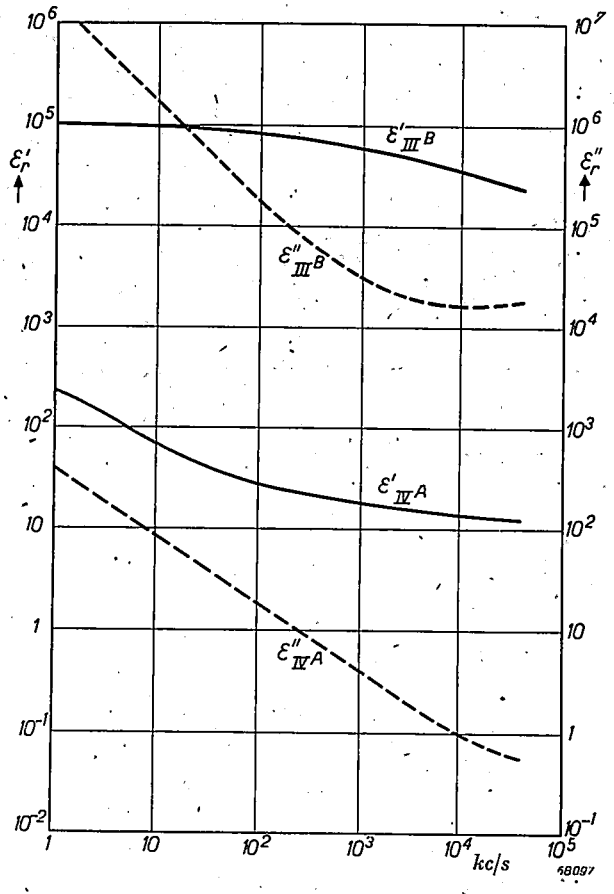


Fig. 10. The real and the imaginary parts of the dielectric constant, ε_r' and ε_r'', as functions of the frequency, for Ferroxcube III^B and IV^A.

wave in vacuo. Since it therefore depends upon the dimensions of the object at what frequency the losses begin to become noticeable, one speaks of dimensional absorption.

Taking a thickness of 1 cm, corresponding roughly to what is often used in practice, it can be calculated for Ferroxcube III, with the aid of the values of μ' and of ϱ (and thus of ε'') given in table III, that the losses in question should become perceptible at a frequency of 2.5 Mc/s. Actually, Brockman, Dowling and Steneck [21]) found that with a thickness of 1 cm Ferroxcube III shows additional losses even at much lower frequencies.

The explanation for this behaviour given by these investigators is that formula (3) does not apply here, because the condition $\varepsilon' \ll \varepsilon''$ is not satisfied for Ferroxcube III in this frequency range. Skin effect is entirely absent. It is, however, quite possible that in this case the induction lines in the ferromagnetic material deviate from their regular course, this being accompanied by losses. The criterion for this is that a length d', satisfying a formula similar to that given in (3), viz:

$$d' = \frac{\lambda}{2\pi}\sqrt{\frac{2}{|\mu_r|\,|\varepsilon_r|}},$$

is of the same order as or less than the smallest dimension of the object. Since d' is equal, except for a numerical factor, to the wavelength of the electromagnetic wave in the material, and therefore, for the occurrence of these losses, it must bear a certain relation to the dimension of the object, Brockman, Dowling and Steneck speak of dimensional resonance. For Ferroxcube III the value of d' is found to be approximately equal to 1 cm already at a frequency of 1 Mc/s.

If in a given practical case difficulties are to be feared as a consequence of this dimensional resonance, the simplest solution is to avoid the effect entirely by using Ferroxcube IV instead of Ferroxcube III. Not only is the value of ϱ much larger (ε'' much smaller) than that for Ferroxcube III, so that eddy-current losses become perceptible only at much higher frequencies, but moreover the dielectric constant ε' is already fairly small below 100 kc/s, so that within reasonable frequency limits no trouble need be feared from dimensional resonance effects.

Summary. Ferroxcube materials are ferromagnetic oxidic materials specially developed for use at high frequencies. Their suitability for such applications lies in the fact that owing to their high resistivity eddy-current losses do not as a rule occur.

As regards the chemical composition these materials belong to the group of ferrites crystallizing with the spinel structure. The saturation magnetization of ferrites is closely related to their crystallographic structure. According to Néel the values of the saturation magnetization found experimentally can be explained by assuming that an interaction exists between magnetic moments on crystallographically non-equivalent lattice sites, which tends to give those moments an antiparallel orientation.

Although eddy-current losses do not usually occur in Ferroxcube, residual losses, given by the expression $\tan\delta/\mu'$, may under certain conditions be considerable. It is therefore highly important to know the behaviour of μ' and $\tan\delta$, or of the real and the imaginary parts of the complex permeability $\mu = \mu' - j\mu''$, for all possible values of induction and of frequency. In this paper the behaviour of these quantities as a function of frequency is discussed, both for very small and for somewhat larger values of the induction. Further, the temperature dependence of the initial permeability is discussed. The behaviour of the imaginary part of the initial permeabilily at high frequencies is shown to bear a relationship to the ferromagnetic resonance absorption; this relationship and the behaviour of the permeability in the presence of stronger fields lead to the conclusion that the magnetization process in ferrites differs from that taking place in metallic ferromagnetic materials.

Finally attention is drawn to the remarkable dielectric properties of ferrites and the extra losses apt to arise therefrom (dimensional resonance).

[21]) F. G. Brockman, P. H. Dowling and W. G. Steneck, Phys. Rev. 77, 85, 1950.

FERROXDURE, A CLASS OF NEW PERMANENT MAGNET MATERIALS

by J. J. WENT*), G. W. RATHENAU, E. W. GORTER and G. W. van OOSTERHOUT.

621.318.12:538.246.2

Present permanent magnet materials are in general metallic alloys mostly containing nickel or cobalt or both. Ferroxdure, the class of new permanent magnet materials developed in the Philips Research Laboratories at Eindhoven consists of oxidic ceramics which do not contain any nickel or cobalt. The interesting and typical properties of these materials and some applications of Ferroxdure are treated in the three parts of this article.

I. MOST IMPORTANT PROPERTIES OF FERROXDURE

Introduction

Ferromagnetic oxides have been the object of extensive research in the Philips Laboratories for many years. The reader of this journal has had ample opportunity to get informed about one class of those materials known by the name of Ferroxcube. As was already often stated in former publications [1, 2]) Ferroxcube materials are characterized by possessing a high electric resistivity together with a high magnetic permeability. Because of the latter property Ferroxcube materials belong to the class of the magnetically soft materials. Ferroxcube is extremely well suited for use in high frequency inductances, high frequency transformers etc. [1]) [2]).

During the last two years another class of ferromagnetic oxides has been studied extensively at the Philips Research Laboratories at Eindhoven. These oxides are magnetically hard, i.e. they can be used as materials for permanent magnets. The remarkable fact about these oxides is that they do not contain any nickel or cobalt, in contradistinction to most other permanent magnet materials. This fact is of great economic importance as these two metals are expensive and at present difficult to obtain. The representatives of this class of new materials have been given the name of Ferroxdure.

Characteristics of magnetically soft and hard materials

While magnetically soft materials are mainly characterized by their permeability and by the losses they introduce when they are placed in an alternating magnetic field, quite different properties are characteristic for magnetically hard materials. Therefore, before entering into details we wish to give a very short survey of the quantities which are frequently used in dealing with magnetically hard materials. This is most conveniently done with the aid of *figure 1a* [3]). In this figure the magnetization J [$4\pi J$] is given as a function of the effective field H_i [4]). As is well known, the magnetization of a ferromagnetic material is in general not uniquely determined by the effective field. This means e.g. that if a ferromagnetic is magnetized to its saturation magnetization, and subsequently the effective field is brought to zero in a continuous way, a certain magnetization remains, the remanent magnetization J_r. The effective field in the direction opposite to the remanent magnetization that must be applied in order to reduce the magnetization to zero is called the J-coercive field (often also the coercive force); it is indicated by the symbol ${}_JH_c$. In the particular case shown in fig. 1*a* ${}_JH_c$ is of the same order of magnitude as J_r [$4\pi J_r$]. A similar coercive field, the B-coercive field denoted by the symbol ${}_BH_c$, can be defined, which according to a completely analogous definition is necessary to reduce the

*) Now with N.V. K.E.M.A., Arnhem (Netherlands).

1) J. L. Snoek, Philips techn. Rev. **8**, 353, 1946; J. L. Snoek, New developments in ferromagnetic materials, Elsevier's Publishing Comp. Amsterdam, New-York 1947; J. J. Went and E. W. Gorter, Philips techn. Rev. **13**, 1951 (No. 7). The latter article is in the following quoted as A.

2) W. Six, shortly to appear in Philips techn. Rev.

3) For a more detailed discussion of most quantities see e.g. J. J. Went, Philips techn. Rev. **10**, 246, 1948.

4) The effective field is the difference between the external field H and the demagnetizing field NJ/μ_0 [$N'J$] where N [N'] is the factor of demagnetization in the direction of the field. Unless explicitly stated differently, in the following H will be used to denote the effective field.

Throughout this article formulae are given using the rationalized Giorgi system of electromagnetic units (see also Philips techn. Rev. **10**, 79, 1948). For the convenience of those readers who are not (yet) familiar with this system, formulae in unrationalized cgs units (Gauss system) are given between square brackets. In the figures scales for both systems of units are given.

It is to be remembered that in the Giorgi system B and J are measured in volt sec/m^2 or weber/m^2 (1 Wb/m^2 = = 10 000 gauss) and H in ampère/m (1000/4π A/m = = 79.6 A/m corresponding to 1 oersted).

In vacuum $B = \mu_0 H$, in which $\mu_0 = 4\pi/10^7$ henry/m. Instead of H (in A/m) often the value of $\mu_0 H$ (in Wb/m^2) is given ($\mu_0 H$ = 1 Wb/m^2 corresponding to 10 000 oersted in the cgs system). In a ferromagnetic material the relation $B = \mu H$ holds, in which $\mu = \mu_r \mu_0$; μ_r, the relative permeability is a dimensionless number (equal to the number indicated by μ in the cgs system).

The value of BH, which has the dimension of an energy per unit volume, is measured in (Vsec/m^2) × (A/m) = VAsec/m^3 = joule/m^3 (1 J/m^3 corresponding to 126 gauss × oersted).

magnetic induction $B = \mu_0 H + J$ [$B = H + 4\pi J$] to zero. In fig. 1*b* a demagnetization curve [5]) is given for a ferromagnetic for which the coercive field is such that $\mu_0 {}_J H_c$ [${}_J H_c$] is much smaller than the remanent magnetization J_r [$4\pi J_r$]. It is seen that in contradistinction to fig. 1*a*, the difference between ${}_J H_c$ and ${}_B H_c$ is very small. As will be pointed out in greater detail below, the figures 1*a* and 1*b* are valid for the cases of Ferroxdure and "Ticonal" G (a well-known permanent magnet material of very high quality [6])) respectively.

Neither the remanent magnetization nor the coercive field are the most adequate quantities to describe the quality of a permanent magnet material. It has been customary to characterize the usefulness of such a material by the maximum value $(BH)_{max}$ which the product of the magnetic induction B and the effective field H can attain on the demagnetization curve. The material with the greatest value of $(BH)_{max}$ will be able to create a given field in an air gap of a given magnetic circuit at the cost of the smallest volume of magnetic material [7]). It will be clear that both a large remanent magnetization and a large coercive field are favourable for obtaining a large value of $(BH)_{max}$.

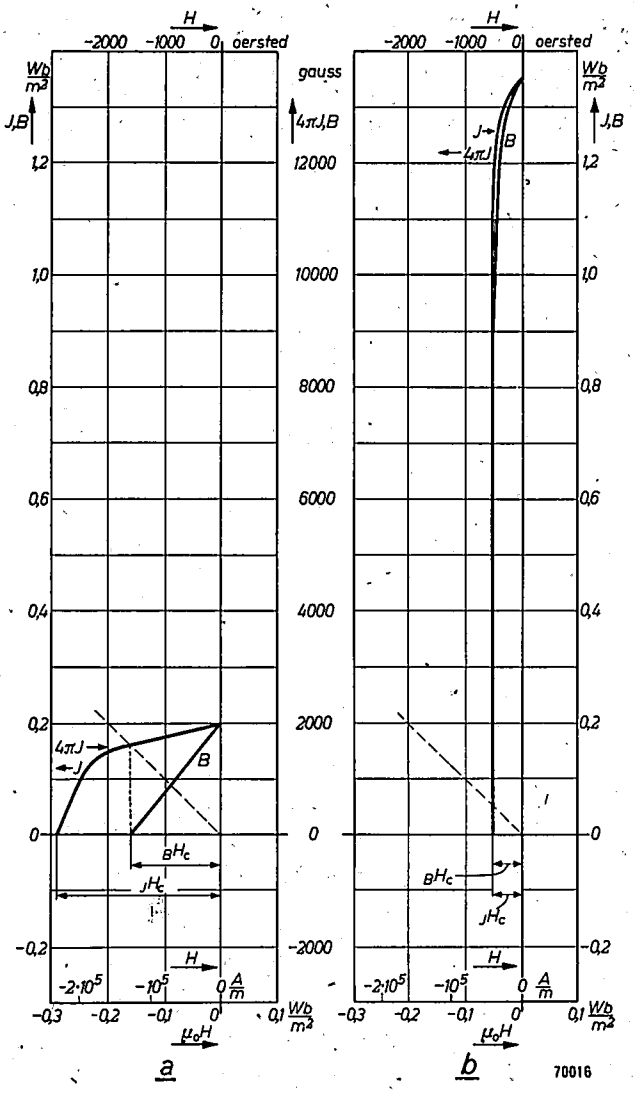


Fig. 1. Demagnetization curves for Ferroxdure (*a*) and "Ticonal" G (*b*). In *a*, J_r [$4\pi J_r$] and ${}_J H_c$ are of the same order of magnitude and thus a large difference between ${}_J H_c$ and ${}_B H_c$ exists. This is not the case in *b*.

Survey of the most important properties of Ferroxdure

The name Ferroxdure denotes a class of sintered, oxidic, magnetically hard materials. In this article the discussion will be limited to three of these materials. Two of these are oxides with compositions $BaFe_{12}O_{19}$ and $BaFe_{18}O_{27}$ respectively. The third material is that which at the moment is available on the market under the name of Ferroxdure, and from now on when speaking simply of Ferroxdure we mean this material. It consists mainly of the oxide $BaFe^{III}_{12}O_{19}$. This oxide has a non-cubic, viz. hexagonal, crystal structure and has *one* axis of easy magnetization parallel to the hexagonal axis (i.e. the crystal anisotropy is uniaxial).

The uniaxial character is essential for a large coercive force in Ferroxdure, just as it was essential for the Ferroxcube materials to be composed of cubic ferrites for obtaining a low coercive force (high permeability).

The saturation magnetization of Ferroxdure is rather low compared e.g. with that of "Ticonal" G: the values at room temperature are about 0.42 Wb/m² [4200 gauss] and 1.41 Wb/m² [14100 gauss] respectively.

In a magnetically hard material consisting of crystals with one axis of preferred magnetization oriented at random like Ferroxdure, the remanent magnetization can in general be shown to be half of the saturation magnetization (cf. e.g. the article quoted in footnote [3])). Thus the value of the remanent induction of Ferroxdure at room temperature is rather low compared with that of "Ticonal" G, namely 0.20-0.21 Wb/m² [2000-2100 gauss] and 1.34 Wb/m² [13400 gauss], respectively. The coercive force ${}_J H_c$, however, is extremely high, namely $\mu_0 {}_J H_c$ = 0.2-0.32 Wb/m² [${}_J H_c$ = 2000-3200 Oe], as compared with $\mu_0 {}_J H_c$ = 0.06 Wb/m² [${}_J H_c$ = 600 Oe] for "Ticonal" G.

In fig. 1*a* and 1*b* the demagnetization curves for these two materials are given. The high coercive force of Ferroxdure combined with the rather low remanent induction causes the large difference which is found between ${}_J H_c$ and ${}_B H_c$. The value

[5]) By the demagnetization curve we mean that part of the complete magnetization curve which is lying in the upper quadrant on the left.

[6]) See e.g. B. Jonas and H. J. Meerkamp van Embden, Philips techn. Rev. **6**, 8, 1941.

[7]) See e.g. A. Th. van Urk, Philips techn. Rev. **5**, 29, 1940. As will be pointed out later, this statement is not so strictly true as is suggested here.

of $(BH)_{max}$ for Ferroxdure, of about 6800 J/m^3 [0.85×10^6 gauss oersted] is about 6 times lower than the $(BH)_{max}$ value for Ticonal G, which is up to 45 800 J/m^3 [5.7×10^6 gauss Oe]. This does not mean, however, as will be pointed out below, that the volume of magnetic material needed to create a certain field strength in an air gap of given volume is 6 times as large as that needed when using "Ticonal" G. Notwithstanding the larger volume of Ferroxdure needed, this material often will offer an economic advantage because it does not contain critical raw materials. Besides the economic advantage of Ferroxdure this material offers new technical possibilities. Its very high electric resistivity allows it to be used in high frequency applications as e.g. in Ferroxcube core transformers that need premagnetization, as deflection magnets for radar tubes or for the adjustment of the temperature coefficient of the permeability of a Ferroxcube core (cf. [2]). Its extremely high coercive force and the resultant strong resistance to demagnetization allow a novel approach to the design of magnetic circuits, which will be discussed in a subsequent paper to be published in this journal. Besides the above applications the following ones should be mentioned specifically:
Loudspeaker circuits (see *fig. 2*),
Magnets for fixing one object to another,
Magnets for oil filters,

Fig. 2. Two Ferroxdure magnets as used in loudspeaker circuits magnetized in the direction of the axis. Note the disc-shaped form of the magnets which is most suitable for a material with high coercive force and relatively low remanence. The upper magnet is lifted by the field of the lower one. The external diameter of the rings is 8 cm.

Magnets for focussing electron beams,
Electric generators and motors (see also *fig. 3*),
Magneto-mechanic couplings.

The physical background of the various properties of Ferroxdure will be discussed in greater

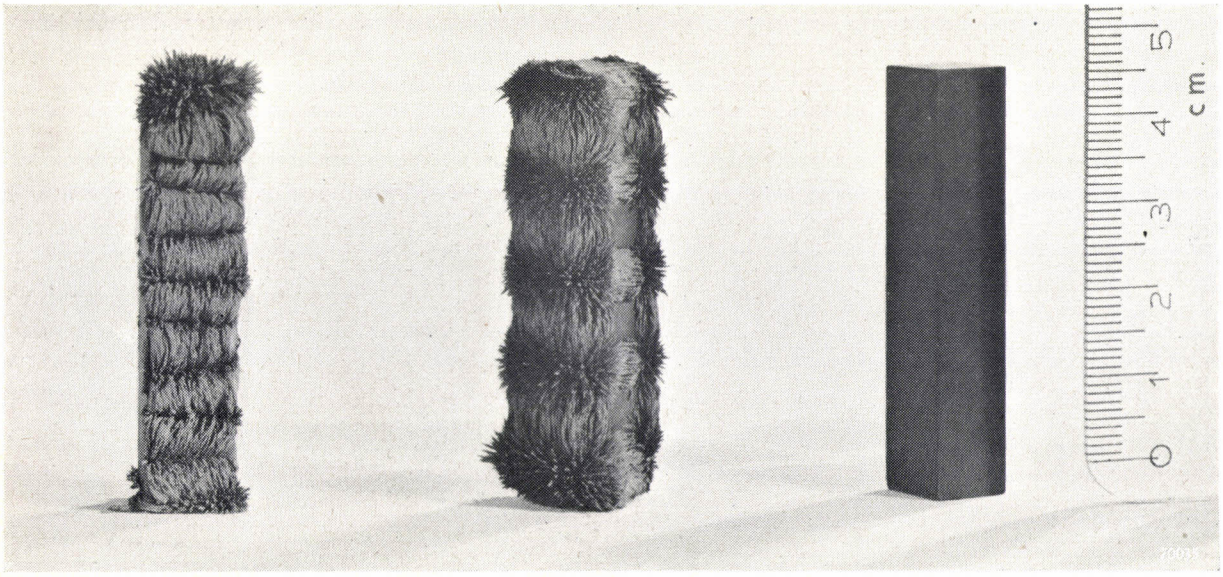


Fig. 3. Rod-shaped Ferroxdure magnets which have been magnetized normal to a vertical face in such a way that on this face alternately north and south poles are formed. The ten poles on the specimen at the left and the five poles on the specimen in the centre are made visible by iron powder. Due to the low value of the permeability and the high value of the coercive force Ferroxdure is specially suited for producing strong alternating poles at a small distance from each other without the necessity of using a large extension of the specimen parallel to the direction of the magnetization.

detail in part II of this article. In order not to complicate the treatment more than necessary we shall discuss the physical properties of the main component of Ferroxdure only and of the other oxide already mentioned above, with composition $BaFe_2^{II}Fe_{16}^{III}O_{27}$, the crystal structure and physical properties of which are very similar to those of $BaFe_{12}O_{19}$. By considering also this second oxide various measurements and theoretical conclusions obtain additional confirmation.

In the first section of part II the magnetic anisotropy will be treated and it will be shown that the most important part is the (magnetic) crystal anisotropy, which is extremely high. This large anisotropy in combination with the relatively small remanent magnetization accounts for the very high coercive force of Ferroxdure which will be discussed next. At the end of part II the following points will be discussed consecutively: the remanent induction, the demagnetization curve and the value of $(BH)_{max}$, and finally the electric resistivity.

In part III of this article the crystal structure of the two oxides $BaFe_{12}O_{19}$ and $BaFe_{18}O_{27}$ will be discussed in greater detail. It will be seen that a close relation exists between the crystal structure and the value of the saturation magnetization of these two oxides.

II. PHYSICAL BACKGROUND OF SOME PROPERTIES OF FERROXDURE

Magnetic anisotropy

It was already mentioned in the introduction that the magnetic anisotropy, i.e. the existence of directions of easy or preferred magnetization, is of the greatest importance for the value of the coercive field. Three different factors contribute towards facilitating magnetization in certain preferential directions: the (magnetic) crystal anisotropy, the stress anisotropy and the shape anisotropy. As will be shown in greater detail below, in the oxides treated in this paper the crystal anisotropy is large and can easily be measured. As regards the stress anisotropy: in A (cf. I, footnote [1])) it was shown that within a small region of the material it is proportional to the product of the linear magnetostriction constant in the direction of stress and the internal or external stress. In a sintered aggregate of non-cubic crystals large internal stresses may be expected due to an anisotropic coefficient of thermal expansion. The magnetostriction constant of Ferroxdure is not zero. Since, however, we have reason to believe that it will not greatly exceed 10^{-5}, and since it is very difficult to make a quantitative estimate of the contribution of the stress anisotropy (which in fact might be appreciable) to the total anisotropy, we will tentatively leave out the stress anisotropy from our further discussion: it will be seen below that, indeed, all relevant phenomena can be explained satisfactorily by taking into account the magnetic crystal anisotropy only.

The shape anisotropy can easily be shown to be small compared with the crystal anisotropy because of the low value of the saturation magnetization.

The crystal anisotropy

For both hexagonal oxides mentioned above, the crystal anisotropy was measured on single crystals. It was found in both cases that the hexagonal axis is the direction of easy magnetization. The so-called magnetocrystalline energy density, which describes the tendency for a preferential orientation, can be written as

$$E_c = K_1 \sin^2 \Theta + K_2 \sin^4 \Theta + \ldots \quad (1)$$

E_c is the energy (per m^3 [cm^3]) necessary to turn the magnetization from the easy direction ($\Theta = 0$) over an angle Θ. The quantities K_1, K_2 are constants, the so-called anisotropy constants. For our purposes it is sufficient to take into account the first term of the series only, so that we have one anisotropy constant K_1, for which we shall write K. Thus formula (1) reads:

$$E_c = K \sin^2 \Theta . \quad (2)$$

The measurements of the anisotropy constants for the two oxides were performed as follows.

In the case of $BaFe_{18}O_{27}$ single crystals were obtained from the melt, by H. P. J. Wijn and Y. Haven of this laboratory, that were large enough to allow the measurement of the magnetization curves in the direction of the hexagonal axis and along different directions perpendicular to it [1]). The hexagonal axis appeared to be the only direction of preferred magnetization, as no anisotropy within the basal plane could be detected at room temperature.

[1]) G. W. Rathenau and J. L. Snoek, Philips Res. Rep. 1, 239, 1945-'46.

The field strengths used were not large enough to saturate the crystal along the difficult direction. The crystal anisotropy was calculated by extrapolating the magnetization in the difficult direction to higher field strengths and measuring the area between the magnetization curves in the easy and difficult directions, this area being the work necessary to turn the saturation magnetization from the easy to the difficult direction of magnetization.

For $BaFe_{12}O_{19}$ no single crystals were available with dimensions large enough to allow the measurement of the magnetization curves to be performed in exactly the same way. The following procedure was employed. As the basal plane is well developed for these oxides, small crystals obtained from polycrystalline material could easily be fixed on a glass plate in such a way that the hexagonal axes were perpendicular to the plate, as was confirmed by X-ray diffraction. From the results of the measurements performed on these specimens (*fig. 1*) the anisotropy constant could be calculated [2]). In *fig. 2* the values of K for the two oxides are plotted against temperature. In *table I* the values of K at room temperature for $BaFe_{12}O_{19}$ are compared with those for some well-known ferromagnetics.

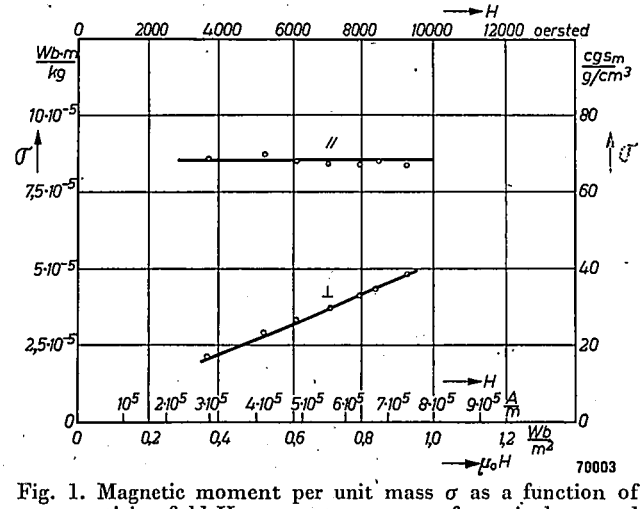


Fig. 1. Magnetic moment per unit mass σ as a function of a magnetizing field H at room temperature for a single crystal of $BaFe_{12}O_{19}$. Curve indicated by $\parallel$: H is parallel to the easy axis. Curve indicated by $\perp$: H is perpendicular to the easy axis.

Table I. The value of K at room temperature for different materials [3]).

Material	K J/m³	[erg/cm³]
$BaFe_{12}O_{19}$	3×10^5	$[3 \times 10^6]$
Fe	0.15×10^5	$[0.15 \times 10^6]$
Co	4×10^5	$[4 \times 10^6]$

[2]) It was assumed that the anisotropy within the basal plane may be neglected as in $BaFe_{18}O_{27}$, which seems a reasonable assumption, considering the very close resemblance of the crystal structures and the experience with other hexagonal crystals, such as cobalt.

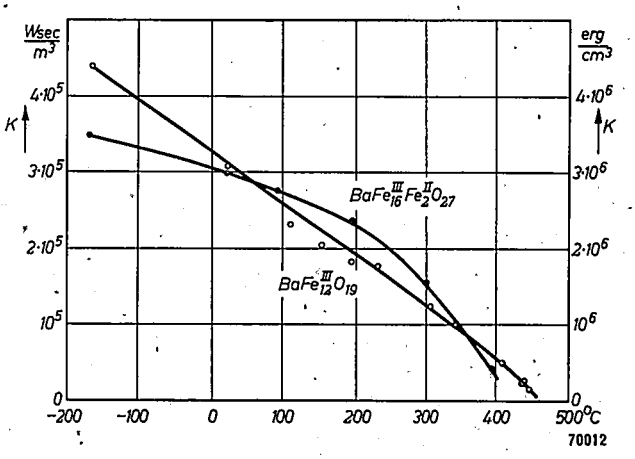


Fig. 2. Anisotropy constant K as a function of temperature.

Another method of measurement is to allow a crystal to perform torsional oscillations in a strong magnetic field in such a way that the axis of preferred orientation oscillates in a plane containing the direction of the magnetic field. By varying the field and extrapolating to fields of infinite strength the anisotropy constant K as well as the saturation magnetization can be calculated from the period of the oscillations. The results are in approximate agreement with those given in fig. 2 [4]).

Coercive force

There are many factors which determine in a complicated way the coercive force in the general case of a massive ferromagnetic. As Ferroxdure is a sintered material consisting of more or less small crystals with an extremely high crystal anisotropy, which have random orientation, a somewhat simpler situation may be expected [5]). In order to simplify the argument we shall start with rather crude assumptions about the demagnetization process and estimate on this basis the coercive force. After comparison with experiment we shall correct our model so as to obtain better agreement with reality. It will be seen that the value of the coercive force obtained with the aid of the first assumptions gives a limit above which the coercive force can in no way be raised.

Consider a ferromagnetic single crystal saturated in its easy direction. We now assume that changes in magnetization occur only by rotation of the magnetization from the easy direction, and ask how strong a field in the direction opposite to the

[3]) For iron, which has cubic magnetic anisotropy, equations (1) and (2) cannot be applied. The quantity K given for iron in table I refers to the difference in magnetic energy in the [100] and [111] directions respectively.

[4]) These measurements were carried out by J. Smit of this laboratory.

[5]) It should be noted that a sample of the material as a whole is isotropic and that only the small crystals of which the sample is composed are anisotropic.

magnetization must be to turn the magnetization over an angle of 180°. If the magnetization is turned over an angle Θ from the easy direction, the total energy density is:

$$E_t = K \sin^2\Theta - HJ_s \cos\Theta .$$

The equilibrium angle Θ_e for a given field H is found by making E_t minimum with respect to Θ:

$$\partial E_t/\partial\Theta = 0;$$
$$2K \sin\Theta \cos\Theta + HJ_s \sin\Theta = 0,$$

This equation has two solutions. Only the first one, being $\sin\Theta_e = 0$, corresponds to a stable equilibrium. The other one,

$$\cos\Theta_e = -HJ_s/2K,$$

describes an unstable equilibrium. From this equation it can be seen that the absolute value of the field strength necessary to turn the magnetization from the first stable position $\Theta_e = 0$ over a very small angle Θ (so that $\cos\Theta \approx 1$) is $2K/J_s$. The field strength decreases with increasing angle of deviation. Thus on applying a field $2K/J_s$ the magnetization jumps discontiuously in the direction of the field. The field strength $H = 2K/J_s$ is called the "turn-over" field strength.

In order to find the relation between this turn-over field strength and the coercive force of an actual sample of $BaFe_{12}O_{19}$, to which oxide we shall limit our discussion, the fact should be taken into account that such a sample is composed of many particles, the easy axes of which are oriented at random.

It can be shown that in this case the coercive force should lie at about K/J_s. In *fig. 3* the

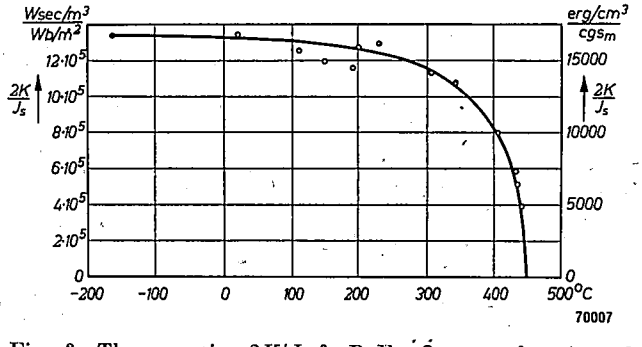


Fig. 3. The quantity $2K/J_s$ for $BaFe_{12}O_{19}$ as a function of temperature.

quantity $2K/J_s$ is plotted against temperature for $BaFe_{12}O_{19}$. It is seen that up to a temperature of 250 °C the coercive force estimated on the assumption that only rotational processes occur lies near $\mu_0 H = 0.75$ Wb/m² [$H = 7500$ Oe]. The experimental value of the coercive force, though very high, is appreciably below this value, so that our model must be modified as being too simple.

The reason for the failure of our calculation is that we did not take into account the formation and the movement of Bloch walls. The stability of a Bloch wall within a ferromagnetic particle with one direction of preferential magnetization depends on several factors [6]). Let us assume that the particle has been magnetized to saturation. Its poles will create a demagnetizing field. By the formation of a wall the demagnetization energy can be diminished by an amount that is the product of a demagnetization factor, which depends on the shape of the particle, the square of the intensity of magnetization and the volume of the particle. If also an external field is applied in the direction opposite to the direction of previous magnetization the formation of the wall will reduce the energy further by an amount which is proportional to the intensity of magnetization, to the field strength and to the volume of the particle. The energy which has to be furnished in order to create the wall is determined by the "effective exchange" energy $(J_s^2/J_0^2)kT_c$, (where T_c is the Curie temperature) k is Boltzmann's constant and J_0 the saturation magnetization at zero absolute temperature), which tends to keep the magnetic moments aligned parallel within the wall [7]) and by the energy of crystalline anisotropy which tends to keep the magnetic moments within the wall away from the directions of difficult magnetization. Furthermore it is proportional to the area of the wall.

In order to show that in isolated particles of $BaFe_{12}O_{19}$ in the absence of an external field wall formation will occur only at relatively large crystal sizes, the critical diameter for wall formation at room temperature is given in *table II*, together with the maximum turn-over field strength $2K/J_s$, for the same materials which have been compared

Table II. The critical diameter d_0 for wall formation in isolated spheres and the turn-over field strength $2K/J_s$.

Material	d_0 microns	$2\mu_0 K/J_s$ in Wb/m²	$[2K/J_s]$ [oersted]
$BaFe_{12}O_{19}$	1.3	1.66	[16600]
Fe	0.028	0.018	[180]
Co	0.24	0.56	[5600]

[6]) C. Kittel, Rev. Mod. Phys. **21**, 541, 1949.

[7]) The "effective exchange" energy describes formally the tendency for parallel alignment of magnetic moments on the same sublattice. This tendency originates from the actually existing negative superexchange interaction between magnetic moments on different sublattices (cf. part III).

in table I [8]). The equation for spheres of diameter d_0 reads roughly [6]):

$$\frac{1}{4}\cdot\frac{1}{3}\cdot\frac{J_s^2}{\mu_0}\cdot\frac{4\pi}{3}\cdot\frac{d_0^3}{8}=\frac{\pi d_0^2}{4}\sigma_w$$

$$\left[\frac{1}{4}\cdot\left(\frac{4\pi}{3}\right)^2\cdot\frac{d_0^3}{8}J_s^2=\frac{\pi d_0^2}{4}\sigma_w\right],$$

from which follows:

$$d_0=\frac{18\sigma_w\mu_0}{J_s^2}\left[d_0=\frac{9\sigma_w}{2\pi J_s^2}\right].$$

If the expression for the wall energy σ_w is substituted [9]) one arrives at:

$$d_0=36\pi\mu_0\frac{\sqrt{K}}{J_s}\sqrt{\frac{kT_c}{aJ_0^2}}\left[d_0=\frac{9\sqrt{K}}{J_s}\cdot\sqrt{\frac{kT_c}{aJ_0^2}}\right]. \quad (3)$$

Here K is the anisotropy constant, and a the mean distance between neighbouring ions belonging to equivalent sublattices (cf. part III).

In the expression (3) for d_0 only the factor $\sqrt{K}/J_s$ is temperature-dependent. It follows from *fig. 4*, which gives this quantity for $BaFe_{12}O_{19}$ as a function of temperature, that the critical diameter for $BaFe_{12}O_{19}$ increases with temperature up to a region near the Curie point. This behaviour is different from that of most traditional materials, for which $\sqrt{K}/J_s$ decreases with temperature. The increase with temperature of d_0 means that at a given field strength in an aggregate of $BaFe_{12}O_{19}$ crystals a smaller number of Bloch walls will be present at higher temperatures and that especially those Bloch walls disappear which have the highest energy and can therefore be most easily moved by an external field. Considering the mobility of one particular Bloch wall in a crystal that is imperfect due to inclusions and lattice defects as well as due to internal tensions the following must be borne in mind. The force driving the wall is proportional to the magnetization J_s on which the external field acts, and the resistance against the movement of a wall is due to local variations of the wall energy and increases with the energy of anisotropy and the intensity of magnetization. It thus follows that the mobility of one particular wall in an external field depends on temperature as $K^{-1/2}$ (fig. 2). Obviously the model of isolated spheres is not strictly applicable to Ferroxdure. However, Ferroxdure is a sintered material, the density of which is about 10% below that calculated for the case of the absence of pores (X-ray-value) [10]). We believe that the interplay of the decrease of the number of walls and the increase of the mobility of a particular wall with temperature is mainly responsible for the maximum in the curve of the coercive force $_JH_c$ vs temperature as shown in *fig. 5*.

As to the actual crystal size for which wall formation can be avoided and accordingly a high coercive force can be obtained, it is seen from table II, that the rough calculation arrives at dimensions of about one micron. It has been verified experimentally by grinding a coarse-grained polycrystal-

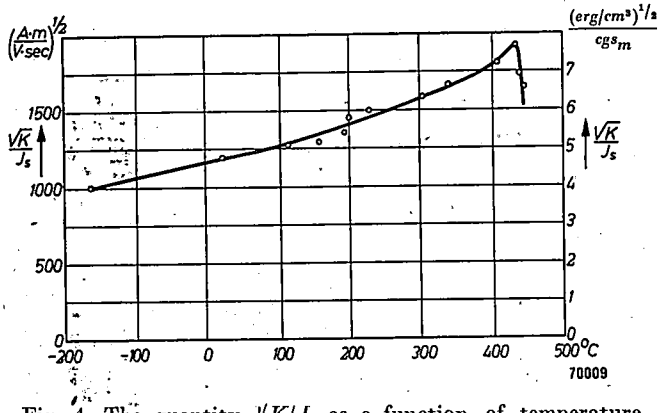


Fig. 4. The quantity $\sqrt{K}/J_s$ as a function of temperature for $BaFe_{12}O_{19}$.

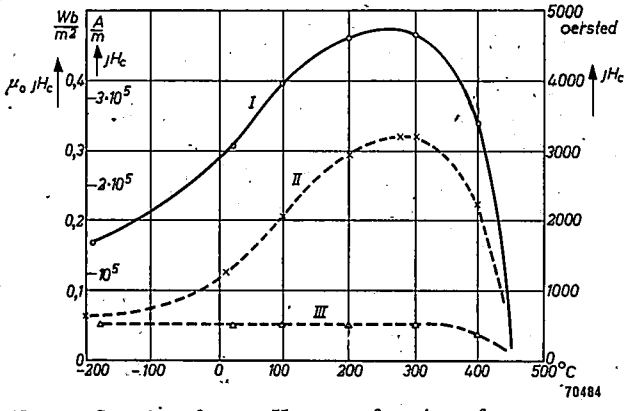


Fig. 5. Coercive force $_JH_c$ as a function of temperature. The drawn line applies to a very fine-grained sintered specimen of $BaFe_{12}O_{19}$, the broken lines to samples which contain larger crystals. In the latter cases the actual values of the coercive force are 10 times smaller than is read on the scale. The lowest curve was found for a very coarse grained sample of Ferroxdure (largest extension of crystals several mm's).

[8]) The comparison with iron can be of qualitative value only as equations (1) and (2) are strictly speaking not applicable to the case of iron, which has a cubic magnetic crystal anisotropy. See also note [3]).

[9]) The expression given in reference [6]) has been altered by taking into account that the exchange energy varies with temperature as J_s^2/J_0^2.

[10]) Calculated by dividing the weight of the unit cell by its volume, $(2W_m/N_A)/\frac{1}{2}ca^2\sqrt{3}$, where c is the cell dimension along the c-axis, a is the cell dimension along the a-axis, W_m is the molecular weight and N_A is Avogadro's number.

line sample of Ferroxdure of approximate composition $BaFe_{12}O_{19}$ to different grain sizes (cf. *table III*) that the coercive force at room temperature increases rapidly when dimensions of one micron are approached.

Table III. Coercive force of Ferroxdure for various grain sizes.

	$\mu_0 {}_JH_c$ in Wb/m²	$[{}_JH_c]$ [Oe]
polycrystalline specimen; crystal diameter > 300 μ	0.0050	[50]
specimen ground; diameter of particles about 100 μ	0.0075	[75]
specimen ground; diameter of particles about 3 μ	0.1240	[1240]

In sintered specimens of $BaFe_{12}O_{19}$ the largest coercive forces are obtained in samples which are fired in such a way as to avoid as much as possible the growth of large crystals.

Remanence

The remanent magnetization of a hard magnetic material consisting of crystals with one direction of easy magnetization oriented at random should be one half of the saturation magnetization. This has been found to be true for sintered Ferroxdure of the approximate composition $BaFe_{12}O_{19}$ at temperatures up to 400 °C, as can be seen from *fig. 6*.

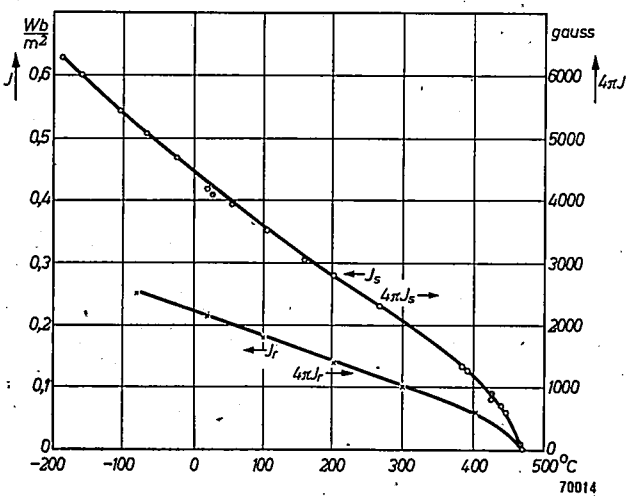


Fig. 6. The remanent magnetization J_r of sintered Ferroxdure of the approximate composition $BaFe_{12}O_{19}$ and the saturation magnetization J_s of $BaFe_{12}O_{19}$ of about the same density as functions of temperature. Note that the relation $J_r = \frac{1}{2}J_s$ holds well.

The value of the remanent induction $B_r = J_r = \varrho\sigma_r$ $[B_r = 4\pi J_r = 4\pi\varrho\sigma_r]$ depends on the apparent density (ϱ) of the material. (In this formula σ_r is the saturation magnetic moment per unit mass in Wb m/kg [gauss cm³/gram] and hence independent of the density.) The X-ray density [10]) of the material is 5300 kg/m³ [5.3 g/cm³] and the highest induction would be obtained with material having this maximum density. This density, however, cannot be obtained without the growth of crystals appreciably larger than one micron, so that the highest remanent induction can only be obtained at the cost of a lower coercive force and vice versa. Therefore a compromise is sought and the density is kept below 5300 kg/m³ [5.3 g/cm³].

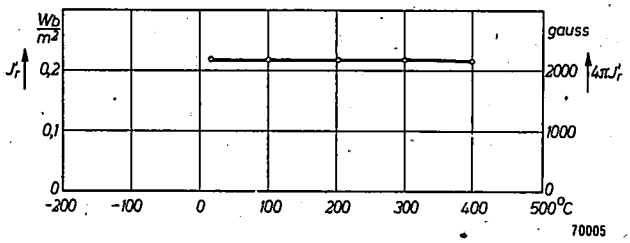


Fig. 7. Remanent magnetization at room temperature J_r' of a sample of Ferroxdure which starting from the remanent state at room temperature has been heated to a temperature plotted as abscissa and cooled to room temperature again.

Owing to the very high coercive force at elevated temperatures, i.e. the high resistance to demagnetization, the remanent magnetization of a rod of Ferroxdure generally returns to its original value after heating to temperatures up to 400 °C. *Fig. 7* shows the remanent induction of a rod with demagnetization factor $N = 0.008$ $[N' = 0{,}1]$ which was magnetized at room temperature, then heated to the temperature plotted as abscissa and measured at room temperature. During and after the heat treatment it was not exposed to a magnetic field except its own demagnetizing field. Thus the remanence can be varied reversibly by a factor 4 (see fig. 6 and 7) by varying the temperature between between 20 °C and 400 °C. This effect can be utilized in temperature measuring and regulating devices, provided the demagnetizing factor is not too large.

Demagnetization curve and value of $(BH)_{max}$

Fig. 8 shows the *B-H* curve for a material with fairly large values of the coercive forces ${}_JH_c$ and ${}_BH_c$. In *fig. 9* the remanent induction is plotted as a function of a demagnetizing field which is applied and removed before measurement. It is seen that demagnetizing fields up to $\mu_0 H = 0.16$ Wb/m² [$H = 1600$ oersted] do not lower the remanence by more than 1%. As a consequence Ferroxdure magnets can be magnetized outside the magnetic circuit in which they are to be used, notwithstanding the fact that they are thus exposed to stronger demagnetizing fields.

The $(BH)_{max}$ values for Ferroxdure lie in the range of 6400-7200 J/m³ [0.8×10^6-0.9×10^6 gauss Oe]. These values are not large in comparison with e.g. those for "Ticonal" G with $(BH)_{max}$ values up to 45 800 J/m³ [5.7×10^6 gauss Oe]. However, as was already mentioned in part I, the $(BH)_{max}$ value should not be used without some reserve as a criterion for the volume of permanent magnet material necessary to obtain a given field strength in an air gap of given volume. This is due to the fact that the demagnetizing field caused by an air gap (and thus also the stray field) depends on the magnetic permeability of the permanent magnet material, the demagnetizing field being stronger the higher the permeability. As the relative permeability of Ferroxdure is only about 1, as against a value of

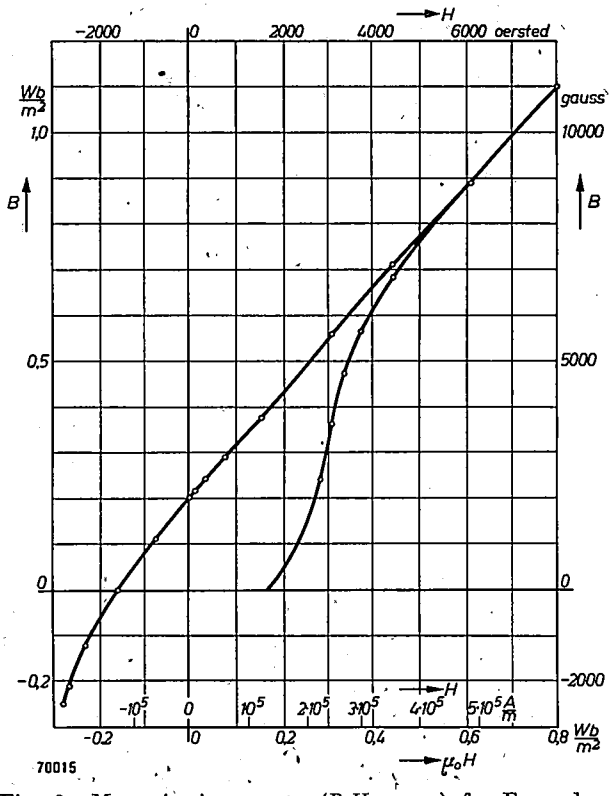


Fig. 8. Magnetization curve (B-H curve) for Ferroxdure.

about 4 for "Ticonal", Ferroxdure compares more favourably with "Ticonal" G than is apparent from the comparison of the $(BH)_{max}$ values. The working point of the material shown in fig. 8 (corresponding to $(BH)_{max} = 6650$ J/m³ [0.83×10^6 gauss oersted]) is $B = 0.0925$ Wb/m² [925 gauss], $\mu_0 H = -0.09$ Wb/m² [$H = -900$ Oe].

Owing to the rather low remanent induction together with the very high coercive force of Ferroxdure, the design of a magnetic circuit with this material will be quite different from that for traditional materials. The permanent magnet in a Ferroxdure circuit has to be disc-shaped (see fig. 2

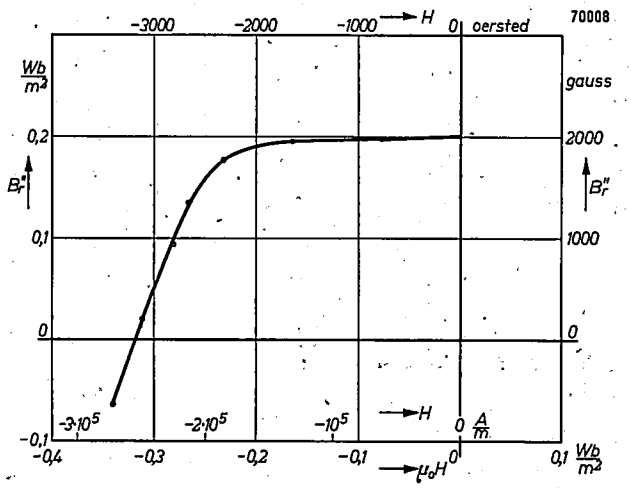


Fig. 9. Remanent induction B_r'' of a sample of Ferroxdure which, starting from the remanent state, has been subjected to a demagnetizing field H during some time.

of part I), while in circuits using traditional materials generally a more rod-shaped magnet has to be used. For certain loudspeaker circuits this proved to be no disadvantage, since the loudspeaker cone itself has a certain lateral extension.

Electric resistivity

The results of measurements of the resistivity of Ferroxdure as a function of temperature are reproduced in *fig. 10* [12]). The electric resistivity of Ferroxdure can be seen to be very large, larger than 10^6 Ωm [10^8 Ωcm] at room temperature. This can be understood from the fact that $BaFe_{12}O_{19}$ contains only trivalent iron ions [11]).

Ferroxdure will therefore prove to be a very

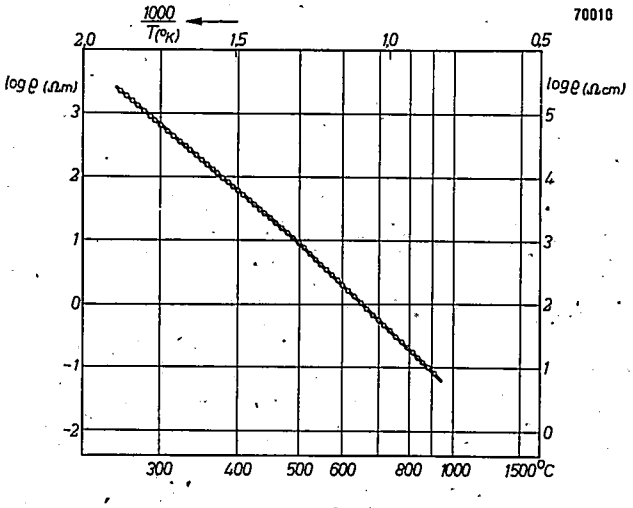


Fig. 10. The logarithm of the resistivity ϱ for Ferroxdure of the approximate composition $BaFe_{12}O_{19}$ as a function of absolute temperature. Plotted is $\log \varrho$ vs $1000/T$.

[11]) J. H. de Boer and E. J. W. Verwey, Proc. Phys. Soc. 49, extra part, 59, 1937.

[12]) The measurements were carried out by J. Smit of this laboratory.

suitable material for high frequency applications, especially in combination with Ferroxcube, e.g. in cases where it is desired to subject this material to a polarizing field. One application of a polarizing field is that a low or negative temperature coefficient of the permeability may thus be obtained (see A and also reference [2]) of part I).

The compound $BaFe_{18}O_{27}$, which contains both ferric and ferrous ions, shows a much higher conductivity than $BaFe_{12}O_{19}$.

A summary of the properties of Ferroxdure is given in *table IV*.

Table IV. Some approximate data on Ferroxdure at room temperature. The composition of the essential phase is $BaFe_{12}O_{19}$. The material, being a sintered ceramic, is brittle and can be ground.

Electric resistivity	$>10^6$ Ωm [10^8 Ωcm]
Saturation magnetic moment per unit mass	8.75×10^{-5} Wb m/kg [70 gauss cm³/g]
Density { X-ray value	5300 kg/m³ [5.3 g/cm³]
Density { actual value	approx. 4800 kg/cm³ [approx. 4.8 g/cm³]
Remanent induction	approx. 0.205 Wb/m² [approx. 2050 gauss]
J-coercive force $\mu_0 {}_JH_c$ [${}_JH_c$]	0.24 Wb/m² [approx. 2400 Oe]
B-coercive force $\mu_0 {}_BH_c$ [${}_BH_c$]	approx. 0.145 Wb/m² [approx. 1450 Oe]
$(BH)_{max}$	approx. 6800 J/m³ [approx. 0.85×10^6 gauss Oe]
Working point	$B \approx 0.1$ Wb/m² [$B \approx 1000$ gauss] $\mu_0 H \approx -0.085$ Wb/m² [$H \approx -850$ Oe]
Temperature coefficient of remanent induction	−0.2 % per °C
Temperature coefficient of induction in working point	approx. −0.15 % per °C
Curie temperature	450 °C

III. SATURATION MAGNETIZATION AND CRYSTAL STRUCTURE

General

The saturation magnetization is of great interest from a practical as well as from a more fundamental point of view. Indeed, it has already been mentioned that an important quantity like the remanent magnetization is closely related to the saturation magnetization. As regards the theoretical aspects, from a study of the saturation magnetization it will be possible to get an insight into the mechanism which is responsible for the magnetic behaviour of the oxides which are dealt with. It will be found that this mechanism is completely analogous to that which prevails in the Ferroxcube materials, which were treated in detail in A (cf. I, [1]) (non-compensated antiferromagnetism). Because of this analogy between Ferroxdure and Ferroxcube we shall start with recalling some facts about the latter materials.

Ferroxcube materials are ferromagnetic oxides which chemically belong to the group of cubic ferrites, the crystal structure of which is the so-called spinel structure. In this structure there are two non-equivalent crystallographic positions available for the metal ions, the so-called octahedral and tetrahedral sites, which form the octahedral and tetrahedral sublattices respectively. The interaction between two magnetic ions is dependent on the kind of sites they occupy. So even if a ferrite with only one kind of magnetic ions (situated on both sublattices) is considered, one has to distinguish between three different interactions, which we shall call briefly: tetrahedral - tetrahedral, tetrahedral-octahedral and octahedral - octahedral interactions, meaning the interaction between a magnetic moment on a tetrahedral site with another one on a tetrahedral site etc. Néel was the first to point out the importance of these three interactions and he generalized the existing "molecular field" theory of ferromagnetism for the case of these different interactions [1]). From the experimental data regarding the values of the saturation magnetization at zero absolute temperature [2]) and regarding the temperature dependence of the paramagnetic susceptibility above the Curie point $\chi = J/\mu_0 H = \mu_r - 1$ [$\chi = J/H = (\mu-1)/4\pi$] Néel deduced that a strong tendency exists for anti-

[1]) L. Néel, Ann. Physique **3**, 137, 1948.

[2]) In this part when speaking of the saturation magnetization we always mean the saturation magnetization at zero absolute temperature.

parallel alignment of magnetic moments on neighbouring non-equivalent sites and a much weaker tendency for antiparallel alignment of magnetic moments on neighbouring equivalent sites. In order to explain roughly the value of the saturation magnetization of most of the single ferrites, for which the conditions are most simple, it is sufficient to take into account the predominant negative tetrahedral-octahedral interaction only. We thus obtain in this simple case for the value of the saturation moment M_s expressed in Bohr magnetons per molecule the formula:

$$M_s = (M_s)_{tetr} - (M_s)_{oct},$$

where $(M_s)_{tetr}$ and $(M_s)_{oct}$ are the sum of all magnetic moments in one "molecule", situated on tetrahedral and octahedral sites respectively. From this the saturation magnetic moment per unit mass in Wb m/kg [gauss cm³/gram], σ_0, is found according to the formula:

$$\sigma_0 = \mu_0 \frac{M_s \mu_B N_A}{W_m} \quad \left[\sigma_0 = \frac{M_s \mu_B N'_A}{W_m}\right], \quad . \quad (1)$$

in which N_A is Avogadro's number, μ_B is the value of the Bohr magneton, viz. 9.27×10^{-24} Am² [9.27×10^{-21} erg/gauss] and W_m is the molecular weight [3]).

We see from the foregoing that Néel's theory is of a formal character in so far as it presupposes the existence of the interactions between the magnetic moments on the various lattice sites and adjusts the sign and magnitude of the interactions so as to be in accordance with the experiments. However, the mere existence of such interactions can not be understood in a more or less "elementary" way, as these interactions must be fundamentally different from the (positive) interaction between the magnetic moments in a ferromagnetic metal. As is well known, the interaction in these metals, the so-called exchange interaction, is of a short range nature, i.e. it is negligible between magnetic moments which are not on neighbouring sites. However, in the ferrites with spinel structure, the magnetic ions are always more or less separated by large oxygen ions and thus are much too far apart to permit of an appreciable direct exchange interaction. Besides this, the exchange interaction is also impeded by the shielding effect of the oxygen ions.

[3]) $N_A = 6.025\times10^{26}$ molecules per kg molecule, $N'_A = 6.025\times10^{23}$ molecules per gram molecule. The factor μ_0 may also be suppressed, in which case σ_0 is expressed in Am²/kg. The two possibilities correspond to the two possible ways of defining a magnetic moment, viz., by a "volume moment" m_v (energy $= m_v H$) or by an "area moment" m_A (energy$_{\text{mutual}} = m_A B$).

Néel already pointed out that a rather intricate mechanism, proposed by Kramers [4]), might be responsible for the above-mentioned interactions between two magnetic ions separated by a non-magnetic anion. In this mechanism, called super-exchange, the anions play an essential role. A more quantitative theory based on these ideas was given by Anderson [5]).

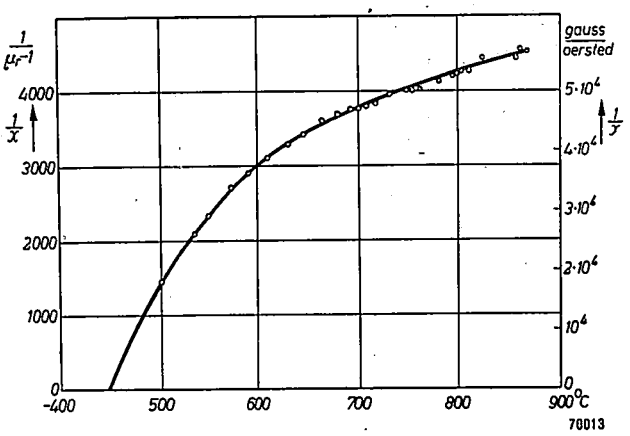


Fig. 1. The reciprocal of the paramagnetic susceptibility $1/\chi$ of $BaFe_{12}O_{19}$ as a function of temperature. It should be noticed that in a normal ferromagnetic above the Curie point this curve should be a straight line according to the Curie-Weiss law. If a straight line would be drawn tangentially to the $1/\chi$ vs. T curve in the Curie point ($1/\chi = 0$) it is seen that $1/\chi$ is smaller and thus χ is larger than if a Curie-Weiss law would hold. This may be made plausible by the assumption that at the Curie point only the weakest interactions lose their ordering influence whereas the stronger interactions still cause a local ordering of the magnetic moments, which is destroyed only at higher temperature. Thus with increasing temperature the number of "free" magnetic moments increases, by which increase the susceptibility falls off more slowly than in a normal ferromagnetic above the Curie point.

Anderson's theory explains the negative sign of the interaction. The theory shows that for given distances between the anion O and the first magnetic ion $Me^{(1)}$ and between the anion and the second magnetic ion $Me^{(2)}$ the interaction is the larger the nearer the angle $Me^{(1)}$-O-$Me^{(2)}$ is to 180°. The interaction will be smallest if the angle $Me^{(1)}$-O-$Me^{(2)}$ is 90°. As regards the dependence of the super-exchange interaction on the distances $Me^{(1)}$-O, O-$Me^{(2)}$, it can be said that it will decrease rapidly with increasing distances. If the dependence on distance is taken into account by neglecting interactions involving Me-O distances larger than 3 Å one can estimate the strength of the interaction between two magnetic ions from the given angles and distances. Thus a useful rule of thumb is obtained.

[4]) H. A. Kramers, Physica **1**, 182-192, 1934.
[5]) P. W. Anderson, Phys. Rev. **79**, 705, 1950.

The preponderance of the interactions of magnetic moments on non-equivalent lattice sites in the ferrites can according to this rule be ascribed to the fact that the angle Me_{tetr}-O-Me_{oct} is of the order of 120°, whereas the angles Me_{tetr}-O-Me_{tetr} and Me_{oct}-O-Me_{oct} are of the order of 80° and 90° respectively.

From this it is suggested that also for compounds more complex than the ferrites, it should be possible to determine very roughly the various interactions without an intricate analysis of the experimental data, once the crystal structure (and thus the various angles and distances) of the compound is known and use is made of the above-formulated rule of thumb. In this way it is possible to make an estimate of the saturation magnetization.

The value of the saturation magnetization as well as the temperature dependence of the magnetic susceptibility (*fig. 1*) of the two oxides to be treated in this part of the paper lend strong support to the assumption that also in these materials there exists a negative super exchange interaction between the magnetic ions.

Crystal structure of $BaFe_{12}O_{19}$

It was already mentioned that the main component of the permanent magnet material which is available at the moment under the name of Ferroxdure is a ferromagnetic phase with hexagonal symmetry, the formula of which is $BaO.6Fe_2O_3$ or $BaFe_{12}O_{19}$. It was shown by Adelsköld [6]) that this compound is isomorphous to the mineral magnetoplumbite, which has a composition near to $Pb(Fe,Mn)_{12}O_{19}$. The crystal structure of this mineral was determined by the same author and is shown in *fig. 2a* (see page 207). The structure given here is slightly idealized, i.e. small deviations of the parameters (coordinates of the ions in the unit cell) of the iron and oxygen ions from the simple values used in this figure have been neglected. The figure shows that the oxygen ions form a hexagonal close-packed lattice, some sites of which are occupied by the Ba-ions. The same argument is valid if the Ba-ions are replaced by strontium or lead ions. These three ions have ionic radii which do not differ appreciably from the radius of an oxygen ion. This is shown in the following table.

The ferric ions are found in the interstices of the oxygen lattice, like the metal ions in the spinel structure. A closer inspection of the crystal structure shows that there are five non-equivalent lattice

Table I. Ionic radii according to V.M. Goldschmidt.

ion	radius in Å
O^{2-}	1.32
Ba^{2+}	1.43
Sr^{2+}	1.27
Pb^{2+}	1.32

sites for the ferric ions. In order to be able to calculate the strength of the various interactions from experimental data, one would have to generalize the molecular field theory of ferromagnetism for the case of five sublattices, which would be much too complicated. However, if the crystal structure is known, we can apply our rule of thumb for the strengths of the interactions. In order to facilitate comparison with the spinel structure, which thanks to former publications may be more familiar to our readers, the spinel lattice has been drawn also in a slightly idealized form in fig. 2*b*, with a vertical orientation of the [111] direction.

A comparison shows that in these idealized structures a large part of the ionic configuration in the $BaFe_{12}O_{19}$ lattice is identical with that of the spinel lattice, as was already pointed out by Adelsköld. In the layers between these spinel "blocks" the oxygen lattice contains a Ba^{2+} ion. In the same layer as the Ba^{2+} ion a ferric ion (◎) appears surrounded in an unusual way by a trigonal bipyramid of five oxygen ions.

The ferric ions on octahedrally surrounded sites belong to three different crystallographic positions (◉, ◍ and ◉) so that together with the tetrahedral position (○) there are five non-equivalent crystallographic positions as shown in the fig. 2*a*.

With the aid of our rule of thumb based on Anderson's theory we can now assign a direction (up or down, at 0 °K) to the magnetic moments of the various ions provided that we arbitrarily fix the direction of the moment of the ion in the ◎ lattice (e.g. up). Now both adjacent moments in the ◉ lattice will point down. The ◎-oxygen-◉ interaction is namely large, because the angle ◎-oxygen-◉ is large (of the order of 140°), whereas the competing ◉-oxygen-◉ interaction, which would tend to orient the ◉-moments mutually antiparallel, will be small because the angle ◉-oxygen-◉ is near 90° (of the order of 80°). Continuing according to this principle we obtain the directions of the moments as given in fig. 2*a*.

Thus for the complete unit cell 16 moments pointing up and 8 pointing down are found i.e. a resultant spontaneous magnetization of 16 − 8 = 8 moments.

[6]) V. Adelsköld, Arkiv för Kemi, Mineralogi och Geologi 12A, No. 29, 1, 1938.

As each ferric ion has a moment of 5 Bohr magnetons, for the saturation magnetic moment M_s a value of 40 Bohr magnetons per unit cell is obtained, which according to formula (1) is equivalent to $\sigma_0 = 12.5 \times 10^{-5}$ Wb m/kg [100.5 gauss·cm³/g] [7]). In *fig. 3* the measured values of the saturation moment are plotted against temperature [8]). By extrapolating this curve towards 0 °K we arrive at a value of $\sim 13.75 \times 10^{-3}$ Wb m/kg [$\sim$110 gauss·cm³/gr] for the saturation moment at zero temperature, which, given the crude assumptions involved in our rule of thumb and the uncertainty in the extrapolation, is in satisfactory agreement with theory [9]).

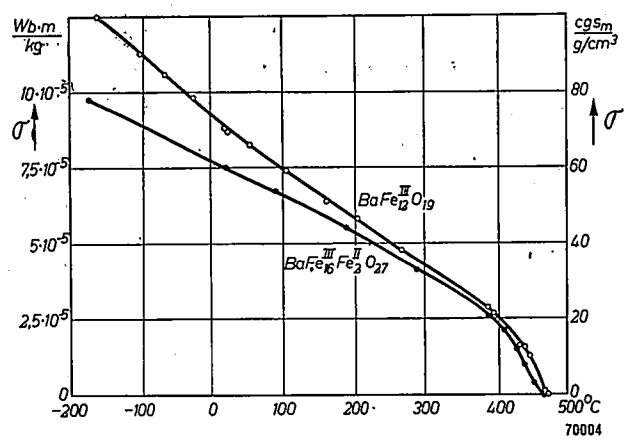


Fig. 3. Saturation magnetic moment σ as a function of temperature.

The low value of the remanent magnetization of Ferroxdure can now be understood from a more fundamental point of view as being due to the fact that in this material the magnetic behaviour is determined by the tendency for antiparallel alignment of the ionic magnetic moments, which will always give rise to a rather low saturation magnetization and thus to a low remanent magnetization.

Crystal structure of $BaFe_2^{II}Fe_{16}^{III}O_{27}$

The second oxide that will be considered here and which also is magnetically hard has a crystal structure very similar to that of $BaFe_{12}O_{19}$.

The chemical composition of this oxide is $BaO.2FeO.8Fe_2O_3$ or $BaFe_2^{II}Fe_{16}^{III}O_{27}$. As may be seen from fig. 2*c* the structure of this compound, which was determined by P. B. Braun of this laboratory [10]), can also be described as consisting of spinel "blocks" separated by layers containing Ba-ions. These layers are identical with those in the structure of $BaFe_{12}O_{19}$. The difference between the two structures is that in $BaFe_2^{II}Fe_{16}^{III}O_{27}$ the spinel "blocks" are higher than in $BaFe_{12}O_{19}$ (cf. fig. 2*a*), because two "molecules" of $Fe^{II}Fe_2^{III}O_4$ (magnetite) are introduced into each block.

The arrangement perpendicular to the hexagonal axis is thereby slightly altered as compared with the arrangement of the analogous ions in $BaFe_{12}O_{19}$. The position of the ferrous ions cannot be determined by X-rays, but in order to obey the electrostatic valency rule [11]) they should be located somewhere midway between the layers containing barium. Assuming a preference for an octahedral position like in magnetite, they should be situated in ◍ or ◍ sublattices, both of which have moments pointing up.

Each Fe^{2+} ion having a moment of 4 Bohr magnetons the spontaneous magnetization of this oxide obtained with the aid of our rule of thumb is $(20—12)5 + 4 \times 4 = 56$ Bohr magnetons, or: $\sigma_0 = 12.4 \times 10^{-5}$ Wbm/kg [99.6 gauss cm³/g].

This value should be compared with the experimental value $\sigma_0 = 10.6 \times 10^{-5}$ Wb m/kg [85 gauss cm³/g].

An illustration of the striking sensibility of the magnetic behaviour to changes in the crystal structure and an additional confirmation of the usefulness of our rule of thumb can be given by considering yet another compound.

A potassium compound exists with the formula $K_2O.11Fe_2O_3$ or $KFe_{11}O_{17}$, which has a crystal structure very similar to that of $BaFe_{12}O_{19}$. This compound, however, is non-ferromagnetic, even at very low temperatures.

The crystal structure, given in fig. 2*d*, was proved by Adelsköld to be the same as that of $Na_2O.\ 11Al_2O_3$ or $NaAl_{11}O_{17}$ (the so-called "β-alumina"), the structure of which was determined by Beevers and Ross [12]). The lattice is identical with that of $BaFe_{12}O_{19}$ except for the arrangement in the layer containing the large ion i.e. the Ba resp. K ion. In $KFe_{11}O_{17}$ the ◎ position is unoccupied and the three oxygen ions present in the $BaFe_{12}O_{19}$ structure have been replaced here by one oxygen ion situated midway between the ferric ions in the ◍ positions. This slight difference in structure explains the large difference in magnetic behaviour. In fact, in

[7]) In applying formula (1) it should be taken into account that one unit cell of $BaFe_{12}O_{19}$ contains two "molecules".

[8]) Since the materials under consideration show an extremely large magnetic anisotropy, a very high field strength is necessary to measure the saturation magnetization in a direction of difficult magnetization. Therefore small single crystals were taken, which were measured in the direction of easy magnetization. The measurements were performed with an apparatus described in reference [1]) of part II. The error is of the order of two percent.

[9]) In making the extrapolation one should take into account that according to Nernst's theorem the slope of the magnetization vs. temperature curve has to become zero near 0 °K.

[10]) The results of this determination will be published in greater detail elsewhere.

[11]) In a stable ionic structure the valency of each anion is exactly or nearly equal, with opposite sign, to the sum of the electrostatic bonds to it from the adjacent cations.

[12]) C. A. Beevers and M. A. S. Ross, Z. Kristallogr. **97**, 57, 1937.

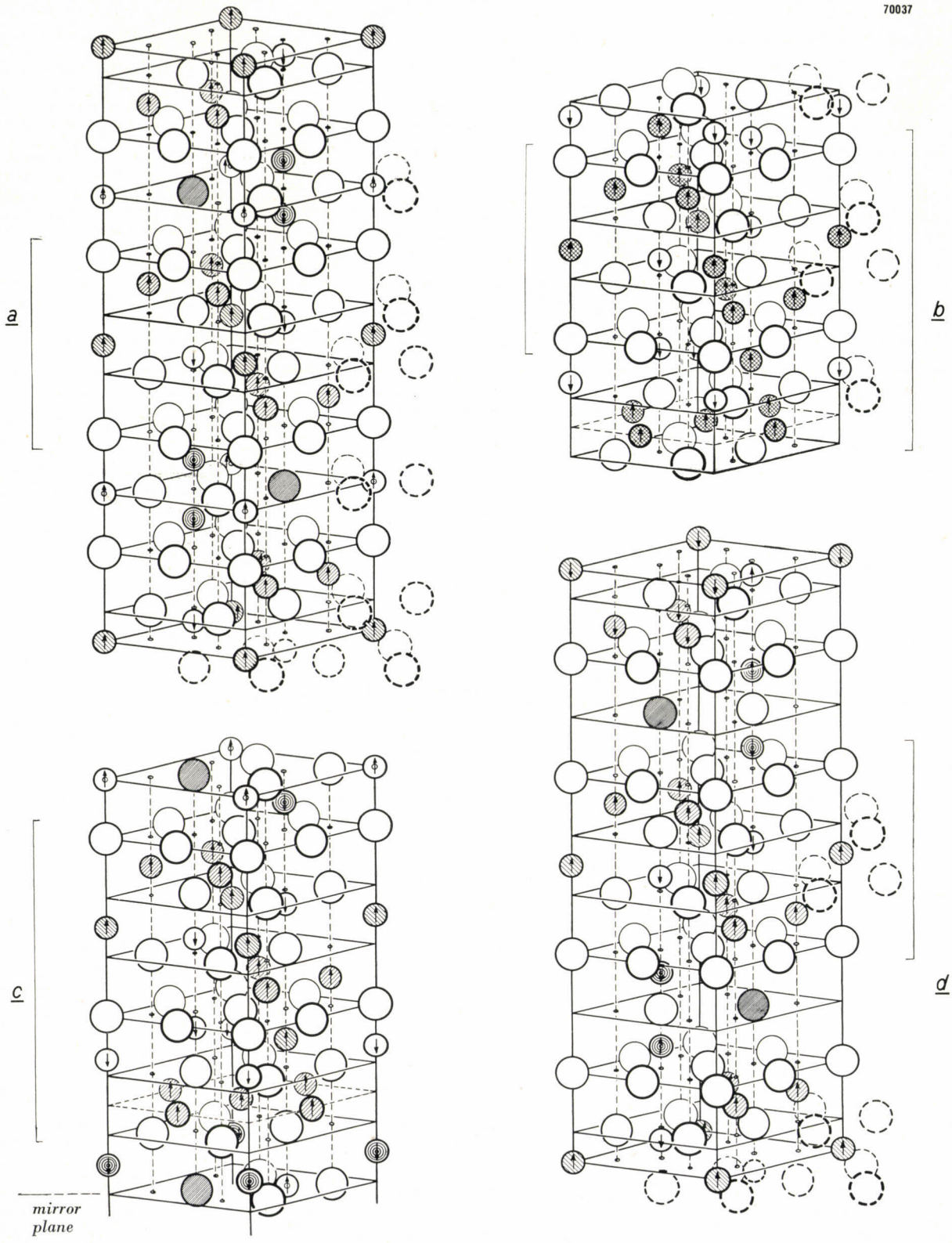


Fig. 2. *a*) Crystal structure of $BaFe_{12}O_{19}$.
b) Spinel structure, with a vertical orientation of the [111] direction.
c) Crystal structure of $BaFe_2^{II}Fe_{16}^{III}O_{27}$ (only half of the unit cell, which possesses a mirror plane, is shown).
d) Crystal structure of $KFe_{11}O_{17}$.

The large unhatched spheres are the oxygen ions, the large hatched ones are the Ba or K ions. The small spheres are the Fe ions: these are hatched in different ways in order to indicate the different crystallographic positions. All the structures are slightly idealized. In parts of the structures *a*, *c*, *d* the positions of the ions are identical to those in the spinel structure *b*. Such parts are indicated by long brackets. The arrows indicate the direction of the magnetic moment of the ions, the direction for one arbitrary ion in each structure being supposed to point upwards.

$BaFe_{12}O_{19}$ the predominant ⊙-oxygen-◉ interactions caused the moments in the ◉ lattices to be parallel; in $KFe_{11}O_{17}$ because the ⊙ lattice is empty only the ◎-oxygen-◉ interaction has to be considered, and as the angle is 180°, this interaction will be strong. The result is that the layer containing the K-ion will be a mirror plane for the directions of the magnetic moments, and the resultant moment of each half unit cell will be cancelled by that of the other half unit cell.

The result is an antiferromagnetic behaviour, in accordance with experiment. The antiferromagnetic Curie temperature may be expected to be very near to the Curie temperature of $BaFe_{12}O_{19}$. Experimental indications supporting this conclusion have been found.

Summary of parts I, II, III. The material manufactured at present under the name of Ferroxdure is an oxidic ceramic material of the approximate composition $BaFe_{12}O_{19}$. It is a magnetically hard material which does not contain any cobalt or nickel and which thus offers great economic advantages. It is characterized by possessing an extremely high value of the coercive force and rather low remanence and saturation magnetization. The extremely high value of the coercive force is due to a large crystal anisotropy. The combination of the two properties, large coercive force and low saturation magnetization means that the material has a large resistance to demagnetization and thus is very well suited for applications where a large resistance to demagnetization is of importance. In the second place these properties open new ways for the design of permanent magnets. The $(BH)_{max}$ value of Ferroxdure is rather low. Its specific electric resistivity is high so that it is very well suited for high frequency applications.

The low value of the saturation magnetization and of the remanent magnetization can be understood to be due to the fact that the mechanism which determines the magnetic behaviour of Ferroxdure is non-compensated antiferromagnetism. The strength of the various interactions between the magnetic ions can be estimated theoretically if the crystal structure is known in detail and if use is made of a rule of thumb, based on Anderson's theory of superexchange interaction. In this way the value of the saturation magnetization of Ferroxdure can be estimated and it is shown that the theoretical value thus obtained does not differ much from the value obtained from experiment.

VOL. 13 No. 8, pp. 209-236 FEBRUARY 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

VELOCITY-MODULATION VALVES FOR 100 TO 1000 WATTS CONTINUOUS OUTPUT

by B. B. van IPEREN.

621.385.831:621.396.615.142.026.443

In the generation of decimetric and centrimetric waves a great advance has been made by replacing the density modulation of the electron stream as used in transmitting valves by velocity modulation which leads at a further stage in the valve to density modulation. The interaction between the electric field and the electrons, which in valves of conventional construction has to take place over distances of tenths of a millimetre, can then take place over distances of some centimetres. This construction makes it possible to generate very much greater powers. In this article it will be explained how, with velocity-modulation valves for 3 to 15 cm wavelength, a continuous output of 100 to 1000 watts can be obtained.

Nowadays there are various special types of valves for generating and amplifying centrimetric waves. In addition to the triodes specially constructed for centrimetric waves — called disk-seal triodes, in which direct density-modulation of the electrons is applied — there may be mentioned the velocity-modulation valve [1]), the reflex klystron and the multireflex valve [2]), and the travelling-wave tube [3]). In these valves the electrons are velocity-modulated and this subsequently leads to density modulation. This principle is applied also in the magnetron, though in a somewhat more complicated manner [4]). Whilst all these valves can be used as oscillators, the disk-seal triode, the velocity-modulation valve and the travelling-wave tube are also useful as amplifiers.

Which type of valve is to be employed depends entirely upon the requirements it has to meet for a particular purpose. As far as oscillators are concerned, to which this article will mainly be restricted, for small outputs (up to, say, 1 watt) reflex klystrons are used on account of their simple construction and operation, whilst also disk-seal triodes and special types of velocity-modulation valves are sometimes employed. For very high outputs (10 to 1000 kW), which are generated only in short pulses, exclusively magnetrons have been used hitherto. It is true that larger continuous outputs can be obtained with magnetrons, but these valves are difficult to tune and, moreover, the heating of the cathode by the returning electrons is apt to lead to instability. For this purpose, however, other suitable valves are available, such as the multireflex valve and the velocity-modulation valve.

The principles of the velocity-modulation valve have already been dealt with at some length in this journal [1]), and it was thereby pointed out that the theory expounded only roughly explained the operation of the valve. In this article, following

1) For a treatise on the principles of the velocity-modulation valve see F. M. Penning, Velocity-modulation valves, Philips Techn. Rev. **8**, 214-224, 1946.

2) F. Coeterier, The multireflection tube, a new oscillator for very short waves, Philips Techn. Rev. **8**, 257-266, 1946.

3) R. Kompfner, Wireless World **52**, 369-372, 1946. For the working of this valve as oscillator see also B. B. van Iperen, The helix as resonator for generating ultra-high frequencies, Philips Techn. Rev. **11**, 221-231, 1950.

4) A review of the working of various modern valves for centimetric waves is to be found in an article by F. M. Penning, Transmitting valves for decimetre and centimetre waves, Communication News **10**, 109-118, 1950.

a brief recapitulation of the principle of the valve, further consideration will be given to the practical realization of the principle and to some theoretical aspects of importance in the construction of the valves.

Elementary theory

The principle of the velocity-modulation valve can best be explained with reference to the diagram in *fig. 1*, where a stream of electrons is assumed to pass through two pairs of grids *a-b* and *c-d* in succession. Let the velocity with which the electrons reach the grid *a* be denoted by v_0 and the voltage required to give them this velocity by V_0. Now, when an

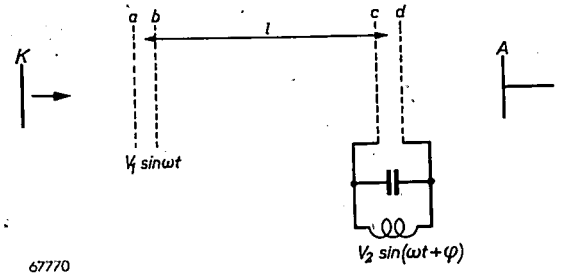


Fig. 1. Schematic representation of a velocity-modulation valve. The electrons emerging from the cathode K arrive with a velocity v_0 at the buncher or input gap bounded by the grids *a* and *b*, between which an alternating voltage is applied. Here the electron beam is modulated in velocity, so that in the drift space between the grids *b* and *c* the beam undergoes a density modulation. When passing through the catcher or output gap between the grids *c* and *d* the beam gives off energy to the output circuit coupled to those grids. Finally the electrons are collected at the anode A.

alternating voltage $V_1 \sin \omega t$ is applied between the two grids *a* and *b* (the buncher or input gap), each electron is accelerated or retarded according to the instant at which it passes through this pair of grids. In the drift space, between the grids *b* and *c*, which are placed a fair distance apart, electrons which started later but have been accelerated between *a* and *b* have an opportunity to overtake other electrons which started earlier but which have suffered retardation. The result of this is that the density of the stream of electrons which at *a-b* was constant in time shows farther on periodical variations with the fundamental frequency ω. As the stream passes through the catcher or output gap between the pair of grids *c-d* its A.C. component induces an equally large alternating current in the circuit connected to the grids *c* and *d* (the output circuit). When this circuit is tuned close to the frequency ω, so that for that frequency it has a high impedance, then the stream of electrons imparts energy to that circuit. Thus the valve acts as an amplifier. For the voltage between the grids *a* and *b* to be induced with as little loss of energy as possible it is necessary to connect also these grids to a tuned circuit (input circuit). When this circuit is electromagnetically coupled with the output circuit an oscillator is obtained.

Since we shall be dealing here particularly with oscillator valves, it is especially important to consider the efficiency that can be reached with a construction such as that referred to above. First a distinction has to be made between the electronic efficiency, η_e, — that part of the D.C. power fed to the valve which is converted into H.F. power in the electron stream — and the overall efficiency η_t, defined similarly, but with the losses in the input and output circuits deducted from the H.F. power generated in the stream.

Assuming in the first instance that the circuit losses are negligible, we have only to deal with the electron efficiency, and for a construction such as that of fig. 1 this can be calculated in a fairly simple manner from the formula:

$$\eta_e = -\frac{V_2}{V_0} \sin(2\pi\xi + \varphi) \cdot J_1(\pi a \xi), \quad . \quad . \quad (1)$$

where V_2 is the amplitude of the alternating voltage between the catcher grids, φ is the phase difference between V_2 and the alternating voltage V_1 between the buncher grids, $a = V_1/V_0$ is the modulation depth, ξ the number of cycles of the H.F. alternating voltage required for an electron with velocity v_0 to traverse the distance from buncher to catcher, and J_1 is the Bessel function of the first kind and the first order.

Formula (1) is derived as follows. After having passed the buncher grids at a moment t_1, an electron has an energy

$$eV = eV_0 + eV_1 \sin \omega t_1 = eV_0 (1 + a \sin \omega t_1), \quad . \quad (2)$$

in which e is the charge of the electron. Taking v_0 and v_1 as representing the velocity of the electron before and after passing the grids, then $eV_0 = \frac{1}{2}mv_0^2$ and $eV = \frac{1}{2}mv_1^2$. Substitution in (2) gives:

$$v_1 = v_0 (1 + a \sin \omega t_1)^{\frac{1}{2}}.$$

Denoting the time of arrival of the electron at the second pair of grids by t_2 and the distance between the two pairs of grids by l, then

$$t_2 = t_1 + \frac{l}{v_0} (1 + a \sin \omega t_1)^{-\frac{1}{2}}.$$

Assuming that $a \ll 1$, thus that the depth of modulation is small, we may write:

$$t_2 = t_1 + \frac{l}{v_0} (1 - \tfrac{1}{2} a \sin \omega t_1). \quad . \quad . \quad . \quad . \quad (3)$$

If the alternating voltage across the catcher grids is $V_2 \sin(\omega t + \varphi)$ then, on passing these grids, the electron loses

energy to an amount

$$E = -eV_2 \sin(\omega t_2 + \varphi),$$

which energy is imparted to the output circuit in the form of oscillation energy. Substituting (3) in this expression, we find for the energy given off by the electron:

$$E = -eV_2 \sin\left(\omega t_1 + \varphi + \frac{\omega l}{v_0} - \frac{\omega l a}{2v_0} \sin \omega t_1\right) \quad . \quad . \quad (4)$$

To arrive at the total amount of energy imparted to the output circuit per second, the values of E found for all electrons passing through in a period of one second have to be added. Thus the power yielded by the stream of electrons is:

$$P_{tot} = -i_0 V_2 \cdot \frac{1}{2\pi} \int_0^{2\pi} \sin\left(\omega t_1 + \varphi + \frac{\omega l}{v_0} - \frac{\omega l a}{2v_0} \sin \omega t_1\right) d(\omega t_1)$$

$$= -i_0 V_2 \sin\left(\varphi + \frac{\omega l}{v_0}\right) J_1\left(\frac{\omega l a}{2v_0}\right),$$

where i_0 is the current density in the stream.

Now $\omega l/v_0 = 2\pi\xi$, and since $i_0 V_0 = P_0$ is the input power, the electronic efficiency is:

$$\eta_e = \frac{P_{tot}}{P_0} = -\frac{V_2}{V_0} \sin(2\pi\xi + \varphi)\, J_1(\pi a \xi).$$

The factors occurring in eq. (1) are now to be chosen such that η_e is the maximum obtainable. This means that the following conditions have to be satisfied:

(1) $J_1(\eta a \xi)$ must be a maximum; such is the case for $\eta a \xi = 1.84$, whereby $J_1(\eta a \xi) = 0.58$.

(2) V_2/V_0 has to be a maximum; this means that this quotient has to be made equal to unity, since for $V_2 > V_0$ some electrons when reaching the catcher grids would be turned back, in which case the foregoing theory no longer holds; a closer investigation shows that the efficiency is then greatly reduced.

(3) $\sin(2\pi\xi + \varphi) = -1$, thus, when n is a whole number, $2\pi\xi + \varphi = (n - \frac{1}{4})2\pi$, from which it follows that $\xi = n - \frac{1}{4} - \varphi/2\pi$.

Provided these three conditions are satisfied, the efficiency is 58%.

These conditions call for some further explanation. (3) can always easily be satisfied because ξ can be given any desired value by chosing a suitable initial velocity v_0 for the electrons. From condition (1) it follows that $a = V_1/V_0 = 1.84/\pi\xi$, from which the value of the modulation voltage is determined. Further, the output $P_{tot} = V_2^2/2Z$, where Z represents the impedance between the catcher grids. If $P_0 = i_0 V_0$ (i_0 being the current in the beam of electrons) is the input power, then in the ideal case — when according to (2) $V_2 = V_0$ — the output is $P_{tot} = 0.58\, P_0$, so that Z has to be given the value $V_0^2/1.16\, P_0 = V_0/1.16\, i_0$. For the case of an amplifier, therefore, with the proper choice of V_1, V_0 and Z, the maximum H.F. power can be derived from the beam of electrons.

In an oscillator valve, on the other hand, the value of V_1 cannot be directly adjusted. For such a valve the optimum adjustment is reached by a suitable choice of the feedback factor, i.e. the ratio $K = V_1/V_2$, which can indeed be directly adjusted, as will presently appear. For adjustment to the maximum electronic efficiency again in the ideal case $V_2 = V_0$, and thus we find: $K = V_1/V_2 = V_1/V_0 = a = 1.84/\pi\xi$.

In practice, however, it is not so much a mattar of the maximum electronic efficiency but rather the maximum overall efficiency. In the first place the losses in the input and output circuits cause the overall efficiency η_t to be smaller than the electronic efficiency η_e. If the buncher and the catcher have the same impedance Z_k then the circuit losses amount to $(V_1^2 + V_2^2)/2Z_k$. Therefore, to obtain a high overall efficiency it is of importance to make Z_k as large as possible.

Moreover, the maximum of η_t need not occur with the same value of the feedback factor K as does the maximum of η_e.

This is readily understood when considering the condition under which the valve just begins to oscillate. In that case the output power from the electron stream is just equal to the circuit losses, so that

$$i_0 V_2 J_1(\pi a \xi) = (V_1^2 + V_2^2)/2Z_k,$$

and since with a small amplitude $J_1(\pi a \xi) \approx \frac{1}{2}\pi a \xi$, this becomes:

$$\tfrac{1}{2} i_0 V_2 \left(\pi\xi \frac{V_1}{V_0}\right) = V_2^2 \frac{1 + K^2}{2Z_k},$$

or

$$i_0 = \frac{1 + K^2}{K} \frac{V_0}{\pi\xi Z_k}.$$

This expression is a minimum when $K = 1$, in which case

$$i_0 = \frac{2V_0}{\pi\xi Z_k}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

For $K = 1$, therefore, the current at which the valve begins to oscillate is lowest. Just above this minimum current determined by (5) the valve therefore oscillates only when K is made equal to 1, at which value in this case also the (here very small) overall efficiency is a maximum. For the value of K at which the electronic efficiency is maximum we found, on the other hand, $K = 1.84/\pi\xi$, which usually is much less than unity. In general the value of K for a maximum total efficiency will lie between 1 and $1.84/\pi\xi$.

Adjustment of the feedback

Adjustment of the feedback factor — the quantity $K = V_1/V_2$ — of two coupled circuits is not a special problem typical of the velocity-modulation

valve but also arises with a triode oscillator working with two tuned L-C circuits. The behaviour of such systems has already been thoroughly investigated [5]), but since the adjustment of the feedback factor is less generally known this problem will be briefly explained here.

Let us assume that the input and output circuits are inductively coupled (*fig. 2*) by a mutual inductance M. As is well known in radio engineering, the behaviour of such a system depends upon the coupling coefficient $f = M/\sqrt{L_1 L_2}$ (not to be confused with the feedback factor K), the quality factors Q_1 and Q_2 and the two natural frequencies ω_1 and ω_2 of the separate circuits. For the case in which we are interested we may for the time being assume that $1/Q_1 = 1/Q_2 = \delta$, while it is further assumed that $\delta \ll f \ll 1$.

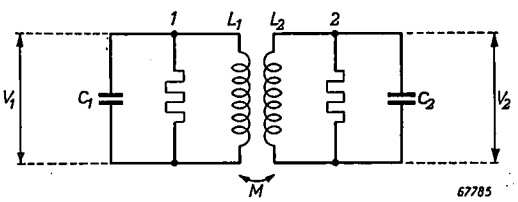


Fig. 2. Two inductively coupled circuits representing schematically the input circuit (*1*) and the output circuit (*2*) of a velocity-modulation valve used as oscillator. With a suitably chosen difference between the natural frequencies of the two circuits it is possible to adjust the feedback factor V_1/V_2 to any desired value, provided the coupling coefficient $f = M/\sqrt{L_1 L_2}$ between the circuits is large enough.

Now three cases are to be considered:

(1) The naturel frequencies of the two circuits are equal, thus $\omega_1 = \omega_2$.

(2) The relative difference between the natural frequencies is very great compared with the coupling coefficient, thus $|\omega_1 - \omega_2|/\omega_1 \gg f$.

(3) The relative difference between the natural frequencies is of the same order as the coupling coefficient.

In the case where $\omega_1 = \omega_2 = \omega_0$ the system has two natural frequencies: $\omega_a = \omega_0 (1 + f)$ and $\omega_b = \omega_0(1 - f)$. At the frequency ω_a the voltages V_1 and V_2 are in phase, thus $\varphi = 0$, whereas at the frequency ω_b they are in anti-phase ($\varphi = 180°$). Further, from considerations of symmetry it follows that in both cases $V_1 = V_2$, so that the feedback factor $K = 1$. This can be represented schematically as in *fig. 3*.

In the second case, where $|\omega_1 - \omega_2|/\omega_1 \gg f$, there are also two natural frequencies but there they are almost equal to the natural frequencies of each of the circuits separately. At the natural frequency $\omega_a \approx \omega_1$ the value of V_1 becomes large, while the oscillation of the second circuit is only weak, so that V_2 remains small ($K \gg 1$). At the other natural frequency, $\omega_b \approx \omega_2$, the situation is just the reverse ($K \ll 1$). The phase relations remain the same as in the first case. Schematically this is as represented in *fig. 4*.

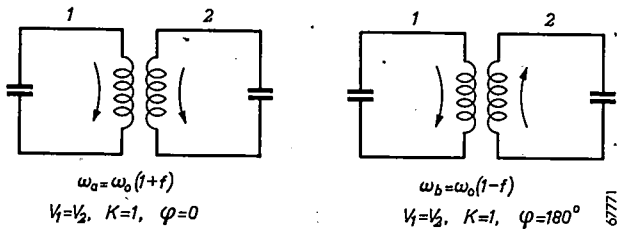


Fig. 3. Schematic representation of the mode of oscillation of the system according to fig. 2, for the case where the two circuits are tuned to the same frequency. At the higher of the two natural frequencies the two circuits are in phase, and at the lower one they are in anti-phase. In both cases the feedback factor K is equal to unity.

In a velocity-modulation valve usually K is required to be less than unity, but not very small. This can be achieved by choosing for $|\omega_1 - \omega_2|$ a value between those of the extreme cases. This brings us to the third case mentioned above, where $|\omega_1 - \omega_2|/\omega_1$ is of the same order as f. The manner in which K depends upon the detuning $(\omega_1 - \omega_2)/\omega_1$ between the two circuits is shown in *fig. 5*, from which it appears that any desired value of K can be reached by a suitable choice of $\omega_1 - \omega_2$ (provided, however, that the above-mentioned condition that the coupling coefficient f is large compared with $1/Q = \delta$, is satisfied).

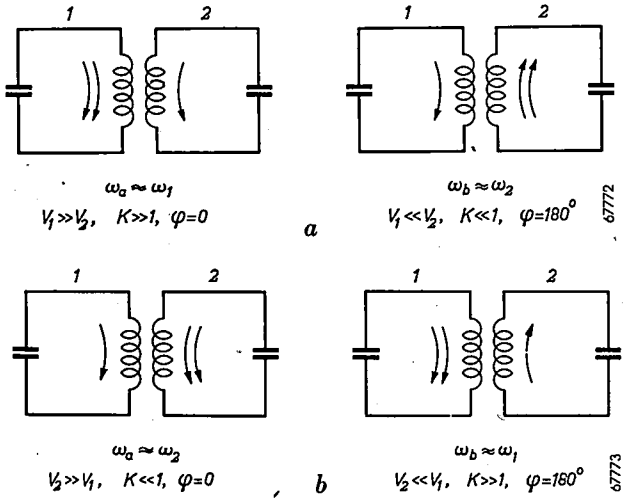


Fig. 4. Schematic representation of the mode of oscillation of the system according to fig. 2 when the natural frequencies of the circuits differ greatly. Here again there are two natural frequencies, but at one of these $K \gg 1$ (very high feedback factor) whilst at the other $K \ll 1$ (very low feedback factor). In the case (*a*) the natural frequency of the first circuit is much greater than that of the second one, in the case (*b*) it is much smaller.

[5]) See, e.g., B. D. H. Tellegen, Philips Res. Rep. 2, 1-19, 1947.

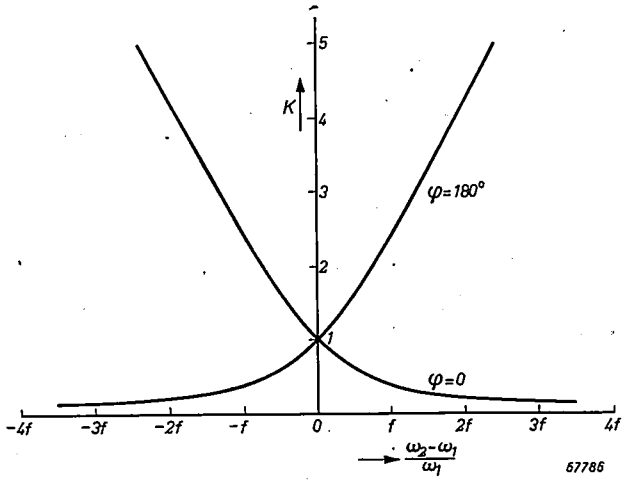


Fig. 5. Relation between the feedback factor K and the detuning of the circuits, for the system of fig. 2. For each value of the relative detuning $(\omega_2 - \omega_1)/\omega_1$ there are two natural frequencies, corresponding to the two curves in the diagram. When with one mode of oscillation, e.g. that where the circuits are in phase ($\varphi = 0$), K is greater than 1, then with the other mode of oscillation ($\varphi = 180°$) K is less than 1, and vice versa. f is the coupling coefficient between the circuits.

General principles of the construction

These considerations are useful for a general insight into the working of a velocity-modulation valve. When, however, the principle is applied for generating energy at a very short wavelength it appears that the conception from which we started is quite inadequate, because account has to be taken of factors which have not yet been included. This can best be explained by first examining more closely the practical construction of a velocity-modulation valve.

If, as was supposed, grids were used in the path of the electron beam then these could have only a small surface, because in the oscillator circuit they form a capacitance which adversely affects the resonance impedance. Owing to the bombardment by the electrons these grids reach a very high temperature, so that the input power has to remain extremely small. For high outputs, therefore, it is preferable to choose a solution whereby each pair of grids is replaced by two hollow cylinders placed in line with each other and with a gap between them. These are respectively the input gap and the output gap. For the oscillator circuits use is made of cavity resonators of such a shape and of such a material that the highest possible impedance is reached.

The valve has to answer a number of other requirements too. In the first place it must be possible to vary the natural frequency of the cavity resonators, since the valve should preferably be tunable, and moreover, as already seen, with a view to adjusting the value of the feedback factor K the natural frequencies of the two cavity resonators should be adjustable with respect to each other. Further, in order to obtain the highest possible output, the tube should be capable of large dissipations.

These latter requirements lead to the valve being so designed that the wall of the cavity resonators serves at the same time as the vacuum envelope. *Fig. 6* gives a cross-sectional (schematic) view of

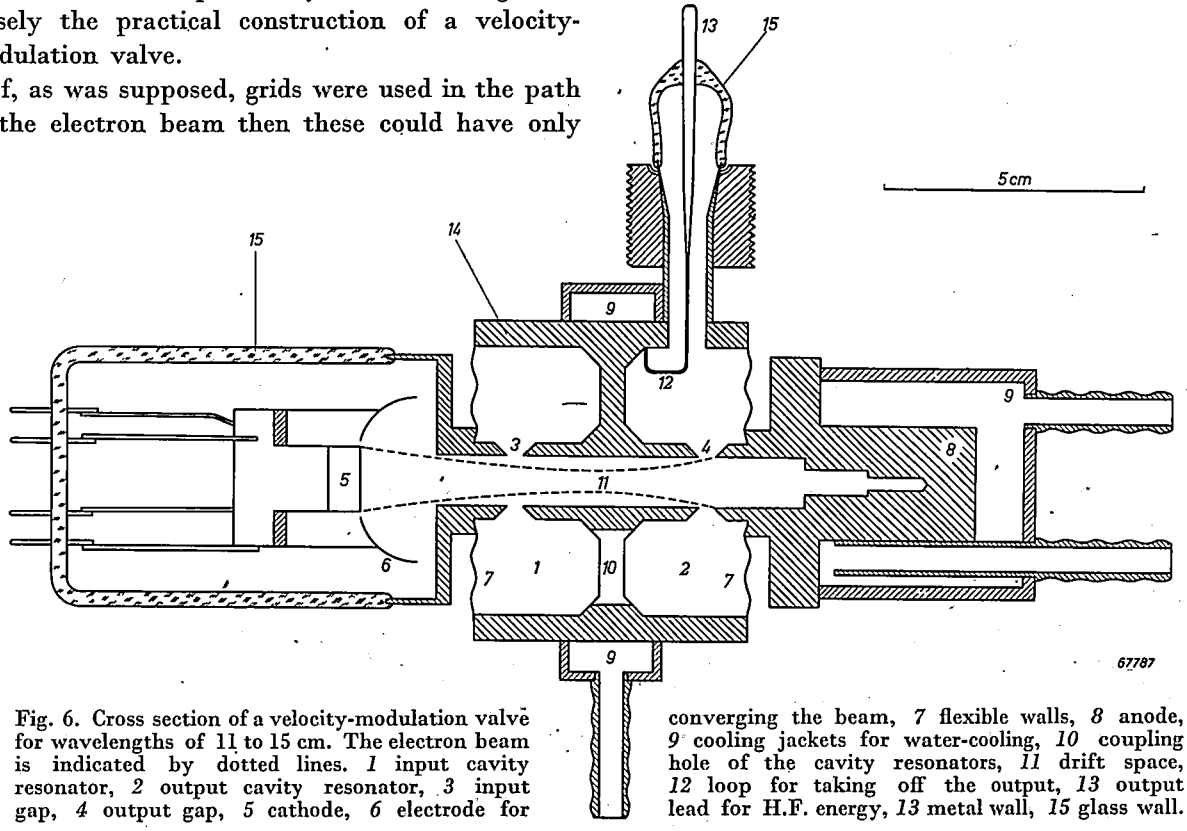


Fig. 6. Cross section of a velocity-modulation valve for wavelengths of 11 to 15 cm. The electron beam is indicated by dotted lines. *1* input cavity resonator, *2* output cavity resonator, *3* input gap, *4* output gap, *5* cathode, *6* electrode for converging the beam, *7* flexible walls, *8* anode, *9* cooling jackets for water-cooling, *10* coupling hole of the cavity resonators, *11* drift space, *12* loop for taking off the output, *13* output lead for H.F. energy, *13* metal wall, *15* glass wall.

a valve constructed in this way; for a detailed explanation see the subscript. The shape of the electron beam is indicated by dotted lines. The natural frequency of the cavity resonators can be varied by changing the width of the gaps, this being possible owing to one of the side walls of each of the cavity resonators being made thin and thus flexible. Since the gap edges form the major part of the capacitance of the cavity resonator, by only slightly changing the gap width a considerable change in wavelength can be obtained.

Effects not taken into account in the elementary theory

As we have seen, under certain assumed conditions the simple theory predicts an efficiency of 58% when disregarding the losses in the cavity resonators, which in practice may well be considerable. In the construction of a valve as described above it is not easy to realize to a good approximation the conditions assumed in the elementary theory. To make this clear we have to consider more closely the manner in which the electrons traverse the system and the interaction between the electron beam and the electromagnetic field.

Influence of mutual repulsion of the electrons

So far it has been assumed that the electrons travel in straight paths, and this is indeed possible of attainment to a good approximation when a very strong magnetic field is placed in the direction in which the electrons are travelling. Such a solution, however, is rather impracticable and preference is therefore given to a valve without magnetic field. Under the influence of their mutual repulsion the electron paths which are originally parallel in a beam will begin to diverge, the more so the stronger the current at a particular electron velocity. Consequently, with a cylindrical tube of a given diameter and length, and with a certain initial velocity of the electrons, a limit is set to the current that can be sent through the tube without causing electrons to be lost at the wall of the tube.

It is better, therefore, to start with a convergent beam of electrons. A convergent beam reaches a minimum diameter at a certain point and then begins to diverge; from the theory it follows that the shape of the beam is symmetrical with respect to this minimum. It is easily seen that a beam with the greatest possible current can be sent through a given tube when the minimum diameter of the beam is half-way along the length of the tube (*fig. 7*). From the theory it follows further that the maximum current which can pass through the tube in this way is proportional to the square of the ratio of the tube diameter to the tube length.

The theory outlined above shows that it is desirable to have the highest possible beam current for a given voltage. In order to maintain the necessary voltages across the input and output circuits

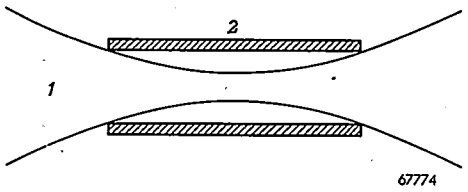


Fig. 7. Owing to mutual repulsion of the electrons, a beam which initially is convergent passes through a minimum diameter and then diverges. The shape of the beam is symmetrical with respect to the minimum. When the beam is sent through a round tube of a certain length and diameter it can be given the greatest current strength without electrons colliding with the wall by ensuring that the minimum diameter of the beam is reached half-way along the tube. *1* = electron beam, *2* = round tube.

it is necessary to supply the power equivalent to the circuit losses. Increasing the current then increases the H.F. power in the beam and this increase is entirely available as output. Thus the efficiency of the tube increases with the beam current.

Influence of the gap field on the paths of the electrons

The divergence of the beam is not the only factor restricting the current strength. Owing to the deflecting action of the H.F. field in the input gap we have to keep well below the limit just mentioned. To explain this let us assume for the sake of simplicity that before reaching the bunching field the beam is parallel. Now let us follow an electron travelling along a path at a certain distance from the axis of the tube (*fig. 8*) and assume that it passes the centre of the field at the moment when the alternating voltage between the gap edges has the maximum value. The radial force to which this electron is subjected during the first half of its passage through the gap will then be equal and opposite to that acting upon the electron during the second half of its passage. To a first approximation, therefore, this electron does not undergo any change in direction (path *1*, fig. 8). The same applies when the H.F. field has its maximum negative value.

Now suppose, however, that the electron passes through the centre of the gap just when the alternating voltage is passing through zero from "accelerating" to "retarding". During the first half of the time spent in the gap the electron is deflected towards the axis, and because of the reversal of field the deflection will be in the same

direction during the second half of the transit. Thus this electron undergoes a change of direction towards the axis (path *2*).

The reverse applies to an electron passing through the field when the voltage is passing through zero from "retarding" to "accelerating". Such an electron is deflected outward, the more so the nearer its path lies to the outer edge of the beam, i.e. the closer the electron passes along the gap edges (path *3*). This deflection may easily reach such proportions that the electron strikes against the wall of the drift space before it has passed the output gap.

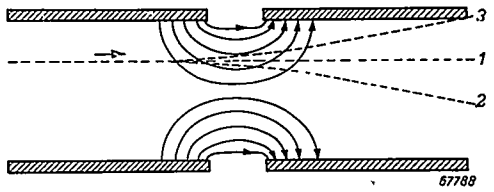


Fig. 8. An electron passing through the H.F. field in the gap of the input cavity undergoes a deflection which varies according to the time that the electron passes that gap.
1 path of an electron passing at the instant that the field has its maximum accelerating or retarding force.
2 path of an electron passing at the instant that the field is changing from the accelerating to the retarding phase.
3 path of an electron passing at the instant that the field is changing from the retarding to the accelerating phase.

Now this is most unfortunate, since these outwardly deflected electrons happen to be the most useful. This can be explained as follows. Electrons passing through just in front of the one we have been considering here have undergone a retardation (negative velocity modulation) and electrons immediately following it undergo an acceleration. On their way to the inductor the electrons of the group in question thus draw closer together and form a so-called bunch. The valve works best when this bunch of electrons enters the output gap at the moment that the H.F. voltage is in its maximum retarding phase, so that the group of electrons in question undergoes a strong retardation. If, owing to deflection in the input gap, part of this group of electrons strikes against the wall of the drift tube, then naturally this reduces the efficiency of the valve. For this reason care has to be taken to ensure that when passing the input gap the beam is fairly narrow. It will presently be shown how this undesired effect can be further counteracted by a special construction of the output gap.

Energy exchange between gap field and electrons

From the foregoing it would seem that the tubes through which the electron beam is sent should be given the largest possible diameter, that is to say, as large as is compatible with the construction of a cavity resonator with a suitable impedance at a certain frequency.

It appears, however, that on other grounds we are very limited in the choice of the tube diameter. To explain this it is necessary to reexamine the elementary theory. There we assumed a pair of grids with an alternating voltage $V_1 \sin \omega t$, and when calculating the electronic efficiency it was assumed that an electron, after passing through the pair of grids with an energy eV_0 at the instant t_1, had an energy $e(V_0 + V_1 \sin \omega t_1)$. This is only exact if the transit time of the electron between the two grids is very short compared with the cycle of the H.F. voltage. If that is not the case then the energy transferred to the electron is always less than $eV_1 \sin \omega t$. It may be proved that for $V_1 \ll V_0$, the energy variation can always be represented as $\beta eV_1 \sin \omega t_1$, where β is a constant smaller than unity. This applies not only for the case where we have a field between two grids but also for the case where the electron passes through a field of a different form, as for instance the H.F. field in the input gap of a valve constructed as in fig. 6. For this latter case, however, β is dependent on the distance between the path of the electrons and the axis of the tube.

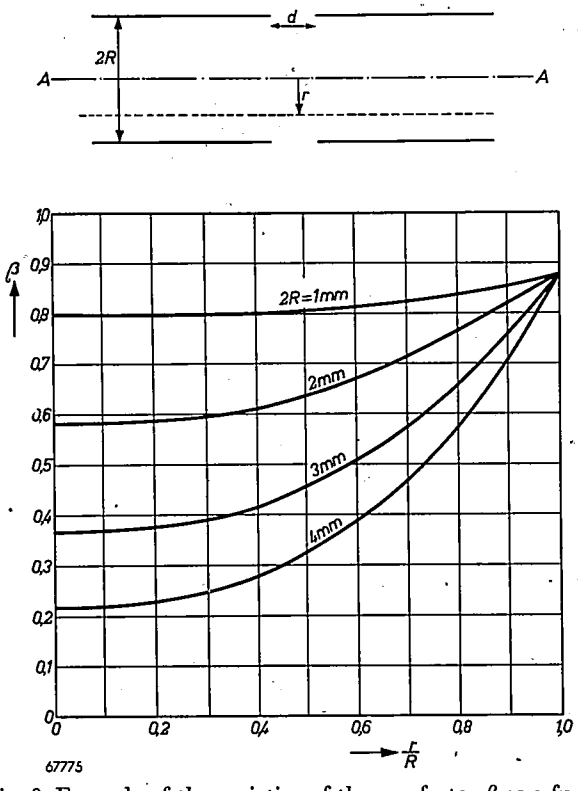


Fig. 9. Example of the variation of the gap factor β as a function of the distance r from the electron path to the axis AA of the tube. The diagram applies for the case of an accelerating voltage $V_0 = 6000$ V, a wavelength $\lambda = 3$ cm, a gap width $d = 0.5$ mm and for different tube diameters $2R$.

An example of the variation of β is to be found in *fig. 9*, from which it is seen that for a not too small value of β one is limited to small tube diameters. The importance of a high value of β can easily be understood: if the energy of the electron is to be changed by an amount eV_1 by means of the H.F. field then this requires an alternating voltage V_1/β, thus a power $V_1^2/\beta^2 Z_k$. Something similar applies also for the output gap. It may therefore be said that the effect in question apparently reduces the impedance of the circuits by a factor β^2.

The fact that β depends upon the distance of the electron path from the axis of the tube has also an adverse effect upon the efficiency of the valve, in that the electrons do not all undergo the same velocity modulation, so that the modulation voltage cannot be made optimum for all electrons in the beam at the same time. The same holds in respect of the interaction with the output circuit, the electrons in the middle of the beam always giving off less energy than those on the outside.

Gap damping

Another phenomenon also arising as a result of the finite transit time of the electrons in the H.F. field is gap damping. By this is meant the following. When an electron beam passes through two grids placed close together and modulating the velocity of that beam, on an average the modulating field does not impart any energy to the beam, since just as many electrons are accelerated as are retarded, whilst the average acceleration equals the average retardation. It appears, however, that this no longer strictly holds if the transit time in the H.F. field is not very short compared with the time of oscillation. In most cases a certain amount of energy is then imparted to the beam by the buncher. This, therefore, constitutes a damping of the input circuit, i.e. an apparent reduction of the impedance of that circuit. Something similar is also the case with the output circuit.

Finally it should be pointed out that in the elementary theory it was assumed that the modulation depth $a \ll 1$, but in practice this is not correct either.

The result of all these effects is that the overall efficiency of velocity-modulation valves remains very much below the calculated value of 58%. We shall not here go further into the question as to how a more exact theory should be formulated. Suffice it to say that in practice an overall efficiency of 25% is to be regarded as very satisfactory.

The electron gun

From the discussion of the static electron paths we have seen that the beam entering the oscillator system is required to have a certain angle of convergence. A method by which such a beam is obtained has been suggested by Pierce [6]) and will be briefly described as follows.

First imagine that we have a diode consisting of two concentric spheres with the outer one acting as a cathode. When a positive voltage is applied to the inner sphere then, for reasons of symmetry, the electrons coming from the cathode move towards the anode in straight lines. Let us now focus our attention upon a circular part of the cathoden surface, from which a convergent beam of electrons is moving towards the anode (see *fig. 10*). If we

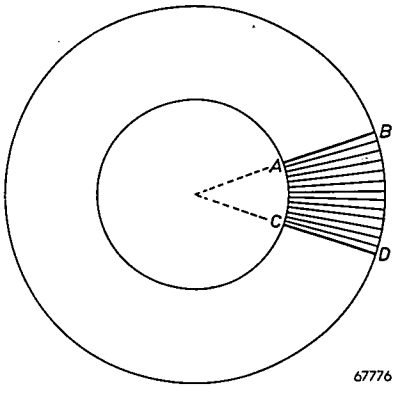


Fig. 10. In a diode consisting of two concentrical spheres, the outer one of which acts as a cathode, the electrons follow straight paths. If only a circular piece *BD* of the outer sphere is permitted to emit electrons then the electron paths within the cone *ABCD* remain rectilinear provided steps are taken to ensure that the potential distribution along the conical surface is the same as that obtaining when the whole of the outer sphere is emitting.

remove the other electrons, for instance by making the rest of the cathode surface non-emitting, then the rectilinear motion of the electrons is disturbed, since the potential distribution in the space between cathode and anode has been changed. Now it is possible to make the electrons in the cone still follow the original straight paths by making, in some way or other, the potential distribution along the conical surface again equal to the original distribution. This is done by setting up metal electrodes of a suitable shape close to the beam. The shape of these screens can be determined experimentally in an electrolytic tank, and they may also be

[6]) J. R. Pierce, J. appl. Phys. **11**, 548-554, 1940. The fact that the problem here is quite different from that in other valves in which narrow electron beams are used, such as cathode-ray tubes, is due to the much higher beam current required.

calculated [7]). In practice it is found to be sufficient if two electrodes are used, the first one at about cathode potential and the second at anode potential, the shape of the latter electrode not being very critical. As second electrode, therefore, the front of the oscillator system, where the beam enters, serves the purpose.

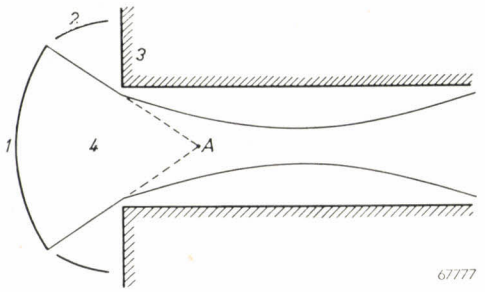


Fig. 11. Shape of the electron beam in a non-oscillating velocity-modulation valve (for the sake of clarity the transverse dimensions are exaggerated). A conical electron beam *4* (converging towards *A*) leaves the cathode *1* provided with a screen *2*. Owing to the divergent action of the hole in the oscillator system *3* the beam becomes less convergent, reaches its minimum diameter half-way along the system and then begins to diverge.

It is to be noted that the beam emerging from the cathode has to be more convergent than that ultimately required, since the opening through which the beam enters the oscillator system acts as a negative lens. Provided there are no oscillations, the shape of the beam is then as represented in *fig. 11*, where its width has been exaggerated for the sake of clarity. In the valve the screen round the cathode is fixed to a can in which the cathode is mounted in the right position with respect to that screen. In valves designed for wavelengths below 10 cm this unit is fixed, and insulated, to a ring fitting in a recess on the front of the oscillator system. In this way the electron gun — the combination of cathode and screen — can be exactly centered with respect to the oscillator system. A photograph of some guns for the various valves to be discussed is given in *fig. 12*.

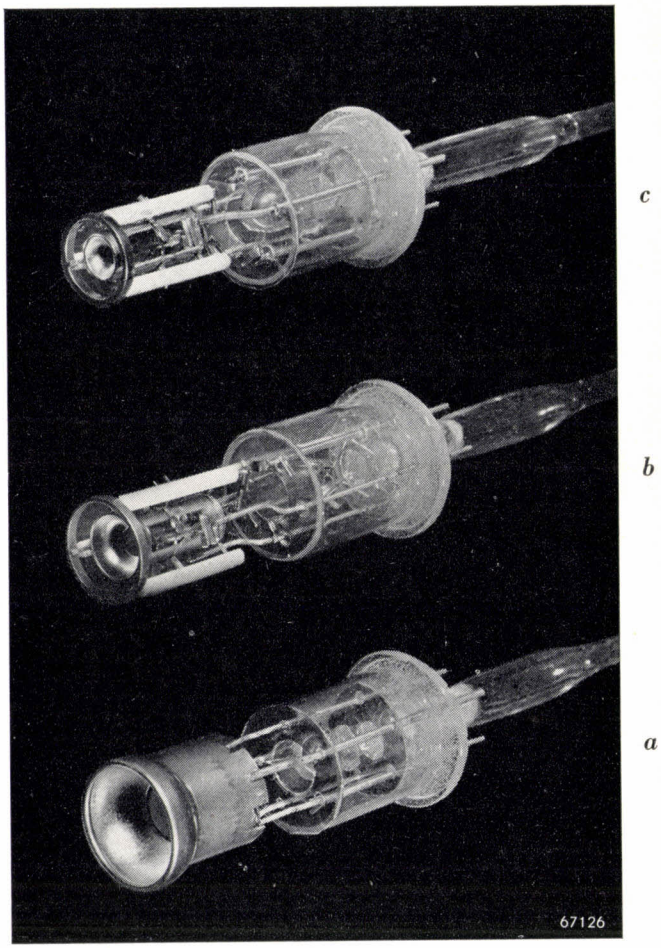


Fig. 12. Photograph of some electron guns as used in the valves described in this article. From *bottom* to *top*: *a*) gun of a valve for wavelengths of 11 to 15 cm and of a valve for 10 cm, 1200 watts; *b*) gun of a valve for 5.5 cm to 7 cm; *c*) gun of a valve for 2.9 cm to 3.6 cm. The glass cylinders between the gun and the bridge prevent the getter (barium) from being precipitated on undesired places.

The cathode

The fact that electron guns can be constructed in the simple way described above is due entirely to the use of the L-cathode, a new type of cathode already fully described in this journal [8]). At first the construction of the cathode was a difficult problem, because a valve of optimum dimensions involves comparatively high voltages (for one of the types to be described later, even as high as 15 kV), whilst moreover for smaller wavelengths the beam diameter, and thus the cathode diameter, has to be made smaller and smaller, resulting in great current densities (up to 2.4 A/cm^2 for the smallest type constructed).

It is therefore not possible to use the normal oxide-coated cathode in the gun construction described here. Owing to the high tension alone the inevitable bombardment with positive ions (formed from residual gases) would destroy the oxide coating of the cathode within a few hours. Apart from that, the current density required is too great for the cathode to last a reasonable length of time.

There are, it is true, various techniques for circumventing such difficulties. In larger types of

[7]) According to a method indicated by E. G. Dorgelo (not published).

[8]) J. H. Lemmens, M. J. Jansen and R. Loosjes, A new thermionic cathode for heavy loads, Philips Techn. Rev. **11**, 341-350, 1950.

valves, for instance tungsten and tantalum have been used as cathode material, but owing to the high working temperature the heating of these metals gives rise to great difficulties which can only be overcome by rather complicated constructions. (It is to be borne in mind that the cathode has to be a circular, almost flat, disc.) A second objection is that these cathodes require a high heater power. Furthermore, this solution is not practicable in valves for very small wavelengths owing to the great current densities required.

The L-cathode, on the other hand, fully meets the requirements. Notwithstanding the great densities of current it has a satisfactory life, and it can also withstand bombardment with high-velocity ions. It needs heating to a temperature only slightly higher than that of a normal oxide-coated cathode, so that a comparatively low heater power suffices.

Feedback and tuning of the cavity resonators

What has been said in the foregoing about the feedback in the case of L-C circuits applies in the main also for two coupled cavity resonators. Here the coupling is brought about via a hole in the wall between the cavities, and since in velocity-modulation valves this hole is situated at a point where there is mainly a magnetic alternating field, this corresponds to an inductive coupling. The coupling coefficient f is fixed by the size of the coupling hole. In practice f is about $\frac{1}{2}\%$, whilst for the unloaded cavity resonators $Q = 1/\delta$ is about 2000 to 4000, so that the condition $f \gg \delta$ is satisfied. It is true that the quality factor of the output cavity resonator is reduced by the loading, but this does not essentially change the behaviour of the system in regard to the variability of the feedback factor.

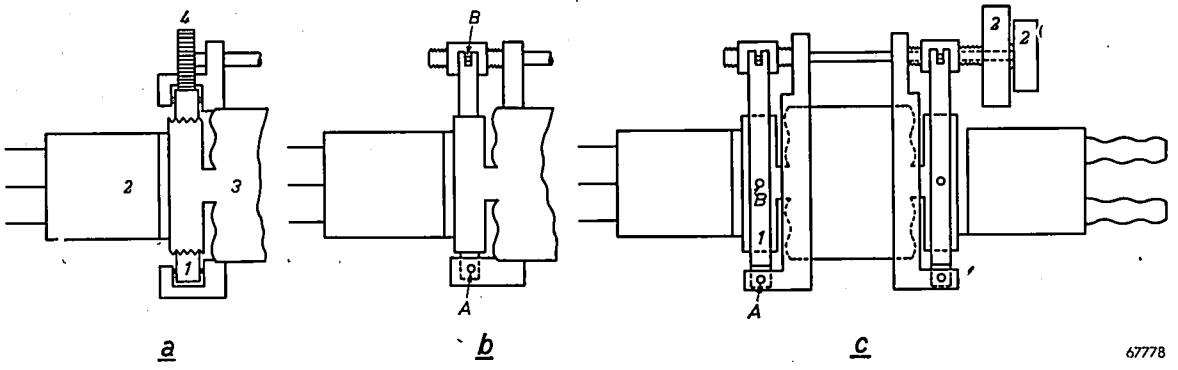


Fig. 13. *a*) Schematic representation of a tuning device operated with the aid of a threaded ring fitting round the movable part of the tube. Rotation of the ring imparts an axial movement to the cathode, respectively anode, part of the valve (here only the cathode end of the valve is shown). *1* ring toothed on the outside and rotated by a small gear wheel *4*. *2* cathode part, *3* middle part of the valve.
b) Tuning device working on the lever principle. The cathode part is pivoted to the middle part at *A* and moved by some means or other at *B*. Here the movable part describes a circular motion instead of an axial one.
c) New tuning mechanism consisting of a ring (*1*) pivoted to the middle part at *A* and likewise pivoted to the cathode part at *B* (two diametrically opposed points). The motion is now axial to a sufficient degree of accuracy. The rings are moved by turning two knobs *2* each serving a cavity resonator. The spindle for one resonator has left-hand thread and that for the other right-hand thread. See further fig. 14.

For adjusting the feedback factor and the wavelength of the oscillations it must be possible to vary the frequency of each of the cavity resonators. To that end provision has to be made for accurately adjusting the width of the gaps in the input and output cavities. It has already been shown in fig. 6 how the parts of the cavities at the cathode and anode ends of the valve are connected with the middle of the valve by means of a diaphragm. Thus the valve consists of three parts: the cathode, the anode and the middle parts, which have to be movable axially with respect to each other.

An obvious means of bringing this about is indicated in *fig. 13a* (only the cathode end has been drawn, the situation being the same at both ends). A threaded ring is fitted round the part to be moved, the position of this ring with respect to the middle part being fixed by means of a groove, so that when the ring is turned the cathode part is displaced with respect to the middle part. Such a device, however, is difficult to make without leaving some play, and there is much friction.

Another frequently applied device, with lever action, is represented in fig. 13*b*. This is very simple but it has the drawback that the slope of the moving part with respect to the middle part of the valve is changed, so that this system can only be used for very small variations of the gap width.

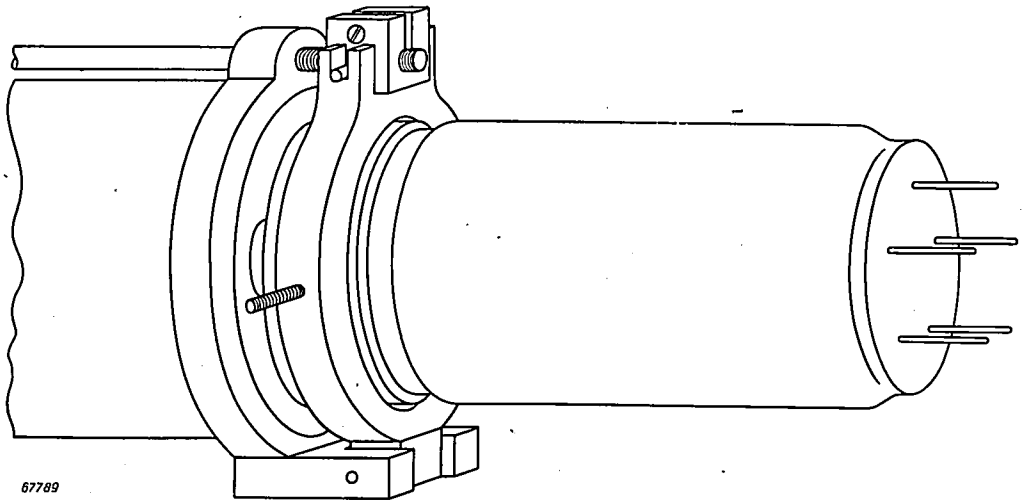


Fig. 14. Sketch showing one end of a valve with the mechanism according to fig. *1*3*c*.

The construction applied for the valves to be described combines the advantages of the two systems just mentioned (see fig. 13*c* and *fig. 14*). Here the lever is in the form of a ring placed round the movable cathode part and connected with it at two pivoting points. The ring is likewise pivoted at the bottom in a junction piece connecting it with the middle part of the valve. Thus a movement is obtained which is of sufficient axial accuracy for the desired range. The two cavity resonators are operated by means of two knobs mounted on concentric spindles, as shown in fig. 13*c*. Right-hand thread is used for the adjustment of one resonator and left-hand thread for the other, so that when both knobs are turned in the same direction the frequency variation brought about in the two cavities is also in the same sense, which is most convenient.

Improved construction of the output gap

When studying the paths followed by the electrons we saw that in the buncher space the electrons underwent a change of direction, such that the most useful electrons were deflected outward and might well strike against the wall of the drift tube and thus be lost.

In some of the types of valves to be mentioned later, this possibility has been greatly reduced in the following way. The electrons are allowed to undergo deflection in the H.F. field in the manner described, so that without further steps being taken they would impinge on the wall of the drift tube. In this wall, however, and also in that which bounds the catcher field of the other side, longitudinal

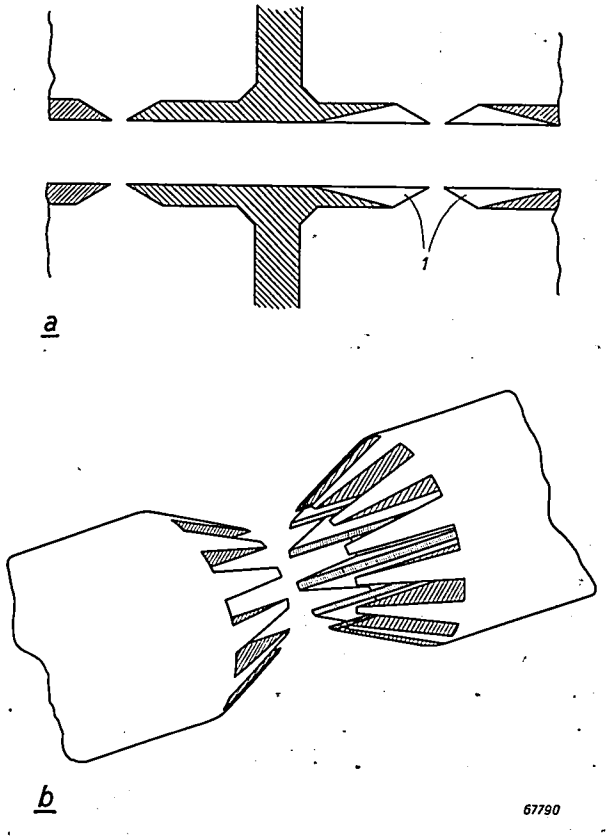


Fig. 15. *a*) Sketch of the output gap in a valve provided with longitudinal grooves *1* in the wall of the drift tube in order to increase the effective beam current. *b*) Schematic representation of the shape of the grooves in a valve as in (*a*).

grooves have been made through which the electrons can for the greater part proceed further on their way and pass through the catcher field. As *fig. 15* shows, the edges of the output gap are thus given the shape of a toothed ring. With this construction it appears to be possible to increase the beam current considerably and at the same time to improve the efficiency of the valves, so that with the same voltage a much larger H.F. output can be obtained.

Some experimental types of valves

Four experimental types of valves have been constructed in the Philips Laboratory.

A cross-sectional diagram of one of them has already been given in fig. 6. For the second type the essential dimensions of the valve were halved, thereby making it suitable for frequencies twice as high. These two types have both been made with the walls of the output gap grooved in the manner just described, so as to increase the effective beam current. The first type of valve is suitable for generating wavelengths of 11 to 15 cm, the greatest output — about 500 watts — being obtained at about 12 cm. With this adjustment the valve voltage is 9 kV and the beam current 220 mA, giving an overall efficiency of 25%. As the wavelength is increased the voltage has to be reduced in proportion to the square of the frequency. Also the beam current then drops, so that the output diminishes rather tapidly. In this range the valve works with an elecrron transit time ξ between the input and the output gaps of $1\frac{3}{4}$ cycles. At wavelengths below 12 cm the electrons can be given a transit time of $2\frac{1}{4}$ cycles. The output of this valve is via a coaxial terminal, the inner conductor of which passes out through a glass cap. A coaxial line can be connected to this system.

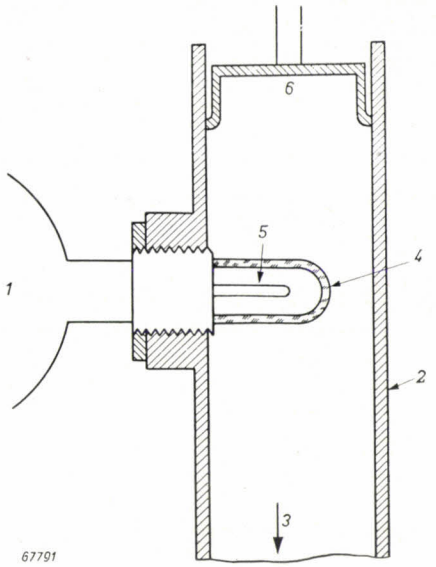


Fig. 16. Schematic representation of the device for extracting energy from a velocity-modulation valve by means of a coaxial output terminal protruding into a rectangular wave guide. *1* valve, *2* wave-guide, *3* direction in which the energy is conducted, *4* glass cap, *5* inner conductor of the coaxial output terminal, *6* plunger.

The second type is suitable for wavelengths of 5.5 to 7 cm, giving at 6.1 cm a maximum output of over 300 W (voltage 8.5 kV, beam current

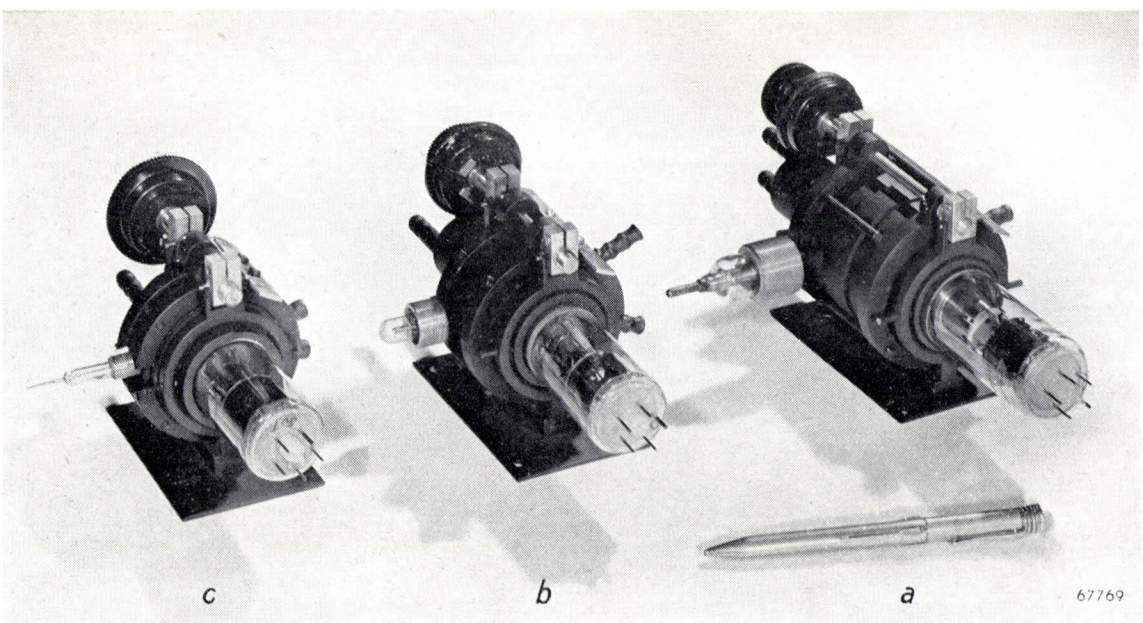


Fig. 17. Some velocity-modulation valves for different wavelength ranges: *a*) valve for $\lambda = 11$ to 15 cm, max. continuous output 500 W; *b*) valve for $\lambda = 5.5$ to 7 cm, max. continuous output 300 W; *c*) valve for $\lambda = 2.9$ to 3.6 cm, max. continuous output 100 W.

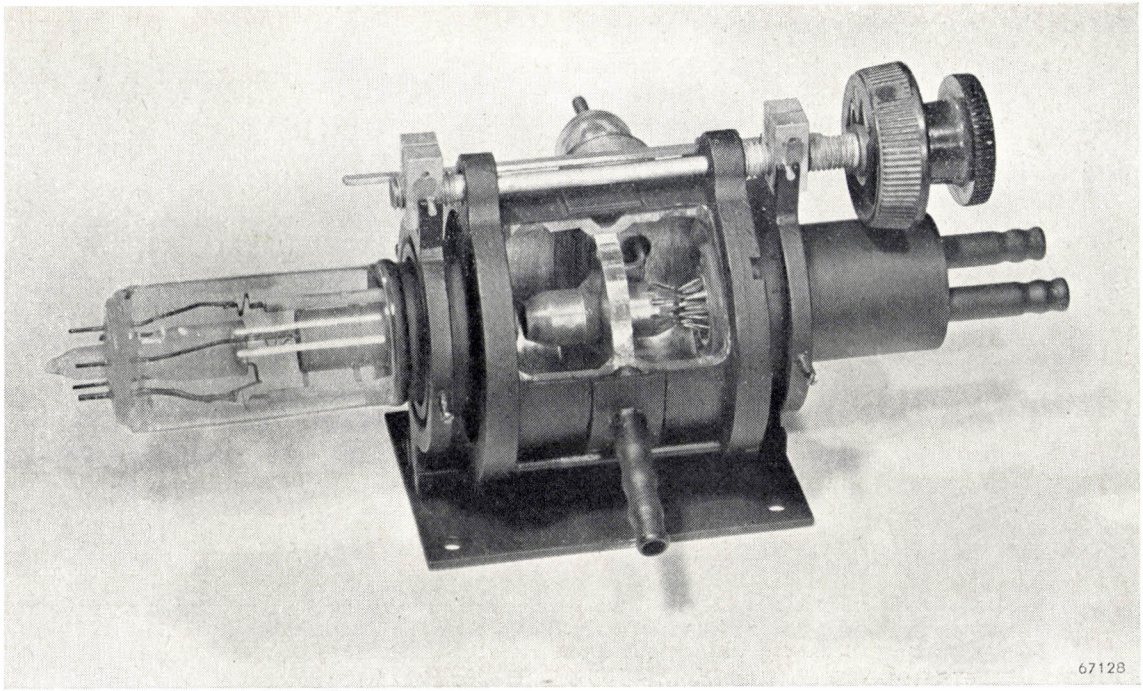
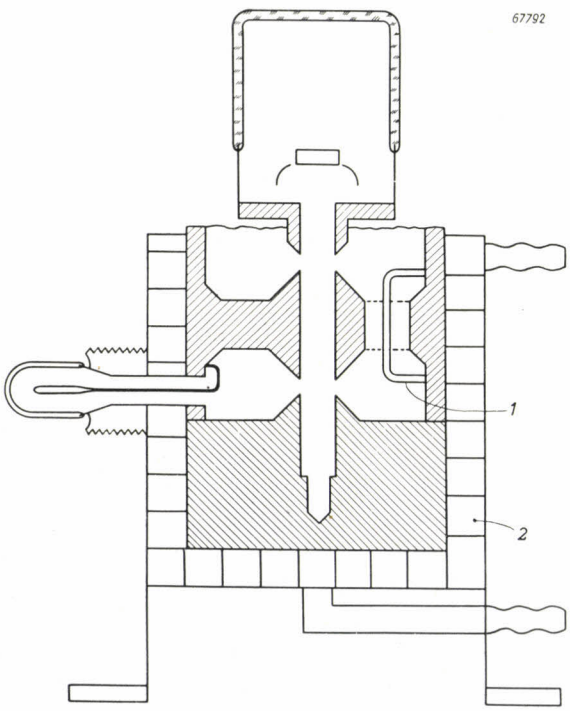
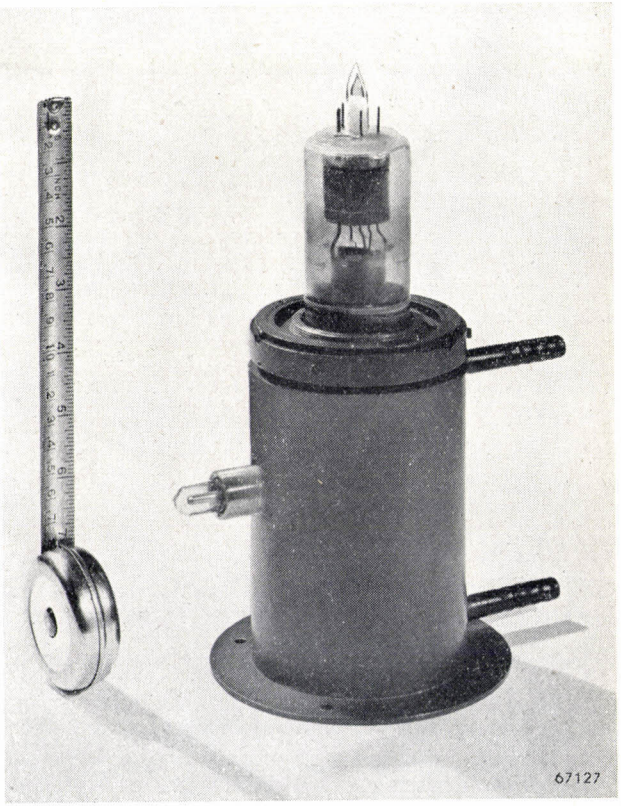


Fig. 18. Velocity-modulation valve for $\lambda = 11$ to 15 cm, with part of the wall of the cavity resonators removed, showing the toothed edges of the output gap.

175 mA, efficiency 20%). Here the inner conductor of the coaxial output system is not led out but is terminated inside the glass cap. This can be coupled into a rectangular wave-guide, as represented in *fig. 16*. Photographs of the complete valves are given in *figs. 17a* and *b*, while *fig. 18* shows a valve with part of the wall removed.





a *b*

Fig. 19. Schematic cross section (*a*) and photograph (*b*) of a velocity-modulation valve for 10 cm waves and 1200 watts continuous output. *1* feedback bracket, *2* cooling jacket with helical partition.

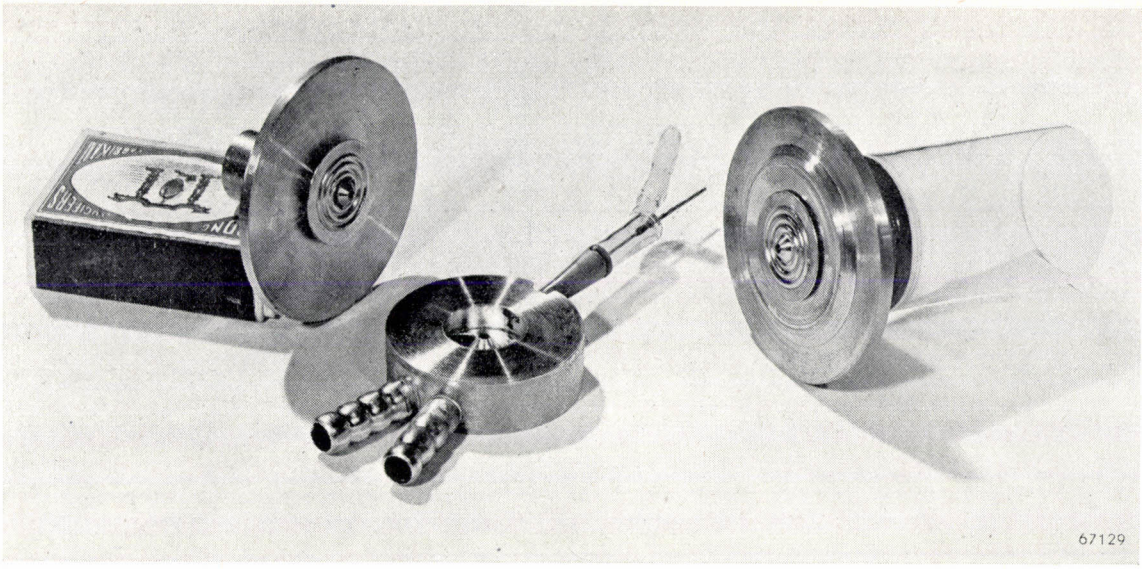


Fig. 20. The three parts making up a velocity-modulation valve for 3 cm waves. From right to left: cathode part with one of the flexible walls, middle part with output terminal and cooling jacket, anode part with the other flexible wall; the cooling jacket for the anode is put on later.

A third type of valve, without tuning, has been constructed for higher outputs than those mentioned above. The output cavity resonator has a fixed resonance, while the frequency of the input cavity can be varied only slightly for correct tuning. Further in this type the wall between the cavities has been made extremely thick (15 mm), with a view to improving the dissipation of the drift tube. As a consequence the mutual coupling of the cavity resonators cannot be effected by means of a hole alone — this hole would act as a wave-guide the cut-off wavelength of which would be much smaller than the wavelength at which the valve works. A bracket is therefore passed through the hole, forming a kind of coaxial feedback system (*fig. 19a*). The valve is enclosed over its entire length in a cooling jacket, to remove the heat produced by the H.F. losses in the cavity resonators — some hundreds of watts. On a wavelength of 10 cm, this valve, a photograph of which is reproduced in fig. 19*b*, yields an output of 1200 watts at a voltage of 15 kV and a beam current of 400 mA (efficiency 20%). Here the output system is similar to that used in the 6 cm valve, i.e. coaxial into a wave-guide (not passing through the glass envelope). In this valve the advantage of the L cathode is very striking, the filament current required being no more than 30 watts.

The fourth type of valve has been designed for wavelengths of about 3 cm. Its construction does not differ essentially from that of the other types. *Fig. 20* shows the three parts (cathode, middle, anode parts) of which the valve is made up, whilst in fig. 17*c* a photograph is given of the complete valve. This is suitable for wavelengths of 2.9 to 3.6 cm, yielding on 3.1 cm an output of 100 watts at a voltage of 6.6 kV, with a beam current of 150 mA (efficiency 10%). For constructional reasons the inner conductor of the coaxial output system is brought out, but here again the energy can be taken off via a rectangular wave-guide.

In this valve, too, the features of the L-cathode show to full advantage. The current density at the cathode surface under top load is 2.4 A/cm^2. With a former version of the L cathode used under these conditions a life of the order of 1000 hours was reached, and the latest version promises to have an even longer life.

Summary. Following a brief treatment of the elementary theory of the velocity-modulation valve, consideration is given to the problem of using the principle of velocity modulation for generating centimetric waves. It appears that in the construction of valves for large outputs the assumptions upon which the elementary theory is based are not sufficiently exact and that account has to be taken of some effects which had been disregarded in the theory. The adverse influence of one of these effects, the deflection of the electrons in the H.F. field of the input gap, can be reduced by giving the output gap a suitable shape. Further, by applying the L cathode, it is possible to simplify the construction of the electron gun. A description is also given of a new device for tuning the cavities. The article is concluded with a discussion of four experimental types of valves and the results obtained with them.

A THROUGH SUPERGROUP FILTER FOR CARRIER TELEPHONE SYSTEMS ON COAXIAL CABLE

by J. F. KLINKHAMER.

621.392.52:621.395.44

Coaxial cables have been used in various countries for some considerable time in conjunction with carrier telephone equipment. Their importance in this field lies in the fact that hundreds of telephone channels can be superimposed on one pair of coaxial tubes. On the other hand, the provision of the necessary flexibility, in respect of the interconnection of blocks of channels at the nodal points of the cable network, has given rise to very severe technical problems. This difficulty arises partly from the multiplicity of channels, which was the raison d'être of the coaxial system, and partly from the requirements laid down by the C.C.I.F. at a rather early stage of development of coaxial technique. The through supergroup filter, described below, has been developed as a solution of some of these interconnection problems.

The transfer problem

During the past 15 years the long distance telephone plant of most countries has been undergoing a rapid expansion, in which carrier telephone systems have played an ever-increasing part. The growth of the carrier telephone network can be seen both in respect of the total route length of carrier cables and in the number of channels provided by each cable; although the number of conductors per cable has decreased, this has been more than compensated by the very large increase in the number of channels which can be provided on each pair of conductors. This figure has risen from 4, 12, 17 and 24 channel systems in the past [1]) to 48 and 60 channel systems on modern quadded carrier cable [2]) and to 600 or 960 channels on a coaxial system.

To discuss this latter trend of carrier technique is beyond the scope of this paper [3]). We are concerned here only with the transfer problems arising from the multiplicity of channels per pair, which occurs typically in the case of coaxial cable. *Fig. 1* shows as an example a line chart of a typical carrier system, in order to illustrate the above general statements. Suppose that between two main centres, c.q. the cities A and B, there is a carrier system capable of providing at least 1200 channels, of which the major part, say 720 channels, is required for communication between A and B, the remaining 480 channels being required for transit traffic, i.e. for communication from A via B to C and b_1 or from B via A to a_1. These transit channels must, therefore, be extended in station B to station C or b_1, and in station A to a_1.

If each of the A-B circuits were provided on the cable by means of a separate pair of conductors, then the circuits could be extended or terminated by simply connecting these conductors in any desired arrangement at A and B. No great difficulty is experienced if this method of connection is used on multipair carrier cables with, say, blocks of 48 [4]) or fewer channels per physical 4-wire circuit (one pair "go" and one pair "return"). Under these circumstances, each set of two pairs can be branched off or connected through in a relatively simple way, and complete blocks of channels, e.g. 48 channels, can be considered as units for traffic purposes.

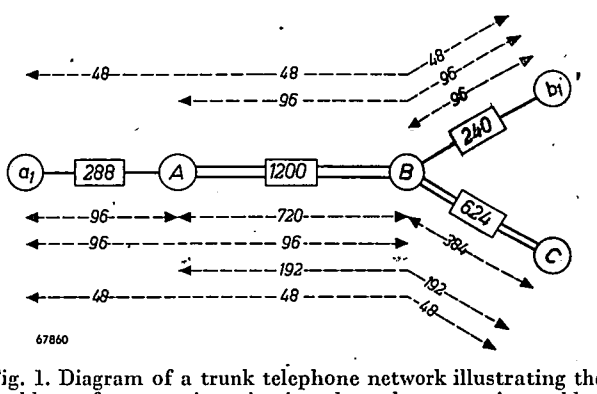


Fig. 1. Diagram of a trunk telephone network illustrating the problem of connecting circuits through on carrier cables. Between a_1 and A there are 288 channels available, between A and B 1200, etc. The numbers between the arrow points on the dotted lines indicate the distribution of channels to meet the traffic requirements.

[1]) D. Goedhart and J. de Jong, Carrier-wave telephony, Philips Techn. Rev. **6**, 325-333, 1941.

[2]) G. H. Bast, D. Goedhart and J. F. Schouten, A 48-channel carrier telephone system, Parts I and II, Philips Techn. Rev. **9**, 161-170, 1947, and **10**, 353-362, 1948/1949.

[3]) An article on these subjects will be published in one of the forthcoming issues of this journal.

[4]) Telephone administrations usually round off the number of channels required to multiples of 12, 48 (as here) or 60, according to the transmission system under consideration.

In this way a reasonable approximation to the ideal arrangement is obtained, in which full use is made of all the available circuits on all sections of the network, as shown in *fig. 1*. When, however, coaxial cables are used, with say 600 channels on each pair of coaxial tubes, it is obvious that such a simple method can no longer be employed. Blocks of 600 channels are generally too large for approximating the traffic requirements between all switching centres.

Before proceeding to discuss the transfer problem on coaxial systems, the methods of modulation used with such system will be summarised, since this explanation is required in order that the technical problems associated with the transfer of small blocks of channels may be fully appreciated.

Modulation processes used on coaxial systems

The modulation processes used on coaxial systems are shown briefly in diagrammatic form in *figs. 2* and *3*. Fig. 2 is a simplified block diagram of the coaxial carrier equipment in a terminal station, whilst fig. 3 shows the various stages of modulation in the form of a frequency allocation diagram.

The subscriber is connected at point *I*. The speech frequencies lying in the band 0-4 kc/s are translated, by means of modulators and their associated filters, to the band of, say, 104-108 kc/s. This channel and 11 others — these latter lying in the bands 100-104, 96-100 kc/s etc. — are combined together at point *II* to form a group of 12 channels, which is thereafter treated as a unit.

The group is next transposed from the basic frequency range 60-108 kc/s [5]) by means of a second stage of modulation, say to the frequency band 312-360 kc/s. This group is then combined

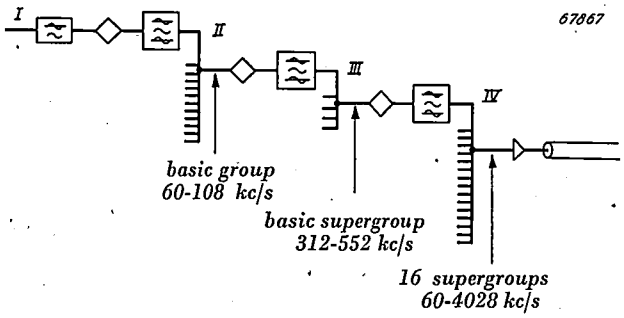


Fig. 2. Simplified block diagram of the modulation equipment in a coaxial carrier system. The subscriber is connected via the exchange at *I*. Twelve voice frequency channels modulate carriers and combine at *II* to form a group in basic position (60-108 kc/s). Five groups modulate five different carriers and combine together at *III* to form a supergroup in basic position (312-552 kc/s). Ten or sixteen (future developments may increase this number) supergroups each modulate different carriers and combine at *IV*. This combination of supergroups is then transmitted via the outgoing coaxial tube. low-pass filter, band-pass filter, modulator, and amplifier.

[5]) The position of the basic groups has been fixed by international agreement, in accordance with the recommendation of the Comité Consultatif International de Téléphonie (C.C.I.F.) that it should occupy either the band 60-108 kc/s or the band 12-60 kc/s, for international, and if practicable, for national traffic. The system described in this article uses a basic group occupying the 60-108 kc/s position, but all the considerations given here apply mutatis mutandis to a system using the other basic group. The division into groups in general, and also the division into supergroups mentioned later in the text, are both based upon C.C.I.F. requirements.

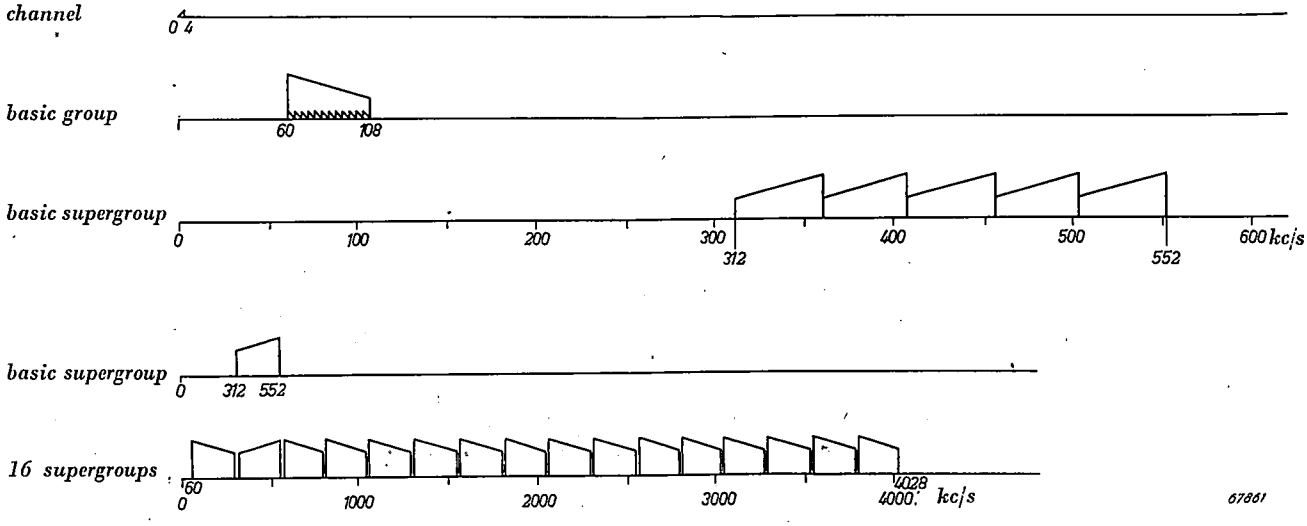


Fig. 3. Frequency allocation diagram showing the modulation processes of fig. 2. The first three stages show the positions occupied in the frequency spectrum by a voice frequency channel, a group of twelve channels in basic positions and a supergroup of five groups in basic position. The last two lines show, on a reduced frequency scale, the supergroup in basic position, and the combination of 16 supergroups sent to line. The ascending lines indicate the direction of increasing voice frequency in the channels, so that it can be seen for each case whether the upper or lower sideband has been chosen.

at *III* with four other groups, which have been translated to the bands 360-408, 408-456 kc/s etc. This combination of five groups is called a supergroup and is also subsequently treated as a unit.

The supergroup is translated from its basic frequency range (312-552 kc/s), by means of a third stage of modulation, to the frequency range it must ultimately occupy on the coaxial line. At *IV* several supergroups are combined together, say 10 or 16, and, after amplification, are transmitted to the coaxial cable. The final positions occupied by the supergroups on the line are such that there is a space of, in most cases, 8 kc/s — the width of two channels — between adjacent supergroups. The reason for this space will be explained later in the text.

The through group filter

Let us now return to the transfer problem: taking a relatively simple case of the general type shown in fig. 1, assume that there are only 12 transit circuits to be extended from a cable *AB* to a cable *BC*.

It can be seen from this diagram that the 12 channels (which form a complete group on both cables *AB* an *BC*) occupy the same frequency range at both points *III* and *IV*. It might be asked whether the points *III* and *IV* could be directly connected together without the modulation equipment shown.

The impracticability of this may be seen from *fig. 5*: The group filter (GF_4 in fig. 4), which selects the transit group, in this case group 4, from the supergroup concerned, has an attenuation/frequency characteristic as shown in fig. 5*b*. The filter not only passes group 4 unattenuated, but also passes parts of the adjacent groups 3 and 5 with varying degrees of attenuation. If the points *III* and *IV* were directly connected together, then those parts of the adjacent groups 3 and 5 passed by the group filter would leave section AB and reach the junction of *III* and *IV*, on either side of the basic frequency range. After remodulation of the group containing the transit channels, say to the position of group 3′ in another supergroup, for transmission to *C*, the vestiges of the previously adjacent groups are further

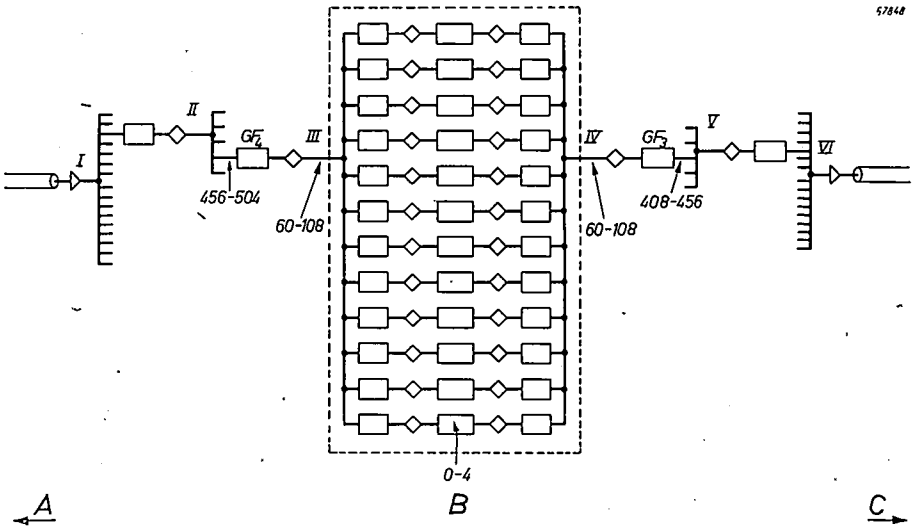


Fig. 4. Schematic block diagram showing how the fourth group of the third supergroup from station *A* is connected through at station *B* to station *C* as the third group of the fifth supergroup, when all the other groups terminate at *B*. If the "through group" is completely demodulated and remodulated, all the equipment between *II* and *V* is required.

One obvious solution is for these 12 incoming channels, as well as all other channels terminating in *B*, to be demodulated there into separate audio frequency channels; the transit channels may then be remodulated for transmission to *C*. At *B*, therefore, there must be two complete sets of terminal equipment, arranged back to back. *Fig. 4* shows, in block schematic form, the equipment required for the 12 through circuits.

attenuated by the group filter GF_3, but are still at a relatively high level at the point *V*, where they will fall within the frequency bands of groups 2′ and 4′ of the supergroup formed at *B* (see figs. 5*g* and *h*). Conversations in a number of *A*-*B* channels may therefore be overheard on *B*-*C* channels, at such a level as to be intolerable.

The above effect can be attributed to the shape of the filter characteristic in the transition band,

i.e. the region between the pass band and the stop band. If these filters had a characteristic more nearly resembling that of an ideal filter (small constant attenuation over the pass range, rising steeply to a high value of attenuation in the attenuating range) then points *III* and *IV* could

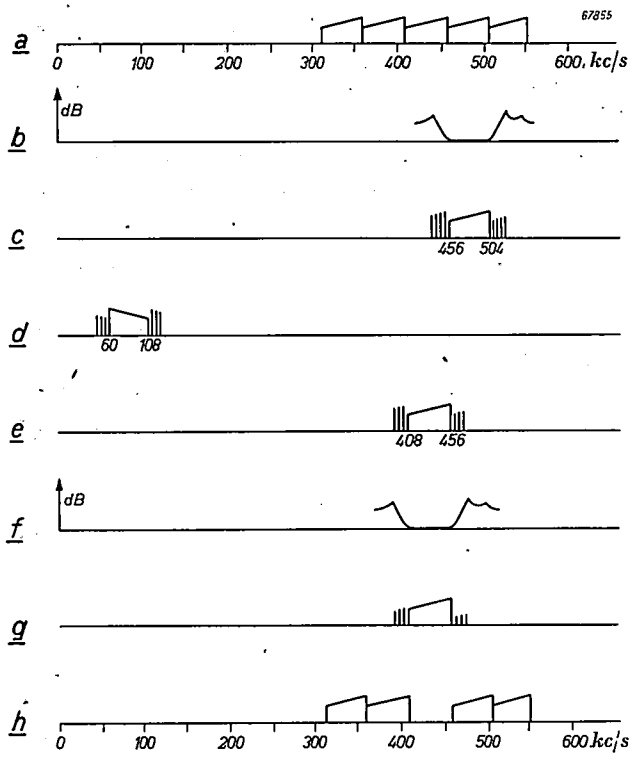


Fig. 5. This diagram illustrates the impracticability of connecting points *III* and *IV* of fig. 4 directly together. *a*) The supergroup at point *II*. *b*) the attenuation/frequency characteristic of the group filter GF_4. *c*) Group 4 after traversing the group filter GE_4. The vertical lines on each side of this group denote vestiges of the adjacent groups (3 and 5) which pass through GF_4. *d*) The through group demodulated to the basic position at point *III* of fig. 4 and, if both are connected, also at point *IV*. *e*) The group at the input of the group filter GF_3. *f*) Attenuation/frequency characteristic of the group filter GF_3. *g*) The group, transferred as group 3, at point *V*, is combined with four other outgoing groups (*h*), to form a supergroup.

be connected direct together and the intermediate apparatus (which ensures that the transit group is free from vestiges of other groups) could be omitted.

As will be seen later, it is not a practical proposition to make the characteristics of all the group filters conform to such a high standard of performance. One of the principal advantages resulting from the use of a multiple modulation system is, in fact, the relative simplicity of the group filters.

It is possible, however, in the case of a transit group, to supplement the effect of the two group filters by means of a special filter, having a characteristic more closely approximating to the ideal. If this latter filter has a pass range of 60-108 kc/s, it can be connected direct between points *III* and *IV*, in place of the equipment represented by that part of fig. 4 enclosed within the dotted line. Since every group required for transit connection would be demodulated to occupy the basic frequency range 60-108 kc/s, only one type of this filter would be required.

Such "through group filters" have been constructed [6]). Although it is impracticable to construct an ideal filter, it is possible to make a filter which approximates to the ideal; but the difficulties increase in proportion to the mid-band frequency at which the filter must operate, the ratio of the pass band to the transition band and the dissipation factor of the coils and capacitors. In the case in question the requirements are rather severe. The pass band is 48 kc/s wide, whereas the transition range between the pass and attenuating regions is only 0.7 kc/s. This figure of 0.7 kc/s is the same as that for the channel filters, which only pass a band nominally 4 kc/s wide and the majority of which work at relatively low frequencies. The attenuation required in the stop bands of the 48 kc/s wide filter is at least 70 db.

It would not be practicable, with components or techniques available at present, to construct a group filter answering these requirements for operation in any of the five frequency bands 312-360, 360-408, 408-456, 456-504, or 504-552 kc/s. Filters having a pass band corresponding to the basic group of 60-108 kc/s, with steep transition characteristics, can, however, be made. For this purpose filters with quartz crystals are employed. These filters will be dealt with presently, but it will be useful first to extend the line of thought developed.

Through supergroup filter

When it is desired to transfer at station *B* more than one group, say two or three groups, two or three through group filters, respectively, are used. When however it is desired to connect through a complete supergroup, the possibility of omitting some of the modulation equipment has again to be considered. From *fig. 6* it can be seen that at points *II* and *V* the supergroups all occupy the basic supergroup band of 312-552 kc/s. As in the case of the group filter already mentioned, the supergroup filter *SF* will pass portions of the adjacent supergroups with little attenuation, preventing thereby the point *II* from being connected direct to point *V*. There is, however, as before,

[6]) C. F. Floyd, Journal of the Institution of Post Office Electrical Engineers **15**, 57-62, 1947, (July); E. F. S. Clarke and T. F. Reed, ibidem **15**, 109-114, 1947 (October).

the possibility, at least in theory, of supplementing the normal supergroup filter *SF* by means of a special "through supergroup-filter". Such a filter would replace all the apparatus contained

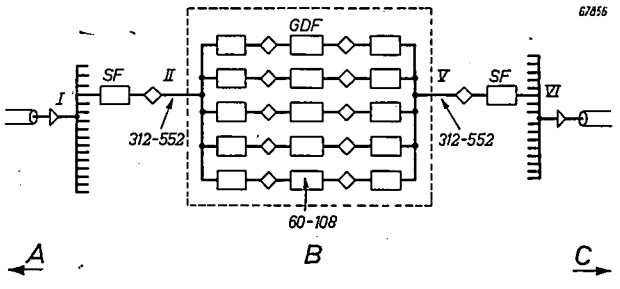


Fig. 6. A complete supergroup (No. 4) connected by means of five through group filters *GDF*, supplementing the action of the supergroup filter *SF*. One single filter, however, in place of the equipment within the dotted line, can serve the same purpose.

within the dotted line on the block diagram of fig. 6. The action of such a filter is shown in the form of a frequency allocation chart in *fig. 7*.

The design of a through supergroup filter, fulfilling all the necessary requirements, is a difficult task. It is true that a much wider gap is available

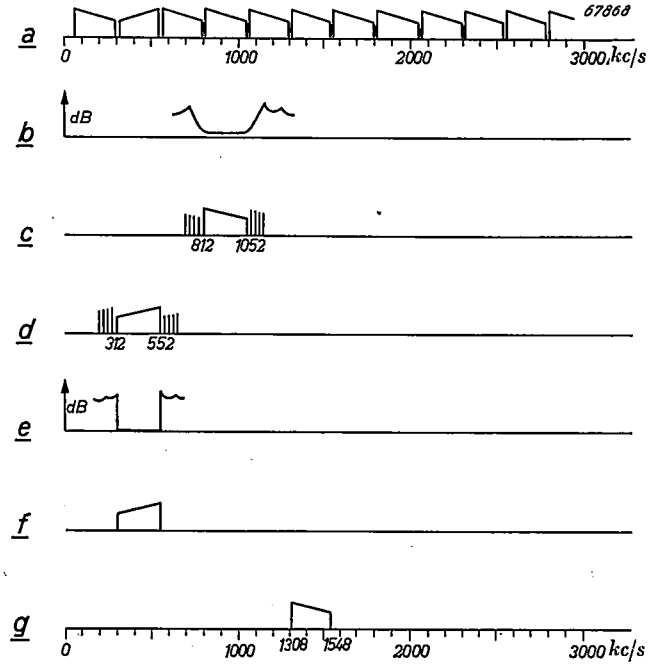


Fig. 7. A supergroup (No. 4) connected through by means of a through supergroup filter. *a*) Incoming supergroups at point *I* of fig. 6. *b*) Attenuation/frequency characteristic of the supergroup filter *SF* for the fourth supergroup. *c*) The "through" supergroup extracted by *SF* with vestiges of the adjacent supergroups. *d*) The through supergroup modulated to the basic position (point *II* in fig. 6). *e*) The attenuation/frequency characteristic required for a through supergroup filter. *f*) The through supergroup at point *V* with the vestiges of adjacent supergroups removed. *g*) The through supergroup, transposed to its new position on the outgoing coaxial cable (point *VI*).

between adjacent supergroups for the transition bands between the pass and attenuating regions, than in the case of the through group filter. (This gap, which is 12 kc/s on either side of the supergroup in the basic position and 8 kc/s between all other supergroups, was in fact left as a provision for the use of through supergroup filters.) Offsetting this advantage, however, is the fact that the filter must be capable of passing a band 240 kc/s wide as well as working at the relatively high frequencies of 312-552 kc/s.

Moreover, through supergroup filters have to fulfil other requirements which can best be summarised from a British Post Office specification.

The insertion loss, i.e. the attenuation caused by the insertion of a filter between a voltage source and a consumer, in the pass band may not vary, as a function of frequency, by more than $\pm$ 0.25 dB, over the whole of the pass band; within the frequency range of each group (48 kc/s wide) this variation may not even exceed $\pm$ 0.125 dB. The discrimination (i.e. the difference between the attenuation in the pass and suppression regions) has to be at least equal to the values shown in *fig. 8*. At 308 kc/s and 556 kc/s in the transition regions, the discrimination must be at least 40 dB. When two sinusoidal signals, having frequencies lying in the pass band, are applied at levels of 30 dB below a milliwatt to the input of the filter, the filter being terminated at both input and output in 75 ohms, then the intermodulation products due to non-linearity of the filter elements must be at least 75 dB below the output level of either of the two signals.

The underlying reasons for these British Post Office requirements are as follows.

The high discrimination at 308 kc/s and 556 kc/s is required because line pilots are transmitted in the gaps between some supergroups for purposes of test and regulation [7]). Such line pilots — sinusoidal signals whose input level and frequency are accurately maintained — are transmitted over the whole distance from the sending to receiving terminal stations, i.e. from the modulating equipment at the sending terminal to the equipment in the receiving terminal where the supergroups are separated by demodulation. After demodulation of a supergroup any adjacent line pilot will appear as one of the frequencies 308 kc/s or 556 kc/s. When a supergroup having an adjacent line pilot is transferred by the through supergroup filter method,

[7]) This is another reason for leaving an 8 kc/s gap between adjacent supergroups.

these frequencies of 308 kc/s or 556 kc/s might coincide, after remodulation, with one of the line pilots used on the outgoing cable. Consequently, the incoming pilot after demodulation has to be sufficiently attenuated in the through supergroup filter to avoid any such leakage which might ensue and cause an intolerable error in, or the possibility of beats with the outgoing pilot.

The requirements with respect to the attenuation variation in the pass band are of greater importance for the present consideration. The reasons for these requirements are based on considerations somewhat different from those already discussed in connection with the cable network of fig. 1.

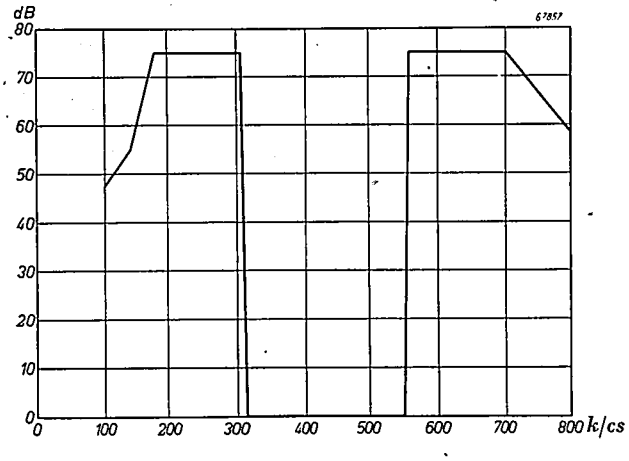


Fig. 8. Minimum discrimination to be given by the through supergroup filter outside the band-pass range (312-552 kc/s).

In this network an example was given in which the majority of the circuits were carrying traffic from B to A (720) and from B to C (384), whilst a relatively small number of transit circuits were connected through from A to C (192). Now let us consider an example of the opposite nature. Suppose that in fig. 1, instead of the numbers of channels stated, the majority of the traffic is between A and C, and that only a few channels, say about 60, are required for traffic from B to A, and a similar number for B to C. B is thus considered as a minor intermediate station on the route AC. Under these circumstances, it would be preferable to connect the cable AB to the cable BC direct, and to extract at B the one supergroup required (*fig. 9*). In order that the same frequency band can be used for this supergroup in both sections of the route, a band suppression filter, having a stop band in the relevant frequency range, must be connected in the position shown by the dotted block in fig. 9 [8]). Such a filter would allow the frequency bands used for transit channels to pass with little or no attenuation. The C.C.I.F. requirements in respect of supergroup spacing are, however, such that a filter of this type cannot be made for the majority of the supergroups [9]).

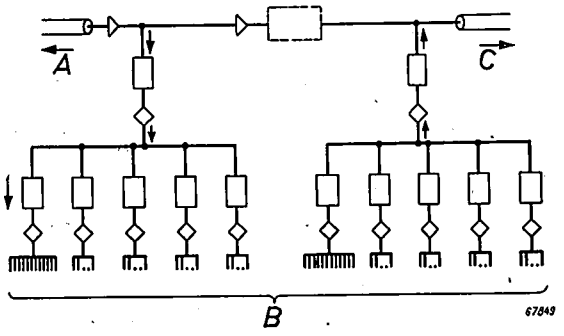


Fig. 9. When at station B all channels except one supergroup have to be connected through, it would be advantageous to join the incoming and outgoing coaxial cables, via a supergroup stop filter (shown dotted). It is not yet feasible, however, to make such a filter for other supergroups than Nos. 1 and 2.

One not altogether attractive but practicable solution to this problem is to demodulate all the supergroups at B, so that they occupy the basic supergroup frequency band of 312-552 kc/s. All the supergroups carrying transit traffic, i.e. all except the A-B supergroup to be extracted, are passed through separate through supergroup filters, and then remodulated to their former positions in the frequency spectrum. The same frequency band is used for the A-B and B-C supergroups; it is obvious that circuits are not connected via a through supergroup filter but terminate at B in two separate sets of terminal equipment (see *fig. 10*).

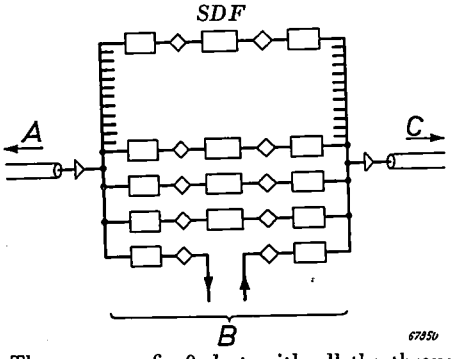


Fig. 10. The same as fig. 9, but with all the through supergroups routed via through supergroup filters *SDF*.

[8]) There are some other possibilities, such as the practice of "branching" or "echelon working" (to borrow a term from Baudot telegraphy) in which the frequency range of the channels that are extracted at an intermediate point is not used again for any other channel for the remaining part of the coaxial tube. Such solutions will, however, be left out of consideration here, as stated in the beginning of the article.

[9]) One might suggest to translate the frequency band of the complete set of supergroups on the coaxial tube bodily, so that a given supergroup may be brought into a frequency range such that a band stop filter may be constructed. For technical reasons, however, which are irrelevant to this paper, this is not practicable.

Thus at a station where some of the channels have to be extracted, the majority of the circuits pass through the station via supergroup modulators and through supergroup filters. On coaxial backbone routes, therefore, on which there are a number of minor intermediate stations or small branches at each of which a few circuits have to be terminated, the important long-distance circuits have to pass through a succession of through supergroup filters.

A multiplicity of through supergroup filters and supergroup translation equipment will also occur in coaxial systems which are arranged to give alternative routing, by means of a triangulated network of cables, between towns of fairly uniform size.

From this explanation it will be seen that the severe requirements for through supergroup filters arise partly from their use in cases where they are not the logical solution of the problem, but where they have to be used because no other solution is available.

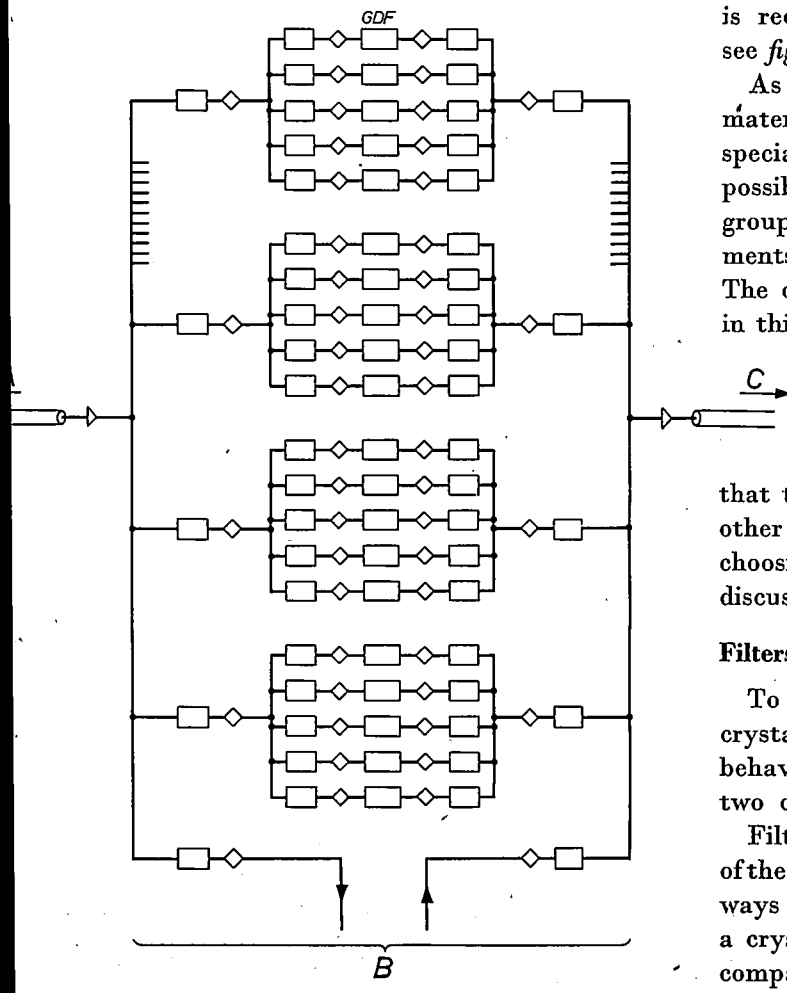


Fig. 11. The method which had to be employed, using through group filters (*GDF*), before a through supergroup filter was available. This method requires about five times the equipment necessary with the through supergroup filter method shown in fig. 10.

Because variations of attenuation in the pass bands of a number of similar filters in tandem are additive, it will readily be understood why the variations of attenuation in the pass range of the through supergroup filter must be kept within such narrow limits. Intermodulation in the filters is also cumulative (though in a different way), which further justifies the corresponding requirements of the specification.

As far as is known no through supergroup filter has hitherto been made which satisfies the above specification in respect of this variation of attenuation in the pass band. When a supergroup had to be extracted at an intermediate point, the only solution available in the past, if waste of frequency band (and revenue) were not permissible, was to demodulate the entire traffic on the coaxial cable down to basic groups. The C.C.I.F. contemplated the necessity of this operation at several points along the length of a typical long distance circuit. This method requires about five times the apparatus that is required for the through supergroup method, see *fig. 11*.

As a result of the use of Ferroxcube as a core material for filter coils, and of the application of special methods of computation, it has now become possible to design and construct a through supergroup filter which completely fulfils the requirements of the above mentioned B.P.O. specification. The construction of such a filter will be discussed in this paper in general terms.

In order to fulfil the requirement that a through supergroup filter must have a characteristic approximating as closely as possible to the ideal, it might be thought that the best solution would be to use quartz or other piezo-electric crystals. The reasons for choosing an alternative solution will now be discussed.

Filters with quartz crystals

To a first approximation a piezo-electric quartz crystal placed between two electrodes, P and Q, behaves electrically as a combination of a coil and two capacitors connected as shown in *fig. 12*.

Filter networks, having sections containing units of the type shown in fig. 12, can be designed in various ways and every such unit can then be realised by a crystal. Such crystal filters have the advantage, compared with those using coils and capacitors, that the elements in the equivalent circuit of the crystal have only small loss angles (Q of between 1000 and 300,000), so that a close approximation to the ideal characteristic can be obtained. Further-

more, if the crystals are properly ground, the equivalent inductance and capacitances are practically independent of changes of temperature. Because of these advantages there is an important field of application for crystals in the manufacture of filters, despite their higher cost and more fragile nature compared with the more usual coil and capacitor combination.

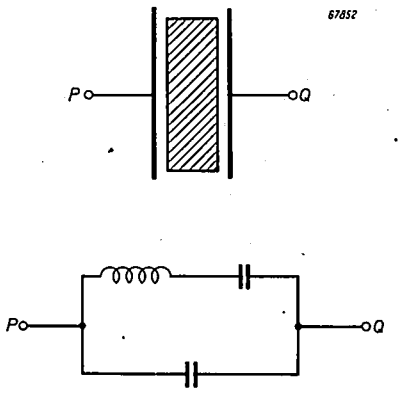


Fig. 12. A crystal clamped between two electrodes behaves electrically as a combination of an inductance with two capacitances as shown above.

Several types of through group filters have been constructed using quartz crystals. In these filters, as is generally the case with filters required to have a relatively wide pass band, a difficulty peculiar to the use of quartz crystals has to be overcome. A crystal slab can vibrate in more than one mode and it is not always possible to suppress undesired modes. If such modes occur, the simple circuit of fig. 12 no longer holds, and an equivalent circuit composed of many more elements is in fact obtained. Each undesired mode of vibration within the pass band gives rise to a narrow, sharp peak of attenuation at a particular frequency. It is extremely difficult to eliminate these peaks from the pass band when crystals are used in wide-band filters.

Some of the through group filters at present in use have this disadvantage, which, however, has been partially overcome by arranging that the undesired peaks of attenuation lie centrally in the space between adjacent channels of the group. Between two adjacent channels of a group there is an unused frequency band of 700 c/s, and the unwanted peaks of attenuation, being narrower than 600 c/s, do not cause any interference.

This artifice cannot, however, be used in through supergroup filters, because the prescribed pass band is much wider and at a higher frequency; the undesired attenuation peaks of the crystal under these circumstances are also much wider, with the result that some of the channels in a supergroup could not be used. The gaps which would have to be left, to avoid the resonance peaks, would conflict with the purpose for which the through supergroup filter is intended, namely to preserve the full flexibility of the system, without loss of channels and the practical complications arising therefrom.

Design of the through supergroup filter

From the above it is seen that until a few years ago it was extremely difficult to meet the requirements for a through supergroup filter, because of the undesired resonances when using crystals, and the excessive losses when using coils and capacitors. The use of the new magnetic ceramic material Ferroxcube for the cores of filter coils, however, has now provided a very satisfactory solution to the problem of building a coil-and-capacitor through supergroup filter. Coils having a Q of 350 at 550 kc/s can easily be manufactured using this core material. The best Q obtainable with coils of comparable dimensions in the past was 200. The loss angle of normal mica capacitors is small enough for use in through supergroup filters.

It was necessary to use the frequency parameter method, rather than the more usual image impedance or stencil methods, in designing the filter. A detailed discussion of these methods will not be attempted in this paper [10]), but it should

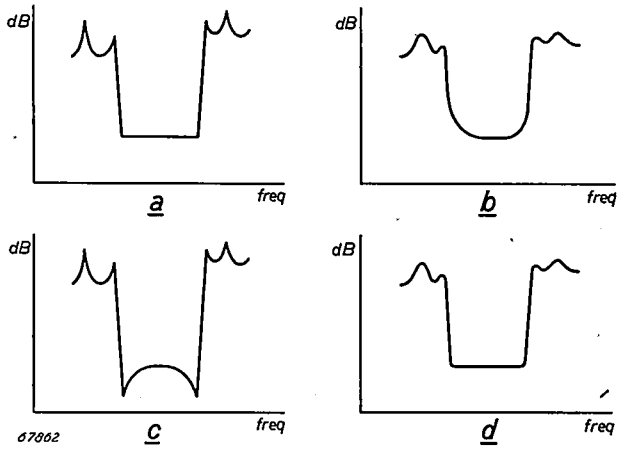


Fig. 13 *a*) Attenuation/frequency characteristic of a filter, designed by the image impedance method, with ideal, loss-free, coils and capacitors, and terminated by the image impedance. *b*) As (*a*) showing the effect of using practical components and terminating with a constant resistance. *c*) Attenuation/frequency characteristic of a filter designed by the frequency parameter method showing over-equalisation obtained with loss-free components. *d*) As (*c*) showing the effect of using practical components. Note the sharp corners at the edges of the pass-band range.

[10]) An explanation of these methods with particular reference to the calculations for the through supergroup filter dealt with in this article, will appear in Philips Research Reports.

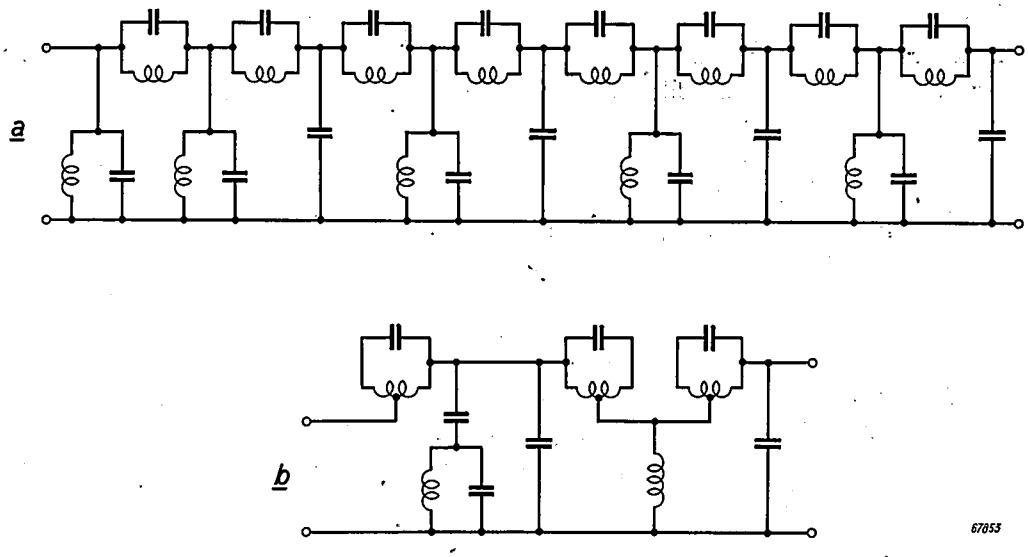


Fig. 14. *a*) Circuit diagram of half of the through supergroup filter. *b*) Circuit diagram of the supplementary pilot stop filter.

be noted that the latter two methods result in a filter in which there are no abrupt changes of attenuation, i.e. the change of attenuation from the pass band to the transition region is represented as a smooth curve on the attenuation/frequency characteristic, and covers a relatively wide frequency range. A filter can be designed, using the image impedance or stencil methods, having sharp changes of attenuation, but only at the cost of using a large number of coils and capacitors (see *fig. 13a* and *b)*. It is possible, however, to design a filter, using the frequency parameter method, and assuming loss-free components, having a characteristic in which there is over-equalisation near the edges of the band (fig. 13*c*). It can then be arranged to use this excess equalisation so that with components having predetermined loss angles, the pass-band characteristic becomes flat, with sharp corners (fig. 13*d*).

In the course of the filter calculations for this specific case, the attenuation requirements were distributed over three correlated filter networks, as follows.

Disregarding initially the required discrimination at the pilot frequency of 556 kc/s, a through supergroup filter was designed having the desired pass band of 312-552 kc/s. The variation of the attenuation over the whole of the pass band of this filter was not greater than $\pm$ 0.125 dB, whereas the discrimination between the pass and attenuation bands was 40 dB. A schematic representation of this filter is shown in *fig. 14a* and its measured attenuation/frequency characteristic in *fig. 15*.

Two of these networks in cascade, separated

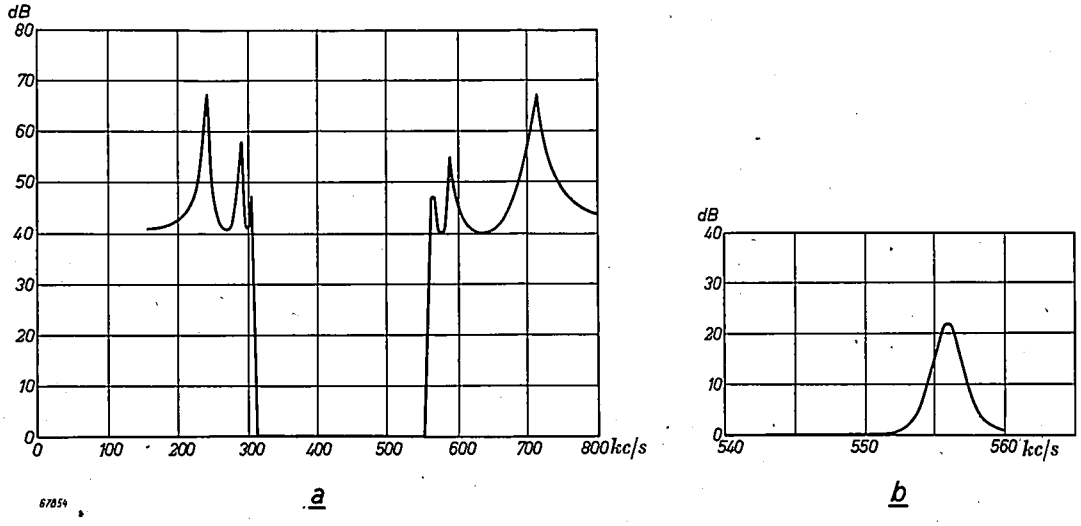


Fig. 15 *a*) Attenuation/frequency characteristic of half of the through supergroup filter. In this figure and in fig. 16 the attenuation in the pass band has been taken equal to 0 *dB*, see also fig. 20. *b*) Attenuation/frequency characteristic of the pilot stop filter.

either by an amplifier [11]) or by a resistance pad, in order to avoid interaction loss caused by the somewhat irregular impedance of the filters, fulfil all the requirements for a through supergroup filter except for the discrimination in the vicinity of 556 kc/s, which is about 20 dB too small. This requirement is met by an additional pilot stop filter which has a constant attenuation in the 312-552 kc/s range, but has a peak of attenuation of about 20 dB at 556 kc/s. The circuit diagram and attenuation/frequency characteristic of this filter are shown in figs. 14*b* and 15*b* respectively. The fully drawn curve of *fig. 16* shows the attenuation/frequency characteristic of the complete filter, whilst the dotted curve on the same figure shows the minimum specification requirements (see fig. 8), for purposes of comparison.

During further development it became apparent at an early stage that a choice had to be made for

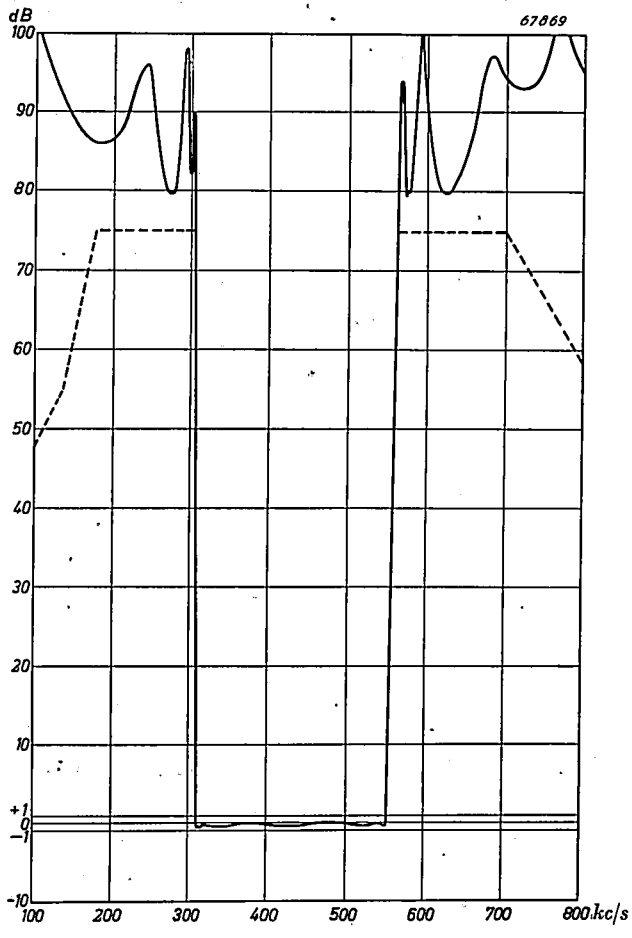


Fig. 16. Attenuation/frequency characteristic of the complete through supergroup filter. The dotted lines indicate the required minimum discrimination, see fig. 8. The specified minimum attenuation in the stop-bands is exceeded by 5 dB or more at all frequencies except in a very narrow frequency range (hardly perceptible on this diagram, because of the scale), from 560-561 kc/s.

[11]) By means of this and other amplifier stages incorporated in the filter, it is also possible to arrange for the "filter" to have a gain (negative pass-band attenuation).

practical reasons between sacrificing about 10 dB of the discrimination in the narrow range between 560 kc/s and 561 kc/s or introducing a further stage of amplification (or alternatively a special transformer). For the filter described in this paper the

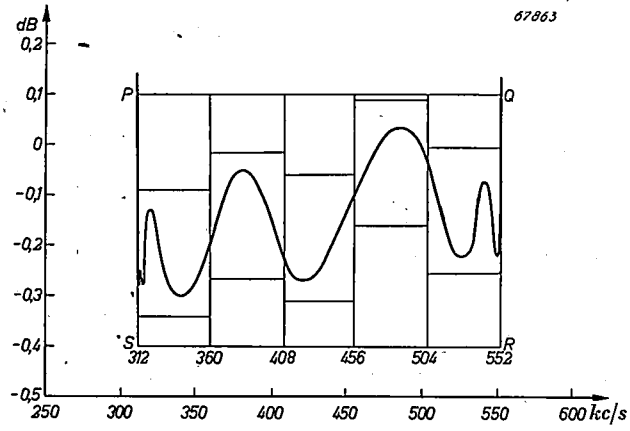


Fig. 17. Attenuation of the through supergroup filter in the pass band. The attenuation scale has been enlarged, so that the variations in attenuation can be seen clearly. The diagram shows that the attenuation variation as a function of frequency remains well within the specified limits, both for the whole supergroup (rectangle *PQRS*) and for each of the groups separately (5 smaller rectangles.)

first solution was chosen, with the result that in the 560-561 kc/s band the requirements of the specification are not fully satisfied [12]). In the rest of the attenuation range however, the requirements of the specification are everywhere exceeded by at least 5 dB.

In order to show how the requirements for the pass-band range have been met, this part of the attenuation/frequency characteristic has been reproduced in *fig. 17* with a greatly enlarged attenuation scale. From this figure it can be seen that the total variation of attenuation in the pass band does not exceed 0.5 dB, since the curve lies wholly within the rectangle *PQRS*. The filter thus fulfils the pass-band requirements. It can also be seen from this figure that the total attenuation variation in each of the five groups (the small rectangles) does not exceed 0.25 dB and thus the requirement that the attenuation should not vary by more than $\pm$ 0.125 dB in any one group is satisfied.

In addition, it may be mentioned at this point that the intermodulation products are all at least 80 dB below the output level of the original signals when these latter are applied at an input level of 30 dB below 1 mW.

[12]) As a consequence, signals of audio frequencies 0-1000 c/s of the channel 560-564 kc/s of the adjacent supergroup will be able to penetrate without sufficient attenuation into a corresponding channel in the outgoing cable. This limited crosstalk, which is worst in the vicinity of 500 c/s, is not very important. If considered necessary, the other alternative might still be adopted.

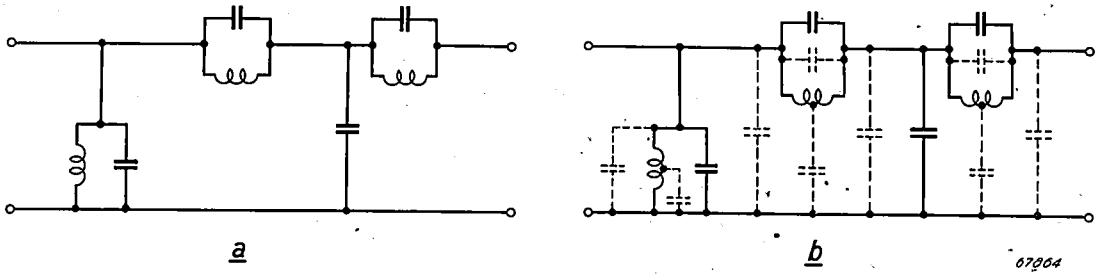


Fig. 18 *a*) Schematic diagram of a typical section of a filter. *b*) The same as (*a*) but including the stray capacitances.

Owing to the closeness with which the impedances of the network had to approximate to the calculated values, and in addition the relatively high working frequencies, the stray capacitances of the wiring, windings etc., had to be taken into account in the design. *Figs. 18a* and *b* show a typical network with and without the most important stray capacitances. The stray capacitances shown are (1) self capacitances of the coils, (2) capacitance between coils or capacitors and earth, and (3) capacitance between wiring junctions and earth.

The way in which all such stray capacitances have been reduced to values which can be tolerated, is in principle quite simple. From the number of possible alternative configurations in which the various circuit units in a filter can be built up, a circuit design has been chosen in which all stray capacitances appear across theoretical capacitances. The value of the actual mica capacitor is then so chosen that the combined capacitance is equal to the theoretical design value. The design of the filter must obviously be such that the theoretical capacitances required are in all cases higher than the stray capacitances. In the three filters mentioned above, therefore, it is necessary to have a capacitor in parallel with each inductance (case 1) and in addition a capacitor between each wiring junction and earth (case 3). From fig. 14 it can be seen that these requirements have been met [13]). The requirement that there should be a capacitor in parallel with the stray capacitances of each coil to earth (case 2), is met for coils which have one terminal connected to earth, if condition (1) is fulfilled. For those coils which have neither terminal earthed, the method normally used in high frequency measuring technique is employed. The coil is enclosed in a "live" i.e. not earthed screen connected to one of the coil terminals, *fig. 19*. The stray capacitance C_1 between the coil and the live screen can be regarded as part of the self-capacitance of the coil, and treated as for case (1). The capacitance between the live screen and earth can be regarded as being concentrated at the terminal to which the screen is connected and can therefore be treated as for case (3).

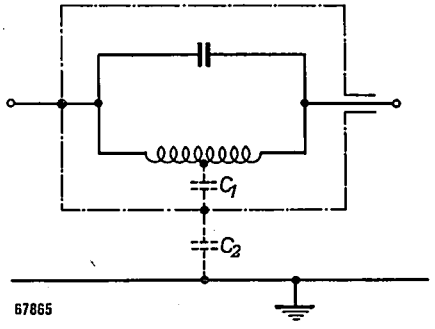


Fig. 19. When a coil is not earthed, the stray capacitances to earth are rendered innocuous by mounting the coil in a live screen connected to one of the coil terminals and dealing with the capacitances C_1 and C_2 separately as described in the text.

The filter panel [14])

Fig. 20 shows, in schematic form, the way in which the three filter networks are combined, together with amplifying stages and pads, to form the complete through supergroup filter.

The amplifying stages are identical, and each comprises a pair of EF 91 valves in parallel, in a circuit with negative feedback. Consequently, if one valve fails, the gain is reduced by only 0.5 dB; in that case the change of gain is uniform throughout the entire frequency band and the attenuation/frequency characteristic shown in fig. 15 is quite unaffected except for the change in datum value.

[13]) The coil in fig. 14*b*, which appears to be an exception to this statement, can be regarded as an approximation to a negative capacitance, which was found to be theoretically necessary at this point.

[14]) The construction of the prototype filter, electrically designed in the Physical Research Laboratory at Eindhoven was undertaken by Dr. Latimer and Messrs. Jackson and Snelling of the Mullard Electronic Products Ltd. London.

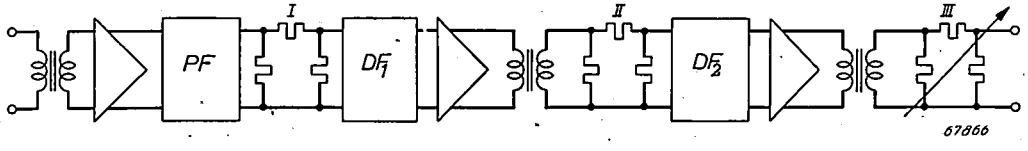


Fig. 20. Block diagram of the complete through supergroup filter. In addition to the two half filters DF_1 and DF_2, and the pilot stop filter PF, there are three amplifiers and three pads (I,II,III). The available gain of the amplifiers is such that the attenuation over the whole of the pass band can, if desired, be made negative. This satisfies a requirement of the specification, not previously mentioned, that it should be possible to adjust the attenuation in the pass band in steps of 1 dB from 7 ± 0.5 dB and from -2 ± 0.5 dB.

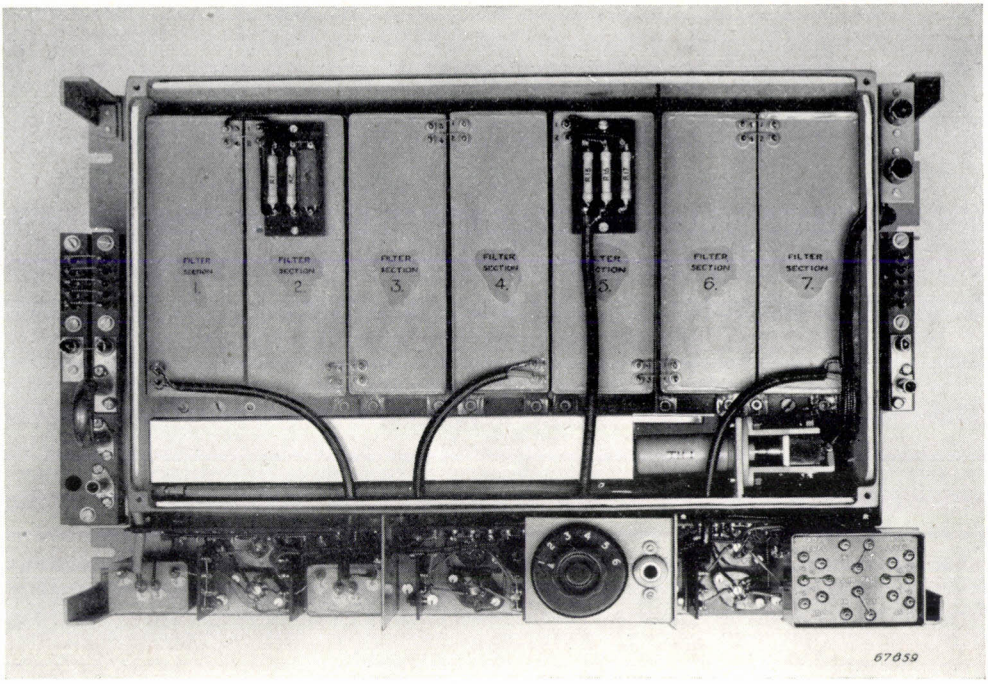


Fig. 21. The through supergroup filter, with dust cover and oven lid removed. The filter networks have been split up into seven convenient units, the mechanical construction and mounting of which is identical to that used for the filters of the 48-channel system described previously (see note [2]). Underneath these seven units can be seen the heater and thermostat assembly. Immediately below the oven are the amplifiers, transformers and pads. The variable pad (*III* in fig. 20) is at the extreme bottom right hand corner. The attenuation of this pad can be varied by altering the soldered connections.

The pad *I* has been introduced because both the pilot stop and band-pass filter sections have been calculated on the basis of termination with a purely ohmic resistance and should be terminated in practice with a fairly good approximation thereof, i.e. with a high return loss. Pad *II* has been introduced for the same reason, because the impedance looking back into the output terminals of the transformer of the preceding amplifier is not a sufficient approximation of an ohmic resistance. The total gain can be regulated over a range of 5 dB by means of the variable pad *III*.

Temperature variations cause variations in the attenuation/frequency characteristic. A variation of temperature from 10 °C to 40 °C causes a variation of attenuation of about 0.5 dB in the pass range of the filter, and unless the temperature were controlled, the limit of $\pm$ 0.125 dB allowed within a group would be exceeded. The filter networks are, therefore, enclosed in a simple thermostatically controlled oven, in which the temperature is kept at $45° \pm 2°$ C.

In conclusion, *figs. 21* and *22* show the complete filter, with and without dust cover. Some features, visible on these figures, are mentioned in the subscripts.

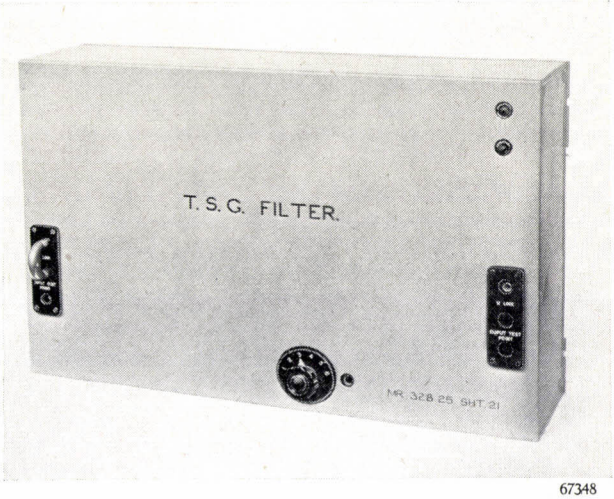


Fig. 22. Front view of the through supergroup filter, with dust cover in position. At the top right-hand corner, the two thermostat indicating lamps can be seen. One of these lamps lights up when the element is on and the other when it is off. The screen-grid current of any of the amplifying valves can be measured by selecting the particular valve required by means of the rotary switch at the bottom of the panel, and plugging a meter into the socket alongside the switch.

Summary. Coaxial carrier telephone systems are capable of transmitting 600 or 960 channels over two coaxial tubes (one "go" and one "return"). The carrier circuits are arranged in supergroups of 5 groups, each group comprising 12 channels.

In order to meet the needs of operational flexibility, facilities must be provided for the transit connection of one or more groups or supergroups at intermediate stations on a coaxial trunk network. In some cases it may even be necessary to connect all the channels through, except for one supergroup. One solution to this problem is to demodulate all the supergroups to the basic supergroup position of 312-552 kc/s. A through supergroup filter at this point permits transit working without further demodulation, and thus avoids the use of considerable additional modulation equipment. The specification which such a through supergroup filter has to meet is given, and is followed by a description of a prototype filter constructed using coils with Ferroxcube cores and mica capacitors. As a result of the very small losses of these components, and the application of the frequency parameter methods in the design of the filter, all the requirements of the specification have been met by a filter occupying a comparatively small space. (The use of quartz crystals for such a filter is not advisable on account of the risk of undesirable crystal resonances in the pass range.) The prototype through supergroup filter was built up from two identical band-pass filters and a supplementary pilot stop filter, separated by amplifiers and pads. The filter sections are housed in a temperature-controlled oven, to reduce the variation of attenuation with temperature.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1957: C. J. Bouwkamp: On Sommerfeld's surface wave (Phys. Rev. **80**, 294, 1950, No. 2).

A critical discussion of some recent papers by Kahan and Eckart on Sommerfeld's surface wave in electromagnetic wave propagation over a plane earth.

1958: J. M. Stevels: The relation between the dielectric losses and the composition of glass (J. Soc. Glass Technology **34**, 80-100, 1950, No. 158).

Discussion of the theory of dielectric losses in glass. Losses are expected due to (1) relaxation phenomena caused by the network modifiers, (2) relaxation phenomena caased by the network itself and (3) vibration of the constituents of the network and of the network modifiers. The dependence of each kind of losses on the frequency and the temperature is discussed. The theoretical considerations are compared with the experimental results and the dependence of the various kinds of losses on the composition is discussed.

1959: F. A. Kröger and J. van den Boomgaard: The fluorescence of magnesium germanate activated by manganese (J. Electrochem. Soc. **97**, 377-382, 1950, No. 11)

Magnesium germanate activated by manganese may contain the activator in different valencies. When this valency is higher than two, fluorescence excited by ultraviolet occurs in seven narrow bands decaying with the same time constant. The variation of this decay constant and the intensity of fluorescence with temperature is explained by the assumption that the relative contribution of radiative and non-radiative transitions is determined by a Boltzmann equilibrium regulating the population of various excited states.

1960: J. A. Haringx: The rigidity of corrugated diaphragms (Appl. sci. Res. A **2**, 299-325, 1950, No. 4).

When the corrugated diaphragm is replaced by a fictitious flat plate of similar properties it is possible to derive a linear differential equation for the deflection. The coefficients of this equation, however, vary in a complicated way and its solution for the pressure-loaded diaphragm is given only for thick and for thin sheets separately. For thick sheets the profile of the corrugation appears to be inessential, whereas for thin sheets it is necessary to distinguish between trapezoidal, triangular and arc-shaped corrugations. By an obvious device the results for thick and for thin sheets are fitted together, so that the deflection can be determined also for the intermediate range of medium sheet thickness. The final results of the present calculation are compared with measurements carried out by others and are found to be in satisfactory agreement with the experiments.

It is to be remarked that, compared on the basis of small deflections, the introduction of corrugations into the sheet leads to a considerable increase of rigidity of the diaphragm. The prevailing assertion that the flat plate is more rigid than the corrugated diaphragm holds only for large deflections, because of the non-linearity between the load and the deflection of the flat plate.

1961: J. L. Snoek and E. J. Haes: Laboratory apparatus for the production of a steady flow of very pure hydrogen (Appl. sci. Res. A **2**, 326-328, 1950, No. 4).

An apparatus is described, containing as a diffusor element a system of twelve thin-walled, nickel tubes in parallel, which, upon being heated to 950 °C and exposed to an excess pressure of about one atmosphere, yields a steady flow of about one quarter of a litre per minute of the purest hydrogen at atmospheric pressure. The apparatus is rugged and simple and has a lifetime of several hundreds of hours.

R 148: C. J. Bouwkamp: On Bethe's theory of diffraction by small holes (Philips Res. Rep. **5**, 321-332, 1950, No. 5).

In this paper on the diffraction of electromagnetic waves by a small circular hole in an infinite plane conducting screen, Bethe derives the diffracted field from fictitious magnetic charges and currents in the hole. A careful examination of Bethe's theory reveals that the electric field should be discontinuous in the hole, which is inconsistent. The reason is that Bethe's expression for the magnetic current density is erroneous, as is shown in the present paper for the simplest possible case, viz. the diffraction of a plane-polarized electromagnetic wave impinging on the hole in the normal direction. Though the error in the current density does not invalidate Bethe's formulae for the distant field, the correct electric field in and near the hole differs appreciably from that predicted by Bethe.

R 149: J. L. Meijering: Segregation in regular ternary solutions, I (Philips Res. Rep. **5**, 333-356, 1950, No. 5).

The segregation in ternary phases is examined from a thermodynamic point of view, the comments being confined to regular solutions. Pertinent information is already gained by analyzing the spinodal curves (boundaries between metastable and unstable concentration regions). The parameters a, b and c, defining the mutual energy interactions of the three component pairs, are varied independently of each other. Their different combinations are divided into eight characteristic categories. Taking the sequence $a \geqslant b \geqslant c$, closed miscibility gaps appear above $T = a/2R$ if $b-c>a>0$, and above $T=0$ if $\sqrt{-c}>\sqrt{-b}+\sqrt{-a}$. An expression is given for the ternary critical point in which such a gap vanishes on heating. Complicated spinodal curves are found when $b+c>a>0$.

R 150: H. C. Hamaker and Th. Hehenkamp: Minimum-cost transformers and chokes, I (Philips Res. Rep. **5**, 357-394, 1950, No. 5).

Three equations play a part in the theoretical design of a transformer: (1) the "price equation", specifying the price as a function of the geometrical dimensions, (2) the "power equation", expressing the apparent power in terms of the dimensions and of the magnetic-flux density and electric-current density, and (3) the "losse equation", giving the losses in terms of the same set of variables. Assuming the power and the losses prescribed, the problem is to find a minimum in the price as given by the first of these calculations, the other two serving as auxiliary conditions. This problem is solved and the solution is casted into a form ready for immediate application. The resulting tables with full instructions for use have been collected in the appendix. It turns out that the minimum-cost design depends on the ratio of power to losses, so that an increase of both in the same ratio does not alter the dimensions of the transformer, a conclusion unacceptable in practice. Frequently a design is not restricted by the losses, but rather by a permissible maximum of the temperature, or of the peak magnetic-flux density. It is discussed how such other characteristics can theoretically be brought into play. Readjustments lead to make-shift solutions which are adopted only by lack of a rigorous theory. Some light may be thrown on their practical value by a comparison with alternative designs in which transformer characteristics other than the losses have been prescribed.

R 151: P. Cornelius: Proposals and recommendations concerning definitions and units of electromagnetic quantities (modifications) (Philips Res. Rep. **5**, 395-396, 1950, No. 5).

Next to the symbol $j = JV$ for the magnetic dipole moment (in Wb·m) the symbol $m = MV$ for the magnetic area moment (in A·m²) is introduced and in analogy therewith, next to the symbol $p = PV$ (in C·m) for the electric dipole moment, a symbol $f = FV$ (in V·m²) for the "electric area moment" (V volume in m³, J magnetic polarisation in Wb/m², M magnetisation in A/m, P electric polarisation in C/m², F "electrisation" in V/m).

VOL. 13 No. 9, pp. 237-268 MARCH 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

DELTA MODULATION, A NEW MODULATION SYSTEM FOR TELECOMMUNICATION

by J. F. SCHOUTEN *), F. de JAGER and J. A. GREEFKES.

621.396.619.16

For the transmission of information such as speech or music the aim is to design a system possessing the optimum properties, having regard to the properties of the transmitting channel on the one hand and, on the other hand, the characteristics of the signal to be transmitted. Recently systems have been developed in which a quantization of the signal is applied both in amplitude and in time. Such a system is described in this article. Its essential feature is the employment of a special kind of negative feedback. Compared with other quantizing modulation systems the apparatus has been greatly simplified.

Signal and interference

A remarkable development is taking place in the technique of telecommunications, in that speech and music are being transmitted by means of a pattern of unit signals, such as already known in telegraphy.

The desire for such a drastic departure from the usual methods of transmission, such as amplitude modulation for instance, has arisen from the fact that it is becoming more and more difficult to ensure interference-free reception at very great distances. In modern carrier-telephony systems the signal transmitted by cable over a distance of 100 km is attenuated 240 db, i.e. a power ratio of 10^{-24} times. This attenuation, which is in the order of the ratio of the power emitted by the sun to that of a pocket-torch lamp, does not in itself constitute any difficulty: in transcontinental communications such attenuations may reach some tens of thousands of decibels, but they are compensated by means of a series of repeaters (relay stations) installed at intervals along the route. But, besides being attenuated, the signal is also distorted. The distortion may consist, for instance, in noise, crosstalk from neighbouring channels, and variations in the level of the signal. These effects, in themselves only small, act along the whole transmission circuit, so that the ultimate distortion of the signal received is a cumulation of very small effects.

When a table is covered with dust it can be cleaned by brushing the dust off, for we know what is the table and what is the dust. However, at the receiving end of the communication system referred to it is impossible to distinguish between signal and interference. This means that the "transmission dust" cannot be brushed off, and so with the usual systems of communication cumulation of interference is unavoidable.

There are, however, communication systems where the interference can indeed be distinguished from the signal, and with such systems, when intermediate repeater stations are used, a practically interference-free signal can be received even at very great distances. Before dealing with these systems we shall discuss some other modern methods of transmission which have, as it were, pointed the way to achieving the ultimate object of suppressing interference. These methods are illustrated in *fig. 1*.

How cumulation of interference is counteracted

Amplitude modulation on a carrier (fig. 1*b*) of the signal to be transmitted (fig. 1*a*) can be replaced by amplitude modulation on a series of equidistant pulses: pulse-amplitude modulation (fig. 1*c*). An audio or video signal is given as

*) Of Philips' Telecommunication Industries, Hilversum.

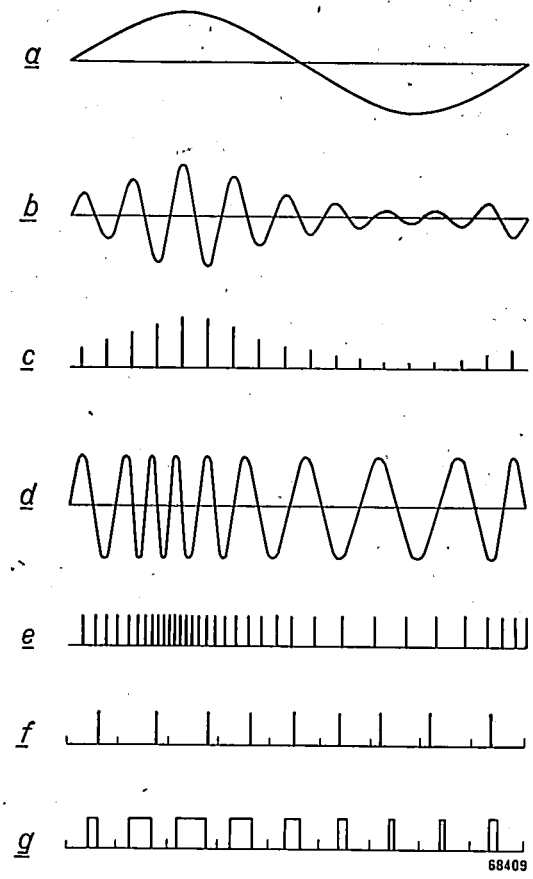


Fig. 1. Diagrammatic representation of various non-quantized modulation systems. *a*) The modulating signal. *b*) Carrier amplitude modulation. *c*) Pulse-amplitude modulation. *d*) Carrier frequency modulation. *e*) Pulse-frequency modulation. *f*) Pulse-position modulation (note the marked time intervals within which each pulse can move). *g*) Pulse-width modulation.

a continuous function; the time is the independent variable, while the dependent variable is the sound pressure, respectively a voltage at the pick-up. It appears that if the value of this voltage is known at equidistant instants the total signal, i.e. including all instantaneous values between those instants, can be wholly reconstructed, provided the frequency bandwidth of the signal is not greater than half the rate of sampling. Thus, the wider the range covered by the frequencies of the signal to be transmitted, the greater is the required sampling rate.

In pulse-amplitude modulation this principle is applied directly. In fact this method is analogous to carrier-amplitude modulation; only the continuous carrier is replaced by discrete pulses. For a frequency band of, say, 3400 c/s to be transmitted, a pulse-repetition frequency of about 8000 c/s is employed. As far as the separation of signal and interference is concerned this method does not, however, offer any advantage over carrier modulation.

Almost simultaneously with pulse-amplitude modulation, frequency modulation (fig. 1*d*) made its appearance in telecommunication technique. With this method it is not the amplitude but the frequency of the carrier that is modulated. Here, too, the continuous carrier can be replaced by a series of pulses, so that pulse-frequency modulation (fig. 1*e*) is obtained [1]). A special form of this is pulse-position modulation (fig. 1*f*), also called pulse-time modulation, whereby the signal is built up from pulses all having the same amplitude and width, called unit pulses. The information is contained in the time pattern of these pulses, such that each pulse occupies a certain place in the time interval allotted to it, according to the instantaneous value of the signal [2]). This offers an entirely new possibility of counteracting interference: if the effects of interference in the transmission channel are sufficiently limited so that they do not severely mutilate the unit pulses then the pattern is retained and can be restored by regenerating the partly mutilated pulses at the receiving end.

Still it is not possible to reach a complete separation between signal and interference, so that some interference remains and is cumulative. The interference in the transmission channel acts in the first instance on the amplitude of the signal transmitted. If the pulses were quite rectangular in shape complete separation of the signal from interference would be possible, provided the interference does not exceed a certain value. Since, however, the slope of the edges of the pulses always has a finite value, there is still some variation in the time at which a pulse is received, i.e. in the position of the pulse in its interval. The resultant interference cannot be separated from the signal and so there is again a cumulative effect.

There is much less interference with pulse-position modulation than with pulse-amplitude modulation, and there is much less interference in the case of carrier-frequency modulation than in that of carrier-amplitude modulation. Frequency modulation and pulse-position modulation have led to the fundamental conception that it is possible to reduce the effect of interference by increasing the frequency band of the high-frequency signal transmitted. This possibility, however, is not efficiently utilised

[1]) Pulse-frequency modulation is something that has existed in nature for millions of years, viz. in the transmission of signals via the nerves of animals. When a nerve is excited a series of electrical pulses of equal amplitude are propagated along the nerve; the number of pulses per second (up to some hundreds) is a measure for the strength of the stimulus.

[2]) Pulse-position modulation has been mentioned in this journal before; see C. J. H. A. Staal, An installation for multiplex-pulse modulation, Philips Techn. Rev. **11**, 133-144, 1949.

in pulse-position modulation, nor in its modified form of pulse-width modulation (based on the principle that the width of the pulses is varied according to the instantaneous value of the signal, fig. 1g).

Pulse-position modulation with the aid of rectangular pulses would, as already stated, give complete suppression of interference, but the frequency band required for undistorted transmission of these signals is infinitely large.

Quantization of the signal

For pulse-position modulation unit signals are employed. This implies a sort of quantization: only discrete values (in this case only the unit value) of the pulse amplitude are permitted. The information is determined, however, by a continuously variable pattern of these pulses, in the sense that any particular pulse may be at any point within the continuous time range allotted to it. One can now go a step further and transmit only discrete values — say, the multiples of a certain unit — of the continuous scale of instantaneous values of the signal. Something similar is regularly done in everday life when expressing length in kilometres, metres or millimetres, temperature in degrees or tenths of a degree, time in hours, minutes or seconds. The consideration behind this inaccurate quotation is the fact that in practice a limited accuracy, depending on the object in view, always appears to be sufficient.

The steps in which the instantaneous value of an audio signal is measured can likewise be chosen with a sufficient fineness to obtain a certain quality of reproduction. The loss in fidelity is offset by the advantage gained in being able to eliminate completely the active effects of interference, provided it is kept below a certain limit.

This can be illustrated with the following example, for which we shall revert to the method of pulse-position modulation. Complete quantization of this modulation means that in the time interval allotted to it each pulse can occur only at a discrete number of fixed places, as illustrated in *fig. 2*. If the interference brings about a displacement of the pulse positions which is less than half the difference in time between two of the discrete places, then, in spite of this interference, at the receiving end one knows exactly what pulse position, thus what signal, is intended. The interference can therefore be entirely separated from the signal and completely eliminated, provided it does not exceed a certain threshold value.

Direct application of the quantization process to this or to one of the other modulating systems mentioned thus gives the desired suppression of interference. But, as will be shown later, from the point of view of efficiency of the transmission system it is uneconomical to proceed in this way.

Special methods of modulation have therefore been developed for transmission of quantized signals. One of these methods is that of delta modulation, which will now be discussed.

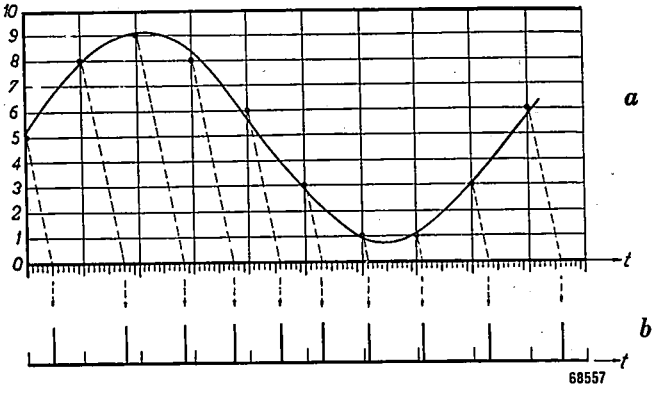


Fig. 2. Quantization of pulse-position modulation. *a*) The instantaneous signal value at the beginning of each interval is approximated by a quantized value. This is "translated" into the series of transmitted pulses (*b*) by a corresponding pulse position in the interval. In this diagram the quantization is carried out in 11 steps. Thus for each pulse there are 11 discrete positions available in its interval. This system is not applied in practice because it is not effective.

Delta modulation

Principle

Delta modulation is a pulse modulation system, pulses being produced at the sending end at equidistant instants. Not all pulses are, however, transmitted. The question whether a pulse is to be transmitted or not is decided in the following way. An "echelon curve" is formed from the pulses in a manner to be explained later. This echelon curve is compared with the modulating signal (*fig. 3b*). If at the moment when a new step has to be added the value of the modulating signal is greater than that of the step-shaped signal, then a positive step is added, but if the value of the modulating signal is smaller a negative step is formed. Thus the echelon curve oscillates and approximates to the modulating signal, being the quantized form thereof. The formation of a positive step is arranged to initiate the transmission of a pulse, whereas the formation of a negative step is not and results in a gap in transmission at that moment in time. What is then transmitted is the series of pulses shown in fig. 3*a*.

At the receiving end the step-shaped approximating signal is built up again from the series of pulses received and finally converted into a con-

tinuous signal giving the reproduction of the original modulating signal. It is clear that if the interference in the transmission channel does not mutilate or displace the pulses as such beyond recognition then this interference can be completely eliminated.

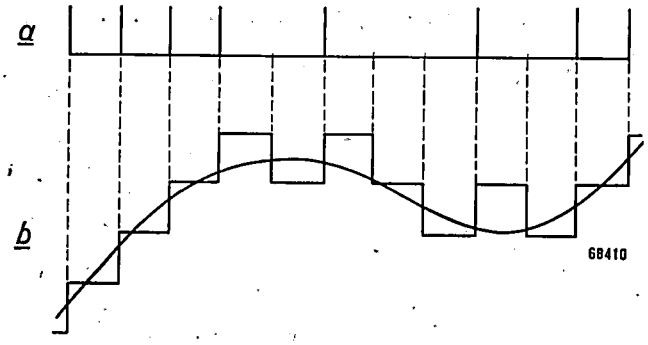


Fig. 3. Principle of delta modulation. The series of pulses emitted by the transmitter (*a*) is generated after comparison of the modulating signal (the continuous curve in *b*) with a step-like approximating signal with equidistant steps built up in the transmitter. When the approximating signal is smaller than the modulating signal a pulse is emitted, but if it is larger the pulse is suppressed. At the receiving end the approximating signal is formed from the series of pulses by integration and from that signal the modulating signal is reconstructed by smoothing out the steps.

This is only a very sketchy exposition and in practice the method is not so simple as that, but anyhow it gives the essence of the method, which is recapitulated in the following concise form.

With the system of delta modulation, from a series of equidistant pulses transmitted, which may have the value 1 or 0, at the receiving end a signal is built up which has to form the nearest possible approximation of the original modulating signal. This approximating signal is constructed not only at the receiving end but also at the transmitter. At the sending end the original signal can therefore always be compared with an approximation similar to that which would be obtained with the aid of the quantized pulses at the receiving end. If the approximating signal is smaller than the input signal then a pulse is sent out by the transmitter, but if the approximating signal is larger than the input signal the next pulse is suppressed.

The "auxiliary receiver" at the sending end, which supplies the approximating signal, is in its simplest form an integrating network. The output voltage from this network is compared with the original signal in a difference meter, a simple resistance network, followed by an amplifier. Thus a sort of inverse feedback is applied in the transmitting circuit. This feedback, however, differs from the usual inverse feedback in two respects. First, the feedback circuit is closed only at the moments when it has to be decided whether a pulse is to be transmitted or not. Second, the voltage acting on the feedback circuit is not proportional to the difference voltage but dependent only upon the polarity of the difference voltage. This inverse feedback is virtually quantized in position and amplitude.

In *fig. 4* a block diagram is given representing the circuit of the transmitter. We see here a pulse generator, *I.G.*, which yields a series of equal pulses with a repetition frequency of, say, 60 000 or 100 000 pulses per second. These pulses have to pass through a pulse modulator, *I.M.* If the voltage at the point *2* is negative then *I.M.* gives a pulse having the same sign as the incoming pulse; if, on other hand, the voltage at *2* is positive then a pulse with the opposite sign is sent to point *3*. The pulses coming from the pulse modulator *I.M.* are sent to an integrating network *I* (i.e. the "auxiliary receiver"). Here each positive pulse will raise the output voltage a certain amount and each negative pulse will lower it by the same amount. In this way an echelon curve or approximating signal *4* is obtained. In the difference meter *V.M.* this signal is compared with the modulating signal *5*, the output voltage always indicating the difference between the two signals and thus controlling the pulse modulator. The next pulse is therefore positive when the approximating voltage at that instant is lower than the input signal and negative when it is higher. The series of positive and negative pulses generated in this manner are

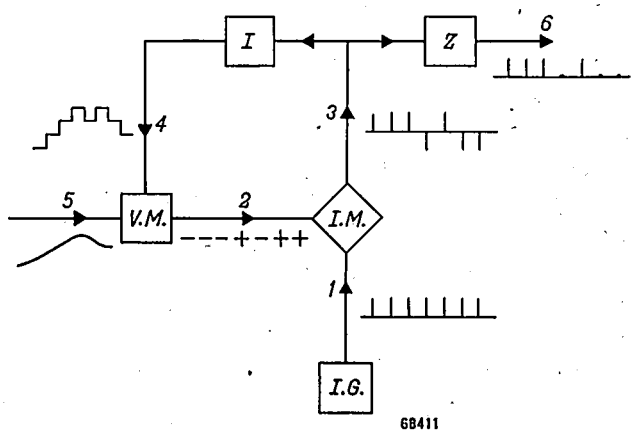


Fig. 4. Block diagram of the transmitter. The pulse generator *I.G.* supplies a series of identical equidistant pulses *I* which are passed on to the transmitter by the pulse modulator *I.M.* either unchanged or with opposite polarity. The action of the pulse modulator is governed by the polarity of the difference voltage *2* obtained in the difference meter *V.M.* by comparison of the modulating signal *5* and the step-like approximating signal *4* formed from the series of pulses *3* by integration in *I* ("auxiliary receiver"). The transmitter suppresses the negative pulses and passes the signal *6* to the aerial or the transmission line.

passed not only to the auxiliary receiver but also to a transmitter Z (for instance a radio relay system) which suppresses the negative pulses and transmits the positive ones.

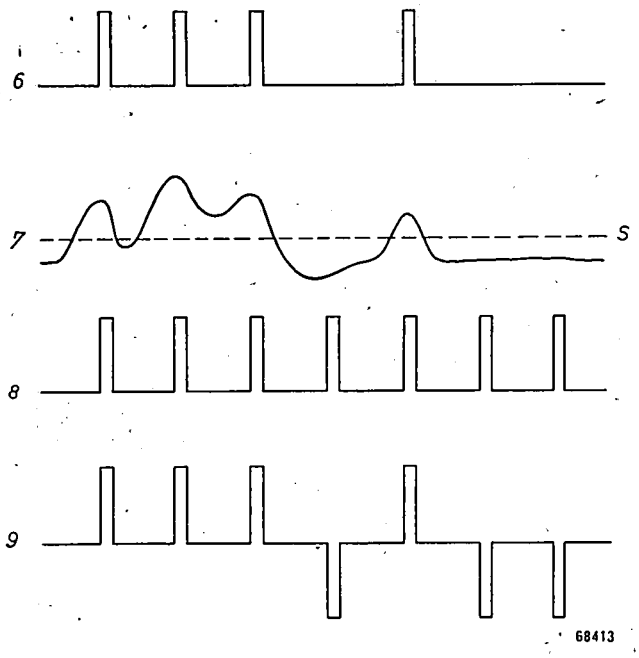


Fig. 5. The series of pulses *6* emitted by the transmitter of fig. 4 becomes distorted in the course of transmission, assuming, for instance, the shape of the signal 7. This signal influences a pulse modulator in the receiver in such a way that, provided the pulse amplitude exceeds a certain limit *s* (mostly chosen equal to half the original pulse amplitude), the pulses *8* produced by a pulse generator are passed on with the correct polarity. If the signal 7 lies below the level *s* the pulses produced by the generator are replaced in the pulse modulator by pulses of the opposite polarity. In this way the series of pulses *9* is obtained.

Fig. 5 shows to what changes the pulses on their way through the aether are subject. The pulse pattern as transmitted corresponds to *6*, while owing to the interference the pattern as received is more or less distorted, resembling, for example, that depicted by 7. Now in order to reconstruct from this pattern the step-shaped approximating signal, in the receiver (*fig. 6*) a pulse modulator is employed which in principle functions in the same way as the pulse modulator in the transmitter: the pulse (*8*) coming from a pulse generator *I.G.* is passed through unmodulated if the incoming pulse exceeds a certain value *s* (for which, as a rule, half the height of the undistorted pulse is taken), but if the incoming pulse is smaller than that value then the pulse modulator *I.M.* gives a pulse of the opposite sign. At the output of the pulse modulator a combination of pulses (*9*) is obtained which is identical to the signal *3* (see figs 4, 5 and 6). (The pulse generator has to be synchronized with the incoming pulses, the apparatus required for this having been omitted from the diagram in fig. 6 for the sake of simplicity.)

When this series of pulses is applied to an integrating network, *I*, the same step-shaped signal (*10*) is obtained as was generated at the transmitter (signal *4*). By passing this signal through a low-pass filter *F* the steps are smoothed out into a continuous signal *11* closely resembling the original input signal.

In this manner interfering signals of at most half the pulse amplitude can be permitted in the course of transmission without being perceptible in the signal.

Sound reproduction by the delta modulation system

Obviously the step-shaped approximating signal (and thus also the continuous output signal derived from it) will always show some small deviations with respect to the original signal. If the modulating signal is sinusoidal then at the receiver a sinusoidal signal is obtained against a background of noise, the so-called quantizing noise, due to the approximation of the instantaneous values of the speech signal by discrete values. The higher the rate of sampling, however, the less is this quantizing noise. In practice, with a pulse frequency of 40 000 pulses per second there is still a perceptible grating quality in the speech signal transmitted, but with a pulse frequency of 100 000 pulses per second there is practically no noticeable quantizing noise. Due to this high pulse frequency a much wider band is required for the transmission than is needed in the case of direct modulation of the signal on a carrier, since the bandwidth required is at least

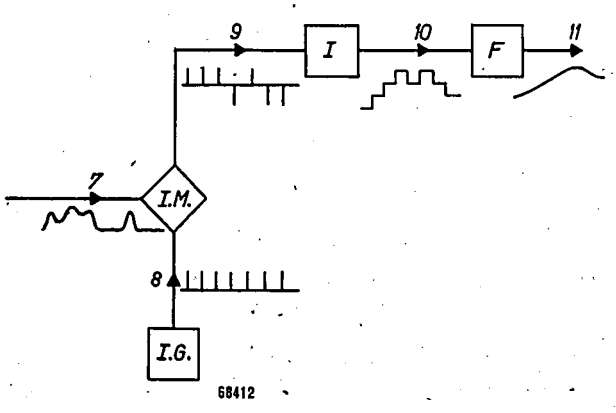
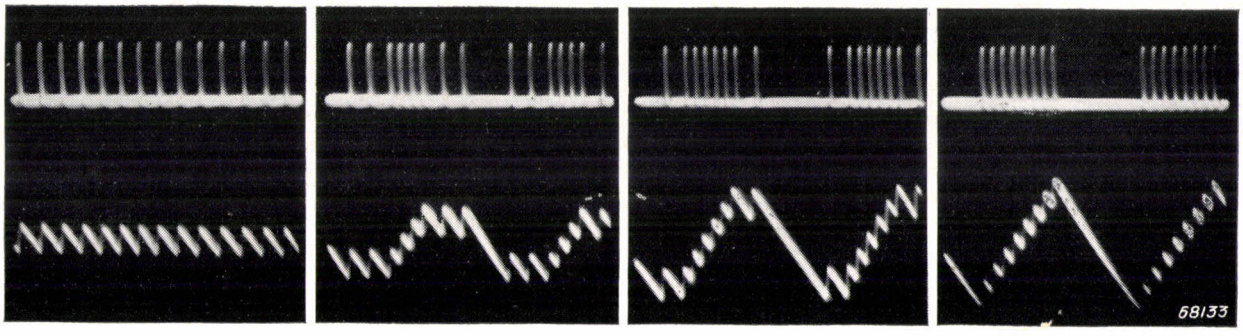


Fig. 6. Block diagram of the receiver. The distorted signal 7 influences the pulse modulator *I.M.* in the manner described under fig. 5. The pulses *8* produced by the pulse generator *I.G.* may be passed on with the correct or opposite polarity and the signal *9* thus produced is converted by an integrator *I* into the echelon curve *10*. From this the continuous-wave signal *11* is formed in the low-pass filter *F*, this signal being the reproduction of the original signal at the input of the transmitter.

equal to half the repetition frequency of the pulses. From the physical point of view this is to be regarded as the price to be paid for the suppression of the interference. Especially for radio relay systems, however, this does not constitute any serious objection, since in this frequency range it is easier to increase the bandwidth than to improve the signal-to-noise ratio.

In practice the pulse modulator is so arranged that only pulses of the values 1 and 0 appear at the output. The input signal is then more or less approximated by a sawtooth instead of an echelon. maximum amplitude has been reached that can be observed at the receiver. Larger input amplitudes will not further modify the approximating signal, and thus at the receiver the same output signal will be obtained. It therefore appears that with delta modulation there is a limitation of the amplitudes to be transmitted. This limitation is reached when the signal voltage assumes a certain slope corresponding to the permanent emission of positive pulses. The higher the frequency of the input signal, the smaller is the maximum amplitude of the output signal, as will be shown below.



a *b* *c* *d*

Fig. 7. The sawtooth-shaped approximating voltage and the corresponding series of pulses transmitted, for an A.F. sinusoidal input signal of 4000 c/s. The rate of sampling of the pulses supplied by the generator is 64 kc/s. *a*) The amplitude of the sine curve is zero; the pulses emitted are alternately 1 and 0. *b*) and *c*) The amplitude is gradually increased, as a result of which the pattern of the series of pulses is changed. *d*) The amplitude has by now become so large that the system reaches complete modulation, and even overmodulation. Even larger amplitudes will not produce any other pulse pattern.

This does not, however, alter the principle of the system.

In *fig. 7* some oscillograms of such an approximating signal are reproduced, together with their corresponding series of pulses. These approximating signals built up from unit pulses are smoothed out at the receiver with the aid of a low-pass filter, as already mentioned, and then give a reproduction of the (in this case sinusoidal) input signal. The frequency of the input signal in the case under consideration is 4000 c/s, the rate of sampling is 64 000 pulses per second. From the four oscillograms it is to be seen how the configuration of the pulses changes with increasing amplitude of the signal transmitted. In fig. 7*a* the input signal is zero. The pulses transmitted are then alternately 1 and 0, and since the sawtooth-shaped approximating signal then has the fundamental frequency of 32 000 c/s, thus containing no frequencies lying within the pass band of the filter, the value of the output signal behind the filter in the receiver will likewise be zero. In figs 7*b* and 7*c* the amplitude has been gradually increased, while in fig. 7*d* the

Other methods for transmitting quantized signals

From the foregoing it is seen that with the method of delta modulation two principles are applied: integration of pulses, and application of a special type of inverse feedback. Owing to the inverse feedback the essence of the method lies in the transmitter responding only to what has been changed with respect to the earlier situation. Each pulse transmitted is only a correction of the preceding signal.

Another system of modulation based on the use of quantized signals is that of pulse code modulation [3]). This system works as follows. As in the case with delta modulation, only the presence or absence of a pulse is of importance. To be able to distinguish more than two amplitude levels with the aid of this "all or nothing" principle, the instantaneous values of the message signal are denoted not by one pulse (present or not) but by a code group of 5 or 7 equidistant pulses each of

[3]) See, e.g., W. M. Goodall, Telephony by pulse code modulation, Bell System Technical Journal **26**, 395-409, 1947.

which may have the value 1 or 0. The position of a pulse in the code group determines its significance. This position principle is met with also in ordinary numerical systems (it was first employed systematically by Indian mathematicians about the year 600). In the decimal system all numbers can be denoted with the aid of only 10 symbols, the position of a symbol in a number determining its value (units, tens, hundreds, etc.). With the method of pulse code modulation, instead of the decimal scale, the simpler binary system is used. The position of a pulse in the code group determines whether it will have the value 1, 2, 4, 8, and so on. In this way, with code groups of five pulses, 32 different signal values can be denoted. This is illustrated in *fig. 8*.

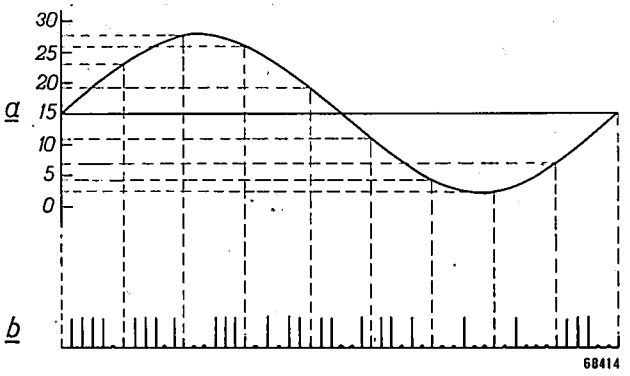


Fig. 8. Principle of pulse code modulation with code groups of 5 pulses. At the beginning of each time interval corresponding to a code group the value of the signal to be transmitted (*a*) is rounded off to one of the discrete values 0 to 31 and "translated" into a pulse pattern (*b*). This pattern is formed by the presence or absence of pulses at the five places marked by dots within the group. A binary system is applied, each place corresponding to a binary value and the measured amplitude thus being expressed according to the binary scale. From left to right the five points in a code group then represent 1, 2, 4, 8 and 16 amplitude "units".

Delta modulation may be regarded as a modified form of pulse code modulation in so far as both systems employ a coding of the signals to be transmitted. Pulse code modulation employs 5 or 7 coding units (sometimes even more), whilst delta modulation, at least in the simple form described here, employs a coding with the aid of one unit, viz. one step in the echelon curve. This is possible on account of the fact that with delta modulation only corrections of the signal are transmitted, whereas in the case of pulse code modulation the complete signal values are transmitted.

With the introduction of pulse code modulation a binary five-unit code came to be applied in telecommunications for the fourth time. In 1837 Cooke and Wheatstone built in London (between Euston and Camden Town) a transmission system of five copper wires connected at the receiving end to a telegraph instrument with five needles registering twenty different letters. In 1926 Raincy obtained a patent for a code system for picture telegraphy which may be regarded as the precursor of the present-day pulse-code modulation systems. The modern telex system, the principle of which was invented by Baudot in 1875, employs a code system usually of 5 units replacing the rather peculiar Morse code of telegraphy (the peculiarity about the Morse code is that the length of the signs is adapted to the frequency in which the signs occur, and therefore has a particular significance).

The systems can be compared only by means of the basic theory concerned with the transmission of information where the concept of information is given a well-defined meaning. This would lead us too far afield, so we shall mention only some of the results reached from the comparison and include in it also the (hypothetical) quantized pulse-position modulation as described on page 239. There are two kinds of interference to be distinguished, the interference which is inherent in the system itself and that which is introduced in the course of the transmission. In the case of non-quantized systems, in theory only a negligible interference (noise) arises in the transmitter and in the receiver, so that if the transmitter and the receiver were directly interconnected the reproduction would be quite free of interference. But, as already pointed out, along the path of transmission there is some interference which is cumulative. If this interference in the course of transmission remains below a certain threshold value then the transmission interference can be entirely eliminated by quantization. As we have seen, quantization means that an "interference" — the quantizing noise — is introduced in the transmitter itself, but this interference does not take part in the cumulation. Moreover, the quantizing noise can be made as small as desired by increasing the band covered by the high-frequency signal transmitted. The transmission interference, however, cannot be reduced.

Quantized pulse-position modulation is characterized by the fact that the threshold value of the transmission interference below which that interference is inactive is relatively low. If the pattern is not to be lost, the pulses may be displaced only slightly in the interval allotted to them. Further it can be shown mathematically that in this system the quantizing noise decreases inversely with increasing bandwidth f.

In coded systems, such as delta modulation and pulse code modulation, it is just the coding that allows of a much greater transmission interference. Furthermore in these two systems the quantizing noise appears to decrease far more with increasing bandwidth, according to a law of $f^{-5/2}$

and of approximately 2^{-f} respectively. This clearly shows the inefficiency, already mentioned, of quantized pulse-position modulation. The same inefficiency is found in all other quantized but not coded systems.

The reason why delta modulation and pulse code modulation are both more advantageous from the point of view of interference suppression than quantized, non-coded, normal systems lies in the use of inverse feedback in the system first mentioned and in the employment of more coding units in the second system. However, also in delta modulation a coding in more units can be applied, for instance by not only judging the difference measurement according to the sign but also dividing the difference into a number of steps and expressing it in the pulse picture transmitted. In this manner the signal-to-noise ratio can be improved for the same bandwidth. It can be shown that in order to obtain the same quality of reproduction fewer units are needed when inverse feedback is applied than when such is not employed.

Properties of delta modulation

Let us now return to the system of delta modulation described. From what has been said above it follows that from the point of view of quantizing noise delta modulation is less advantageous, for the same bandwidth, than pulse code modulation. It appears that in order to obtain a certain quality in the reproduction of speech in the case of delta modulation the band has to be about 50% wider than that required for the same quality with pulse code modulation. On the other hand, however, the apparatus is very much simpler. Not only is there no coding device required, but, since with delta modulation all the pulses are of equal significance, the requirements to be met for the circuit tolerances are less severe; in the difference meter only the polarity of the voltage is determined, so that the amplifiers for the difference voltage need not possess strictly linear characteristics. It is also permissible for a valve adjustment, or for the amplitude of the pulse, to change gradually with time, because in that case the inverse feedback ensures that the signal is not distorted. Finally, the construction of the feedback and integrating networks in the receiver is not very critical; deviations do not give rise to non-linear distortion (as is the case with the networks employed for pulse code modulation) but change only slightly the frequency characteristic of the whole system, which change can be corrected by a simple egalization. In the designing of the complicated quantization apparatus required for pulse code modulation great care has to be taken to ensure that it is linear, whereas for the quantization in the case of delta modulation a simple and essentially linear network is employed.

It would lead us too far to discuss the technical details of the circuit, but mention is to be made of the fact that the frequency characteristic of the network in the feedback circuit can be given an entirely different form for the frequencies above the audio range than for the lower frequencies, thereby reducing the quantizing noise.

Both with pulse code modulation and with delta modulation the possibilities are limited: the signals to be transmitted must not be too large, while at the receiver very small signals cannot be distinguished from the zero signal. In the case of pulse code modulation this limitation is governed by the size of the steps in the 5- or 7-unit code, the smallest signal to be transmitted corresponding to a code group with the "signal value" 1, and the largest corresponding to the transmission of a code group with the signal value 32 (5-unit code) or 128 (7-unit code). Thus the limitation is independent of frequency.

As already seen from fig. 7*d*, the limit in the case of delta modulation, for the large signal values, is reached when the signal voltage has a certain slope corresponding to the permanent transmission of positive pulses. If, however, the signal voltage is so small that notwithstanding the presence of the signal the approximating voltage retains the shape depicted in fig. 7*a*, and thus the pulses are alternately transmitted and suppressed, then nothing is to be perceived of the signal behind the low-pass filter in the receiver. By means of special circuits the smallest signal to be reproduced (threshold value) can be made somewhat smaller than what follows from this reasoning, but the minimum amplitude that can be observed at the receiver always remains finite.

As far as reproduction of the sound is concerned, the important consequence of the signal amplitude being limited owing to the existence of a maximum slope in the case of delta modulation is that the maximum amplitude that can be transmitted is inversely proportional to the frequency. Since in the case considered the threshold value of the signals received does not depend upon the frequency, this means that the dynamic range, i.e. the ratio of the maximum to the minimum amplitudes at the receiving end, becomes smaller as the frequency is raised.

This deviation from the ideal reproduction of the audio signal happens to coincide with a property inherent in the reproduction of speech and music and in the receptive qualities of the human ear, so that as far as its physiological effect is concerned it does not adversely influence the transmission: the maximum amplitude of the sound pressure in the audio spectrum decreases with increasing frequency, as is likewise the case with the dynamic range perception of the ear (especially at high frequencies).

A third quantity, one that does not occur in normal acoustics, is the gamma of the transmission system, by which is to be understood the number of distinguishable steps between the maximum and minimum amplitudes transmitted. With delta modulation the gamma becomes smaller as the frequency increases, and although physiological experiments have shown that in the range of speech frequencies this should lead to a deterioration in the quality of reproduction, in practice this is not evident.

For frequencies above 6000 c/s there is a reduction of the gamma with increasing frequency also in hearing. At these frequencies delta modulation therefore matches the properties of the human ear in two respects.

This brings us to the following highly important point. An optimum transmission system is characterized by the fact that it transmits what is to be transmitted and nothing more than that. When the signal power decreases as a function of the frequency then any possibilities offered by the system for the transmission of large powers at high frequencies are superfluous and undesired. When, as appears to be the case, the human ear perceives smaller dynamic ranges at the higher frequencies, in the optimum system it would be impossible for larger dynamic ranges than desirable to be transmitted. This explains why in the case of delta modulation a relatively small bandwidth suffices. The matching of this system to the peculiarities both of the speech signal and of the human ear partly neutralizes the mathematical deviations from the ideal reproduction.

The deeper significance of delta modulation

Being in essence based upon inverse feedback, delta modulation belongs to the family of regulating systems and servo-mechanisms, but with its quantization both in amplitude and in time it is a member that has been little studied.

As already stated, only corrections of the preceding signal are transmitted, this being brought about with the aid of the integrating network. Mathematical integration, however, would mean that even the distant past would contribute towards the present amplitude. Considering that there is a definite limit set to the time within which earlier signal values of a speech or music signal are related to the instantaneous value, those earlier values may gradually be forgotten. A resistance-shunted capacitor (as is present in our integrating RC network) thus supplies the first useful variation of the purely mathematical integration. A second variation consists in the fact that by adding more elements to the network an attempt is made to make the output signal correspond to the most probable continuation of the signal to be transmitted (that continuation where the first or more time derivatives of the signal voltage do not change with time). All the transmitter need do is then to transmit, via the difference meter, the corrections with respect to this most probable continuation.

This can also be formulated in another way. Each audio or video signal — whether speech, music or picture — is characterized by certain correlation properties, i.e. by a certain relation existing between the successive signal values. Any transmission of the known part of the signal, thus of that part of the amplitude that follows from the correlation properties, is superfluous, for the very reason that it is already known. In any optimum transmission system, therefore, these correlation properties should be discounted at the transmitter and only what is to be regarded as new and individual should be transmitted.

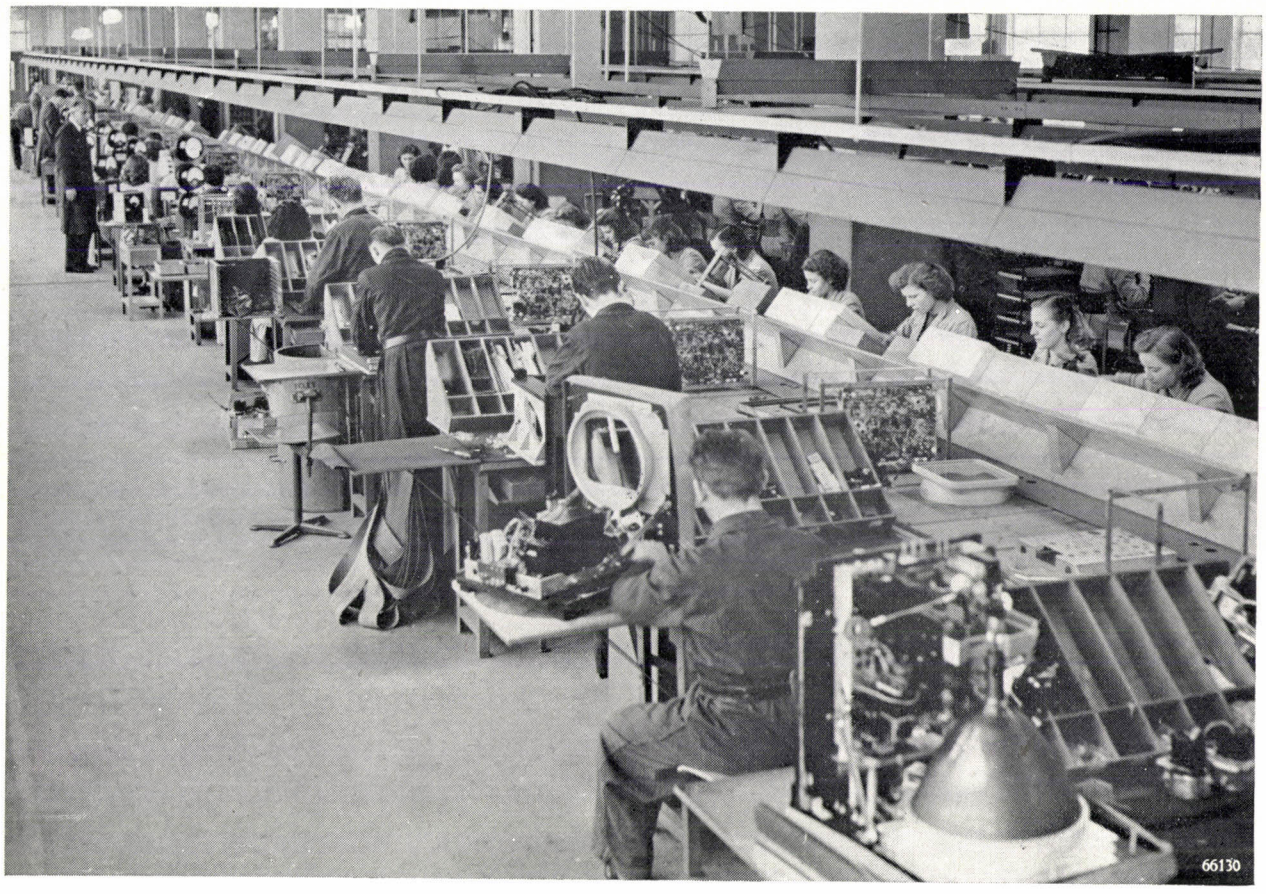
This implies that the signal transmitted in the case of an optimum transmission system possesses the statistical properties of noise, thus having no correlation [4]).

It is the function of the feedback circuit to take into account these correlation properties, which means, therefore, that the properties of the feedback network must be determined by the nature of the signal to be transmitted, on the one hand, and by that of the ear (or eye) on the other hand.

[4]) S. Goldman, Proc. Inst. Radio Engrs **36**, 584, 1948.

Summary. In this article the development of different modulation systems is described, a development which aims at achieving the transmission of audio or video signals as free as possible from interference. The interference occurring in carrier modulation systems and also in pulse-position modulation arises from very small effects of interference being cumulative in the course of transmission. This cumulation can be eliminated by quantizing the signals to be transmitted, both in time and in amplitude. The system of delta modulation, based on this principle, is described and compared with other systems employing quantization (pulse code modulation). It appears that for a good reproduction, using a reasonable bandwidth for the transmitted signal, the apparatus required for delta modulation is relatively very simple, because of coding being applied with the aid of only one unit. This is possible by reason of the fact that with delta modulation account is taken of the correlation properties of the signal, by employing a special kind of inverse feedback in the transmitter, and that the properties of the system are matched with those of speech and music and of the human ear.

THE MANUFACTURE OF TELEVISION RECEIVERS



The first female assembly hand on the right receives a new chassis periodically, mounts on it some parts taken out of trays in front of her, and then passes it on to her neighbour, who in turn mounts on some other parts and passes on the chassis to the next assembly hand, and so on. By the time it reaches the end of the line, on the extreme left of the photograph, the chassis is complete except for the valves and picture tube. It is then placed on a conveyor belt and carried in the opposite direction (in the photograph from left to richt) past a number of points where tests are carried out and any defects traced and corrected. At a certain stage the valves are mounted in their sockets and further tests are carried out. By the time the chassis reaches the assembly hand in the foreground the wooden bottom and the front panel can be put on; here is to be seen also the rubber ring (called a masker) in which, at the next stage, the front end of the picture tube comes to rest. The apparatus in the immediate foreground on the right — the direct-vision receiver type TX 500 U — has the picture tube fitted in.

Further stages, not shown in the photograph, are: testing for picture and sound, running test for a number of hours, completion of the cabinet, and packing of the set.

AUTOMATIC WELDING WITH CONTACT ELECTRODES

by W. P. van den BLINK, H. BIENFAIT and J. A. van BERGEN.

621.791.753.4.002.52

A superficial analysis of a welder's work shows that this depends on two things. The first thing is judgment and experience, enabling the welder to determine how a joint is to be laid, i.e. what kind of joint is to be used, with what type of rod it is to be made, what current is to be used and what the welding speed should be, etc. The second thing is skill in laying the weld — in "holding the arc", keeping up the speed, replacing the rod, "joining up" the new bead, etc. The last-mentioned functions lend themselves to mechanization, which is everywhere desirable for increasing productivity. This mechanization has now become possible through the development of certain types of electrodes together with an automatic welder.

Philips Contact electrodes

The Contact electrodes derive their name from the fact that when welding with a rod of this kind the coating of the rod can be kept continuously in contact with the workpiece, i.e. with these electrodes a joint can be laid by "touch welding".

For a detailed discussion of Contact arc welding and the many aspects of this new method of welding, reference may be made to two articles which appeared in volume 8 of this journal in 1946 [1]). Here we shall only very briefly outline the fundamental principles and the importance of this technique.

It is known that arc welding with normal rods can as a rule be properly carried out only by trained welders. One of the things that a welder has to learn from practical experience is the art of "holding the arc": the arc between the tip of the rod and the workpiece has to be kept at exactly the same length, so that the welding material from the rod runs smoothly and the penetration of the workpiece is uniform. Moreover, it is of importance to maintain the right length of arc because any deviations, however momentary, may have undesirable consequences. If the arc is too long then the arc voltage rises and in the extreme case the arc may even be extinguished. If, on the other hand, the arc is too short, too little heat is developed and in the extreme case the droplets of metal make a short circuit between the welding electrode and the pool, with the result that the rod may even freeze onto the workpiece, as every welder will have experienced in the beginning of his training.

Thus a very steady hand is required, and it is this which puts a strain on the welder and makes his work tiring, whilst often the results are less consistent than one would wish for.

The metal that is deposited from Contact electrodes is contained not only in the core (wire) but for a part also in the coating in the form of a fine powder. This makes the coating much thicker than that of normal rods, and when welding a very deep cup is formed (*fig. 1*). This cup ensures that the right length of arc is maintained while the rod is kept in touch with the workpiece. In this way it is easy to make a uniform weld and there is no risk of freezing, since the length of arc is constant and much greater than the dimensions of the droplets. Thus with Contact electrodes the arc is maintained as it were automatically.

With normal electrodes, i.e. electrodes having a coating of the normal composition, a deep cup can also be obtained, simply by making the coating very thick, but this would be uneconomical, whilst the

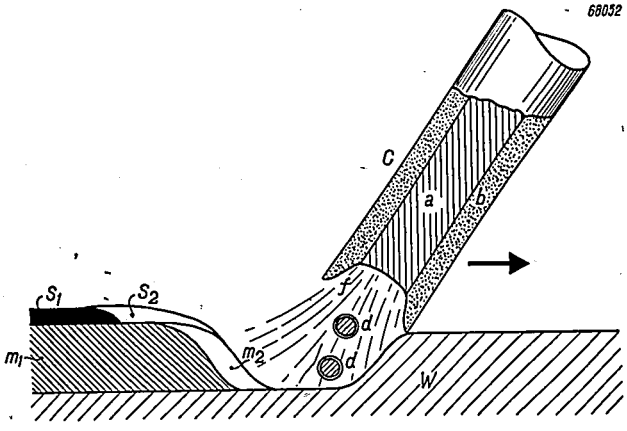


Fig. 1. The melting down of a Contact welding rod. The arrow indicates the direction of travel. *C* welding rod with core *a* and coating *b* (containing iron powder), *W* workpiece, *d* droplets of molten iron, m_1 solid and m_2 liquid metal of the weld, s_1 solid and s_2 liquid slag. The depth of the cup *f* is such as to allow of touch-welding with this electrode.

[1]) P. C. van der Willigen, Contact arc-welding, Philips Techn. Rev. 8, 161-167, 1946; the same author, Penetration and welding speed in contact arc welding, Philips Techn. Rev. 8, 304-309, 1946.

excess of slag-forming material would spoil the quality of the weld. The thickening of the coating of Contact electrodes by incorporating in it a large percentage of the iron that is to be deposited, makes it possible to maintain the optimum ratio of slag to iron. Moreover, there is another advantage of a different nature: Contact electrodes are self-starting. The high iron content makes the coating semi-conducting, so that when it is brought into contact with the workpiece current flows immediately, and owing to the local heating the arc is struck. This dispenses with the troublesome "tapping".

This is not the place to go into the metallurgically favourable properties of Contact electrodes, arising, for instance, from the deep cup protecting the running metal from the atmosphere.

It has just been said that with Contact electrodes the holding of the arc is, as it were, automatically ensured, thanks to the principle of touch welding. All the welder has to do when working with a single rod is to move it along at the right speed for laying the bead. Now it is important to note that the forward travel can also quite easily be made automatic. Furthermore, it is now possible, with the aid of a new apparatus called the Philips Contact automatic welder, for welding to be done entirely automatically with a number of rods in succession. When employing this apparatus welded joints of any length can be made fully automatically.

First we shall deal with the automatic forward travel of the rod and then describe the construction of the automatic welder itself.

Automatic forward travel

In *fig. 2* a sketch is given showing how automatic forward travel can be brought about. This method was actually applied and described some years back by Van der Willigen [2]).

The rod is clamped at a certain angle onto the end of a bar which is set at an angle in the vertical plane and, guided by four insulating rollers, is able to slide down under the force of gravity. Thus the tip of the welding electrode is held against the horizontal workpiece, while the electrode is fed with current via the bar. As the tip of the electrode melts so the bar bearing upon it gradually descends and moves the electrode along parallel to itself. The point of contact between electrode and workpiece, starting at O, thus moves along horizontally according to the rate at which the welding material is deposited until the whole of the rod is consumed. To avoid the arc from being taken over by the clamping device, the bar is, of course, drawn back in good time.

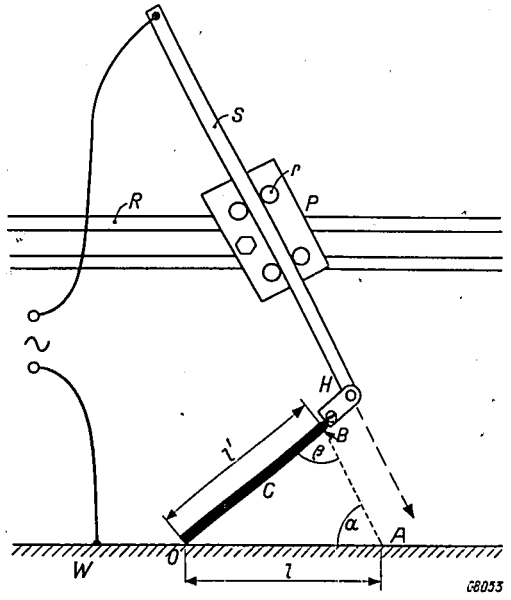


Fig. 2. Set-up for automatic forward travel (for touch-welding). C the Contact welding electrode fixed with adjustable clamp H to the guided bar S. The bar is guided by four insulating, grooved rollers r and by its weight keeps the tip of the electrode pressed against the workpiece W. As the electrode melts down it travels parallel to itself. The mounting plate P, carrying the rollers r, can be fixed to the rail R at any desired angle.

A line drawn through the clamped end of the welding electrode and parallel to the guided bar intersects the workpiece at the point A. $OA = l$ is the length of the bead laid by one rod fully consumed (useful rod length l'). For any welded joint a certain bead length is prescribed, since this determines the thickness of the weld when using a particular kind of electrode (*fig. 3*). The desired bead length can be obtained by a suitable choice of the angles of inclination α and β of the guided bar and the welding

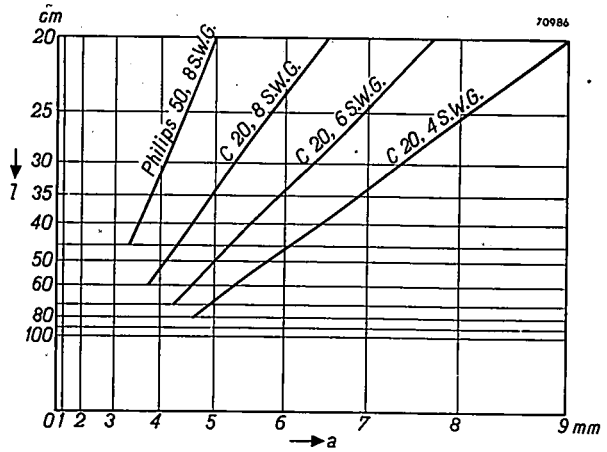


Fig. 3. Diagram for determining the correct bead length l per electrode for a prescribed weld thickness a (in a fillet weld). The lines refer to four different types of welding electrodes.

[2]) P. C. van der Willigen, Contact arc welding, The Welding J., Welding Res. Suppl. 25, 313-320, 1946.

electrode respectively. From fig. 2 it is seen that the condition to be met is:

$$\frac{\sin\beta}{\sin\alpha} = \frac{l}{l'} \quad \ldots\ldots\ldots \quad (1)$$

One of the angles is in the first instance arbitrary. In fact, this is obvious from geometrical considerations: the bead OA to be laid is fixed, but the end B of the welding electrode may lie at any point of the circle with radius l' about the centre O, provided the direction in which the bar is guided is chosen parallel to BA.

In practice it appears that not all combinations of the angles α and β possible according to eq. (1) are equally satisfactory. When for this equation a diagram is plotted from which, with the quotient l/l' and one of the angles given, the other angle can be read, then in that diagram a certain area (i.e. a not sharply defined relation between α and β) can be indicated where the best welding results can be obtained. The position of this area depends to some extent on the type and diameter of the welding electrode, the shape of the weld and the current strength. *Fig. 4* gives such an α-β diagram, in which the optimum area is hatched for a particular case (Contact electrode type C 20, 6 S.W.G. — or C 20-5 — used for a fillet weld made in the flat position, with the maximum permissible current of 350 A). To the left of this hatched area the quality of the weld is adversely affected by the tendency of the slag to run ahead of the arc, whereas to the right of that area the slag has a tendency to lag behind.

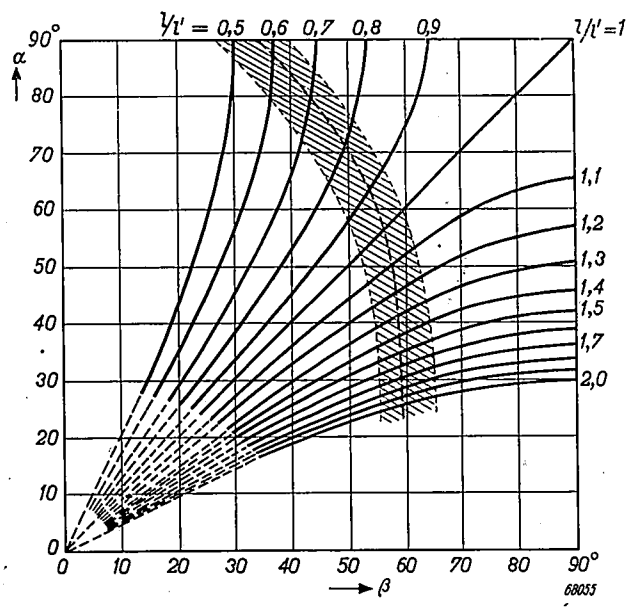


Fig. 4. Diagram for determining the angles of adjustment α and β for automatic welding. The quotient l/l' of bead length and useful electrode length is given, the corresponding curves then showing all combinations of the angles α and β possible for obtaining the desired length of bead. When a fillet weld is to be made in the flat position with electrodes of the C 20, 6 S.W.G. (C 20-5) type, using a current of 350 A, in practice only combinations within the hatched area prove to be satisfactory.

The degree of mechanization of Contact welding achieved by this method of automatic forward travel makes it possible for welding in mass production to be done by less practised welders, or even perhaps by almost unskilled labour, while furthermore the results are more uniform than is usually possible when welding by hand. The fact is that in many cases (depending upon the type of electrode and the current) it takes less than one minute for an electrode to be consumed; the welding speed (rate of travel) is thus very critical if the weld is to be of uniform thickness throughout. With automatic forward travel a constant weld thickness is guaranteed because the rate of travel is determined by the rate of melting of the rod.

Automatic welding of long joints

With the arrangement according to fig. 2 both the holding of the arc and the rate of travel are made automatic, all that is left to be done by hand being the breaking of the arc (drawing back the guided bar) and the starting of a new electrode. We shall now go two steps farther and show how these last two functions can also be carried out automatically.

For laying a long, straight bead requiring a number of welding electrodes, a series of the devices illustrated in fig. 2 can be mounted on an insulated rail set up over the workpiece in line with the joint. The intervals between these automatic welders are chosen equal to the bead length of one rod. Welding rods, e.g. Contact electrodes of the type C 20, 6 S.W.G. or C 20, 4 S.W.G. (continental designation C 20-5 and C 20-6, core diameters 5 mm and 6 mm respectively), are clamped into the holders with their tips resting on the workpiece but insulated from it by a piece of asbestos paper (excepting the first rod). The power supply may provisionally be arranged by connecting each guided bar to the rail with a flexible cable. The rail is connected to one terminal of the supply source and the workpiece to the other. The arrangement is then as sketched in *fig. 5*.

As soon as the current is switched on welding is started with the first electrode (not being insulated by asbestos paper, and the C 20 type of electrodes being self-starting) and the rod is carried along automatically until its clamping device strikes up against the next electrode. As soon as this happens the arc length begins to increase, because

the rod is brought to a standstill while it is still melting. When the arc has been lengthened by a few millimetres (the arc voltage then rising) the atmosphere around the tip of the second electrode will have been sufficiently heated and ionized to

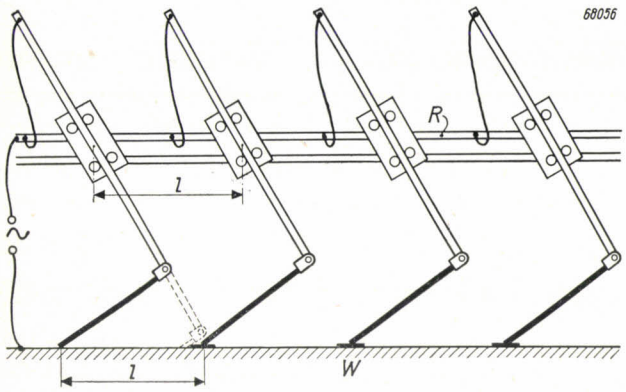


Fig. 5. Diagram of the set-up for laying long welds. A number of welding heads according to fig. 2 are mounted at intervals equal to l on a rail R over the workpiece W. The second, third and following electrodes are temporarily insulated from the workpiece by means of pieces of asbestos paper laid under them. When one electrode has been consumed (dotted position of the guided bar) the next electrode automatically takes over the arc and at that moment the remaining end of the burnt-down electrode has to be drawn up away from the workpiece.

cause the ignition voltage of this electrode to drop below the potential available. Since at the same time the heat melts away the insulating asbestos paper, the second electrode strikes and starts welding [3]). This cycle repeats itself with the whole series of welding rods, so that in this way any length of joint can be welded automatically.

Now we have to consider how the remaining function mentioned above is performed: each electrode holder has to be immediately withdrawn as soon as its electrode has melted away and the next electrode is started. If one holder were not withdrawn in time then the remainder of its electrode would drop further as soon as the next electrode is started, thus making contact with the workpiece again, with the risk of the arc returning from the new electrode to the old one, resulting in an inadequate joint of the new bead and/or even freezing of the clamp to the workpiece. It is therefore highly desirable not to leave the simple function of withdrawing the electrode holder to the care of an operator, but to have this performed automatically too.

[3]) Ordinary paper cannot be used for the insulation instead of asbestos paper because it would burn through too quickly and thus the second electrode would be ignited before the pool from the first one is close enough to it. From the following text it will become clear why this cannot be tolerated.

This can be achieved with the Contact automatic welder, which is a further development of the principle of fig. 2. A photograph of this apparatus is given in *fig. 6*.

The fundamental idea is that the withdrawal of the holder should be brought about by means of the current that begins to flow through the next welding electrode as soon as it is started. One is then sure of each holder being retracted exactly at the right moment, not too soon (before the next electrode has been started) and not too late (when the holder has dropped so far that the arc may return to the old electrode again).

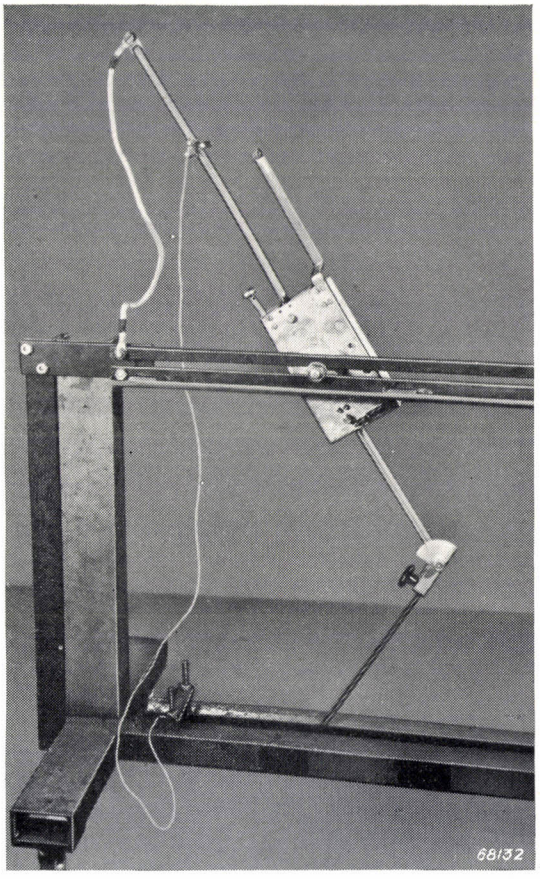


Fig. 6. The Contact automatic welder, mounted in the position for welding.

The Contact automatic welder

The operation of the automatic welder can be described with the aid of the schematic drawing in *fig. 7*, where two "welding heads", connected so as to come into action successively, are mounted on a rail. (Of course more than two welding heads can be mounted in this way.)

The current is not fed to the welding head N direct by the rail but via a terminal c in the preceding head M. As long as the electrode of the head

M is welding, the hinged steel plate p is in the position as drawn, being kept there by a leaf spring b. In this position the terminal c is connected to the rail via the winding of the small electromagnet e.

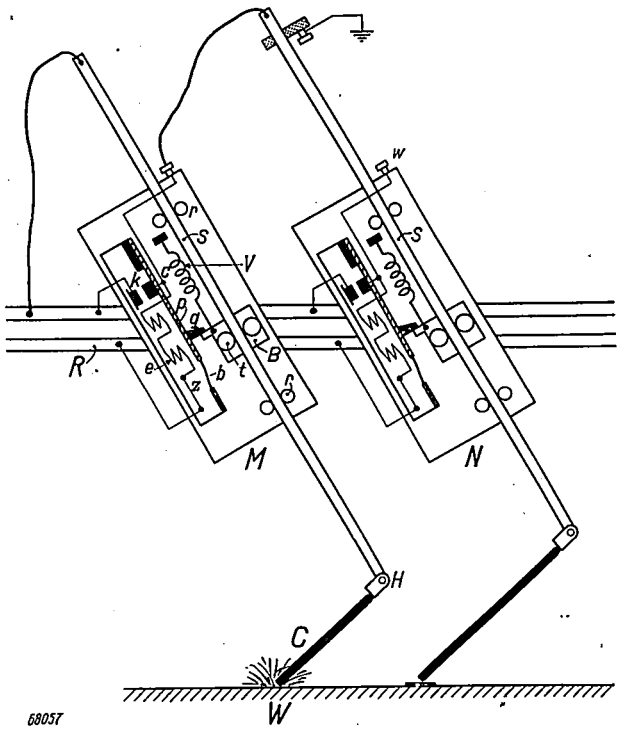


Fig. 7. Schematic drawing of two welding heads M and N, each fitted with a system for automatically retracting the guided bar. The electromagnet e is energized by the current beginning to flow through the welding electrode of the next head. It then attracts the steel armature p, the pawl q thereby releasing a stop, the spring V then clamping and drawing up the bar; at the same time the heavy contact k is closed, thereby shorting the winding of the magnet and allowing the welding current, of say 300 A, to flow from the welding head N via the terminal c and contact k to the rail R. The armature p is held in its new position by the leaf spring b, which has two states of equilibrium. z is a safety fuse.

As soon as current begins to flow through the welding electrode of the head N, this current (at first amounting to only a few amperes) energizes the electromagnet in the head M, thereby attracting the armature p towards it and releasing a pawl which brings a strong spring into action. This spring first clamps the guided bar onto a block by means of two rollers t and then draws up the block with the bar, thus completing the retraction of the remainder of the electrode in the holder of the welding head M. At the same time that the armature p is drawn towards the electromagnet it closes a heavy contact k, thereby shorting the winding of the electromagnet and connecting the terminal c — which now has to carry the full welding current of about 300 A for the welding head N — direct to the rail. The plate p is then again held in its new position by the leaf spring b, which has an unstable middle position, thus tending to spring in or out.

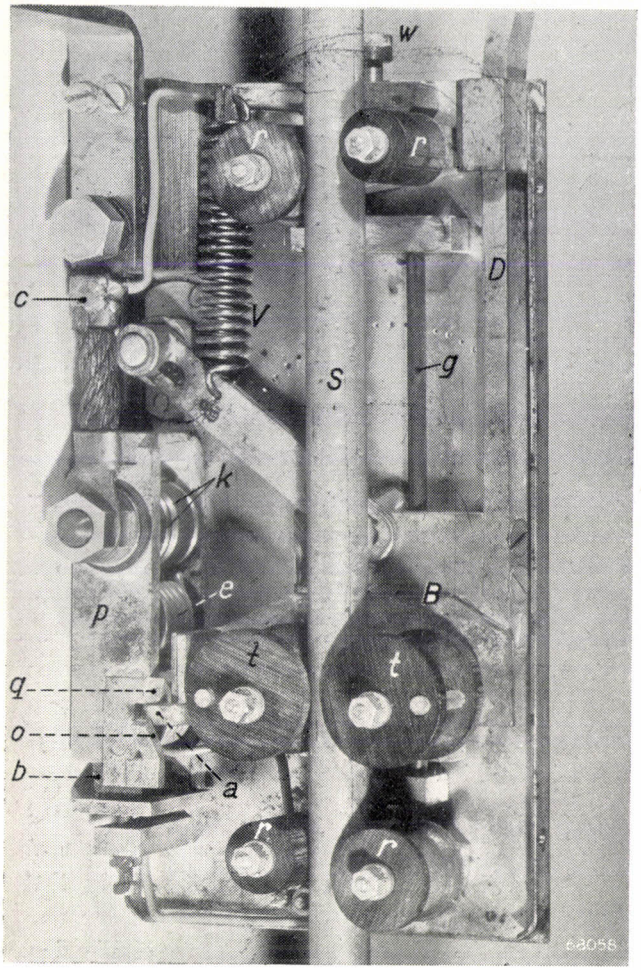
The manner in which the guided bar is clamped and drawn up is illustrated in *fig. 8* and explained in the subscript [4]). In the clamped position the bar can be drawn up still farther for inserting a new electrode in the holder ready for the next welding pass. When the first workpiece has been finished with and the next one is put into position for welding, the spring in each welding head is reset by hand, the guided bar being thereby released and the armature p with pawl q and the contact k returned to its original position. This can likewise be followed in fig. 8.

The terminal c in each welding head N is connected to a contact w on the top of the same head (see figs 6 and 7). Instead of connecting c or w to the next welding head an earthed contact can be mounted on the bar of the head, the electromagnet of this head then being energized as soon as its own bar has dropped a certain predetermined distance. This system is applied to the last welding head of a series, which thus automatically cuts out and stops the welding of the workpiece.

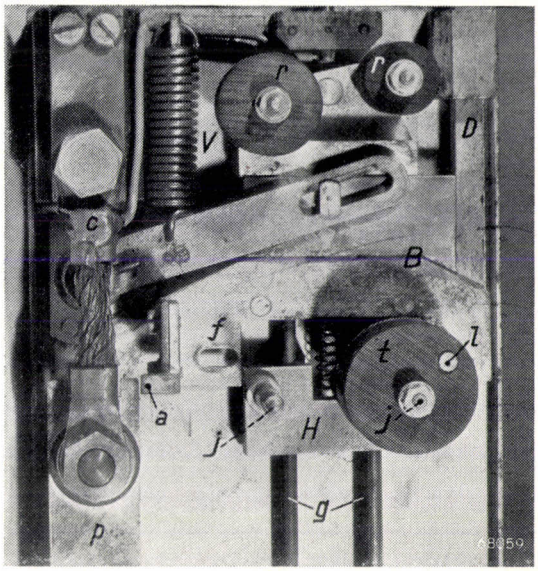
Working with the Contact automatic welder

Obviously the advantages of the automatic welder are most manifest in cases where the same kind of welded joint has to be made over and over again, such as in mass production. The time taken in setting up the rail and positioning the series of welding heads, a job that has only to be done once, is then negligible compared with the time saved in welding. When welding by hand often a great deal of time is wasted in making the joints between one run and the next: when one electrode has been consumed a new one has to be put in the holder, and this often takes so long that in the meantime the pool has set, in which case the slag has to be removed from the end of the last bead, in order to avoid slag inclusions, before starting with the new electrode. When welding automatically with a series of welding heads there is no such interruption, each new electrode taking over the hot bead from the previous one, and when the weld is completed the joints are scarcely perceptible. The useful running time of the welding-current generator is much longer and, as already remarked, the operators do not need so much training, or even practically none at all. Also the operators' time

[4]) Acknowledgement is due to J. B. van der Wal of Philips Welding Rod Factory at Utrecht (Holland) for his cooperation in designing this construction.



a



b

Fig. 8. *a*) Close-up of the opened welding head, showing, i.a., the parts represented in fig. 7. The sliding block *B* carrying the clamping rollers *t* will glide upward along two pins *g* (only one can be seen here), under the tension of the strong spring *V*, as soon as the stop *a* is released by the pawl *q* being attracted together with the armature *p* by the electromagnet. The block *B* can subsequently be lowered again by pressing down the plunger *D*, thereby stretching the spring *V* again. In this motion of *B*, after sliding over the pawl *q*, the stop *a* forces the plate with pawl, via the sloping face *o*, forward into the starting position, which is stabilized by the leaf spring *b* and in which position the block *B* is then again held by the pawl *q*. *b*) Here the sliding block *B* is in the raised position. The guided bar carrying the welding electrode has been removed, as also one of the two clamping rollers *t*, to show the clamping mechanism. The clamping rollers are rotated about the eccentric spindles *j* fixed in the auxiliary block *H*. When the block *B*, to which the auxiliary block is connected by means of a small spring, is pushed right down then *H* strikes against a stop, the spring is compressed and, owing to the downward movement of the slots *f* in which the pins *l* of the clamping rollers are engaged, the two clamping rollers are rotated a little about *j*, such as to draw them apart. The guided bar is then able to slide freely between the rollers. When, however, the pawl *q* is raised and the block *B* starts moving upward, the small spring forces the auxiliary block *H* away from the block *B*, the two clamping rollers are turned towards each other so as to clamp the guided bar, and as *B* and *H* continue their upward movement the bar is carried along with them.

In this position the guided bar can be drawn up further by hand if desired, the frictional force acting upon the clamping rollers then rotating them away from each other and thus reducing the clamping force. The guided bar cannot, however, slide down because then the frictional force acting upon the clamping rollers would cause the latter to rotate towards each other and thus increase the clamping force.

can be fully utilized, since the stubs of the electrodes can be removed and new electrodes put in the holders for the next run while the welding is proceeding. If sufficient power sources are available it is possible for two series of welding heads to be operated simultaneously by one man, either on two workpieces or on the same workpiece if it is of great length. The latter method has the advantage that excessive transverse shrinking can be avoided: the two series can be started from the same point in the middle of the workpiece and continued in opposite directions outward.

The power-supply rail can be set up according to the nature of the workpieces to be welded. If they are easily transported a permanent set-up is recommended, with a specially adapted bench or jig for quickly aligning the workpieces. If the workpieces are so heavy that they can only be lowered onto the bench with an overhead crane then it may be advantageous to be able to swing the rail back by hingeing it onto a wall or frame. In cases where the workpiece cannot be transported, as for instance when welding a floor or a ship's deck, a mobile frame carrying the rail and the welding heads is necessary.

Fig. 9 shows an example of welding a workpiece in a fixed position.

In principle overhead or vertical welding is also possible, with welding heads of a somewhat modified form. The pressure then required on the electrodes can be applied by means of a set of weights with cables passing over pulleys.

The foregoing should suffice to show that the Contact automatic welder is a very flexible machine easily adapted on the one hand to the welding equipment available and, on the other hand, to the requirements of the particular job.

Summary. Since Contact welding electrodes allow of touch-welding, the welder is relieved of the task of "holding the arc", this being done, as it were, automatically. Also the forward travel of the rod and the change-over from one electrode to the next can be carried out automatically, so that welds of any length can be made entirely automatically. For this purpose Philips have developed the Contact automatic welder. This consists of a "welding head" fitted with an oblique bar sliding in its own longitudinal direction and with the welding electrode clamped to its lower end at a given angle. The weight of the bar presses the tip of the electrode onto the workpiece and as the electrode melts down it is carried along parallel to itself. The rate of travel (and thus the thickness of the weld) is adjusted by varying the angles at which the bar and the electrode are set. A series of such welding heads are mounted at appropriate intervals along a rail above the workpiece, and each electrode comes to rest at the point where its bead begins. Each electrode takes over the arc from the previous one automatically as soon as the latter is consumed, the preceding electrode holder then being automatically retracted by means of an electromagnet in the head. In this way welding can be done in mass production with unskilled personnel, while still obtaining very good and uniform welds. The joints between the beads are scarcely perceptible.

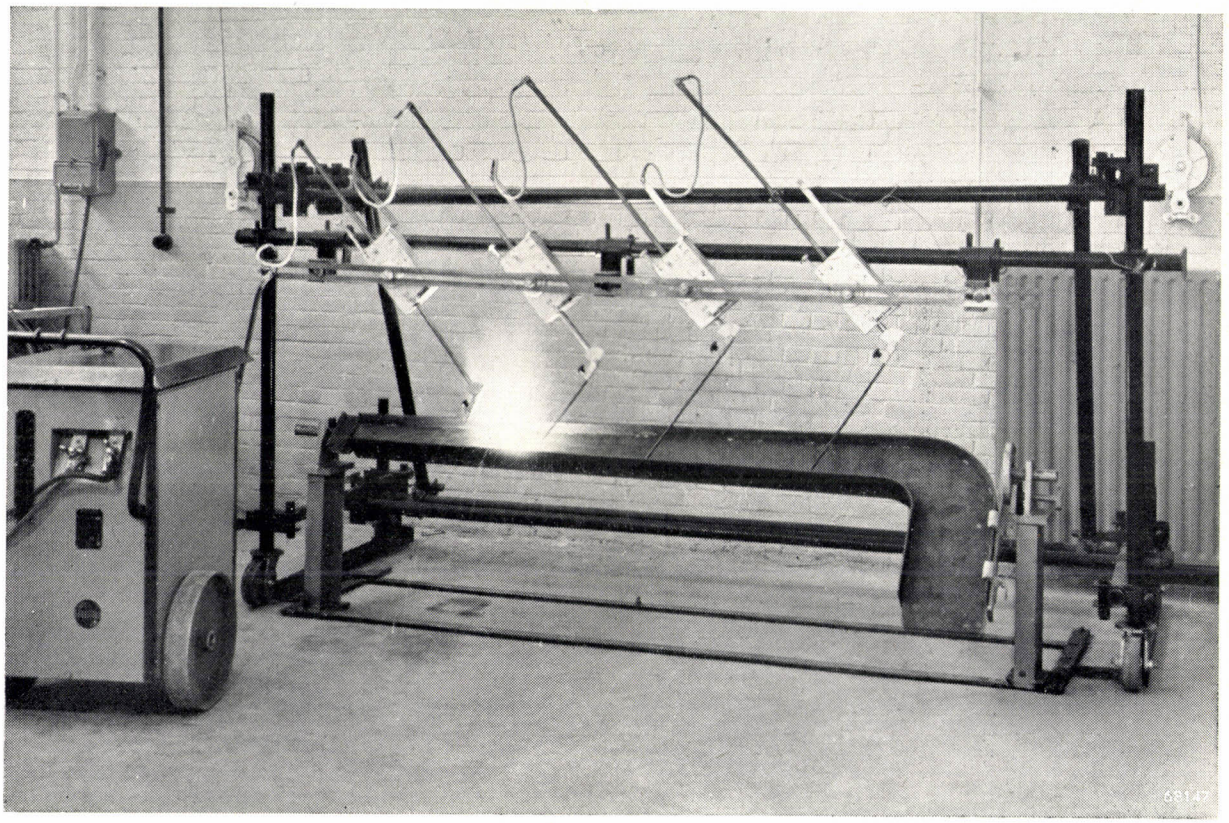


Fig. 9. The Contact automatic welder at work. The electrode of the first welding head has been consumed and the bar drawn up. Welding is being continued with the second welding head.

APPARATUS FOR TESTING TRANSISTORS

by P. J. W. JOCHEMS and F. H. STIELTJES.

621.314.632:621.315.59:621.317.7

In America some years ago a circuit element was invented which, like the triode, is able to amplify a signal. In its most common form this circuit element, called transistor, consists of a block of germanium with one large electrode and two small, pointed electrodes placed in close proximity to each other.

The transistor is still in a stage of development, the ultimate outcome of which is uncertain. The subject, however, is of such importance that Philips have also taken it up energetically. The testing apparatus described here has proved to be very useful in carrying out these investigations.

In the early stage of radio-telegraphy the crystal detector was for many years the most reliable and most used detector, until in the first world war it had to make place for the triode, which has the advantage of being able to amplify a signal at the same time. Some twenty years later, with the advent of centrimetric-wave technique (radar), for which at first there were no suitable tubes available, the crystal detector again came to the fore.

There has been no lack of attempts to develop this detector into a circuit element capable of functioning as an amplifier, like the triode. In America these efforts have been crowned with success by the invention of the transistor [1]). It is not the intention of this article to give a detailed account of the working and uses of the transistor, which can be found in the literature on the subject [2]). Here only a brief summary will be given, before proceeding to describe two types of apparatus with which the main characteristics and the gain factor, respectively, of transistors can be determined.

Principle of the transistor

In its most common form a transistor consists of a semi-conducting crystal, soldered onto a metal plate (the base, B in *fig. 1*), and two metal points bearing resiliently on the crystal. A suitable material for the semi-conducting crystal is germanium with a surplus of conduction electrons ("type N"). One of the metal points is called the emitter (E), because it sends into the crystal a controlling current, and with it charge carriers, while the other is called the collector (C), because it collects the charge carriers introduced with the controlling current.

The contact of each of the metal points with the germanium has a rectifying action, and in the case of type N germanium this allows a positive current to flow much more readily from the point to the germanium than in the other direction.

The working of the transistor can be demonstrated with the aid of the circuit represented in fig. 1. A small potential, positive with respect to the base, is applied to the emitter, such that a current of 0.5 to 1 mA flows in the forward direction. The collector is given a much higher negative potential, such that a current of several milliamps flows in the opposite direction. Connected in series with the collector is a resistance, the load resistor R_2, of say 20,000 ohms (i.e. much higher than the resistance in the emitter circuit).

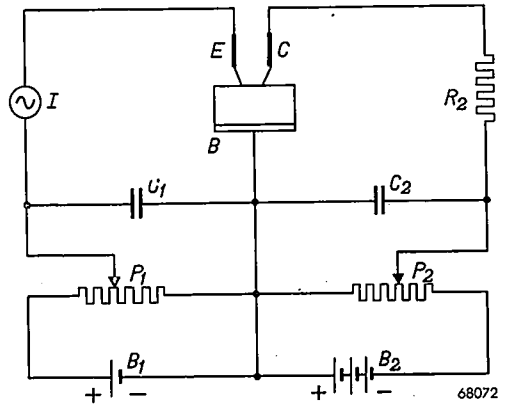


Fig. 1. Amplifying circuit with a transistor (germanium crystal with base B, emitter E and collector C). I source of input current or voltage. R_2 load resistor. B_1 and B_2 direct-voltage sources with potentiometers P_1 and P_2 and bypass capacitors C_1 and C_2.

When the emitter current is slightly varied, e.g. by incorporating an alternating-current source in the emitter circuit, then a variation in current, i.e. an alternating current, occurs likewise in the collector circuit. This is particularly the case when

[1]) J. Bardeen and W. H. Brattain, The transistor, a semi-conductor triode, Phys. Rev. **74**, 230-231, 1948.

[2]) See, e.g., J. Bardeen and W. H. Brattain, Physical principles involved in transistor action, Phys. Rev. **75**, 1208-1225, 1949, and W. Shockley, Electrons and holes in semi-conductors, with applications to transistor electronics, Van Nostrand Co., Inc., New York 1950.

the emitter and the collector are very close together, say at a distance of 0.05 mm. It is even possible for the amplitude of the alternating current in the collector circuit to be greater than that in the emitter circuit. But even without that there may be a considerable A.C. power gain, since the resistance in the collector circuit is much greater than that in the emitter circuit.

Thus, under suitably chosen conditions, the transistor is able to amplify a signal, and when the output is fed back to the input it can be caused to oscillate. So far, it is therefore like the triode. Practical advantages of the transistor over the triode (or other amplifying valves) are: (1) the absence of a heated cathode, (2) the smaller dimensions in which it can be made (*fig. 2*), and (3) its practically

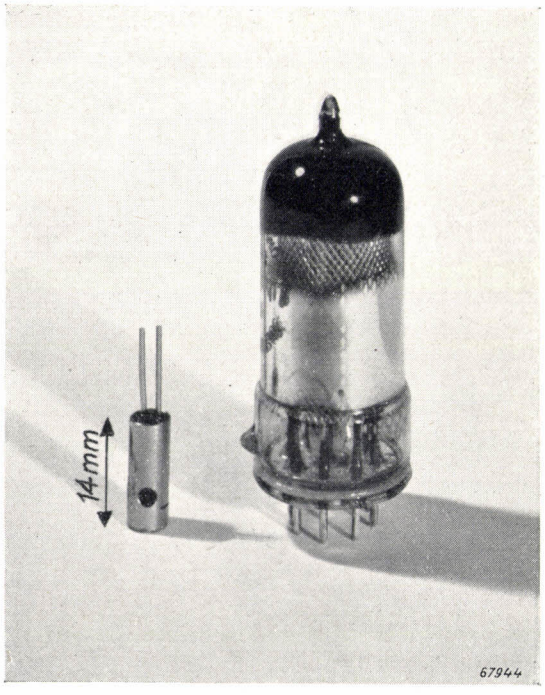


Fig. 2. True-to-scale photograph of a Philips transistor, with an EF 42 pentode for comparison. (It should not be concluded that the transistor and the pentode could replace each other.)

unlimited life under normal use. These advantages make the transistor particularly attractive for those applications where hitherto large numbers of tubes have been employed, e.g. in telephone repeaters and electronic computers. Before that stage is reached, however, there are still some difficulties to be overcome.

Notwithstanding the analogy mentioned between a transistor and a triode, there is an important difference in their working. The triode — provided it is so adjusted that no grid current flows and if it is used at not too high frequencies — does not take up any input power. The input circuit (emitter circuit) of the transistor, on the other hand, has a fairly low impedance and a certain amount of power is absorbed. The consequences of this for the transistor technique are: (1) that it is just as important to speak of power gain as of voltage gain and (2) that account has to be taken of the matching, that is to say, that the internal resistance of the input-voltage source and the load resistance each have in themselves the most favourable value to give either maximum power gain or undistorted output, as required. (As a matter of fact the same applies for valves which are used in such a way that they do need a certain input power, as for instance transmitting tubes, which are mostly so adjusted that grid current flows.)

In what follows, an apparatus is described by means of which the most important characteristics of a transistor can be displayed on the screen of a cathode-ray oscilloscope and the power gain can be measured directly.

Oscillographic recording of transistor characteristics

When dealing with a transistor one has to consider the currents i_e and i_c flowing through the emitter and the collector respectively, and the voltages v_e and v_c between the emitter and collector, respectively, and the base. (With all these symbols the instantaneous values are meant.) Each of these quantities plotted as a function of one of the others forms a transistor characteristic.

In principle, for instance, the emitter characteristic $v_e = \mathrm{f}(i_e)$ at $i_c = 0$ can be traced on an oscilloscope with the aid of a circuit as represented in *fig. 3a*. A variable alternating voltage derived from a transformer is applied in series with the transistor (emitter and base) and a resistor R. The voltage across R is proportional to the current i_e flowing through the transistor. This voltage is applied to the input terminals for the horizontal deflection of a cathode-ray oscilloscope, the voltage v_e (between the emitter and the base) being applied to the input terminals for the vertical deflection. An oscillogram is then obtained of the characteristic $v_e = \mathrm{f}(i_e)$.

When, however, a normal oscilloscope is used for this purpose one is faced with two difficulties. The first is that the voltages v_e and i_eR are too small to give sufficient deflection without being amplified. Also owing to the rectifying action both these voltages have a direct-voltage component. The amplifiers of a normal oscilloscope such as the type

GM 3159 [3]) do not amplify direct voltages, so that this component cannot be represented in the oscillogram. Consequently the position of the coordinate axes — the direction of which is known — remains undetermined. Other types of oscilloscopes, such as the GM 3152 [4]) and the GM 3156 [5]), have only one amplifier, and this again is not a direct-voltage amplifier.

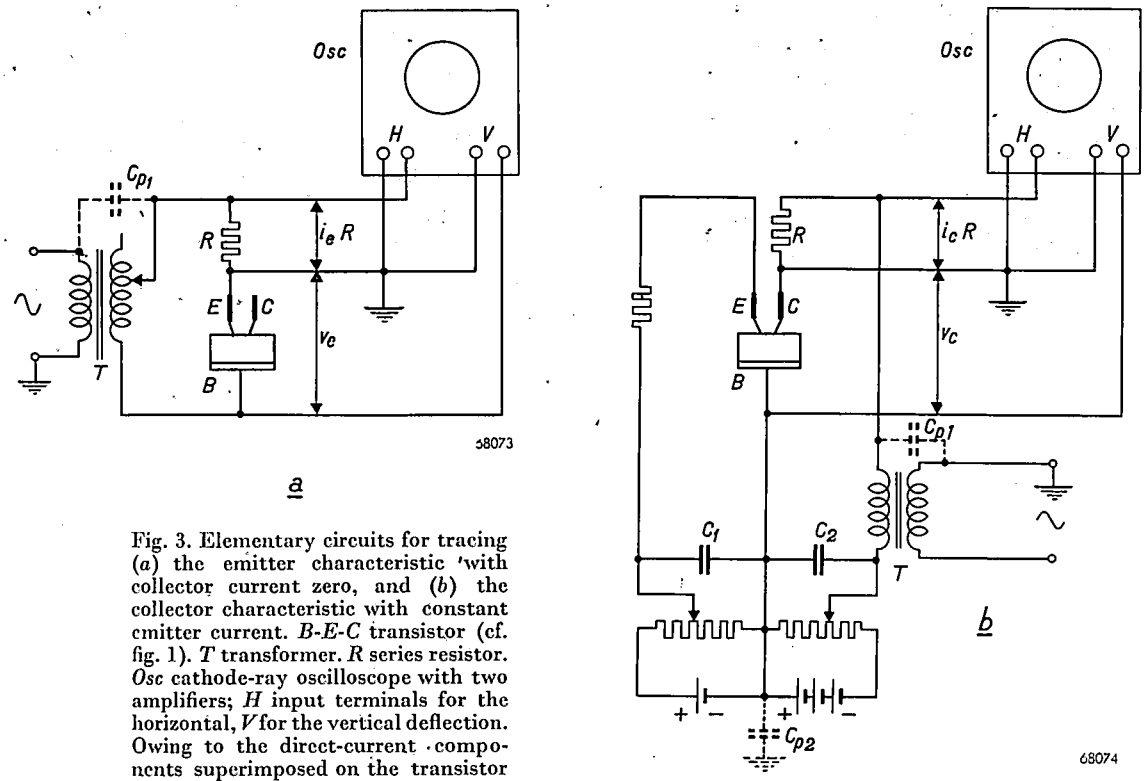


Fig. 3. Elementary circuits for tracing (*a*) the emitter characteristic with collector current zero, and (*b*) the collector characteristic with constant emitter current. *B-E-C* transistor (cf. fig. 1). *T* transformer. *R* series resistor. *Osc* cathode-ray oscilloscope with two amplifiers; *H* input terminals for the horizontal, *V* for the vertical deflection. Owing to the direct-current components superimposed on the transistor currents as a consequence of the rectifying action, the two amplifiers would have to be direct-voltage amplifiers. The stray capacitances (C_{p1}, C_{p2}) arising from the one-sided earthing of the input terminals cause the formation of a loop in the oscillogram (see fig. 4).

The second difficulty arises from the fact that in a normal oscilloscope the inputs of the two amplifiers are asymmetric (earthed on one side) and thus have one common terminal.

In the first place this fact precludes the use of such amplifiers in cases where there is *no* common terminal for the two pairs of terminals whose voltage differences are to be traced as functions of each other. But even when the pairs of terminals have one point in common (in the case of fig. 3*a* the emitter), which is then earthed, interference arises as a consequence of inevitable stray capacitances. The capacitance C_{p1} between the coils of the transformer in fig. 3*a*, for instance, is in parallel with the resistance R. The voltage across R is therefore no longer proportional to the current i_e at every instant, since only part of the alternating-current components of i_e flow through R, and moreover there is a stray alternating current flowing from the non-earthed mains terminal via C_{p1} and R to the other (earthed) terminal. As a consequence of all this, the forward and return lines of the characteristic in the oscillogram do not exactly coincide but form quite a wide loop (see *fig. 4*).

Further, the capacitance of the D.C. sources with respect to earth may form a shunt, as seen from fig. 3*b*. The object of this circuit is to record the collector voltage v_c as a function of the collector current i_c for given values of the emitter current. If, owing to the oscilloscope being connected to the circuit, the collector is earthed, then the stray capacitance C_{p2} of the D.C. sources comes to lie between the collector and the base, whilst the stray capacitance C_{p1} of the transformer T is shunted across the resistance R, from which the voltage for the horizontal deflection is derived.

When the characteristic of a vacuum diode is to be traced it is already known in advance that at low frequencies (where the effects of electrode capacitance and transit time of the electrons are negligible) a properly recorded characteristic will not

[3]) Described in Philips Techn. Rev. **9**, 202-210, 1947.
[4]) Described in Philips Techn. Rev. **4**, 198-204, 1939.
[5]) Described in Philips Techn. Rev. **5**, 277-285, 1940.

show any loop. So if a loop does appear in the oscillogram this can be ascribed entirely to a fault in the measuring equipment, as indicated in the case of fig. 3. In such a case a small loop is not very troublesome. With a new element such as the transistor,

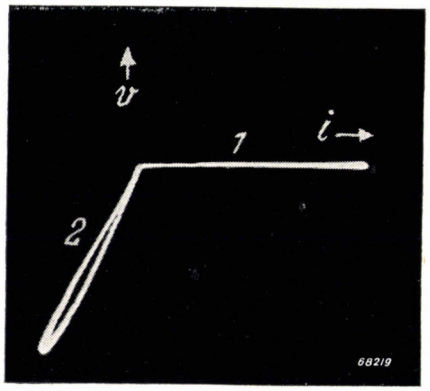


Fig. 4. Emitter characteristic traced by the system according to fig. 3*a*. *1* forward direction, *2* inverse direction (with perceptible looping). There are no coordinate axes.

however, one is not so sure that the formation of loops is due entirely to the apparatus employed. In the case of a breakdown, for instance, — we shall see an oscillogram of this presently — there is a noticeable development of heat which, owing to thermal inertia, may result in a real loop. For these purposes, therefore, the measuring equipment should be quite free of the fault in question.

An oscilloscope quite suitable for the purpose, which has two symmetrical direct-voltage amplifiers, is at the moment still in a stage of development. But a satisfactory solution has been reached by employing existing Philips' equipment, as will be shown in what follows.

Tracing the coordinate axes

The first objection — the absence of coordinate axes — can be met in a simple way by replacing each of the amplifiers of the oscilloscope by an electronic switch, one input of which is not used. The circuit is then as represented diagrammatically in *fig. 5*, where for the sake of clarity the electronic switches have been drawn as ordinary two-pole two-way switches, thus ignoring the fact that they act as amplifiers. This manner of representation — which implies that a direct-voltage component of the input voltages is also transmitted — is permissible because owing to the switching these direct voltages are converted into alternating voltages, of rectangular wave form, which, provided the switching frequency is sufficiently high, are transmitted without any distortion.

The switches S_V and S_H are to be imagined as operating at different frequencies. These frequencies are chosen, for instance, higher than the frequency of the voltage supplied by the transformer T. There are then always intervals where both switches are in the position shown in fig. 5 with fully drawn lines. A part of the characteristic — in the present case $v_e = f(i_e)$ — is then displayed on the screen of the oscilloscope. A moment later S_V, for instance, will have assumed the position indicated by a broken lines. The vertical-deflection voltage is then zero, but not so the horizontal-deflection voltage, so that the spot describes on the screen a small part of the horizontal axis. A little later both switches are in the broken-line positions and the spot is then at the origin. Finally, in the intervals during which S_V is in the fully-drawn position and S_H in the broken-line position a part of the vertical axis is traced on the screen. Thus the oscillogram shows the characteristic (as a broken line), the two coordinate axes (at least for the part corresponding to the sweep of the currents or voltages) and the origin.

An electronic switch suitable for our purpose is the type GM 4580 [6]), the switching frequency of which is variable between 2 c/s and 40,000 c/s (with the latest type, GM 4580/01, up to 50,000 c/s).

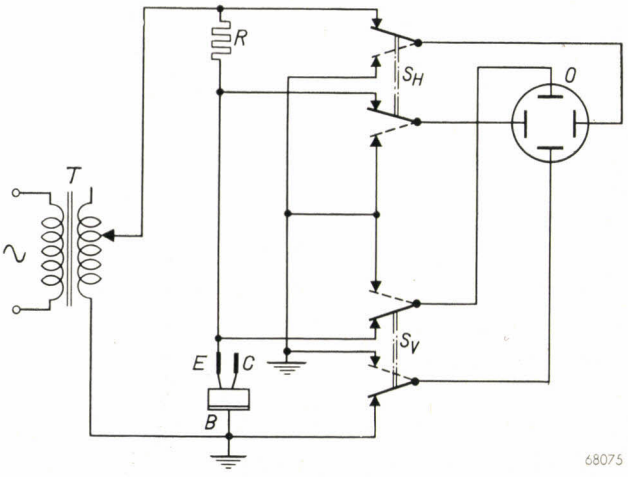


Fig. 5. The switches S_H and S_V represent electronic switches taking the place of the amplifiers in the oscilloscope *Osc* of fig. 3. *O* oscilloscope tube. With this system not only the characteristic but also the two coordinate axes and the origin are traced.

Tracing two voltages without a common terminal

When the inputs of the amplifiers of the oscilloscope have one common terminal, and as a consequence the point of the circuit connected to that terminal is earthed, then, as shown with the aid

[6]) Described in Philips Techn. Rev. **9**, 340-346, 1948.

of fig. 3, stray capacitances have a disturbing influence. Two entirely independent inputs, on the other hand, permit of some point being earthed which will eliminate this influence. If, for instance, the oscilloscope of figs 3*a* and *b* had this property, the point B could be earthed, in which case the stray alternating current originating from the mains and flowing via C_{p1} would no longer flow to earth through R but through the transformer secondary, and then neither C_{p1} would form a shunt across R nor (in fig. 3*b*) would C_{p2} form a shunt across the collector. Moreover, with an oscilloscope having independent inputs it would, of course, be possible to trace the voltage v_V between two terminals *1* and *2* as a function of the voltage v_H between two other terminals *3* and *4*, independently of any voltage between terminal *1* or *2* on the one hand and terminal *3* or *4* on the other hand.

It will be shown that this independency of the two inputs can be obtained when the amplifiers (in this case the electronic switches) have a symmetrical input and are provided with negative feedback in a particular way, as represented in *fig. 6*.

Let us consider first only one of the amplifiers, say A_V. The potentials of the input terminals with respect to earth are denoted by v_a and v_b, those of the output terminals by v_A and v_B. Now:

$$v_a = \tfrac{1}{2}(v_a + v_b) + \tfrac{1}{2}(v_a - v_b)$$

and

$$v_b = \tfrac{1}{2}(v_a + v_b) - \tfrac{1}{2}(v_a - v_b)\,,$$

where $v_a - v_b$ is the input voltage v_V and $\frac{1}{2}(v_a+v_b)$ is the mean potential of the input terminals with respect to earth.

The gain m, which we shall assume to be much larger than unity, is the ratio of the voltage difference $v_A - v_B$ between the output terminals to the voltage difference $v_a - v_b$ between the input terminals. Thus:

$$v_A - v_B = m(v_a - v_b) = mv_V\,.$$

The quantity $v_A - v_B$ determines the magnitude of the vertical deflection on the oscillograms. Thus this deflection — except for a condition to be dealt with presently — is independent of the voltage $\frac{1}{2}(v_a+v_b)$. This means that v_a and v_b need not be regarded as voltages with respect to earth, but that they may equally well be regarded as voltages with respect to some other point having itself an arbitrary potential with respect to earth.

This applies, of course, also to the horizontal deflection if the respective amplifier (A_H; input voltage v_H) is constructed in the same way. It is therefore possible to trace a voltage v_V as a function of a voltage v_H independently of the voltage level of each pair of input terminals, thus without it being necessary for the terminal *1* or *2* to have the same potential as the terminal *3* or *4*.

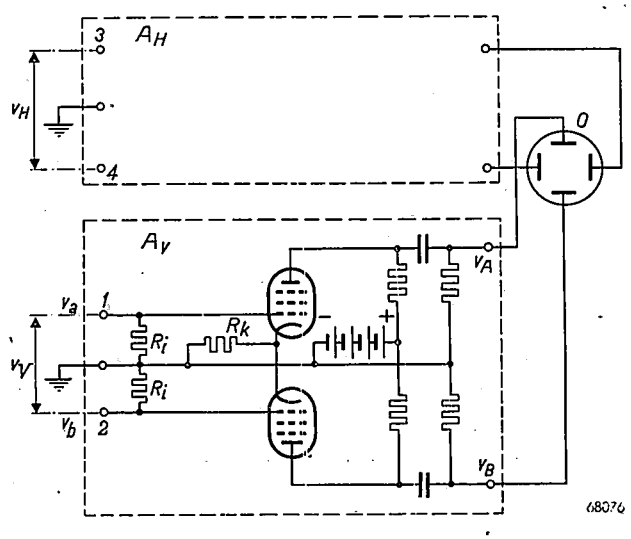


Fig. 6. Oscilloscope tube O with identical amplifiers A_H and A_V, for the two directions of deflection. R_i input resistors, R_k cathode resistor. The oscillogram obtained represents $v_V = f(v_H)$, to a certain extent independently of the voltage across the terminal pairs *1-2* and *3-4*.

So far, however, no account has been taken of the influence that the sum voltage, v_a+v_b, may have not only upon the amplitude of the deflection but also upon the focusing of the oscillogram.

The first effect of this is that $\frac{1}{2}(v_a+v_b)$ might be so large as to cause the valves to be overloaded, distortion then occurring in the amplifier. This is amply counteracted by connecting a large (non-bypassed) resistor R_k in the common cathode lead of the valves connected in push-pull. The sum voltage influences the two cathode currents in the same sense, so that the current through R_k changes according to the sum voltage and a large proportion of $\frac{1}{2}(v_a+v_b)$ is taken up in R_k. In other words, R_k gives a heavy negative feedback for the sum voltage.

The difference voltage, $v_a - v_b$, on the other hand influences the two cathode currents in opposite senses and thus the current passing through R_k remains unchanged. The gain m of the difference voltage therefore remains unaffected by the introduction of R_k.

The voltages from the output terminals are now

$$v^A = n\cdot\tfrac{1}{2}(v_a + v_b) + m\cdot\tfrac{1}{2}(v_a - v_b), \quad \ldots \quad (1)$$

and

$$v_B = n\cdot\tfrac{1}{2}(v_a + v_b) - m\cdot\tfrac{1}{2}(v_a - v_b), \quad \ldots \quad (2)$$

where $n \ll m$.

As is the case without the feedback resistor, the voltage v_A—v_B determining the amplitude of the deflection is thus equal to $m(v_a$—$v_b)$.

However, importance is also to be attached to $\frac{1}{2}(v_A+v_B)$, the mean potential of the output terminals with respect to earth, since this voltage is at the same time the mean potential of the pair of deflecting plates with respect to the (earthed) accelerating anode of the cathode-ray tube. This potential defocuses the electron beam and therefore, if a sharply focused oscillogram is to be obtained, it has to be kept low. This defocusing is the second effect of v_a+v_b referred to above.

From (1) and (2) it is found that

$$\tfrac{1}{2}(v_A + v_B) \doteq n\cdot\tfrac{1}{2}(v_a + v_b)\,,$$

and from this it is seen that in order to minimize the effect under consideration n has to be made small, which is a further reason for having a heavy negative feedback for the sum voltage.

Further particulars about the use of the circuit

The circuit at which we have arrived in the foregoing is that depicted in *fig. 7*. It consists of two electronic switches GM 4580 (*H* for the horizontal, *V* for the vertical deflection) and an oscilloscope tube with accessories. Owing to the symmetrical output of the electronic switches a tube is used which has four symmetrical deflecting plates, e.g. the type DN 9-4.

The most important parts of one of the electronic switches are shown in fig. 7. These comprise two double pentodes EFF 51, each with a cathode resistor R_{kI} and R_{kII}. An auxiliary current, i_{comm}, of rectangular wave form, derived from a multivibrator (not shown here), is sent through each of these resistors. These two auxiliary currents (switching currents) are in antiphase, so that the tubes are alternately cut off by the voltage $i_{\text{comm}}R_k$. The frequency of the current i_{comm} is variable, as already mentioned.

A variable positive or negative direct voltage (E_I, E_{II}) can be applied to one of the control grids of the EFF 51 tubes to shift the oscillograms corresponding to the inputs *I* and *II* with respect to each other [7]), in the horizontal direction in the case of the upper electronic switch in fig. 7 and in the vertical direction in the case of the lower switch. These direct voltages do not as a rule interfere with the circuit to which the electronic switch is connected, but such cannot be said to be the case when the object under test is a transistor. From fig. 7 it is seen that E_{IV} for instance affects the D.C. adjustment of the transistor (though via high resistances). For this reason the variable direct voltage at the two inputs used (*I* in fig. 7) has been made zero (this is the only change that had to be made in the electronic switches GM 4580). At the other two inputs, which are not connected to any other point, use is then made of the variable direct voltage (E_{II}) in order to get the coordinate axes in the right place. This is brought about in the following way.

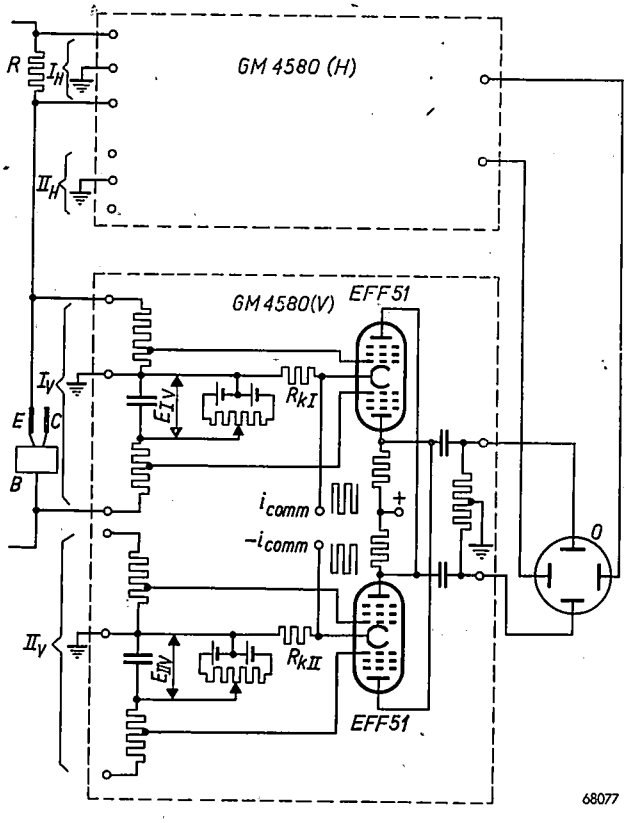


Fig. 7. Oscilloscope tube *O* with two electronic switches GM 4580 (*H* = horizontal, *V* = vertical), each with two symmetrical inputs and two double pentodes EFF 51. R_{kI}, R_{kII} cathode resistors. i_{comm} and $-i_{\text{comm}}$ switching currents. E_{IV} and E_{IIV} variable direct voltages (here the voltages E_{IV} and E_{IH}, not indicated, have to be zero).

The apparatus is started up, but without sending current through the transistor. Owing to the switching each of the electronic switches then supplies an alternating voltage of rectangular wave form, the amplitude of which depends upon E_{II_V} or E_{II_H} respectively. These voltages are not of the same frequency, since the switching frequencies have been differently adjusted. As is easily understood, the oscillogram then shows four dots situated in the corners of a rectangle. The width of the

[7]) See the article quoted in note [6]), pp 345 and 346.

rectangle depends upon the direct voltage E_{II_H} at the electronic switch for the horizontal deflection, while its height is governed by the direct voltage E_{II_V} at the swith for the vertical deflection. By changing these direct voltages the dots can be made to coincide with each other, this then giving the correct adjustment. The fact that for this adjustment E_{II_H} and E_{II_V} are not as a rule zero is due to small differences between the valves. In course of time these differences may vary slightly, so that the adjustment has to be repeated now and again.

Upon a current being sent through the transistor both its characteristic and the two axes appear on the screen, as already described.

An oscilloscope tube is rather more sensitive at the pair of deflecting plates nearest to the electron gun than at the other pair. It is therefore advisable so to adjust the gains of the electronic switches that this difference is just compensated. In what follows it will be assumed that this has been done.

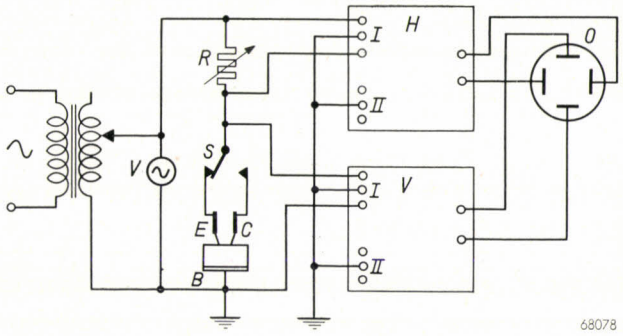


Fig. 8. Circuit for tracing the emitter and collector characteristics.

Some transistor characteristics

Emitter characteristic at $i_c = 0$ and collector characteristic at $i_e = 0$

Fig. 8 represents the circuit with which, according to the position of the switch S, one of the two so-called diode characteristics (the emitter characteristic for collector current zero and the collector characteristic for emitter current zero) can be traced.

As already remarked, both the emitter and the collector contacts have a rectifying action. Thus the diode characteristics show a sharp bend near the origin. It depends upon the value of the resistor R which of the two parts of the characteristic shows up clearly: if the value of R is of the same order as the resistance in the forward direction (*fig. 9a*) then the forward characteristic is clearly shown and the characteristic of the inverse direction almost

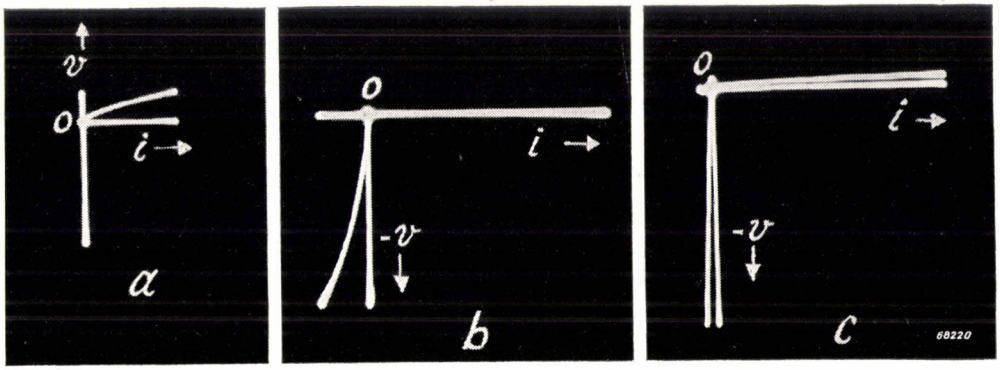


Fig. 9. Diode characteristics of a transistor with different values of the load resistor R: (a) R of the same order as the resistance in the forward direction, (b) R of the same order as the resistance in the inverse direction, (c) intermediate case.

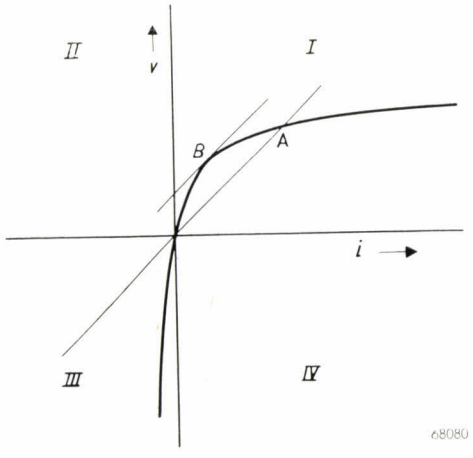


Fig. 10. Diode characteristic of a transistor on an enlarged scale. In the quadrant I lies the characteristic of the forward direction, in quadrant III that of the inverse direction. At the point A where the characteristic is intersected by a line drawn through the origin at an angle of 45° to the axes the resistance of the emitter (or collector) is equal to the value of the load resistor R. At B, where the tangent to the characteristic makes an angle of 45° with the axes, the difference resistance is equal to R.

coincides with the negative v-axis; if the value of R is much greater (fig. 9*b*) then the forward characteristic lies approximately on the i-axis, while the inverse characteristic is clearly seen. Fig. 9*c* represents an intermediate case where both parts of the characteristic can still be distinguished. It is for this reason that the resistor R has been made continuously variable, from 1800 ohms to 200 000 ohms. The value of R can be read from a scale.

In *fig. 10* the characteristic is shown on an enlarged scale. A line drawn through the origin at an angle of 45° to the axes intersects the characteristic at a point A, where the ratio of voltage and current is just equal to R. At the point B, where the tangent is at an angle of 45°, the difference resistance is equal to R. If it is desired to know at what value of current and voltage the resistance, respectively the difference resistance, has a certain value R', then the characteristic has to be traced with $R = R'$ and the points A and B determined along the curve.

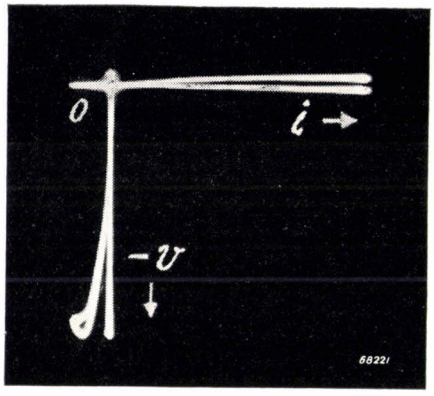


Fig. 11. When the voltage is raised too high breakdown occurs, this being made visible in the oscillogram because when the current is raised beyond a certain value the voltage drops. When measurements are taken with not too low frequencies (say 50 c/s) a loop is formed in the oscillogram.

A suitable value of the alternating voltage for carrying out the test is 100 V (r.m.s. value). Upon the voltage being raised a value is ultimately reached where breakdown occurs. This is clearly seen on the oscillogram (*fig. 11*); when the current is raised beyond a certain level this causes the voltage to drop. A loop is then formed (this time a real one!). The voltage at which this takes place can be read from a voltmeter.

Collector characteristics with the emitter current as parameter

When the collector voltage is traced as a function of the collector current for different values of the emitter current (kept constant in each case), a family of curves is obtained as represented in *fig. 12*.

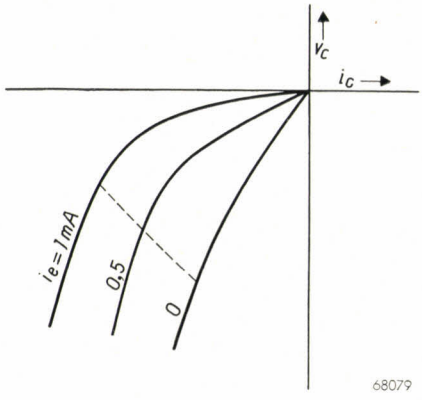


Fig. 12. Collector characteristic $v_c = f(i_c)$ at different constant values of the emitter current i_e.

When the emitter current is varied, say from 0 to 1 mA, the collector voltage and collector current vary according to the broken line drawn in fig. 12. If the scale of v_c is the same as that of $i_c R_2$ (where R_2 is the load resistance in the collector circuit), as is assumed to be the case here, then this line makes an angle of 45° to the axes. The length of this line is proportional to the sweep of i_c and thus a measure for the ratio of that sweep to the sweep of the emitter current: this ratio resembles the mutual conductance of an amplifying tube. With the aid of the apparatus described an oscillogram of this line can easily be produced (*fig. 13*) by recording v_c as a function of i_c with a given alternating current superimposed on the direct current in the emitter circuit.

The length and the position of the line depend, in general, upon the D.C. adjustments of the transistor. Just as it is often desired that the mutual conductance of an amplifying valve varies little

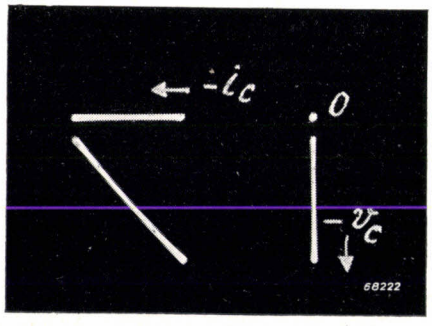


Fig. 13. Collector voltage v_c as a function of the collector current i_c, with an alternating-current source in the emitter circuit.

with the adjustment, so in many cases with a transistor it is desired that the ratio of collector-current sweep to emitter-current sweep varies little with the adjustment. In order to determine the influence

of the collector direct current, this current can be gradually varied by superimposing on the collector current an alternating current with a frequency much lower than that of the emitter current. In

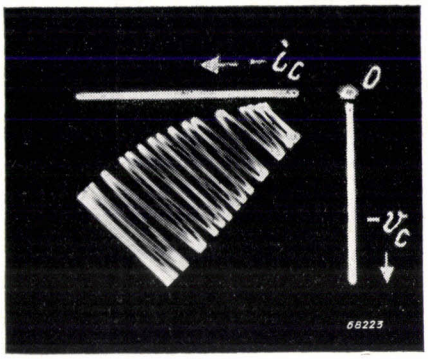


Fig. 14. Oscillogram of the collector voltage as a function of the collector current sweep with a frequency of 50 c/s, while the emitter current sweep with a frequency of 1000 c/s remains within certain limits. The ratio of the two frequencies in this case being exactly a whole number (20), the result is a Lissajous figure.

the oscillogram a series of lines are then traced running at an angle of about 45° (*fig. 14*). To show clearly how this oscillogram is built up, the ratio of the frequencies has been made just equal to a whole number; the oscillogram is then a Lissajous figure. In practice, however, one can work just as well with any ratio of frequencies (provided it is sufficiently large).

The field of characteristics traced in this manner is bounded on the one side by the characteristic $v_c = f(i_c)$ for the highest value of i_e and on the other side by the corresponding characteristic for the lowest value of i_e. As fig. 14 shows, these boundary characteristics run for a part almost parallel and thus the intermediate lines at an angle of

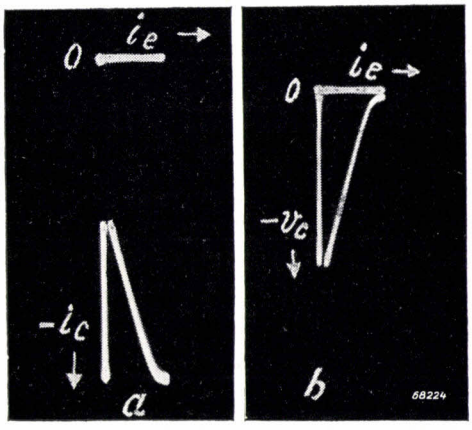


Fig. 15. *a*) Collector current, *b*) collector voltage, both as functions of the emitter current.

45° are approximately constant in length. Therefore within this area the requirement that the ratio of collector and emitter current sweeps should depend little upon the adjustment is satisfied.

Collector current and collector voltage as functions of the emitter current

As previously mentioned, with the apparatus described it is not necessary that the pairs of terminals from which the voltage differences are to be traced as functions of each other should have one common terminal. To demonstrate this, in *fig. 15* two oscillograms are shown which were taken without any common terminal, as may be seen from the circuit diagram in *fig. 16*: these are recordings of the collector current and of the collector voltage as functions of the emitter current.

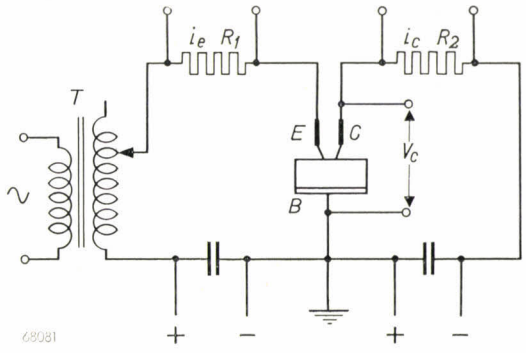


Fig. 16. Circuit for recording the collector current or collector voltage as a function of the emitter current. In one case $i_c R_2$ and in the other case v_c is recorded as a function of $i_e R_1$. It should be noted that R_1 has no point in common with the collector circuit.

The first of these oscillograms shows that within a certain area there is a practically linear relation between i_c and i_e (as also appeared from fig. 14). The second oscillogram shows a remarkable analogy with the characteristic of the anode current as a function of the grid voltage of an amplifying valve.

Transducer gain

In fig. 1 a transistor was represented connected as an amplifier. The voltage at the input is now assumed to be derived from an alternating-voltage source, with an r.m.s. value V_1 and internal resistance R_1. The r.m.s. value of the alternating voltage across the output resistance R_2 will be denoted by V_2.

An important factor is the ratio of the A.C. power developed in R_2 — thus V_2^2/R_2 — to the maximum output power of the alternating-voltage source V_1, i.e. $(\frac{1}{2}V_1)^2/R_1$. In communication technique this

ratio is called the transducer gain k, which is thus written as:

$$k = \frac{V_2^2/R_2}{(\frac{1}{2}V_1)^2/R_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This quantity has a maximum for a given value of R_2 and, with the present transistors, may assume a value of about 100.

If R_1 and R_2 are made variable, and their values are known, then, as (3) shows, k can be determined by measuring V_1 and V_2.

V_1' is used as input voltage V_1:

$$V_1 = aV_1' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

With the aid of the other arm of S_1 the value of R_1 is varied. The tappings on the transformer and on the resistor are so chosen that in the four positions of S_1 the ratio a^2/R_1 remains constant.

The lower arm of a second two-pole switch, S_2, short-circuits an adjustable part of the resistor in the collector circuit, the remaining part being R_2. This short-circuiting takes place across a capacitor

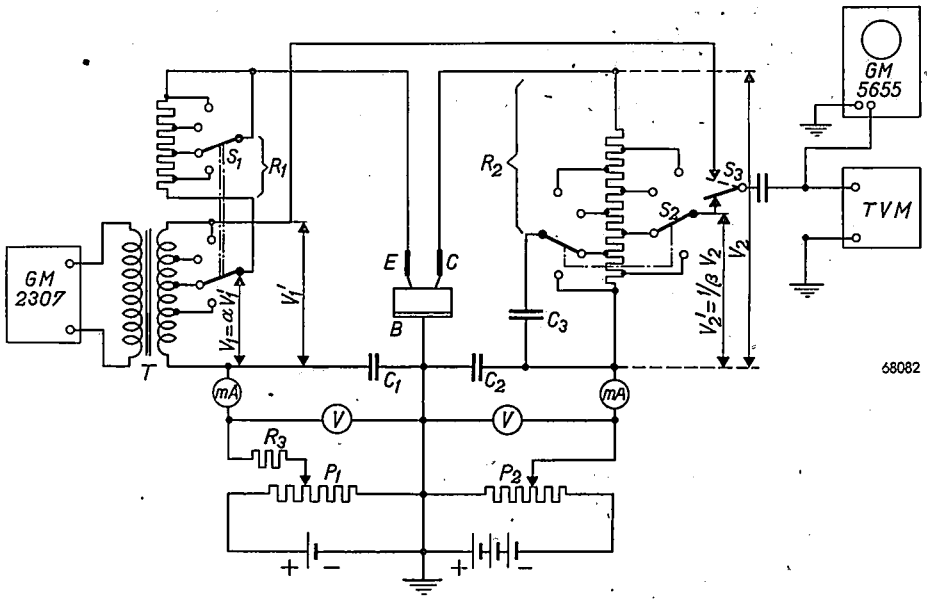


Fig. 17. Circuit for the direct reading of the transducer gain of a transistor (*B-E-C*). T input transformer fed from a signal generator GM 2307 adjusted to, say, 1000 c/s. S_1 switch with which a voltage $V_1 = aV_1'$ is tapped from the transformer and at the same time a certain resistance R_1 is included in the emitter circuit. By means of the switch S_2 the value of the resistance R_2 in the collector circuit is adjusted (C_3 is a capacitor to be regarded as a short-circuit for the alternating current) and at the same time the proportion β tapped from the alternating voltage across R_2. The tappings on the transformer and the resistors have been so chosen that $\beta^2/R_2 = 25\ a^2/R_1 =$ constant.

P_1, P_2 potentiometers for controlling the direct-current adjustments. R_3 high resistance maintaining the direct-current adjustment in the emitter circuit when varying R_1. C_1, C_2 bypass capacitors.

Via the switch S_3 the voltage V_1' is compared with V_2' on the valve voltmeter *TVM*, which has db and mW scales. When with V_1' the reading is 20 db, the gain with V_2' is read directly in decibels.

Our measuring set-up is based on this principle, but by an artifice the value of the transducer gain can be read directly in decibels. The circuit is represented in *fig. 17*.

Here, instead of the voltages V_1 and V_2 being measured, two voltages V_1' and V_2', related in a certain manner to V_1 and V_2 respectively, are compared one with the other (with the aid of the switch S_3 and a valve voltmeter). V_1' is the constant alternating voltage of the whole of the secondary winding of a transformer T fed, for instance, from a signal generator adjusted to 1000 c/s. Via the lower arm of a two-pole switch S_1 a proportion a of the voltage C_3, so that the variation of R_2 does not affect the D.C. adjustment of the collector [8]).

The second arm of S_2 makes contact with a tapping on R_1. The alternating voltage between this tapping and earth is measured with the valve voltmeter and denoted as V_2', i.e.:

$$V_2 = \beta V_2' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The two series of tappings on the resistor are so

[8]) Any change in the direct-current adjustment of the emitter when varying R_1 is counteracted by a resistor R_3 (fig. 17) of a much higher value than R_1, so that the current through R_3 is practically independent of R_1.

chosen that in the four positions of S_2 the ratio β^2/R_2 is constant and exactly 25 times the constant ratio α^2/E_1.

Substitution of (4) and (5) in (3) yields:

$$k = \frac{(\beta V_2')^2/R_2}{\frac{1}{4}(\alpha V_1')^2 R_1} = 4\frac{\beta^2/R_2}{\alpha^2/R_1}\left(\frac{V_2'}{V_1'}\right)^2 = 4 \times 25\left(\frac{V_2'}{V_1'}\right)^2 = 100\left(\frac{V_2'}{V_1'}\right)^2 .$$

If, for instance, we find that V_2' is just equal to V_1' then $k = 100$.

of k in decibels. As we have seen, $V_2' = V_1'$ corresponds to $k = 100$, i.e. 10 log 100 = 20 db, and $V_2 = pV_1'$ thus corresponds to $k = 100\,p^2$, i.e. 10 log $(100\,p^2)$ = (20 + 20 log p) db. The scale on the voltmeter is such that with a voltage V at the terminals the reading is:

$$c \text{ db} = 20 \log (V/V_0) \text{ db},$$

where V_0 is a voltage chosen as zero level. Thus when measuring V_2' the reading is:

$$a \text{ db} = 20 \log (V_2'/V_0) \text{ db},$$

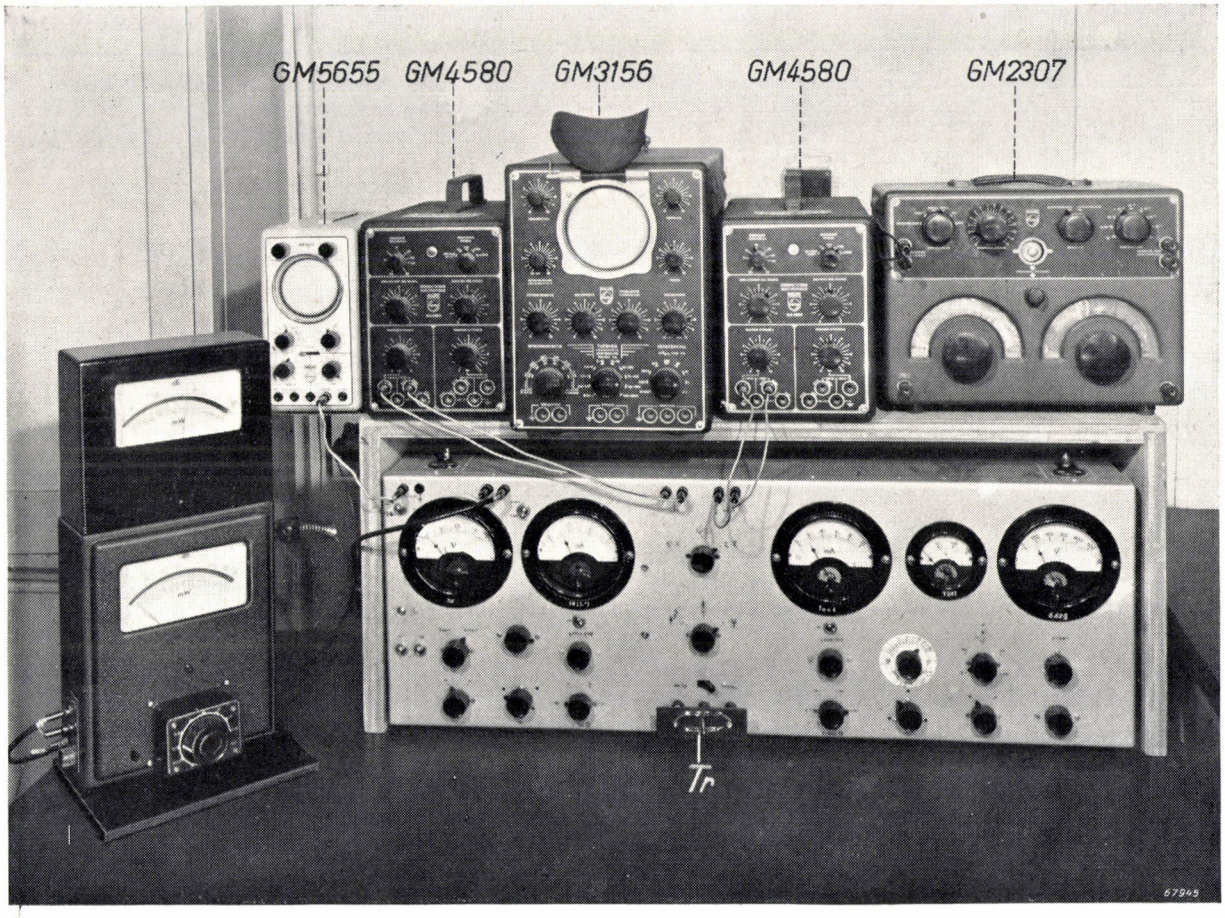


Fig. 18. General picture of the installation. *Tr* transistor under test. To the left and right of the transistor are the controls and meters for the direct-current adjustments and for the resistors in the emitter and collector circuits. Above that, from left to right: checking oscilloscope GM 5655 (cf. fig. 17), electronic switch GM 4580, oscilloscope GM3156, second electronic switch GM 4580, signal generator GM 2307. On the extreme left the valve voltmeter with two moving-coil meters.

When the switch S_3 is put in the position indicated by a broken line the voltage V_1' is applied to the terminals of the valve voltmeter. This meter could be so calibrated as to read 100 at a voltage having a suitable value to serve as V_1' and a value $100\,p^2$ at a voltage of $V_2' = pV_1'$, so that k can be read directly.

We have given preference, however, to readings

and when measuring V_1':

$$b \text{ db} = 20 \log (V_1'/V_0) \text{ db}.$$

The difference between these two readings is:

$$a - b = 20 \log (V_2'/V_1') = 20 \log p .$$

Thus the transducer gain to be measured is:

$$k = (20 + a - b) \text{ db}.$$

When V_1' is chosen such that $b = 20$ db, the reading a gives k directly in decibels.

The moving-coil instrument of the valve voltmeter has, in addition to a scale calibrated in decibels, a scale calibrated according to $V_2'^2$, from which can be read in mW the A.C. output developed in R_2. This output is $V_2^2/R_2 = \beta^2 V_2'/R_2 = \text{const.} \times V_2'^2$ so that when the position of the switch S_2 is changed the calibration of the scale remains the same. Another switch (not drawn) provides for the choice of two sensitivities: in one position the full deflection corresponds to 23 db and 1 mW, in the other position it corresponds to 34 db and about 12 mW. Since it is not practicable to have four scales on one instrument, two moving-coil meters are used, each with one decibel and one mW scale. The change-over from one meter to the other is effected together with the selection of the sensitivity.

When choosing V_1 care has to be taken that no distortion arises. As a check upon this, an oscilloscope is connected in parallel with the terminals of the voltmeter (fig. 17). If the oscillogram shows distortion of the output voltage then V_1' has to be reduced until the distortion disappears.

A picture of the whole set-up is given in *fig. 18*.

Summary. The transistor is a circuit element which, like the triode, is capable of amplifying signals. One particular design consists of a germanium crystal (with an excess of conduction electrons), soldered onto a base electrode, and two metal points (the emitter and the collector) placed close together and resting on the crystal.

Here an apparatus is described for tracing various transistor characteristics and measuring the transducer gain.

For the first purpose two electronic switches, type GM 4580, are employed. The emitter voltage, for instance, is applied to one pair of input terminals of one switch, while the voltage produced by the emitter current in a resistor is applied to one pair of input terminals of the other switch. In the emitter circuit there is also an alternating-voltage source. The output terminals of the electronic switches are connected to the vertical and horizontal deflection plates of an oscilloscope tube, which displays the characteristic (emitter voltage as a function of the emitter current) and also — since the other pairs of input terminals of the electronic switches remain open — the coordinate axes and the origin.

It is shown that — thanks to the symmetrical inputs and the method of applying negative feedback in the electronic switches — it is not necessary to have a common terminal for the voltages that are traced one against the other. When the measuring apparatus is connected this does not, therefore, automatically involve earthing of a particular point of the circuit, which, owing to the inevitable stray capacitances, would lead to the formation of a loop on the oscillogram. On the contrary, one is now free to choose the earthing point of the circuit such as to eleminate the effect of the stray capacitances.

In a similar manner it is possible to trace other characteristics, such as the collector voltage as a function of the collector current with a constant or a varying emitter current, or the collector current as a function of the emitter current, etc.

The transducer gain is measured with a different circuit. The gain is indicated by a meter calibrated in decibels.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1962: K. F. Niessen: On the deviations between theoretical and experimental values of the specific heat of superconductors (Physica **16**, 709-718, 1950, No. 9).

The deviations mentioned in the title are explained qualitatively by the combination of two effects, the more important of which is based on the assumption that the thickness of Heisenberg's superconducting layer, covering part of the Fermi surface, does not drop abruptly to zero at the boundary of the layer. The other effect, which was more obvious but appeared to be of less importance, comes into play when the thickness of the superconducting layer decreases with increasing temperature.

1963: J. de Jonge, R. J. H. Alink and R. Dijkstra: Absorption spectrum and photodecomposition of o-hydroxybenzene diazonium sulphate (Rec. Trav. chim. Pays-Bas **69**, 1448-1454, 1950, No. 11).

The long-wave absorption of o-hydroxybenzenediazonium salts in water is found at 3950 Å, but in 50% sulphuric acid as solvent it appears at 3500 Å. While the photodecomposition of o-hydroxybenzenediazonium sulphate in water leads quantitatively to a derivative of cylopentadiene, decomposition of the same compound in 50% sulphuric acid results for the greater part in catechol. The effect of sulphuric acid, both on the spectrum and the photo-decomposition, may be due to an increase in energy of the quinoid resonance structure of the hydroxybenzenediazonium ion by the acid.

1964: D. Polder: Ferrite materials (Proc. Inst. El. Engrs. **97** II, 246-256, 1950, No. 56).

Survey of electrical and magnetic behaviour of ferrite materials in connection with physico-chemical properties. The manufacture is dealt with and important possibilities for application of different ferrite materials are briefly reviewed. Some of the experimental methods are described for measuring the chemical and magnetic properties in different frequency ranges.

1965*: J. M. Stevels: Expériences et théories sur la facteur de puissance des verres en fonction de leur composition (Verres et Réfractaires **4**, 83-89, 1950, No. 2). (Experiments and theory on power factors of glasses as a function of their composition; in French)

Translated from **R 127**.

1966: J. I. de Jong: Note on the reaction of urea and formaldehyde (Rec. Trav. chim. Pays-Bas, **69**, 1568, 1950, No. 12).

In contradiction to other investigators a purely bimolecular reaction without an initial fall in p_H is found when neutral formol is added to a freshly prepared solution of repeatedly crystallised urea. The decrease in p_H and the high initial reaction rate found by others may be due to small amounts of ammonium salts.

1967: J. D. Fast: Le vieillissement du fer et de l'acier (Revue Métallurgie **47**, 779-786, 1950, No. 10). (Ageing of iron and steel, in French)

For the contents of this paper see Philips Techn. Rev. **13**, 1951 (No. 6).

1968*: Balth. van der Pol and H. Bremmer: Operational calculus based on the two-sided Laplace integral (Cambridge Univ. Press 1950, XIII + 424 pp., 97 figs.).

After a short introduction concerning the person of Oliver Heaviside and his "operational calculus" the Laplace transformation is dealt with at length. Apart from some historical remarks on the one-sided transformation (integration between 0 and ∞) the two-sided Laplace integral ($-\infty$ to $+\infty$) is introduced *ab initio*.

The Fourier integral is dealt with as an introduction to the Laplace transformation. A number of simple examples and elementary rules are given, after which the very essential conceptions of unit function and delta function are introduced. Two chapters are devoted to convergence questions and asymptotic expressions. Then follows among others the treatment of linear differential equations with constant coefficients, *idem* with variable coefficients, step functions and saw tooth functions, differential equations and integral equations. Partial differential equations are dealt with both from the point of the Laplace transformation with one variable and

from that of the "simultaneous" transformation in many variables. The special advantages of the operational calculus come to the fore in deducing unexpected and complex relations between functions and in dealing with transient phenomena in linear networks (filters).

The book, which contains many worked-out examples, is concluded by a list of rules ("grammar") and a list of transformations ("vocabulary").

1969: G. W. Rathenau and J. L. Meijering: Rapid oxydation of metals and alloys in the presence of MoO_3 (Metallurgia **42**, 162-172, 1950, No. 251).

The influence of MoO_3 on the oxidation in air of copper, aluminium bronze, silver, aluminium silver, nickel, nichrome and two heat-resisting steels is measured as a function of temperature. Discontinuities in the curves are correlated with eutectic temperatures measured in mixtures of MoO_3 and oxides of the metals concerned. Metallographic examination of the oxidized specimens reveals some peculiar effects.

Besides by the formation of liquid oxidic phases, the oxidation can be enhanced by MoO_3 in two other ways: incorporation of Mo in the crystal lattice of the oxide skin, and thermodynamic stabilization of Ag_2O by formation of molybdates (cf. also Philips techn. Review **12**, 213-220, 1951, No. 8).

1970: J. Haantjes and F. W. de Vrijer: Flicker in television pictures (Wireless Engineer **28**, 40-42, 1951, No. 2).

See Philips techn. Rev. **13**, 55-60, 1951 (No. 3).

1971: N. W. H. Addink: Quantitive spectrochemical analysis by means of the direct current carbon arc, I. General methods. (Rec. Trav. chim. Pays-Bas **70**, 155-167, 1951, No. 2).

See Philips techn. Rev. **12**, 337-348, 1951 (No. 12).

1972: N. W. H. Addink: Quantitive spectrochemical analysis by means of the direct current carbon arc, II. Biological materials. A possible correlation between the zinc content of liver and blood and the cancer problem (Rec. Trav. chim. Pays-Bas **70**, 168-181, 1951, No. 2).

For the contents of this paper see abstracts No. **1947**.

1973: N. W. H. Addink: The degree of imperfection of crystals (Rec. Trav. chim. Pays-Bas **70**, 202-208, 1951, No. 2).

An attempt is made to arrange crystals of various materials according to their degree of imperfection, derived from the apparent value of N (the Avogadro number). N has been determined for each kind of crystal according to the formula: $N = cM/Vd$, where M is the molecular weight, c a factor related to the number of molecules per unit cell, V the volume of the unit cell and d the density of the crystal. Metals, in accordance with their mosaic structure, show a relatively high value of N and therefore a high degree of imperfection (0.06% of their volume), whilst the lowest value of N has been found for diamond, quartz, calcite and potassium chloride (prepared from a solution), namely $N_{chem} = (6.0228 \pm 0.0014)\ 10^{23}$, as compared with Cohen's and Dumond's value $(6.02378 \pm 0.00011)\ 10^{23}$. Special attention is drawn to the behaviour of perfect crystals of potassium chloride (from a solution) and imperfect crystals of the same substance (from the melt): adsorption phenomena of water vapour enable the thickness of mosaic blocks to be estimated. A value of 6×10^{-4} cm has been found, assuming a monomolecular layer of water molecules on the internal surface.

1974: R. van der Veen: Influence of day-length on the dormancy of some species of the genus Populus (Physiologica Plantarum **4**, 35-40, 1951).

Experiments show that dormancy of poplar trees (Populus) is induced solely by day length. In long days (14 hours and longer) growth and development of new leaves is maintained; in short days (12 hours and shorter) the trees go into dormancy irrespective of the temperature. At 20°C and 29°C all eight species of Populus investigated had stopped their growth in less than 45 short days. After 102 short days all these trees had shed their leaves and had gone into complete dormancy.

1975*: P. J. Bouma: Farbe und Farbwahrnehmung. Einführung in das Studium der Farbreize und Farbempfindungen. (N.V. Philips Gloeilampenfabrieken, Techn. and Scientific Literature Dept. 1951 (XVI + 358 pp., 113 figs.).

German translation of "Physical aspects of colour" by the same author; see Philips techn. Rev. **9**, 158, 1947.

1976*: J. M. Stevels: Expériences et théories sur le facteur de puissance des verres en fonction de leur composition, II (Verres et Réfractaires **5**, 4-14, 1951, No. 1). (Some experiments and theories on the power factor of glasses as a function of their composition; in French).

French translation of **R 158**.

R 152: C. J. Bouwkamp: On the diffraction of electromagnetic waves by small circular disks and holes (Philips Res. Rep. **5**, 401-422, 1950, No. 6).

This paper deals with the diffraction of a plane-polarized electromagnetic wave by a conducting circular disk for the case of normal incidence. Integro-differential equations are derived for the currents induced in the disk. These equations are approximately solved on the assumption that the radius of the disk is small compared to the wavelength. Six terms of a power-series solution in the basic variable ka are derived, where k is the wave number and a the radius of the disk. The scattered field on the surface of the disk is calculated, to the same degree of accuracy, and also the field in the wave zone. An expression is obtained for the scattering coefficient of the disk. Finally, Babinet's principle is applied to obtain the corresponding solution for the diffraction of a plane wave by a circular hole in an infinite, plane, conducting screen (see **R 148**).

R 153: G. Diemer: Passive feedback admittance of disc-seal triodes (Philips Res. Rep. **5**, 423-434, 1950, No. 6).

At microwave-frequencies the inductance of the grid wires in disc-seal grounded-grid triodes plays an important part in the passive feedback from anode to cathode. It is shown that with a well-designed valve geometry this can be used to neutralize more or less the capacitive feedback through the anode-to-cathode capacitance.

R 154: J. M. L. Janssen: The method of discontinuities in Fourier analysis (Philips Res. Rep. **5**, 435-460, 1950, No. 6).

A survey is given of the results arrived at in various publications dealing with the method of discontinuities in Fourier analysis. It is shown that a more general conception is practicable when starting from the Fourier integrals instead of the Fourier series. Then in a simple and natural way also the frequency spectrum of continuous functions may be found by this method. It is investigated in how far rapid changes may be construed as being real discontinuities. Correction factors are derived which take into account the shape of the rapid changes. Several examples are given, whilst, *inter alia*, the method is applied to the problem of the summation of a series, an Euler summation formula then being found.

R 155: W. Ch. van Geel and B. C. Bouma: Variation en fonction de la fréquence de la caractéristique dynamique $i(V)$ du système Al-Al_2O_3-électrolyte (Philips Res. Rep. **5**, 461-475, 1950, No. 6). (Variation of the dynamic characteristic $i(V)$ of the system Al-Al_2O_3-electrolyte as a function of the frequency; in French.)

The current-voltage characteristic of the system Al-Al_2O_3-electrolyte is not a single curve but shows a loop. With increasing frequency the height of the loop (maximum current) decreases. At about 5000 c/s the loop has almost disappeared and the system has lost its rectifying properties. For a newly formed layer, the current in the direction of easy flow increases strongly during the first moments after the application of an alternating voltage. Some seconds are necessary for the current to reach its maximum. The constitution of the layer is not a stable one. For rectification a particular constitution in the layer has to be created and after that the situation is still not stable but changes with the direction of the applied voltage. An effort is made to explain the phenomena observed. It is assumed that the oxide layer contains an excess of aluminium in that part of the layer which is bounded by the aluminium. This part is semi-conducting (excess semiconductor). The other part of the layer is the barrier layer, the thickness of which changes with the direction of the applied voltage and gives rise to the loop. The variation of the thickness is caused by electrolysis in the Al_2O_3 layer.

Another possibility is to assume that the part of the oxide layer near the electrolyte contains a surplus of oxygen and forms a defective semi-conductor.

VOL. 13 No. 10, pp. 269-300 APRIL 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

FLUOROGRAPHY WITH THE AID OF A MIRROR SYSTEM

by W. HONDIUS BOLDINGH. 778.33:771.313.26:616.24-073.75

From the point of view of the universality of the applied sciences the instrumentarium of the roentgenologist has long been one of the most interesting. Its development is related just as much to physics and mechanics as to biology, and there are also quite a number of economic considerations to be taken into account. In one of the most recent roentgenological methods, that of fluorography, also optics play one of the main parts. The application of the well-known Schmidt optical system in this field may well prove to bring about in the future a great change in the practice of X-ray diagnostics.

Introduction

Nowadays there are mainly three methods employed in X-ray diagnostics: fluoroscopy (screening), contact photography and fluorography. In screening, an X-ray image is formed on a fluorescent screen viewed by the physician direct. In contact photography a cassette containing a photographic film takes the place of the screen. In fluorography, with the aid of a camera set up at a distance from the patient, a reduced picture is made of the X-ray image visible on a fluorescent screen.

The principle of fluorography is not much younger than the other two methods, but it is only in the last fifteen years that it has been brought to development. Its field of application (for the present) is mainly mass chest survey in the fight against tuberculosis. It has the characteristic advantage of photography that the image is recorded, whilst with the small size of the recorded picture the heavy expense and the other drawbacks of contact photography are avoided. It is no exaggeration to say that it is this method of fluorography that has made periodical chest examination of large groups of the population a practical possibility.

In the development of the apparatus for fluorography there were two technical problems to solve: how to create the conditions for obtaining the best possible picture quality with the reduced photographs, and how to simplify the operation of the apparatus as far as possible (considering that it may also have to be mobile).

These two problems have already been dealt with at some length in earlier publications in this journal [1] [2]). In particular a description has been given of a mobile installation for fluorography which was developed during World War II by the North American Philips Co. [2]) and attracted considerable attention. The working of this installation had been made so highly automatic that a continuous flow of 200 people per hour could be dealt with quite comfortably. Compared with this installation, the new apparatus for mass chest survey which have subsequently been developed by Philips show no essential changes as regards their operation and the technicalities connected therewith. But as regards the other problem mentioned — the conditions for an optimum picture quality — there has indeed been an important new development. This development, begun some ten years back, has led to the replacement of the previously used camera with lens system (a Fairchild camera with a Kodak Fluoro Ektar objective) by a camera with a mirror system having a much higher light-gathering power.

[1]) A. Bouwers and G. C. E. Burger, X-ray photography with the camera, Philips Techn. Rev. **5**, 258-263, 1940.
[2]) H. J. Di Giovanni, W. Kes and K. Lowitzsch, A transportable X-ray apparatus for mass chest survey, Philips Techn. Rev. **10**, 105-113, 1948.

Without entering into a renewed discussion of the problem of picture quality in fluorography it will be clear that, with a given X-ray intensity, the higher the light-gathering power of the camera the better will be the quality of the picture. It is then possible to use less sensitive fluorescent screens and/or films, which as a rule are finer-grained and thus able to give a sharper image; or else the exposure time can be made shorter, thereby reducing the kinetic blurring caused by the movement of the object (heart, blood vessels and lungs). Furthermore there is the advantage that the examinee receives a smaller dose of X-rays. The superiority of the mirror camera that has now been developed, as compared with the lens camera, is therefore most directly apparent in the exposure time required: whereas with the lens camera this was 0.1 to 0.5 sec, depending upon the thickness of the patient, now, other things being equal, 0.04 to 0.20 sec suffices.

This higher light-gathering power of the mirror camera can also be turned to advantage in quite a different way, namely by reducing the X-ray intensity required for the fluorographs. Accepting the same kinetic blurring as before, one can then manage with a much smaller and less expensive apparatus for generating the X-rays, whilst one can also work with an X-ray tube having a smaller focus and thereby reduce the geometric blurring of the picture.

In this article a description of the new mirror camera will be given, preceded by some remarks about mirror systems in general.

Historical commentary on mirror systems

The use of optical systems with mirrors is by no means new, these having been applied for centuries for astronomical purposes. Mirror systems, at least when they do not also contain refracting surfaces, are free of chromatic aberrations. There are, however, two objections, the first lying in the fact that in order to gather the reflected light, either in an ocular or on a photographic film or plate, it is necessary to place in the path of the rays a constructional element which intercepts part of the light falling on the mirror. The second objection, of a technological nature, was the difficulty to make surfaces of high reflectivity the reflective power of which does not deteriorate after prolonged use.

For astronomers the first objection mentioned is of no great consequence because for their purposes a very small field of vision suffices (less than 1 to 2 degrees) and so only a small part of the incident light need be intercepted. For other purposes, however, as in normal photography, the field of vision required is many times greater, so that very much more light is lost. Taken together with the other objection mentioned, this was the reason why in the last century only lens systems were employed for camera optics. In fact, for a given permissible magnitude of the aberrations, no greater light-gathering power could be obtained with the mirrors than with lenses owing to the loss of light just referred to: the mirrors were not free of image errors such as spherical aberration and field curvature, and on the other hand one has learned to reduce the chromatic aberration of the lenses by using a combination of different kinds of glass.

There are two factors which in the last decade have greatly improved the chances for mirror objectives for non-astronomical purposes. The technique of producing reflectors by evaporating an aluminium coat on glass has been so far advanced that practically no reduction of reflectivity need be feared, and, secondly, a simple method has been found for almost entirely eliminating spherical aberration, the most important image error of a mirror. This latter method, discovered and worked out for astronomical application by B. Schmidt [3]), is based upon the use of a correction plate of a certain profile placed in the centre of curvature of a concave spherical mirror. Thanks to this improvement it became possible to design a mirror system with a greater light-gathering power than is obtainable with a lens system, even when the field of vision is large.

The much more favourable situation thereby created for mirror optics was realized by various designers [4]) and led to a number of practical applications of mirror systems outside the domain of astronomy. One of these, for projection television receivers, was described in this journal some years ago [5]), and in connection therewith also the fundamentals of the Schmidt optical system and the construction of the aforementioned correction plate were dealt with [6]) [7]). There is, therefore, no need for

[3]) B. Schmidt, Ein lichtstarkes, komafreies Spiegelsystem, Mitt. Hamburger Sternwarte Bergedorff **7**, 15-17, 1932

[4]) See, e.g.,:
D. D. Maksutov, New catadioptric meniscus systems, J. Opt. Soc. Amer. **34**, 270; 1944.
A. Bouwers, Achievements in optics, Elsevier Publ. Co., Amsterdam 1946
H. Rinia and P. M. van Alphen, A new method of producing aspherical optical surfaces, Proc. Kon. Ned. Ak. Wet., Amsterdam **49**, 146-149, 1946.

[5]) P. M. van Alphen and H. Rinia, Projection television receiver, Philips Techn. Rev. **10**, 69-78, 1948.

[6]) W. de Groot, Optical aberrations in lens and mirror systems, Philips Techn. Rev. **9**, 301-308, 1947.

[7]) H. Rinia and P. M. van Alphen, The manufacture of correction plates for Schmidt optical systems, Philips Techn. Rev. **9**, 349-356, 1947.

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

us to enlarge upon this here, the more so since the mirror optical system of our X-ray camera is substantially identical with that of the television receiver described. It should be pointed out that it has been no mere accident that mirror optics has found such a ready application for these purposes, more so than for normal photography. With mirror systems much more than with lens systems, serious complications, whereas in the other applications mentioned it is a fairly easy matter to make allowances for it.

Description of the mirror camera

One of the new Philips installations for fluorography is illustrated in *fig. 1*, complete with the high-tension generator and the control desk, which

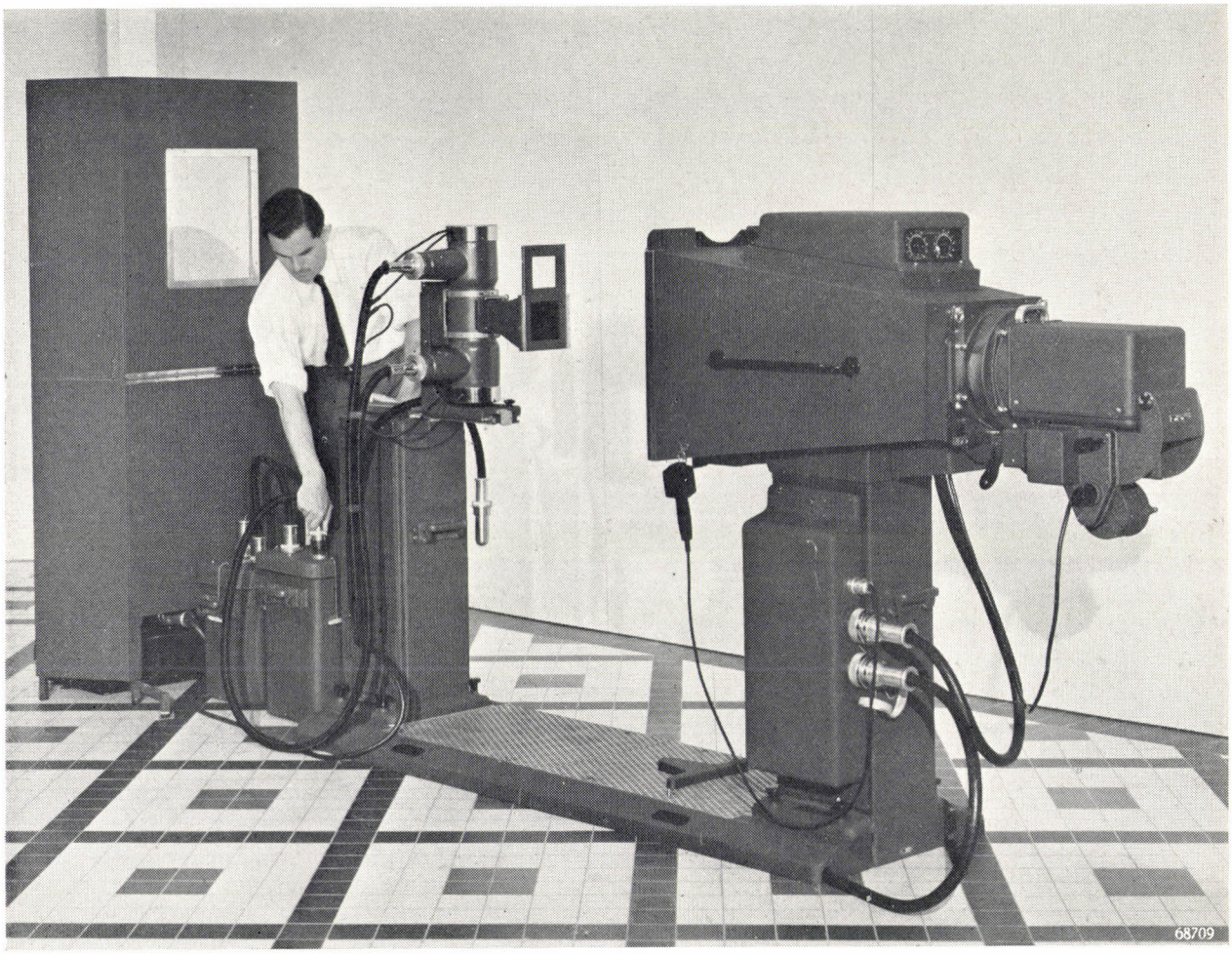


Fig. 1. Apparatus for mass chest survey by fluorography (type No. 11490). On the left the high-tension generator and control desk with metal screen for protection of the operator, in the middle the X-ray tube, on the right a light-tight hood, with fluorescent screen on the left end and the mirror camera on the right end.

the elimination of the principal aberrations can only be complete (or practically so) when the object is at a specific distance. In normal photography a certain latitude has to be allowed in this distance between the object and the camera, whereas in television projection and in X-ray fluorography, as also in astronomy, one works with a fixed object distance. Moreover, with the Schmidt optical system field curvature is not eliminated, which for normal photography would involve rather need not be discussed here again. The patient is placed with his chest close up against the front of the "hood", at the other end of which the mirror camera is attached; *fig. 2* gives a better picture of the hood and camera.

Fig. 3 is a diagrammatic cross section of the hood with camera. The rays of light emitted by the fluorescent screen pass through the correction plate *C*, fall on the concave spherical mirror *S* and are focused upon the sensitive surface of the film strip

F in the focal plane of the optical system. The screen image of 42 cm × 42 cm is thereby minimized $10\frac{1}{2}$ times to a size of 40 mm × 40 mm, for which a reel of non-perforated film 45 mm wide is used. The casettes of the camera can take a reel of film of the normal length of 30 metres, on which about 600 pictures can be taken.

As already mentioned in passing, with the Schmidt system field curvature is not eliminated, in contrast to what one is accustomed to with lens systems. Consequently, during the exposure, the film has to be spherically curved, concentrically with the mirror and with the convex side towards it. The film is given the required form by pressing it from behind against the square film frame with the aid of a spherically curved plate (radius of curvature 103.5 mm). At first sight it may seem surprising that this spherical deformation of the film (the material being stretched in the middle and compressed at the edges) is possible. One condition is that the thickness of the film must lie within certain limits. If it is too thick there is a risk of it tearing, while if it is too thin it may wrinkle. A thickness of 0.15 mm was found to be suitable.

The forward travel of the film, like all the other operating actions, is fully automatic. After the examinee has been placed in front of the camera hood and his identification card (the number on which is photographed together with the image on the fluorescent screen) inserted in a slot in the hood, all the operator has to do it to press in a button. The X-ray tube is thereby automatically brought into action and switched off again after the correct exposure time, the spherical "plunger" is withdrawn, the film travels one picture frame forward and is again spherically pressed out, while a locking device then comes into action preventing another photograph being taken before the identification card has been changed. This and other precautions against possible mistakes, as also the accompanying warning signals, the automatic timer (see figs 2 and 3) and other auxiliairy apparatus will not be gone into here; they are described in the article quoted in footnote [2]),

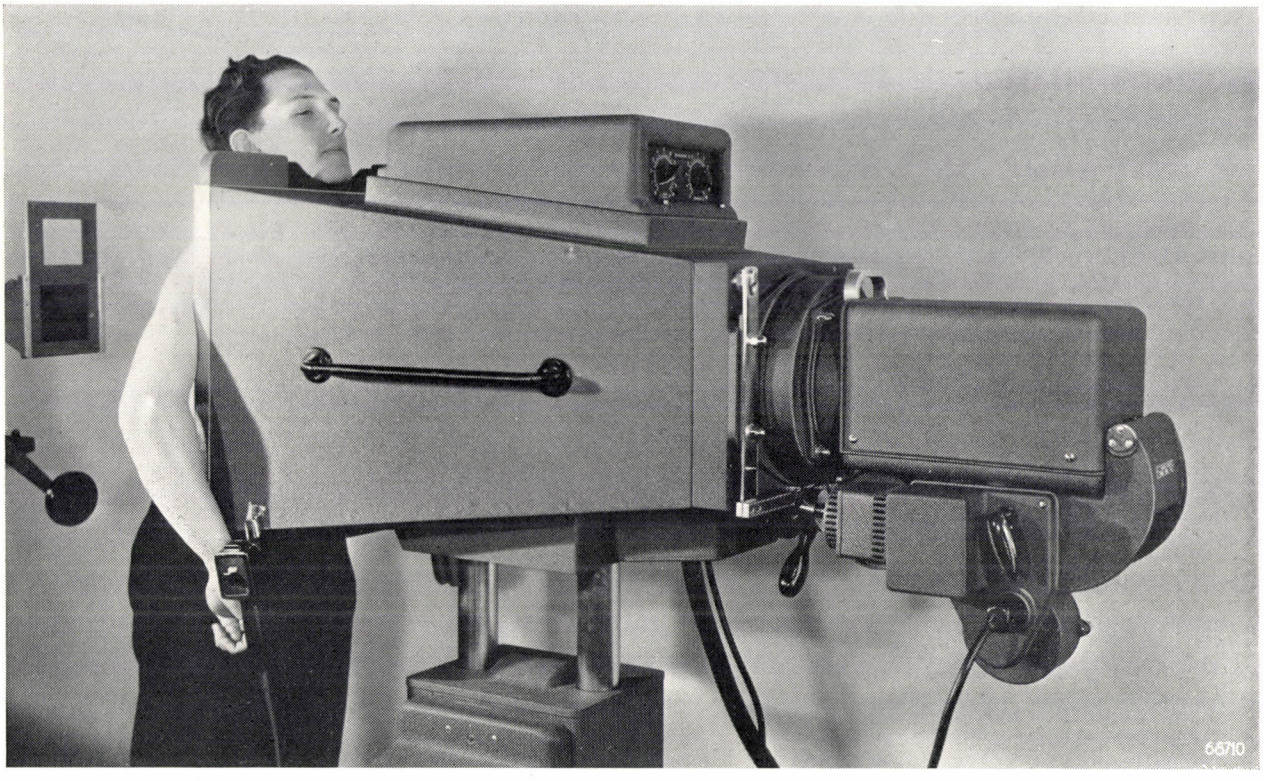


Fig. 2 The hood with mirror camera in use. When the examinee has been placed in the right position and his identification card inserted in a slot in the hood the operator has only to press a button for the photograph to be taken and everything is prepared automatically for the next one. By means of an automatic timer, seen mounted on top of the hood, the exposure is stopped as soon as a certain average density of the film is reached. The operator has only to choose one of three fixed tube voltages, according to the thickness of the examinee, and to adjust the hood and the X-ray tube to the right height.

When a new reel of film has to be put in, the entire film-travel mechanism with the film holder can be turned down. *Fig. 4* shows the camera in this open position.

As to the most important property of the camera — its speed — a rough idea of this has been given above when comparing the exposure time required for lung photographs with that for the lens camera formerly used. It may, however, be desired to express this in terms allowing of a direct comparison with any other cameras. For conventional photographic cameras of normal speed it is customary to quote the relative aperture D/f; D is the diameter of the diaphragm limiting the effective beam of light, f is the focal length. We could, of course, quote this quotient also for mirror cameras, but it is to be noted that here the relative aperture is not a proper measure for their light-gathering power. Since this is a question that has already given rise to many misapprehensions, it is deemed to be of sufficient importance to go into it more closely.

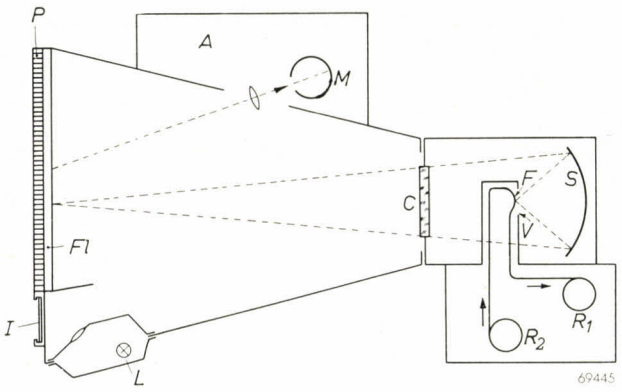


Fig. 3. Diagrammatic cross section of the hood and mirror camera. *Fl* fluorescent screen, *P* grid intercepting the scattered X-rays coming from the examinee, *S* spherical mirror, *C* correction plate, *F* film strip pressed out spherically against the film gate *V*, R_1 and R_2 rewinding and unwinding reels for 30 metres of film; *A* automatic timer with multiplier photocell *M* (this measures the luminance of a certain part of the fluorescing screen and regulates accordingly the time during which the current for the X-ray tube is kept switched on); *I* identification card, *L* lamp for illuminating this card.

The figure of merit of optical systems in general

The point at issue is the illumination of the focal plane and how this is correlated with the "relative aperture".

Let us suppose that the object is a luminous plane with a luminance B, placed perpendiculer to the axis of the optical system, and that the object is a Lambert radiator. The rays which come from a surface element ΔS of the object lying on the axis and which enter the optical system fill a cone the half angle of which is denoted by a (*fig. 5*). The luminous flux gathered by the optical system from the element ΔS is then $\Delta S \cdot \pi B \sin^2 a$. This luminous

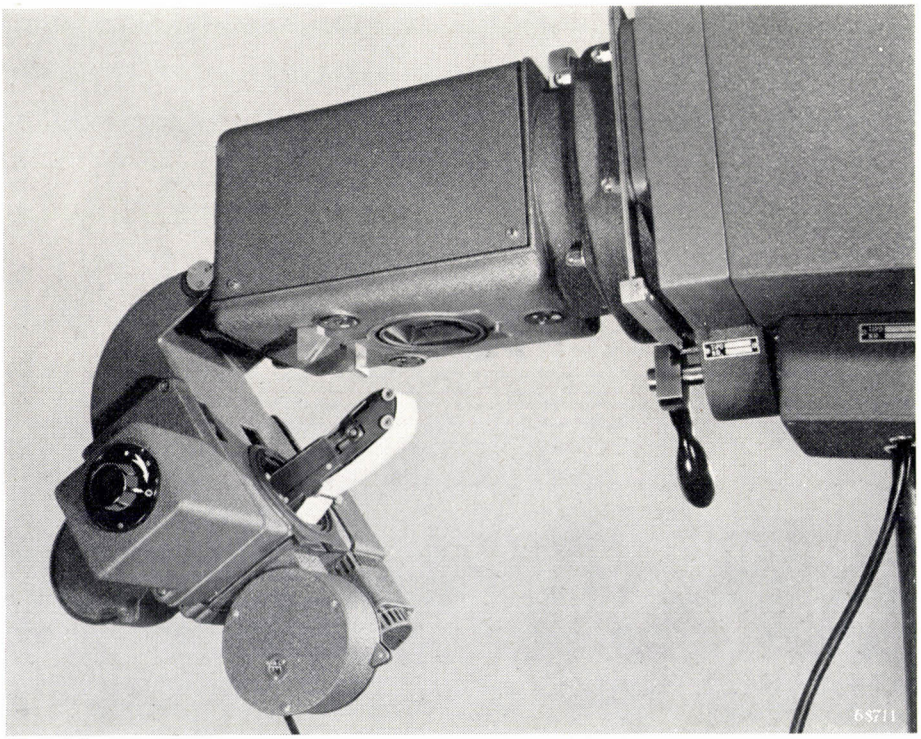


Fig. 4. The mirror camera with the film holder turned down.

flux falls in the image on an area $\Delta S'$, the illumination of which is thus

$$\frac{\Delta S}{\Delta S'} \pi B \sin^2 \alpha \quad \ldots\ldots\ldots \quad (1)$$

Denoting the linear reduction of the projected image — the "reduction factor" — by r, then $\Delta S/\Delta S' = r^2$.

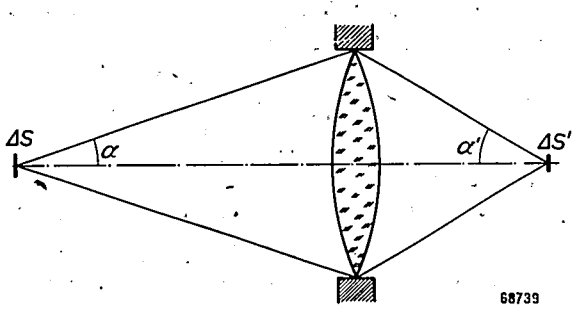


Fig. 5. The limiting diaphragm (entrance pupil) of an optical system (with lenses or mirrors) receives from a surface element ΔS of the object a cone of light with half angle α.

To conform with existing definitions we formulate the illumination on the image plane as

$$\tfrac{1}{4}\pi B \cdot \frac{1}{N^2}, \quad \ldots\ldots\ldots \quad (1a)$$

the quantity $1{:}N$, according to the foregoing, being given by

$$\frac{1}{N} = 2r \sin \alpha \quad \ldots\ldots\ldots \quad (2)$$

This quantity was called the "aperture ratio" of the optical system for the given reduction factor and was adopted in 1950 by the British Standards Institution as a figure of merit of photographic objectives. As shown by the deduction, it indicates directly the "light gathering" power of the optical system determined by the geometrical arrangement.

Often, for example in the formula given by the British Standards Institution, use is made of the half angle α' of the cone of light on the image side falling upon the image plane when the optical system is focused for the reduction factor r (see fig. 5). For a coma-free optical system the "sine law"

$$\frac{\sin \alpha'}{\sin \alpha} = r$$

applies. The value $\sin \alpha' = r \sin \alpha$ is the "numerical aperture" introduced by Von Abbe for microscope objectives (thus applying, just like the aperture ratio, only for a certain reduction factor, in this case the enlargement corresponding to the tube length and indicated on the microscope objective).

The aperture ratio $1{:}N$ can be related to the relative aperture D/f in the following way. Let us first imagine the simple case of a thin, single lens with the diaphragm lying in the plane of the lens (*fig. 6*). Here the lens formulae apply as known from elementary text books. For the object distance v, measured from the object to the diaphragm, thus in this case to the lens itself, we have $v = f(r+1)$, and from fig. 6 it follows that

$$\tan \alpha = \frac{D}{2v} = \frac{D}{2f(r+1)} \quad \ldots\ldots \quad (3)$$

Now in all the cases that can be considered here the half angle α of the light cone is so small that the tangent can be put equal to the sine. Consequently the formula for $1{:}N$ can be developed into [8]):

$$\frac{1}{N} \approx 2r \tan \alpha = \frac{D}{f} \frac{r}{r+1} \quad \ldots\ldots \quad (4)$$

This formula (4), derived for a thin, single lens, may likewise be applied for the composite photographic lenses. It is true that we must then take as the object distance the distance from the object to the first principal plane of the lens, but with these lenses the entrance pupil, i.e. the diaphragm, may be assumed to lie approximately in that plane, so that the formula still holds.

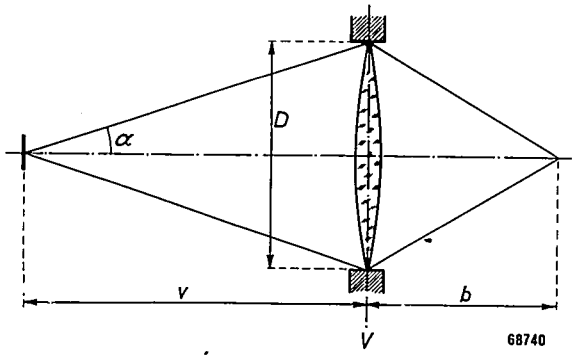


Fig. 6. Path of the rays through a thin lens. V is the first principal plane (which for a thin lens coincides with the second principal plane), v the object distance, b the image distance.

In the case of mirror systems, however, the situation is different. Here the function of the diaphragm is performed by the correction-plate holder placed in the centre of curvature of the mirror and thus

[8]) Without the simplification mentioned, with the aid of the known relation between sine and tangent.

$$\sin^2 \alpha = \frac{\tan^2 \alpha}{1+\tan^2 \alpha}, \quad \ldots\ldots\ldots \quad (5)$$

we find the more general equation

$$\frac{1}{N} = \frac{D}{f} \frac{r}{\sqrt{(r+1)^2 + \left(\frac{D}{4f}\right)^2}} \quad \ldots\ldots \quad (6)$$

The simplification is permissible because in all systems to be considered D/f is not greater than 1.5 (the "f-number" not smaller than $1/1.5 = 0.67$) and the reduction r is at least 5. Then, according to eqs (3) and (4), $(D/4f)^2 \ll (r+1)^2$, which amounts to the same as $\tan \alpha \approx \sin \alpha$ ($\alpha < 7°$).

lying at a distance $2f$ in front of the mirror (which is approximately the position of the first principal plane); see *fig.* 7. In this case the distance from the object to the diaphragm is :

$$v - 2f = f(r+1) - 2f = f(r-1) .$$

Formula (3) for the apex angle of the light cone gathered has therefore to be replaced by:

$$\tan \alpha = \frac{D}{2f\,(r-1)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3a)$$

and formula (4) — for the same simplifying supposition that α is not too large — by:

$$\frac{1}{N} = \frac{D}{f}\,\frac{r}{r-1} . \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

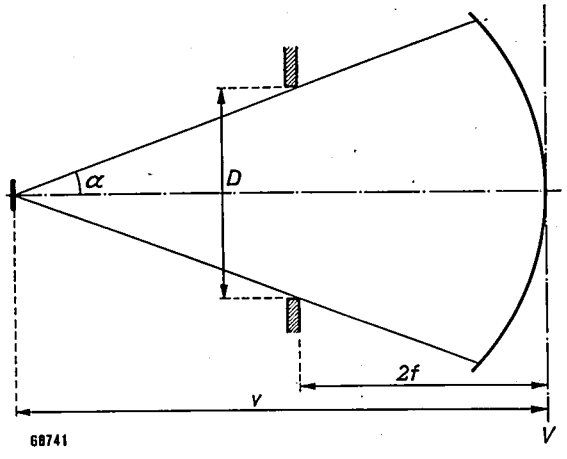


Fig. 7. In a mirror system the limiting diaphragm is a distance $2f$ in front of the first principal plane V. Given the same object distance v and diaphragm diameter D the angle 2α of the cone of light received is therefore greater than that in a lens system.

Now we are in a position to draw the conclusions for the point at issue in this chapter. It is seen that for very large object distances, i.e. very great reductions (where $r + 1 \approx r \approx r - 1$), the formulae (4) and (4a) are both simplified to $1{:}N = D/f$, in other words we find both for lens and for mirror systems the well-known rule that the illumination in the image is proportional to the square of the relative aperture. In the case of relatively small reductions, as applied in X-ray fluorography, D/f alone is not, however, sufficient to indicate the light-gathering power, and especially not when comparing a mirror system with a lens system. Then the reduction factor must certainly be taken into account, and this is best done by quoting the "aperture ratio" $1{:}N$, together with the reduction factor.

Since the equation $1{:}N = 2r \sin \alpha$ gives directly the light-gathering power of the optical system, it is only logical to allow for the obstruction of the film holder by including the "shadow factor" which must always be taken into account with mirror systems and is likewise of a geometrical nature. (The importance of this shadow factor has already been pointed out in our historical commentary). As far as the illumination of the focal plane is concerned the useful cone of the entering light having the apex angle 2α, but being partly intercepted by the blacked-out centre of the correction plate concealing the film holder, is equivalent to a cone of light having a smaller apex angle $2\alpha_{\text{eff}}$ and not intercepted at all. The quantity $2r \sin\alpha_{\text{eff}}$, again written as a fraction with 1 as numerator, will be called the "effective aperture ratio" With this number as basis it is possible to make a true comparison of the light-gathering power of lens systems and of mirror systems.

The speed of the Philips mirror camera

Our X-ray mirror camera has a focal length of 100 mm and a diaphragm diameter (diameter of the correction plate) of 115 mm. For a reduction factor $r = 10.5$ the aperture ratio according to formula (4a) works out to $1 : N = 1{:}0.79$. The shadow factor amounts to about 0.45, which means that 55% of the light is intercepted by the film holder. Therefore the effective aperture ratio is 1 : 1.17. This is to be compared to the aperture ratio of 1 : 1.9 for the lens system formely used, where the reduction factor was 6 (film width 70 mm). The light-gathering power of the mirror camera is therefore $1.9^2/1.17^2 = 2.6$ times that of the lens camera.

Apart from its geometrical light-gathering power also the losses through reflection or absorption in the parts of the optical system have to be considered. In the case of the lens system in question the lens surfaces are coated, so that the losses of light in that objective, including absorptions, are reduced to about 30%. In the mirror camera the losses in the optical system are higher, 20% of the light being lost through reflection from the surfaces of the correction plate and another 20% through absorption in the surface of the mirror. The difference, however, is almost outbalanced by the fact that when using a lens camera 20% of the light is lost through absorption in a plate of lead glass that has to be placed between the fluorescent screen and the camera in order to protect the film against the X-rays penetrating through the screen. With a mirror camera there is no need for such a plate because the X-rays are sufficiently absorbed in the metallic parts of the film holder. (X-rays are now prevented from passing through the hood walls by lining it with a thin layer of lead.)

According ro the deduction given, $(1/N)^2$ multiplied by $\frac{1}{4}\pi B$ and by a transmission factor allowing for the losses of light just mentioned gives at once the illumination on the film. This, however, only holds for the centre of the field of vision. Towards the edges of the image the illumination diminishes through various causes, such as owing to the fact

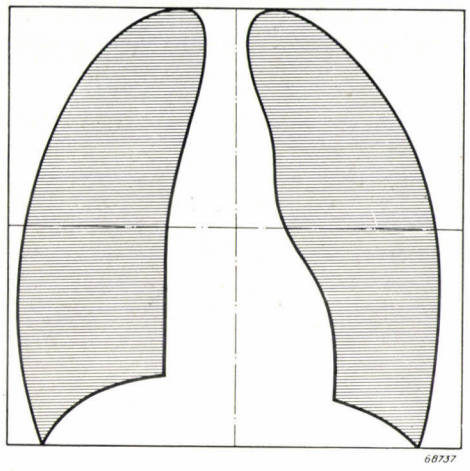


Fig. 8. Position of the lungs in a chest photograph.

that light beams striking the objective at an angle are limited not only by the diaphragm but also by the holders of the lenses or the mirror. This effect, which is known under the name of vignetting, is of less consequence in the mirror system than in lens systems. This is particularly of importance when taking photographs of the lungs, because the physician is less interested in the middle portion of the picture on account of the fact that most affections of the lungs occur in the outer parts. Taking, by way of example, as the representative part a zone at about $\frac{1}{4}$ of the edge of the field of vision (*fig. 8*), with the lens system the illumination

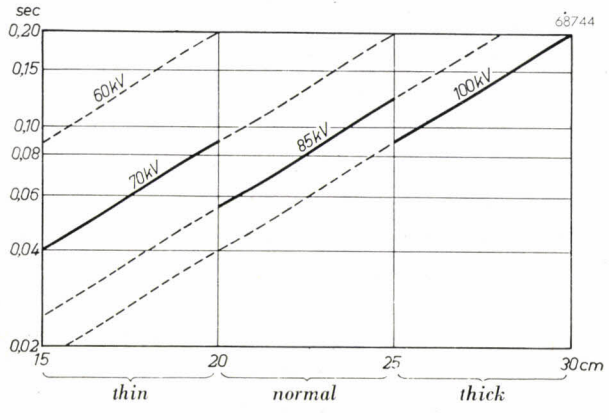


Fig. 9. Exposure times for chest photographs taken with the 45 mm mirror camera, as a function of the thickness of the examinee, for three tube voltages. The power applied to the tube is in all cases 14 kW; the distance from the X-ray focus to the fluorescent screen is 90 cm; fluorescent screen Levy West 48, film Gevaert Scopix G.

is reduced to 45%, whereas in our mirror system it is reduced to 70% of that in the centre.

The luminance B of the fluorescing screen, which together with the light-gathering power of the optical system determines the illumination on the film and thus the exposure time required, depends upon the hardness and intensity of the X-rays and upon the thickness (absorption) of the examinee. For practical purposes it is desired to express the necessary exposure time directly as a function of these two variables. This has been done in the graph of *fig. 9*, plotted from data collected in the periodical chest survey of Philips' personnel at Eindhoven [9]). It applies for a suitably chosen combination of film, fluorescent screen and scattered-ray grid: the same power was fed to the X-ray tube at all voltage steps. It is seen that with the examinees divided into three thickness groups, for which tube voltages of respectively 70, 85 and 100 kV are chosen, all photographs can be taken with exposures lying between 0.04 and 0.20 sec.

The size of the photographs

It has been seen that the mirror camera is constructed for 45 mm film, whereas for the lens camera described earlier a 70 mm film was used. The reasons for this change of film size are as follows.

To judge the relative merits of the two sizes there are three main aspects to be considered:

1) Considering that the basic idea of fluorography is to save material and other expenses, it is only logical to choose the smallest size of picture. This, however, is not the deciding factor when a comparison is made with full-size radiography, because with the 70 mm film the cost of the photographs is already so far reduced that the size no longer plays an important part anyhow. However, in the case of mass chest survey, where full-size pictures are prohibitive, the 45 mm film size offers an appreciable saving in film cost (about 50%) compared with the 70 mm size.

2) In favour of the larger size it might be mentioned that the pictures could then, if necessary, be examined with the naked eye. But for a critical examination both the 70 mm and the 45 mm pictures have to be magnified either with a magnifying glass or with a projector.

3) Finally there is the important point of the quality of the picture. If the picture is excessively small the (relative) definition is affected by the granulation of the film, and that is why we no longer use the 35 mm film which was at first em-

[9]) G. C. E. Burger, not published.

ployed for fluorography and which gave pictures no more than 24 or 30 mm wide. With the size of 45 mm now chosen there is no trouble from granulation provided one of the very fine-grained makes of film is used which are now available on the market. It is true that for a still larger size of picture and also necessary to work with a larger focal length, as is to be understood when considering the size of the field of vision, which is roughly given by the quotient of the image diagonal and the focal length, and with our 45 mm camera amounts to about 25°. The field of vision is limited by the occurrence of

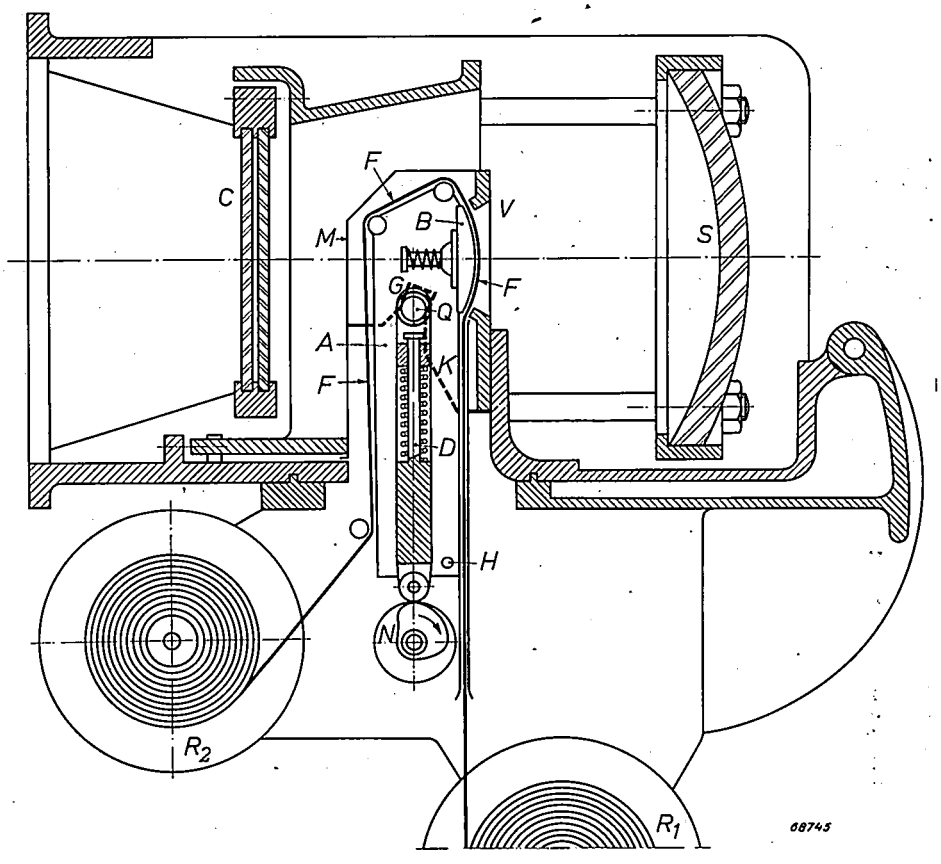


Fig. 10. Longitudinal cross section of the mirror camera. The drawing is still greatly simplified; in particular, all details of the automatic film travel and the locking and signalling devices coupled therewith are omitted. In addition to the parts already discussed, such as the mirror S and the correction plate C — which are shown here in their right proportions and distances — the drawing shows some details of the film holder and the device for spherically pressing out the film. V fixed film gate, onto which is fixed a cap K with oblique slot G; A arm, rotating about a spindle H, carrying the rollers guiding the film F and to which is attached, behind the path of the film, the spherical pressure plate B with ball joint. By the action of a cam disc N and rod D, via a spring, the roller Q is forced upward; as it moves upward this roller slides in the oblique slot G, thereby displacing the entire arm A towards the film gate and pressing the film, via the spherical pressure plate, against this gate with a force of 16 kg, so that the film assumes a spherical shape in the opening. Except for the film gate V the arm A with the film is sealed off light-tight by a metal housing M.

the same sharpness of definition a coarser-grained film could be used, thereby taking advantage of the higher sensitivity of a coarse-grained emulsion and thus allowing a shorter exposure time, but this type of film is not normally available.

It is technically quite possible to construct a mirror camera for larger size pictures [10]). Then it is aberrations of a higher order when the angle is larger. Taking the same permissible field of vision, a 70 mm camera would require a focal length roughly $70/45 \approx 1.5$ times that required for the 45 mm one. And since the correction plate and the mirror would then have to be given a correspondingly larger diameter in order to get the same relative aperture D/f, this would inevitably lead to a much larger and heavier, and thus more expensive, construction than that of the 45 mm size.

It is an open question whether the medical world will consider the somewhat easier examination of

[10]) A mirror camera for 70 mm film was built some years ago in Denmark; see Amer. J. Röntg. Rad. Therapy **59**, 416-419, 1948. A more recent construction of a 70 mm camera is that of A. Bouwers (Fortschr. a. d. Geb. Röntgenstr. **74**, 578, 1951, No. 5).

70 mm pictures of sufficient importance to justify the much greater expense (and less easy transportation!) of a 70 mm camera for mass chest survey.

Construction of the camera

Fig. 10 is a simplified cross-sectional drawing giving an idea of the mechanical construction of the mirror camera described here. It has already been said that we shall not deal here with the provisions made for the forward travel of the film, etc., which do not differ essentially from those found in the lens camera. We shall discuss only some constructional problems inherent in the designing of the mirror camera.

The position of the film with respect to the mirror is highly critical, a displacement of 0.02 mm in the axial direction already being sufficient to cause a perceptible reduction in the definition of the picture. This is related to the very large aperture ratio of the camera (very large angle under which the rays merge upon a focal point). A very robust construction and most precise finishing of the mechanical parts bearing the film and the mirror proved to be not sufficient to guarantee the proper distance: it was found necessary also to eliminate the effect of thermal expansion of those parts. To this end use was made of metals having different expansion coefficients, a measure such as is taken, for instance, in the construction of time-pieces to ensure that the length of the pendulum remains constant. The camera housing, in which the correction plate with a holding ring is mounted and to which the film holder is hinged (at the other end) — see fig. 10 and also fig. 4 —, has been made of an alluminium alloy; the spherical mirror is attached to the supporting flange of the correction plate with four steel rods and has no other further supports (see *fig. 11*).

An axial displacement of 0.02 mm of the film corresponds to a displacement of $r^2 \times 0.02 \approx 2$ mm of the fluorescent screen. The distance from the mirror system to the screen is therefore much less critical and there is no need of any special measures for compensating the thermal expansion of the hood keeping the camera at the correct distance from the screen (fig. 3). To give the latter distance the right value, taking into account the small differences occurring between various specimens of the camera in respect to the position of the film holder, the length of each camera is matched to the fixed length of the hood with the aid of spacing rings. This is done on a sort of optical bench, whereby the path of the rays is reversed to that in actual use: a film strip with a test pattern is put into position and strongly illuminated, the optical system projecting a $10\frac{1}{2}$ times enlarged picture of that pattern onto a screen, the position of which is adjusted until the picture is properly focused. From the final position of the screen it can then be calculated how many spacing rings are required for a given length of hood. At the same time if there are any aberrations in the projected picture the mirror system can be adjusted, for which purpose the holder of the correction plate is fitted with a number of set screws.

It is to be noted that either a blue or a green fluorescing screen can be used at the same distance from the camera. This is not possible with a lens camera because, as already stated, when a lens camera is used a lead glass plate has to be placed between the screen and the camera, and owing to the difference in the index of refraction for the two colours there is then a difference in the optical distance.

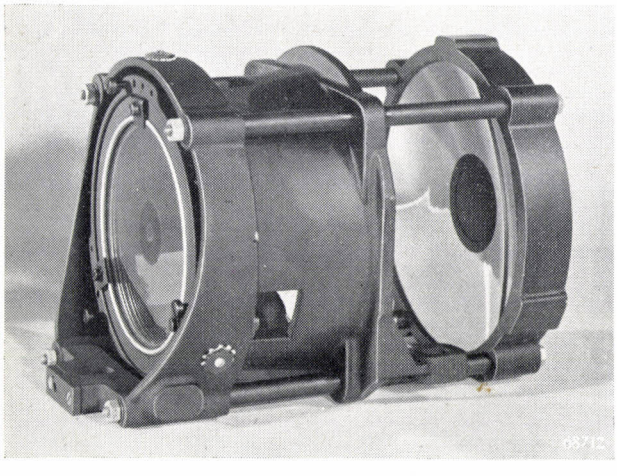


Fig. 11. The spherical mirror (on the right) is attached to the flange carrying the correction plate (on the left) with four steel rods.

A detail of mechanical interest in the mirror camera is the device for the spherical deformaton of the film. The construction of this device and that for the forward travel of the film, which have to be kept as small as possible to limit the shadow factor, is shown in fig. 10. The film holder is in the form of a pivoted arm carrying the convex pressure plate (*B*) and the rollers for guiding the film. The arm is turned away from or towards the fixed film frame by the action of a roller sliding in an oblique slot. This to and fro movement of the roller, which thus releases the film to be carried forward and then presses it out again, is brought about via a cam disc by the small motor which also drives the wheels for the forward travel and the rewinding reel.

Non-perforated film is used in order to make use of its full width as far as possible, but this has the drawback that the possibility of it slipping during transportation is not precluded. Therefore, in order to avoid two pictures overlapping entirely or partly, a safety device is provided which momentarily closes a contact as soon as the film has been carried 50 mm forward. The next picture cannot be taken until this has been done. Should the contact not have been closed at the end of the forward travel then the relay switching on the X-ray tube is blocked and a warning signal is given by a lamp. The forward travel mechanism can then be brought into action again by pressing a button.

After 30 pictures have been taken the apparatus punches a mark in the edge of the film, so that certain parts of the reel of film can easily be identified in the dark room and, if desired, developed separately.

Results

As explained in the foregoing, the new feature about the Schmidt mirror systems lies in the fact that the aberrations of a spherical mirror can be almost entirely eliminated within a relatively large field of vision. Therefore when judging the performance of a mirror camera it is necessary not only to know the light-gathering power of the system (see above) but also to ascertain in how far the picture is indeed free of aberrations right up to the edges. For this purpose, with each specimen of the mirror camera a series of photographs are taken of a test plate put in the position of the fluorescent screen and containing a large number of identical test pictures spaced over the entire area (see *fig. 12*). Each test picture consists of six groups of lines with respectively 6, 12, 18, 24, 30 and 36 lines per cm (*fig. 13*). With the aid of a magnifying glass the photograph is examined, for each of the test pictures, to see which group of lines is just "resolved", the results then being recorded on an inspection form. Should the results obtained make it necessary, the camera can then be readjusted. The results of such a test, which may be taken as being representative for the 45 mm mirror camera, are shown in *fig. 14a*. Fig. 14*b* shows the results for a representative specimen of the lens camera obtained in the same way. It is seen that in the centre of the field both cameras completely resolve all groups of lines but that towards the edges the resolving power of the mirror camera is on the whole better. Whereas, with lens cameras, in the parts of the picture which are most important from the medical point of view the resolution is limited by the inaccuray of the lens optics, with the mirror camera the screen is the limiting factor. The mirror camera is, moreover, practically free of astigmatism, which is rather marked in the case of the lens camera.

From the photograph of the test plate reproduced in fig. 12 it is seen that after the spherically pressed-out film has returned to its plane the picture shows practically no barrel distortion.

Every camera leaving the works is accompanied by a photograph of the test plate made with that camera on the kind of film customarily used for

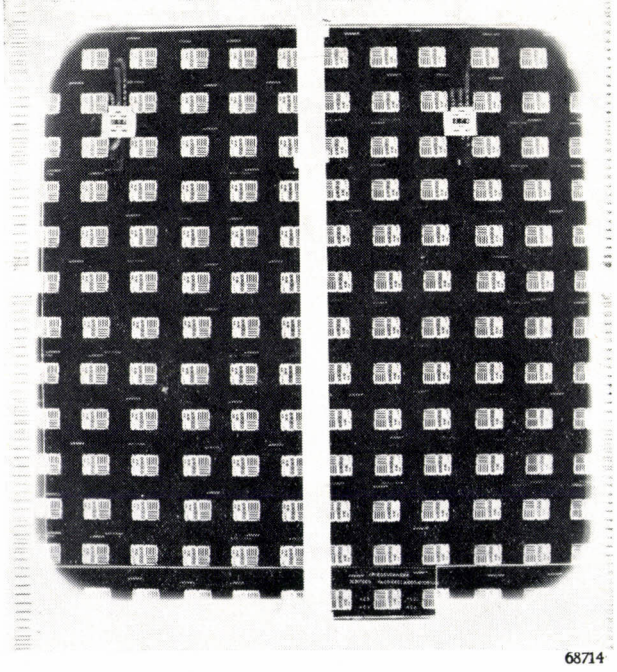


Fig. 12. Test plate with which the resolving power of each camera is tested. It has a large number of test pictures, produced photographically, each containing six groups of lines of different fineness; see fig. 13.

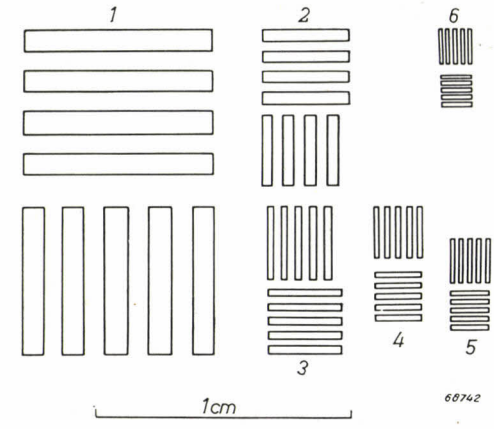


Fig. 13. Each test picture has six groups of lines, some vertical and some horizontal, with densities of respectively 6, 12, 18, 24, 30 and 36 lines per cm.

fluorography. With this photograph it is possible for the adjustment of the camera to be checked at any time, by fixing against the fluorescent screen inside the hood a number of test pictures (likewise supplied with the camera) and photographing these with the fluorescent screen itself as the source of (diffuse) light.

Finally, in *fig. 15* a three-times enlarged reproduction is given of a lung photograph taken with the mirror camera described in this article. The quality of the picture, which is of course not adequately rendered by the autotype reproduction, is almost as good as that of a contact photograph.

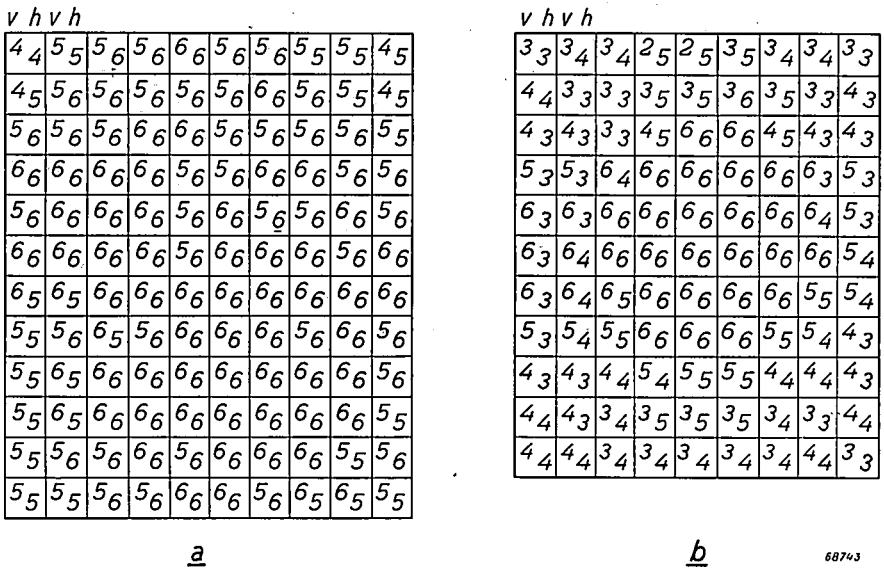


Fig. 14. Result of the testing a) of a representative specimen of the 45 mm mirror camera, *b*) ditto of a lens camera. In the square corresponding to each test picture on the photographed test plate the number is given of the group of lines (vertical and horizontal side by side) found to be resolved when examined with a strong magnifying glass. Naturally for this test a very fine-grained film is used. In the middle of the picture, for both cameras, the resolving power is so high that even the finest group of lines "6" (i.e. 36 lines per mm on the film)is resolved. Towards the edges the resolving power diminishes much less with the mirror camera than with the lens camera. Further, the mirror camera shows practically no astigmatism: the numbers for vertical and horizontal resolving power differ by no more than one unit in any square. With the lens camera the astigmatism in some parts of the field is rather strong (up to three units difference between the vertical and the horizontal numbers).

It can therefore be said that the mirror camera may now make it possible for fluorography to be applied successfully also for other purposes than mass chest survey, replacing full-size radiography for normal roentgenological practice. The resultant saving in the cost of films may play an important part in the budgets, for instance, of large hospitals.

Summary. In fluorography the X-ray image formed on a fluorescent screen is photographed on a reduced scale with the aid of a camera. For this method, which has been universally adopted for mass chest survey, the camera must have a very high light-gathering power, for the sake of good picture quality and/or a simple X-ray apparatus. For this purpose Philips constructed a camera with a Schmidt mirror system. The pictures obtained are 40 mm × 40 mm in size (reduction factor 10.5) and are recorded on non-perforated 45 mm film. A brief description of this mirror camera is given. The forward travel of the film is fully automatic, as is also the entire operation of the installation for mass chest survey. Some typical details of the mirror camera are dealt with more closely, e.g. the device for pressing out the film spherically (in the Schmidt optical system the focal plane is concentrical with the concave mirror), the adjustment of the camera, etc. The choice of the film size and the light-gathering power reached are treated at some length. With such small reduction factors as are employed in this case the aperture ratio 1 : N has to be taken as a measure for the light-gathering power. The "shadow factor" inevitable with mirror systems can also be discounted in this number; one may then speak of the "effective aperture ratio". For the 45 mm mirror camera, with reduction factor 10.5, the effective aperture ratio is 1:1.17, compared with the ratio 1:1.9 for a camera with lens systems formerly used for the same purpose (with a reduction factor 6). The result is that now chest photographs can be made with exposures of 0.04 to 0.20 sec, whereas with the lens camera 0.1 to 0.5 sec was needed. The resolving power of the mirror camera proves to be very high over the whole of the field of vision (about 25°); in the centre it is more than 36 lines per cm on the screen, and at the edges slightly less but still much better than that of the lens camera. Thanks to the very good picture quality obtained with the mirror camera it may now be possible also in normal X-ray work to replace the expensive full-size radiograms by reduced fluorograms.

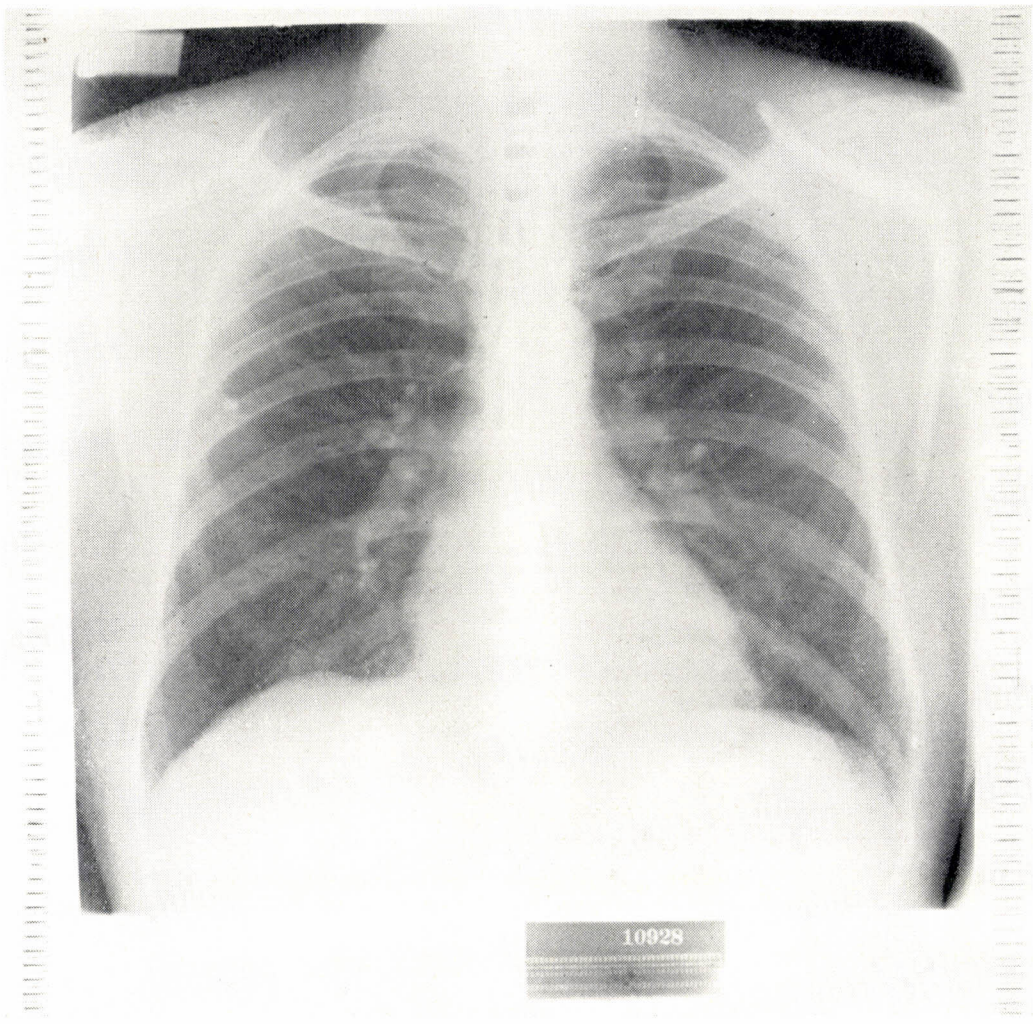


Fig. 15. Reproduction of a lung fluorogram made with the 45 mm mirror camera on Gevaert Scopic G film and then enlarged 3 times. (The reproduction of the original cannot give a quite true idea of the quality.)

GEIGER-MÜLLER COUNTERS

by N. WARMOLTZ.

537.542:539.16.08:621.387.424

The time when Geiger-Müller counters were used only in laboratories is now past. At the present day this instrument is being employed not only for measuring radioactivity in nuclear-physical research and in the production of nuclaer energy but also as an indicator to give warning of danger from radioactivity, for example in civil and military defence. Furthermore, it forms an indispensable part of the equipment of prospectors for radioactive ores. The fact that it has become possible to make with such a sensitive instrument a measuring apparatus sufficiently robust and small for many of these purposes is due in part to the discovery of the influence that certain admixtures in the quenching gas, such as the halogens, have upon the properties of a radiation counter tube.

History of the radiation counter tube

The indication of individual elementary particles, atomic nuclei and radiation quanta by means of the discharge phenomena they can bring about in an electronic discharge tube — in principle consisting of a vessel filled with gas and containing two cold electrodes — dates back to 1908, when Rutherford and Geiger succeeded in counting alpha particles with such a device. The type of counter tube used by those investigators has since then been greatly improved upon, as a deeper insight was obtained into processes taking place inside the tube. In this article the historical development of the counter tube will be reviewed and some new counters produced by Philips as a result of that development will be discussed.

The first counter tube made by the investigators mentioned closely resembled in outward appearance the form of the tubes used today. It was a cylindrical tube with two coaxial electrodes of circular cross section (*fig. 1*), the outer one serving as cathode and the inner one, in the form of a thin wire, as anode. The inter-electrode potential was about 1300 V, and the tube was filled with carbon dioxide or air under a pressure of a few cm Hg. The following is the mode of operation of their counter.

The alpha particles to be counted are admitted into the tube through a very thin mica window or through a cock with a large bore connected to a partially evacuated space in which the specimen of the radioactive substance is placed. As an alpha particle enters the tube it collides with and ionises a number of gas molecules. Under the influence of the electric field the electrons thereby released migrate towards the anode wire and mainly in its vicinity form a number of ions and electrons. All the charged particles thus created are attracted by the field towards the electrodes and produce a pulse across a load resistor; Rutherford and Geiger employed an electrometer for registering this pulse. The next alpha particle entering the tube produces another pulse, and thus the particles can be counted by counting these pulses.

No self-sustained discharge existed in the tube, the number of ions formed by the alpha particle entering the tube being simply multiplied by a certain factor.

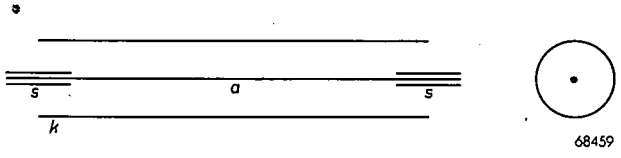


Fig. 1. Representation of a Geiger-Müller counter tube: *a* anode wire, *k* cathode cylinder, *s* screening caps for the ends.

The strength of the pulse depends, i.a., upon the position and the angle of incidence of the alpha particle in the tube, these factors influencing the paths described by the particles. To reduce this dependency Geiger and Rutherford changed the shape of the counter tube, making it spherical instead of cylindrical; a small ball acting as anode was placed in the centre of a semi-spherical cathode. The gas filling was helium under a pressure of about $^1/_3$ atm. and with this gas under such a relatively high pressure it was possible to obtain the desired multiplication of ions at relatively low potentials.

With this apparatus and the electrometers that were available at that time only alpha particles could be counted, the ionisation caused by beta and gamma rays being insufficient in spite of the multiplying action of the gas. In 1913, therefore, Geiger proceeded to raise the pressure of the gas and applied higher voltages via a large resistor in the supply circuit for the tube. In addition, a higher field strength was obtained inside the tube

by giving the electrodes a suitable shape; the cathode was given the form of a sharp point facing a flat plate acting as anode. The gas filling was air under atmospheric pressure; the voltage was 1000 to 1800 volts. As a result of all these changes the initially formed ions and electrons were multiplied to such an extent as to give rise to a self-sustained gas discharge, but (as was essential for the purpose) this discharge possessed the property of quickly quenching itself, so that another pulse could be registered before the next particle entered the tube.

With such a counter also beta and gamma rays could be detected, but it was not possible to distinguish one kind of particle from another because the height of the pulse across the load resistor was determined only by the properties of the self-sustained discharge and not by the nature of the ionising particle. This tube was practically free from directional effects.

The pointed cathode was difficult to make in a reproducible manner and was therefore subsequently replaced by a very small sphere. The anode retained its shape of a flat plate. Still later, Geiger and Müller reverted (in 1928) to the cylindrical arrangement, with air filling (5 cm pressure), and applied voltages which were higher than the starting voltage of a self-sustained discharge. The anode wire was strongly oxidized to facilitate quenching of the discharge.

Thus there arose in principle the construction of the Geiger-Müller counter as used at the present day.

In the years that followed there appeared a great variety of counter tubes, mainly distinguished one from the other by the nature of the gas filling and the processing of the electrodes. In order to judge the effect of the gas filling we have to consider the processes taking place in the gas when an ionising particle enters the tube. It was not until good use had been made of the counter tube for a number of years that a closer insight could be obtained into the nature of these processes, and even at the present day not all the problems arising have been satisfactorily solved.

The use of a gas discharge for detecting and counting particles

Let us consider a cylindrical counter having an anode wire and filled with a gas suitable for counting. The counter is exposed to a constant and not too strong radiation. Through ionisation of the gas molecules a particle entering the counter forms a number of pairs of ions and electrons along its path. The charge formed per cm of the particle's path under a specific pressure depends greatly upon the nature of the particle and its velocity. The total charge is determined by the length of the path in the counter, which length depends upon the dimensions of the counter and also upon the position and angle of incidence of the particle.

When a voltage V smaller than the ionisation voltage V_i of the gas is applied between anode and cathode, via a resistor, then — excepting the losses through recombination — only the ions and electrons formed by the ionisation are collected at the electrodes. Even if the voltage were raised to several times V_i very few new ions and electrons would be formed in the counter because the velocities reached by the electrons are mostly too low. The arrival of the ions and electrons on the electrodes can be observed as a pulse S across the afore-mentioned resistor with the aid of a sensitive amplifier. Given the geometry of the counter and the source of radiation, the pulse height is an indication of the nature and velocity of the particle; if its nature is know then its velocity can be determined, and vice versa.

The higher the voltage applied to the counter, the smaller are the losses through recombination, and thus the greater is the pulse S, but once the recombination has been reduced to a negligible amount, the value of S remains constant even though V is raised further. The voltage at which this takes place depends upon the shape of the tube and the gas filling.

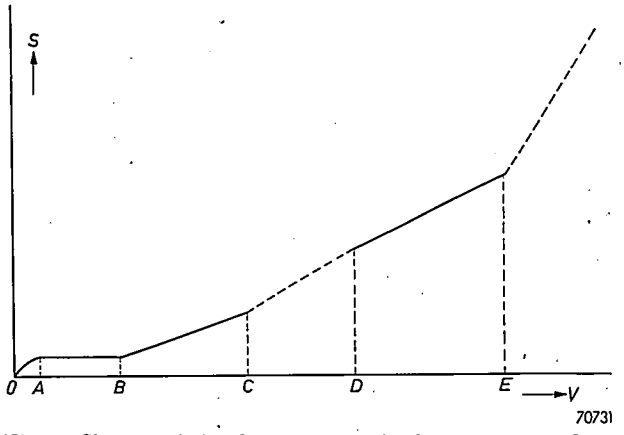


Fig. 2. Characteristic of a counter tube for constant and not too intensive radiation, showing the amplitude S of the pulses produced, as a function of the supply voltage V. In the range OA, owing to recombination, not all the ions formed by an incident particle or quantum are collected at the electrodes, whilst in the range AB such is indeed the case. OAB is said to be the range in which the tube functions as an ionisation chamber. From B to C the pulses are amplified as a result of ionisation of the gas by the electrons formed in the tube ("proportional range"). From C to D other processes begin to play a part, thereby reducing reproducibility. D corresponds to the voltage for a self-sustained discharge, DE is the Geiger-Müller range in which the discharge is quenched after a short time, while beyond E a continuous discharge occurs.

At these low voltages (some tens of volts) the counter tube is said to function as an ionisation chamber. The diagram in *fig. 2* represents the relation between the pulse height and the voltage applied to the counter tube. The part *OAB* of the curve corresponds to the range described above.

While current passes through the tube, the voltage across the tube is of course lower than the supply voltage owing to the voltage drop in the load resistor. In the case of an ionisation chamber, however, this voltage drop is negligible on account of the very small current intensity.

To give an idea of the sensitivity of a counter tube functioning as an ionisation chamber, it is to be noted that in air of 1 atmosphere pressure an alpha particle of 3.3 MeV forms about 5×10^4 pairs of ions and electrons per cm of the path length. Assuming the length of the path traversed by the particle in the chamber to be 1 cm, this means that a charge of 8×10^{-15} coulomb flows to the electrodes. The peak value of the voltage across the series resistor produced by this charge can be roughly calculated by assuming that the whole of the charge is used in charging the capacitance parallel to the resistance. For a capacitance of 10 pF (the value usually given to the input capacitance of an amplifier connected to a counter tube) we then arrive at a voltage of about 8×10^{-4} V. For an electron (beta particle) entering the counter tube with an energy of 1MeV the number of ion-electron pairs formed is about a factor 1000 smaller, and thus the pulse has a value of 8×10^{-7} V. The inevitable noise of the amplifier will amount to, say, 10^{-5} V. It is obvious that an alpha particle can be detected but not a beta particle.

When the tube voltage is raised an increasing amount of excitation and secondary ionisation of the gas molecules is produced by the electrons accelerated in the field, with the result that a greater charge is accumulated at the electrodes and the pulse becomes greater. In this way it is possible to reach an amplification up to 1000 times. Then the range *BC* in fig. 2 is reached and we have the Rutherford and Geiger counter. In this range (as also in the lower tube-voltage ranges) the pulse hight is proportional to the number of ions initially formed by the particle entering the tube. Counters operating particularly in the range *BC* have therefore been given the name of proportional counters; they are distinguished from the ionisation chamber in that they have a multiplication factor largely exceeding unity. What has been mentioned in connection with the ionisation chamber about direction-dependency and the nature of the particles to be counted applies also to these proportional counters. In this range, too, the tube currents are so small that there is hardly any drop in the tube voltage when the tube is conducting.

By raising the voltage higher the range *CD* is reached. Apart from the ionisation by electrons other processes then begin to play a part, especially the release of electrons from the cathode by the positive ion bombardment and by the impact of radiation emitted by excited gas molecules as they revert to their original state. The total amplification becomes very high, but only at the cost of reproducilibity, so that this range is less suitable for quantitative measurements.

When the voltage is higher than that corresponding to point *D* in fig. 2 a self-sustained discharge takes place as soon as the incident particle causes some ionisation in the gas. The voltage corresponding to point *D* is called the starting voltage for a self-sustained discharge. The current strength of the discharge does not depend upon the number of ions initially formed, i.e. not upon the nature of the particle entering the counter, but upon the voltage applied to the tube and also the value of the load resistor, since the voltage drop across this resistor is then no longer negligible, so that it partly determines the tube voltage during the discharge.

For the following particle to be detected it is necessary that the self-sustained discharge should be quenched before that particle enters the counter. As will be shown later, with some gas fillings this quenching takes place spontaneously provided the load resistor has a certain minimum value, but with other counters quenching is brought about by lowering the tube voltage after each discharge by means of a special circuit. Hence these counters also produce pulses. These pulses are mostly of longer duration than those produced by a proportional counter; the multiplication of the initially formed charge, however, is much greater, viz. about 10^7 times. It is this range that is invariably used in Geiger-Müller counters [1]). An electron of 1 MeV can produce pulses of 1 up to 100 V across the load resistor.

When the voltage is raised above the value corresponding to the point *E* in fig. 2 the counter is no longer self-quenching and the self-sustained discharge is very difficult to quench even with the aid of the usual simple circuits. This range is therefore never used for counting purposes.

[1]) In literature, for the sake of brevity, these are often referred to as Geiger counters, but considering their historical development the name Geiger-Müller counter is to be preferred.

The following comments will be confined to the Geiger-Müller counters, the ionisation chamber and the proportional counter being left out of consideration although these are also used for many purposes.

Properties of the Geiger-Müller counter

So far we have spoken only of the amplitude of the pulses produced by amplification of the initially formed charges. Now we shall consider to what extent the number of pulses given by the Geiger-Müller counter is a correct measure for the number of particles entering it. As will presently be seen, generally speaking not every particle entering produces a pulse; owing to various causes the efficiency of the counter is less than 100 %.

In the first place it is found that the tube voltage has little effect upon the efficiency of many Geiger-Müller counters. When plotting as a function of the tube voltage the number of counts per unit time given by a Geiger-Müller counter exposed to constant radiation, a counting characteristic as shown in *fig. 3* is obtained (for a radioactive source of moderate strength). Within the

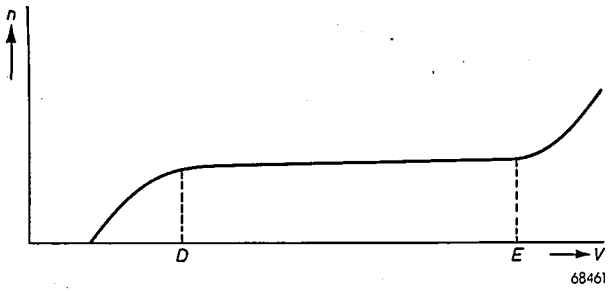


Fig. 3. Counting characteristic of a Geiger-Müller counter Plotted along the ordinate is the number (n) of particles counted per unit time in the case of a constant radiation, and along the abscissa the voltage V in the tube circuit. The range DE (these letters have the same meaning as in fig. 2) is called the plateau of the tube.

range DE, already defined above, the counting rate depends but little upon the applied voltage (often increasing slightly in a linear manner as the voltage is increased). The range DE is called the plateau of the counter tube, and for measuring purposes the supply voltage is usually chosen somewhere within this range. The slope of the plateau is expressed by the relative increase in the counting rate per 100 V increase in applied voltage. For a good counter it is customary to specify a plateau slope smaller than 10% per 100 V. The length of the plateau should be at least, say, 100 V.

In connection with these requirements it is well to take into account the manner in which the counter is to function. Let us take, for example, a counter tube having a plateau slope of 10 % per 100 V. When the tube voltage (of about 1000 V) is stabilized to within 1 volt (as is easily done) then, with a constant radiation, the voltage variations cause a variation of no more than 0.1% in the number of pulses per unit time. This is much less than the uncertainty in the various measurements due to changes in the geometry of the set-up of counter tube and specimen, which often amounts to a few per cent even though the utmost care is taken. When stabilized voltages are employed a plateau slope greater than that mentioned above might therefore suffice, but if the voltage source is not stabilized or the counter is fed from batteries (which always show a gradual voltage variation) then of course a fairly flat plateau is desired. A small slope is also of importance for counters whose plateau becomes displaced in its entirety during their lifetime; with good counters, however, this hardly ever happens.

The small slope of the plateau of a Geiger-Müller counter is due partly to the special form and arrangement of the electrodes. With a counter tube consisting of two plane parallel plates the closer the formation of the ion pairs by the incident particle is situated to the cathode the greater is the chance of a discharge taking place, because then the distance travelled by the electrons and thus the chance of new ion pairs being formed is greater. The higher the voltage the greater is the volume within which the primary ion pair may be formed to give a certain probability of initiating the discharge. Thus, with a constant radiation, the number of particles counted increases with the voltage, and this means a considerable plateau slope. With the cylindrical arrangement of the electrodes and the gas fillings commonly used, however, the area of high field strength is concentrated about the anode wire and the electrons formed by the incident particle ionise the gas only in the vicinity of that wire. Except for the few electrons formed very close to the anode wire or those which occur at the free ends of the counter tube, all the electrons traverse the favourable area and so have the same chance of forming new ion-electron pairs and thus a discharge. Disregarding the very beginning of the plateau, this chance does not depend upon the voltage, and this means a small plateau slope. (With some gas fillings some of the electrons formed at a distance away from the anode wire are "captured"; this reduces the sensitivity of the counter, but the slope of the plateau remains small.)

The fact that, nevertheless, the plateau is not always flat is due to the following causes:

1) The occurrence of secondary discharges. Sometimes the discharge initiated by the incident particle leaves behind in the counting tube so many residual charges or molecules in an excited metastable state (possibly on the surfaces of the electrodes) that sooner or later there is a chance of a second discharge pulse being formed. The higher the volt-

age the greater is the chance of these secondary discharges occurring. Of course the circuit employed is also of importance in this respect.

2) End effects. The higher the voltage the larger is the area at the free ends of the tube where an electron has a certain chance of forming a discharge, as a consequence of which the "sensitive volume" and thus the number of counts is increased.

Provided the radiation is not too strong the efficiency of a Geiger-Müller counter is hardly affected by the intensity of the radiation. It is, however, strongly influenced by the nature of the particles and by the construction of the counter tube. Counters for soft radiation (beta particles with very low energy; soft X-rays) are therefore often fitted with a window having a low absorption factor (mica window) and the composition of the gas filling is so chosen that the particles or quanta have a good chance of ionising the gas molecules [2]). Counters for hard gamma rays, on the other hand, are preferably made with a wall of a heavy metal, for instance lead, in which the gamma quanta have the greatest probability of releasing electrons, which in turn ionise gas molecules. The efficiency of such a counter is much lower.

For very weak and very strong radiations the efficiency of Geiger-Müller counters is, however, no longer constant. This is illustrated in *fig. 4*, where the number of counts recorded per unit time has been plotted as a function of the number of particles (e.g. electrons) entering the counter per unit time. For the sake of simplicity it is assumed that with an average intensity of radiation the efficiency is exactly 100 %, the slope of the line at that point then being 45°. With a low radiation intensity a greater number of counts are recorded than correspond to one pulse per incident particle. This is because a certain number of pulses is always recorded even though no source of radiation has been brought into the vicinity. These are referred to as "background". Obviously this background is of the greatest importance when measuring weak radiations, but still its presence does not preclude the possibility of measuring weak radiations accurately, as will be seen from the following. Let the number of counts due to the background during a certain period be B (this number can be counted) and the number of counts to be ascribed to the radiation source during the same period L. Then the total number of counts registered is $B + L$, from which L can be calculated. Now, owing to the statistical distribution of the instants at which the pulses are produced, in the sum $B + L$ there is an error proportional to $\sqrt{B + L}$. The relative error in L is therefore $\sqrt{B + L}/L$. By extending the counting period this relative error can be made as small as desired, even when B/L is in the order of or exceeds unity (strong background, weak source of radiation). The weaker the background, however, the shorter the test period may be for a prescribed accuracy. For this reason a Geiger-Müller counter should have the least possible background.

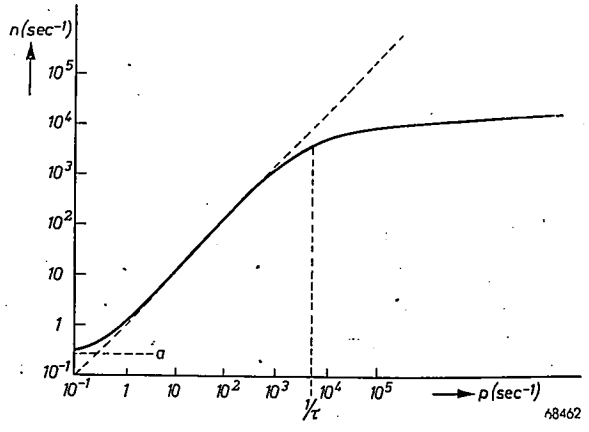


Fig. 4. The number of counts per second (n) produced by the tube, plotted as a function of the number of particles (p) entering the counter per second, for the case of a beta counter with an efficiency of 100% for an avarage radiation intensity. For a low intensity radiation the background (a) limits the accuracy of measurements. At very high intensities the resolving power of the counter sets a limit to the number of coints registered. This resolving power is determined by the finite dead time of the counter τ; when p, and thus n, becomes of the order of $1/\tau$ the curve bends to the right.

The background is mainly due to the ever-present cosmic radiation, but it is often found to have a greater value than that calculated from the intensity of that radiation alone, and this difference cannot always be fully accounted for. The background can be reduced by making the counter volume smaller; for some measurements this may be an advantage, but for others not so, because then the number of radioactive particles picked up is also reduced. Further, the counter can be screened with a lead jacket, the specimen then being placed inside the jacket.

Fig. 4 also shows that when the radiation is of great intensity the efficiency of the counter is reduced. This is due to the Geiger-Müller counter having a finite "dead time" and consequently a limited resolving power. By "dead time" is to be understood the interval of time elapsing between the beginning of a discharge and the moment when the counter tube is again able to respond to an incident particle. It comprises two parts, the dis-

[2]) See for these problems: J. Bleeksma, G. Kloos and H. J. Di Giovanni, X-ray spectrometer with Geiger counter for measuring powder diffraction patterns, Philips Techn. Rev. **10**, 1-12, 1948.

charge time, i.e. the time during which the discharge is taking place, and the recovery time, i.e. the time it takes, after quenching of the discharge, for the slowly moving, positive, gas ions to reach the cathode. During the latter interval of time these gas ions so influence the field that the ionisation brought about by a new particle is unable to initiate a new discharge. The counter does not become sensitive again until the gas ions have approached to within a certain distance from the cathode, but the pulse that can be produced at that moment is smaller than before the discharge. When all the ions have reached the cathode the tube again functions normally. Both the discharge time and the recovery time depend upon the gas filling, the voltage applied and the dimensions of the counter, and the circuit employed (threshold value).

It will be clear that if the radiation is so intense that several particles occur in a period of time equal to or shorter than the dead time, then these particles will not each produce a pulse. Thus the efficiency of the counter is reduced when the average frequency of the particles becomes of the order of the reciprocal of the dead time. In other words, the finite resolving power of the counter tube sets an upper limit to the radiation intensities that can be measured [3]).

From the foregoing it is obvious that the dead time should be kept as short as possible. For most counters it is in the order of one tenth of a millisecond.

The gas filling of a Geiger-Müller counter

As already remarked, the composition and pressure of the gas in a Geiger-Müller counter should be so chosen that the self-sustained discharge following the entry of a particle is quickly quenched, be it with or without the aid of certain circuits.

Counters with statistical quenching

Originally, for the filling of a Geiger-Müller tube gases like hydrogen, nitrogen and air were used, possibly with a rare gas added [4]).

In a counter tube filled with such a gas the ignition is followed by a stationary gas discharge. By employing a large load resistance for limiting the tube current to small values this discharge can be arranged to quench itself after a certain time. As the current passing is very small, so few particles take part in the discharge that its statistical fluctuations are large enough to quench the discharge entirely (cf. *fig. 5*). This can be understood when we come to consider the mechanism by which the discharge is maintained under low pressures. On its way towards the anode an electron formed between the electrodes releases a number of ions and other electrons, the latter doing the same in turn. At the same time a number of radiation quanta

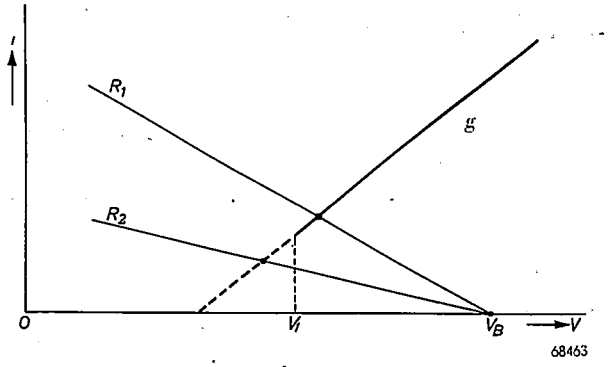


Fig. 5. The current i passing through a statistically quenching Geiger-Müller tube plotted against the tube voltage V (current-voltage characterisitc, line g). A load resistor is connected in series with the tube. With a given load resistor R the voltage at the tube during a discharge adjusts itself to the value corresponding to the point where the line g intersects the "resistance line" for the value R. The latter line intersects the abscissa at the point V_B corresponding to the supply voltage, the slope of the line being inversely proportional to the resistance. For a resistance (R_2) exceeding a given value the point of intersection lies on the broken part of the characteristic (tube voltage $< V_i$); the discharge is then unstable and is quickly quenched. For smaller resistances (R_1) the discharge is stable.

are formed. If these ions and quanta release only just one electron from the cathode the discharge is maintained. If, however, the statistically fluctuating number of particles in one of these processes drops below a certain limit, the succession is interrupted and the discharge quenched.

In fact the duration of the discharge itself shows statistical fluctuations, an example of which is to be seen in *fig. 6*, representing the discharge current in an argon-hydrogen counter as a function of time and showing a number of successive discharges, about 100, traced on the screen of a cathode-ray oscilloscope and photographed. The fluctuations in the duration of the discharge are clearly seen. The two photographs, taken for average discharge currents differing by a factor of about 3, give an idea of the relation between the average duration of the discharge and the average current. This average value of the discharge time forms part of the dead time of the counter, and to keep it small it is necessary to make the current small and thus to give the load resistance a higher value, the latter being

[3]) With low radiation intensity, too, it may happen that one particle follows another in a period of time shorter than the dead time. In practice, however, the loss in efficiency with a low radiation intensity is mostly neglible.

[4]) See, e.g., A. Bouwers and F. A. Heyn, A simple apparatus for counting electrons, Philips Techn. Rev. **6**, 75-79, 1941.

about 10^{10} ohms. This, however, has the drawback that after the discharge has ceased the original voltage across the counter is but slowly restored, so that the recovery time, and thus the dead time, is again lengthened. This effect can be minimized by employing in the place of a simple resistor a valve circuit, whereby a series resistance of about $10^6\ \Omega$

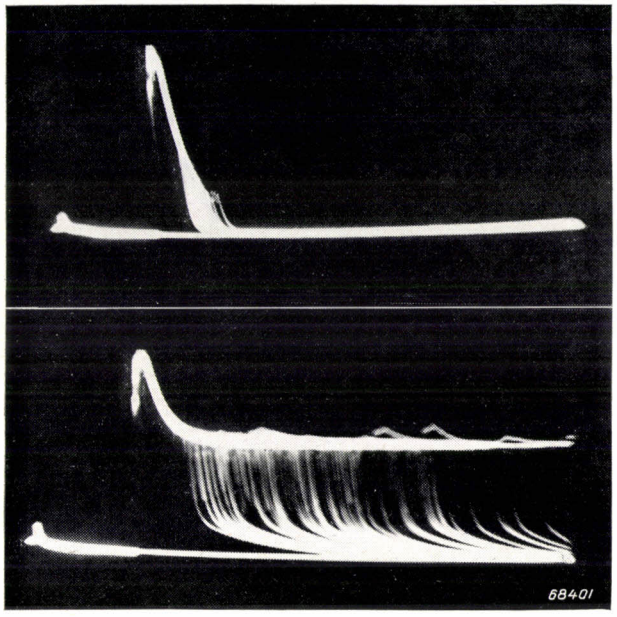


Fig. 6. Current flowing through the tube as a function of time in the case of a statistically quenching counter filled with argon and hydrogen. The oscillogram shows about 100 superimposed sweeps. The discharge always begins at the peak. There are considerable statistical fluctuations in the duration of the discharge. The average duration of the discharge depends on the tube current. The current after the ignition amounts in (*a*) to 0.12 μA, in (*b*) to 0.35 μA. The supply voltage is 870 V; the total duration of the time base sweep is 1.75 msec.

can be used instead of 10^9 to $10^{10}\ \Omega$ (e.g. a Neher-Harper or a Neher-Pickering circuit; see the article quoted in footnote [4]).

The counter tubes described above are occasionally called non-self-quenching counters, but in principle this is not the right name for them since in fact they are indeed self-quenching, though this takes place slowly and varies in duration. It is therefore better to speak of them as slow counters. The voltage range within which these counters operate extends from about 900 V to 1500 V. Examples are the Philips counters for gamma rays, types 18500 and 18501, made of stainless steel with a wall thickness of 0.3 mm and 0.1 mm respectively. The type 18501 can be used also for hard beta rays. The dead time of these counters (connected in their proper circuits) is about 10^{-4} sec. They have an unlimited life and their background (without screening of the tube) is less than 20 counts per minute.

Counters with space-charge quenching

Another type is the so-called self-quenching counter, which when used in series with a not very large load resistor is quenched spontaneously in a practically constant and relatively short time. These counters are also called rapid counters, which is a better name for them. They are characterized by certain admixtures in their gas filling.

In 1937 Trost discovered that an admixture of ethyl alcohol to the argon gas makes the counter tube self-quenching even if the load resistor has a low value. Other organic gases were also found to be useful as quenching gases, e.g. methane, amyl acetate, methylene bromide and tetramethyl lead.

The operating voltage of these counters lies at about 1000 V or higher, which is about the same as that of the slow counters, while their dead time is likewise about 10^{-4} sec but there is no need to use special circuits.

It is generally assumed that in these counters, owing to the presence of the quenching agent, the density of ionisation produced is so great that after the electrons have disappeared from the tube (they travel to the anode wire very quickly) a large space charge of positive ions is left. The field strength near the anode thus drops quickly below the value required for further ionisation and the discharge is quenched. The discharge time is therefore short. As the ions move towards the cathode the field strength at the anode is restored and the counter gradually regains sensitivity for new particles.

If the ions were to release sufficient new particles from the cathode the tube could be re-ignited, but the presence of the quenching gas appears to prevent this. A great drawback attaching to these counters with organic admixtures in the gas filling is that they have a finite and rather short life and during their lifetime vary in sensitivity. Both these effects are due to dissociation of the organic molecules through collision with electrons or to the absorption of photons, since the effect of the fragments of the molecules upon the quenching and the position of the plateau of the counter is different from that of the original molecules. Furthermore, these phenomena may have an adverse effect upon the properties of the counter when they are deposited on the anode or the cathode. The temperature (vapour pressure of the quenching gas) also influences the operation of the tube, causing variations in the plateau and thus in the adjustment and efficiency of the counter. In the course of time a great variety of radicals appear in the tube, until decom-

position has progressed so far that too many substances with unfavourable properties (electron capture, pollution of electrode surfaces) are formed and render the counter unserviceable. The rate at which this takes place depends upon the nature of the quenching gas. With large organic molecules it takes longer than with small molecules, like those of ethyl alcohol, before the fragments formed are so small that quenching becomes ineffective. Further, the life of the counter is influenced by the circuit in which it is used: if the circuit is such as to accelerate the quenching then this gives the counter a longer life. It is therefore advisable to use a quenching circuit also with these rapid counters.

The life of an alcohol counter is in the order of 10^7 to 10^9 counts, whilst its background is in the order of 10 to 20 counts per minute, depending upon its construction.

Halogen counters

In 1937 Geiger and Haxel (German patent 682 657) advised the use of halogens, in whole or in part, for the filling of counting tubes. These counters were claimed to be self-quenching and to have a long life. In the years that followed this idea was worked out further, mainly in the U.S.A. [5]), with the result that now there are very useful self-quenching counters with unlimited life and a low operating voltage.

These so-called halogen counters contain a rare gas, or a mixture of rare gases, with a small quantity of an halogen added. The halogen, of which only chlorine and bromine need be considered (fluorine is too reactive, while the vapour pressure of iodine is too dependent upon temperature), acts as the quenching gas. Dissociation of the halogen molecules in such a tube does not permanently reduce the amount of quenching gas available, because after a time the fragments recombine into complete molecules. This accounts for the practically unlimited life and constancy of halogen counters.

Halogens have a greater destructive effect upon the materials of the tube than the organic gases, but this can be avoided by a suitable choice of materials. Many of the materials commonly used in the construction of counter tubes are, therefore, unsuitable for halogen counters.

As is the case with all self-quenching counters, the discharge time is, practically speaking, not subject to statistical fluctuations, nor is it greatly influenced by the current passing through the tube.

[5]) S. H. Liebson and H. Friedman, Rev. Sci. Instr. **19**, 303, 1948.; S. H. Liebson, Rev. sci. Instr. **20**, 483, 1949.

This is illustrated in *fig. 7*, which should be compared with fig. 6 : a large number of discharges (about 100) have again been recorded on a photographic plate. It appears that, notwithstanding the greatly differing load resistances and voltages, there is practically no variation in the average discharge

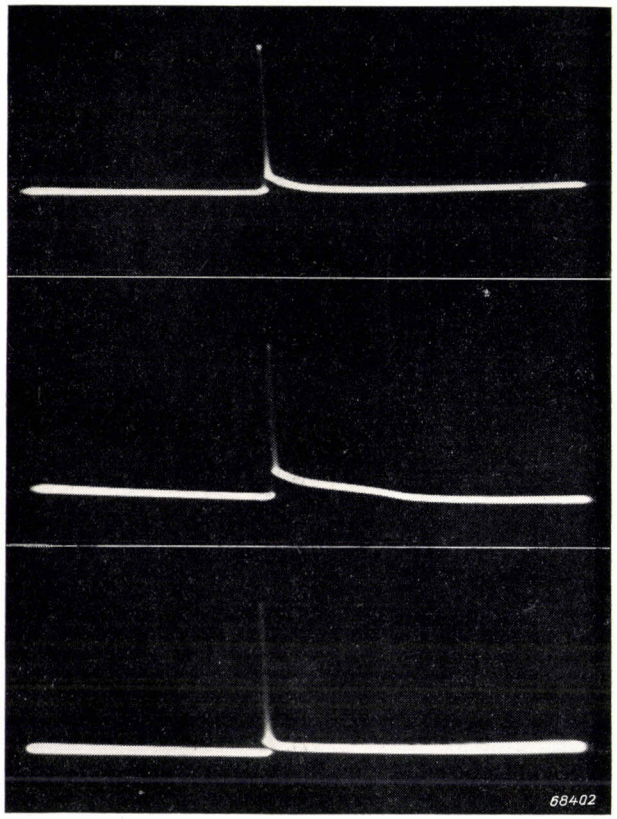


Fig. 7. As in fig. 6 but for a space-charge quenched halogen counter. The load resistor for (*a*) is 1000 MΩ, for (*b*) and (*c*) 10 MΩ .The supply voltage for (*a*) and (*b*) is 350 V and for (*c*) 800 V. Here, again, about 100 counts have been registered in succession. There is no longer any question of statistical fluctuations in the duration of the discharge. The total duration of the time-base sweep is 0.62 msec.

time and that, likewise under these differing conditions, the individual pulses are equal in height. Even when the load resistance used is so small that there is a great tendency to form a continuous discharge (cf. fig. 5) the pulse is still of the same duration. Here there is no indication at all of a gradual transition from pulses of increasing duration to a continuous discharge, as is the case with statistically quenching counters, when the load resistance is reduced.

Low-voltage counters

Until recently the voltage for a normal counter tube was in the order of 1000 V. In many cases a lower operating voltage would be preferable, namely where a battery is the source of supply. In some circuits it is also advantageous if the tube voltage

is no higher than the voltage on the electronic valves employed in the circuit. Of course the working voltage can be lowered somewhat by varying the pressure in the counter, but then the counter usually loses sensitivity.

The great influence of certain admixtures upon the disruptive voltage of a gas has been known for a long time [6]). It appears that in general a lowering of the disruptive voltage may be expected if the gas atoms possess a metastable state where their excitation energy is higher than the ionisation energy of the admixtures. The metastable atoms are then able to return to a lower state by giving off their excitation energy to the foreign atoms, which are thereby ionised. Thus for the same tube voltage more ions are obtained than when there are no admixtures in the gas, and the disruptive voltage is lower. For instance, neon with 0.1% argon has a minimum disruption voltage of about 165 V — for plane parallel iron electrodes — whilst under the same conditions pure neon has a disruptive voltage of 450 V.

It might be thought that this mixture could be used to give a counter tube a low starting voltage for a self-sustained discharge. However, to prevent the postive argon or neon ions, and any remaining metastable neon atoms (which have a long life and are not influenced by the field), from causing reignition the potential must be lowered a long time after each count, which means that the counter is very slow in its action and a special quenching circuit is required. (Such a circuit has in fact been used by Simpson.) If, however, a little halogen is also added to the mixture to act as quenching agent, the counter quenches itself in a short space of time. With this extra admixture the disruptive voltage is slightly raised, but since the quantity of halogen needed for quenching is only very small it is possible to make in this way a self-quenching counter with a working voltage of 250 V at a normal gas pressure (about 10 cm mercury). The halogen molecules themselves can also serve as pick-ups for the metastable excitation energy (neon-bromine mixtures), but then the disruptive voltage is often rather higher than when argon is also added to the gas filling.

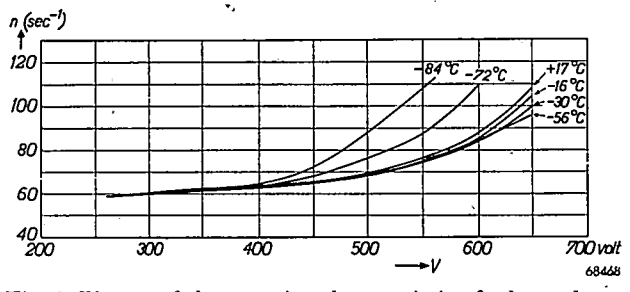


Fig. 8. Plateau of the counting characteristic of a low-voltage halogen counter at different temperatures.

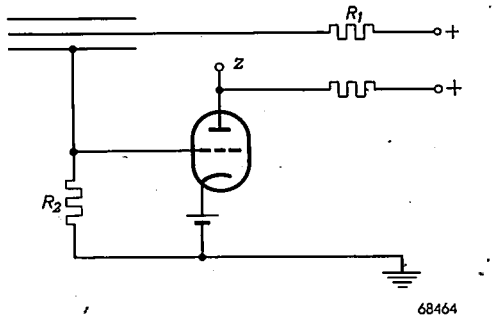


Fig. 9. Circuit with which the plateaus given in this article have been measured. The pulses are taken off at *z*.

For some applications of low-voltage counters the counting characteristic has to be insensitive to relatively large temperature fluctuations. In *fig. 8* the plateau of the counting characteristic is shown for a low-voltage halogen counter, measured at different temperatures, in the simple circuit of *fig. 9*. From this diagram (fig. 8) it is seen that between about —60 and +17 °C the influence of temperature is negligible, the length of the plateau covering a range of 250 V and the slope being less than 5% per 100 V. The plateau likewise remains practically unchanged at higher temperatures up to 100 °C. The shortening of the plateau at temperatures below —70 °C is due to condensation of bromine from the gas mixture.

Halogen low-voltage counters have a practically unlimited life, more than 10^{11} counts having been reached in life tests. The background of low-voltage counters of about the same dimensions as the types 18500 and 18501 is about 25 counts per minute when they are not screened and 14 counts per minute when screened with a 5 cm layer of lead.

When halogen counters are used in one of the normal quenching circuits account has to be taken of the fact that sometimes the capacitance in parallel with the counter influences the quenching.

Counters with an organic gas as quenching agent can only withstand discharge currents of a few microamps, whereas halogen counters can be used in circuits where the current passing through the counter has a mean value of many microamps. (In fact halogen counters can withstand a glow discharge without any adverse effect, although the continuous currents occurring then are even higher.) Such a circuit is often employed in portable counting apparatus and is represented schematically in *fig. 10*. A direct-recording meter is used for high intensities and a head-phone for low intensities. Strong gamma or X-rays of some tens of Roentgen units per hour can quite easily be measured in this way. In such cases counters are often used whose sensitivity has been purposely reduced; see the last paragraph in this article.

[6]) See, e.g. M. J. Druyvesteyn and F. M. Penning, Rev. mod. Phys. **12**, 87-174, 1940.

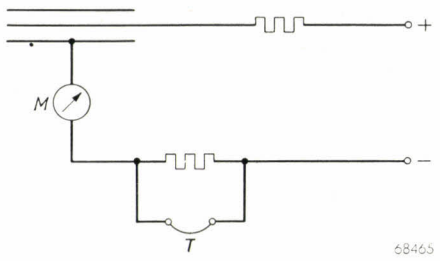


Fig. 10. Simple measuring circuit using halogen counters. *M* microammeter, e.g. 0-25 μA. *T* head-phone.

Some particular types of counters

It has been shown that in the construction of Geiger-Müller counters for various purposes all sorts of problems are involved: the choice of materials for the tube and the anode wire, the composition and thickness of the window, the dimensions of the tube and the composition of the gas filling all have to satisfy different requirements for the various applications. Without going any further into these problems here, a summary is given, by way of illustration, of some prototypes of new counters designed by Philips.

Fig. 11 shows a number of halogen counters. The counter (*a*) for beta and gamma rays, operating at about 1000 V, with an inner diameter of 30 mm and a mica window of 10 μ, has a very long and flat plateau (length about 700 V, slope 3% per 100 V). *Fig. 12* gives the counting characteristic of this tube measured in the simple circuit of fig. 9. The background amounts to 50 counts per minute (unscreened). A smaller type (*b*), diameter 15 mm and window 5 μ thick, is particularly suitable for measuring very soft beta rays and has a background of 18 to 20 counts/min. These two counters, designed for counting beta particles, also have an efficiency of about 2% for gamma rays. With a different gas filling they may also be given a working voltage of about 600 V, but the efficiency for gamma rays is somewhat reduced.

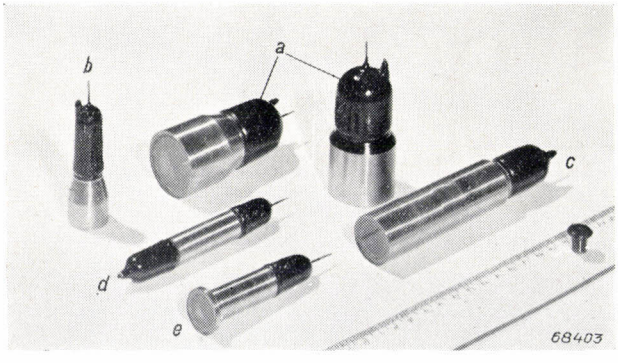


Fig. 11. Some types of halogen counters: (*a*) Counter for beta and gamma rays, 30 mm internal diameter, with mica window 10 μ thick. (*b*) The same counter in a smaller size, internal diameter 15 mm, mica window 5 μ thick. (*c*) X-ray counter, int. dia. 19 mm, mica window 10-12.5 μ. (*d*) and (*e*) Low-voltage counters, for both beta and gamma rays, int. dia. 14 mm, those for gamma rays having a mica window (5 to 10 μ thick).

The tube (*c*) is intended for use in an X-ray diffraction apparatus [2]). It is especially suitable for measuring the Kα radiation of copper, for which wavelength (about 1.5 Å) the gas filling has good absorbing properties. The counter tube described in the article quoted in footnote [2]) contained methylene bromide as quenching agent. When this tube is replaced by the halogen counter (*c*) in fig. 11, with the same working voltage, which has in fact been done in a new model of the diffraction apparatus, a gain in sensitivity of 2.5 times is obtained. This is because the sensitive volume of the tube is confined to a volume around the anode wire, since many of the electrons formed are captured and lost in the rather large quantity of methylene bromide required in that tube for quenching. The very much smaller quantity of halogen required in the new tube intercepts so few electrons that practically the entire volume of the counter is sensitive. Furthermore, in the latter case the gas pressure can be slightly raised, which favours the absorption of the Roentgen quanta and thus also improves the efficiency of the tube. The plateau of this counter has a slope of 3% per 100 V and extends from 1300 V to beyond 2000 V (see *fig. 13*). The 12 μ mica window can, if desired, be replaced by a thinner one, down to 5 μ. The resultant gain in sensitivity is particularly important in measuring the softer chromium radiation. The background amounts to 50 counts per minute.

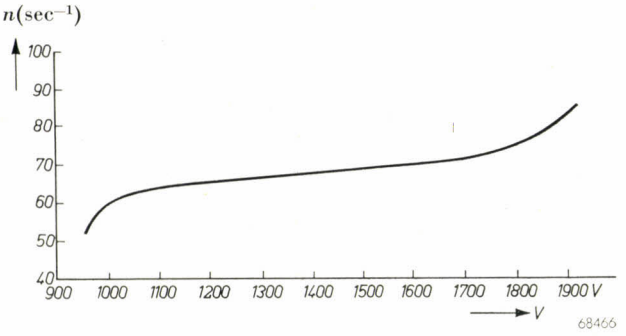


Fig. 12 Plateau of the beta and gamma counter *a* in fig. 11. Slope 3% per 100 V, length about 700 V. The background of this counter screened with 5 cm lead amounts to 25 to 30 counts per minute.

The tube (*d*) is a low-voltage counter for gamma rays and hard beta rays, while a similar tube (*e*) with mica window is intended for softer rays.

For some purposes it is advantageous to have an insensitive counter, for instance when a very high intensity of radiation is to be measured and the

resolving power of normal counters is exceeded. For hard gamma rays it is not at all convenient to render a counter insensitive by enveloping it in a lead jacket, and for portable counters it is quite impracticable. Two gamma-ray counters have therefore been developed like the type (*d*) in fig. 11; these are illustrated in *fig. 14* side by side with the first type. One of these counters (*b*) is 20 times and the other (*c*) 200 times less sensitive than the original counter (*a*). They both operate at a relatively low voltage and have a small background corresponding to their dimensions.

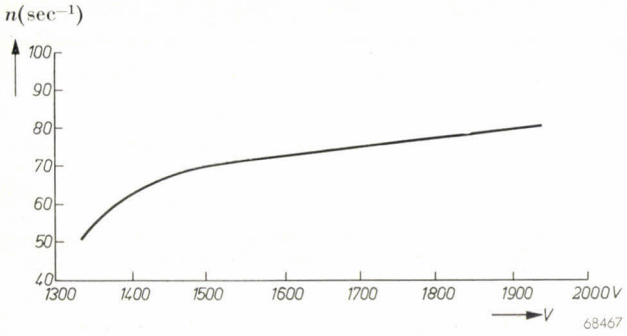


Fig. 13. Plateau of the X-rays counter *c* in fig. 11. This extends to beyond 2000 V and has a slope of about 3% per 100 V. The background is 50 counts per minute.

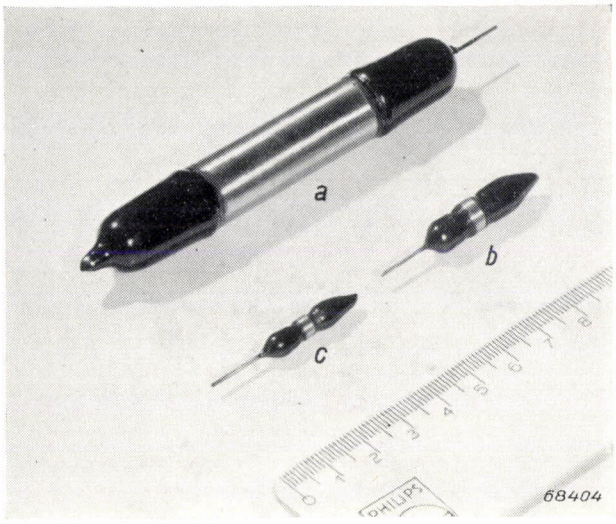


Fig. 14. Insensitive counters for gamma rays. (*a*) The same counter as *d* in fig. 11. (*b*) A counter 20 times less sensitive than *a*, (*c*) a counter 200 times less sensitive.

Summary. The historical development of Geiger-Müller counters from the first original type up to the present-day designs is discussed with reference to the properties of gas discharges, with a brief indication of the various ways in which a gas-discharge tube can be used for counting ionising particles or quanta. Particular consideration is givento the properties of counting tubes in the "Geiger-Müller range" (plateau) and the classification of existing counters into those with statistical quenching and those with space-charge quenching. The properties of the argon-hydrogen counter, an example of a statistically quenched tube, and those of the counter with an organic gas as quenching agent, an example of a counter with space-charge quenching, are gone into more closely. Finally the new self-quenching counters with halogen as quenching agent are described in various constructions and applications, mention being made, among others, of a low-voltage counter with a starting voltage of 250 volts, some types of beta and gamma counters with mica window, and a counter of great sensitivity for X-rays of about 1 to 2 Å wavelength; these counters have a practically unlimited life.

63407

THE STRUCTURE OF GLASS

by J. M. STEVELS.

666.1:539.213.1

Although the material "glass" has now been known for some thousands of years, the variations in which it may be obtained are not by any means exhausted. This is evident from the fact that during the last decades many kinds of glass with new properties have been developed and found application in various technical directions. For the man of science it is gratifying to see that now, after centuries of empiric rule in this domain, in many cases theoretical conceptions as to the structure of glass are serving as a guide to the development of new materials.

Introduction

For many uses of glass, particularly in the field of electrotechnology, the requirements which have to be met nowadays go far beyond the properties that are usually to be found in normal kinds of glass. To illustrate this we have only to take a few cases occurring in the field of activity of our laboratories. For radar transmitting valves a glass is required which shows no excessive dielectric losses in an alternating field with frequencies in the order of 10^{10} c/s (3-cm waves). For the transmitting valves used for normal radio broadcasting at frequencies round about 10^6 c/s (300 metres) the glass must likewise not have too high dielectric losses, whilst it must also have a low softening point for easy manufacture. The development of the cathode-ray tubes for television receivers particularly involves special glass-technical problems. For the small projection-television tubes, in which electrons are accelerated by voltages of 25 kilovolts or higher, a glass is required which is able to withstand lengthy bombardment by the electrons and by X-rays without being subject to discoloration.

It was fortunate that at the moment when such very special requirements as these were placed before the glass technologist some insight had already been obtained into the structure of glass in general. This suggested the lines on which further work could be done to reach the desired results.

The foundations for our present-day knowledge of the structure of glass were laid by Zachariasen [1], who in 1932 wrote a classical article on the subject. The theories he expounded have already been set forth in this journal [2], so that it will suffice to recall them here quite briefly. We shall then proceed to deal with the refinements which Zachariasen's theory has subsequently undergone on various points and which are due mainly to the work of Warren and Weyl in the U.S.A. and of Dietzel and Smekal in Germany. It is due to these refinements in particular that we are not only able to grasp a number of peculiarities in the already known physical properties of glass but are now better prepared to cope with the new requirements of technical science.

Zachariasen's theory

There are a number of oxides, called glass-forming oxides, which may occur in the vitreous as well as in the crystalline state (SiO_2, B_2O_3, P_2O_5). According to Zachariasen, in the crystalline and in the vitreous state these oxides are built up of the same elements, namely polyhedrons (tetrahedrons or triangles) of oxygen ions with the highly charged cations Si^{4+}, B^{3+}, P^{5+} at their centres. The only difference is that in the crystalline state these polyhedrons are arranged regularly, whereas in the vitreous state they are not. A schematic representation of this is given in *fig. 1*.

For the network of oxygen polyhedrons to occur in these two forms it is necessary that the structure of the oxide in the crystalline state satisfies a number of conditions given by Zachariasen, which (see the article quoted in footnote [2])) involve the following.

1. Every oxygen ion must be bound to not more than two positive ions which must be highly charged and small.
2. The number of oxygen ions which surround such a positive ion (forming a polyhedron) must be neither very large nor small (3 or 4).
3. The oxygen polyhedrons adjacent to each other must have common corners (bridging oxygen ions), but no common edges or faces.
4. Each polyhedron must have at least three oxygen ions in common with neighbouring polyhedrons.

[1]) W. H. Zachariasen, J. Amer. Chem. Soc. **54**, 3841, 1932.
[2]) J. M. Stevels, Philips Techn. Rev. **8**, 231—237, 1946.

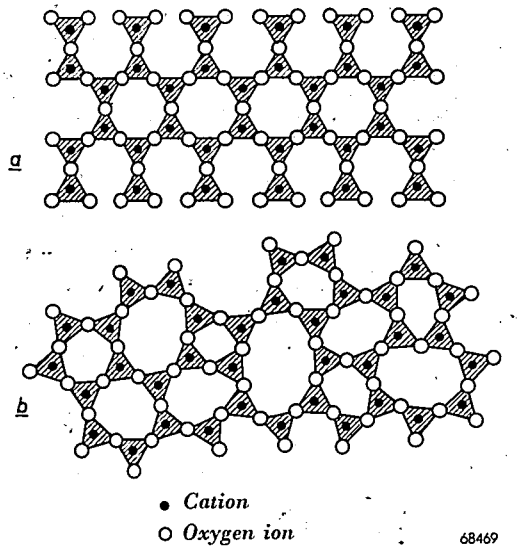


Fig. 1. Two-dimensional representation *a*) of a crystalline lattice A_2O_3 (A is a cation) and *b*) the corresponding network in the vitreous state.

We shall not enter into the arguments on which these rules have been formulated. Suffice it to say that oxides which in the crystalline state satisfy these conditions have a very open structure, so that the polyhedrons need not necessarily arrange themselves according to a periodic pattern when, upon cooling of the melt, the solid phase is formed.

The structure of the many glasses that can be made by fusing these glass-forming oxides together with a large number of metallic oxides, such as CaO, BaO, PbO, Na_2O, is likewise described by the theory of Zachariasen. While the metal ions find a place in the interstices of the open network just mentioned, the added oxygen ions are taken up through a number of oxygen bridges in the network being broken, each bridging oxygen ion being replaced by two non-bridging oxygen ions. Obviously any continued change of this nature will greatly influence the properties of the glass.

Fig. 2. is a diagrammatic representation of the structure of such a glass.

Actually, of course, it is not each metal ion that seeks a place in an interstice of the network. It is better to say that while the melt is solidifying the oxygen chains arrange themselves around the metal ions. The presence of these ions modifies to some extent the form that the network finally assumes, and for that reason they have been given the name of network modifiers. The ions occurring in the centres of the oxygen polyhedrons, and which, therefore, together with the oxygen ions actually form the network, are called network formers.

This simple theory of the structure of glasses was verified about 1938 with the aid of X-ray diffraction photographs, mainly by Warren and his associates [3]).

We shall now proceed to discuss some of the aforementioned refinements of the theory that have been made in the course of the last ten years.

The structure of borate glasses

It will be obvious that the addition of metallic oxides to a glass-forming oxide cannot be continued at will. It has already been pointed out that the breaking down of oxygen bridges greatly influences the properties of a glass, and one can imagine that if this is carried too far the typical vitreous structure of coherent tetrahedrons or triangles ultimately becomes unstable. It can readily be understood that a network completely spatially bonded is no longer possible as soon as the number of contact points (bridging oxygen ions) per polyhedron, denoted by Y, becomes less than two [4]). "Islands" are then formed, which in themselves may consist of a rather large number of polyhedrons. The smaller the value of Y, the smaller are the islands. For $Y = 1$ these islands average two polyhedrons, whilst for $Y = 0$ the "structure"

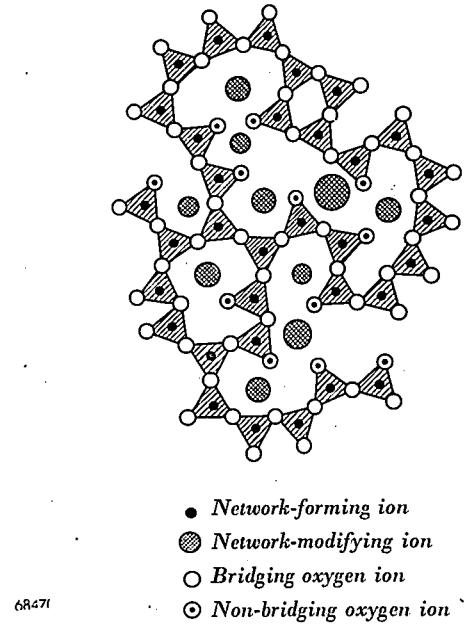


Fig. 2. Two-dimensional representation of the network of a glass containing, in addition to oxygen ions, both network-forming and various network-modifying ions. Non-bridging oxygen ions occur.

[3]) B. E. Warren and J. Biscoe, J. Amer. Ceram. Soc. **18**, 49, 1935; J. Biscoe and B. E. Warren, J. Amer. Ceram. Soc. **21**, 287, 1938.

[4]) J. M. Stevels, J. Soc. Glass Techn. **30**, 31, 1946; Ned. T. Natuurkunde **12**, 257, 1946.

consists of isolated polyhedrons. Obviously the structures with small Y values will have a strong tendency to order themselves, thus readily changing into the crystalline state.

The number of contact points per polyhedron, Y, can quite easily be calculated from the composition of the glass. In this connection the composition is characterized by the ratio R of the total number of oxygen ions to the total number of network-forming ions. For glasses with tetrahedrons as elements, i.e. with network-forming ions having a coordination number of 4 (silicate and phosphate glasses), the relation is [5]):

$$Y = 8 - 2R, \quad \ldots\ldots \quad (1)$$

whilst for glasses with triangles as elements, i.e. with network-forming ions having a coordination of 3, the rule is:

$$Y = 6 - 2R \ldots\ldots \quad (2)$$

Silicate glasses can, indeed, only be formed as long as $R \leqq 3$; in other words, as soon as $Y < 2$ (i.e. as soon as islands begin to form) it is no longer possible for the melt to solidify in the vitreous state. With phosphate glasses it appears that one can go a little farther in adding metallic oxides, as far as $R = 3.2$. Then $Y = 1.6$, which means that it is possible to reach the vitreous state notwithstanding the occurrence of (not too small) islands.

With borate glasses, considering that the B_2O_3 structure consists of triangles, one would expect the addition of metallic oxides to cause formation of islands ($Y<2$) as soon as $R>2$. With borium-containing glasses, however, a peculiar complication arises, in that in such glasses metallic oxides may be taken up according to a mechanism differing somewhat from that described above. When metallic oxides are added to B_2O_3 — for instance Na_2O — the Na^+ ions will find a place as network modifiers but the mechanism of bridge breaking does not take place. The excess of added oxygen is taken up owing to the property of the B^{3+} ion being able to occur in the centre of an oxygen triangle but also in the centre of an oxygen tetrahedron. Thus the network is then built up from both oxygen triangles and oxygen tetrahedrons, while there is not a single non-bridging oxygen ion. Such a structure is represented in *fig. 3*.

It is remarkable that this process of taking up oxygen, whereby a change takes place in the coordination number for some of the B^{3+} ions and the network gains more and more in strength, continues until a certain concentration is reached, which, for the system Na_2O-B_2O_3, is 18 mol. % Na_2O. With greater concentrations of Na_2O the previously described bridge-breaking mechanism comes into action again, non-bridging oxygen ions then being formed. These ranges of concentrations have been named respectively the accumulation region and the destruction region.

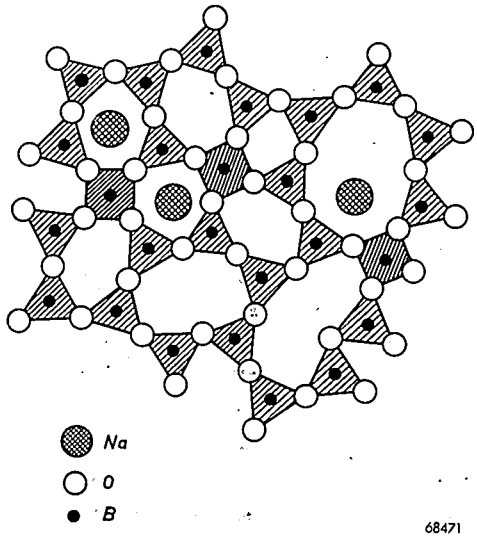


Fig. 3. Two-dimensional representation of a borate glass with small Na_2O content. This is characterized by the absence of non-bridging oxygen ions. (B^{3+} ions intriangular or tetrahedral surroundings are represented in the diagram by a coordination of 3 and 4 respectively in the plane of the drawing.)

The state of affairs outlined here implies that in such a borate glass relatively more metallic oxides can be taken up before island formation arises. In the accumulation region, where no non-bridging oxygen ions occur, Y always equals $2R$ (cf. footnote [5])), so that the average number of contact points per polyhedron increases with increasing R! In the destruction region the ratio of the munber of triangles to the number of tetrahedrons remains constant, such that the average coordination number of the B^{3+} ion $Z = 3.22$. In this region, therefore, the relation (cf. footnote [5]) is:

$$Y = 6{,}44 - 2R \ldots\ldots \quad (3)$$

Island-forming ($Y<2$) will then not take place until $R > 2.22$. Actually it appears that most borates may become vitreous up to $R = 2.4$, so that here again, just as in the case of phosphate

[5]) Denoting the average number of non-bridging oxygen ions per polyhedron by X and the total number of oxygen ions per polyhedron (the coordination number) by Z, then, as may be verified by simply counting the ions, $X+Y=Z$ and $X+\frac{1}{2}Y=R$, from which it follows that $Y=2Z-2R$.

glasses, a small degree of island-formation is no hindrance to vitrification.

The foregoing is illustrated graphically in *fig. 4*.

Of particular importance in practice are the consequences that the changing of the coordination number and the accompanying strengthening of the network have for the physical properties of this group of glasses. By way of example, in *fig. 5* the coefficient of thermal expansion of the pure sodium-borate glasses is plotted as a function

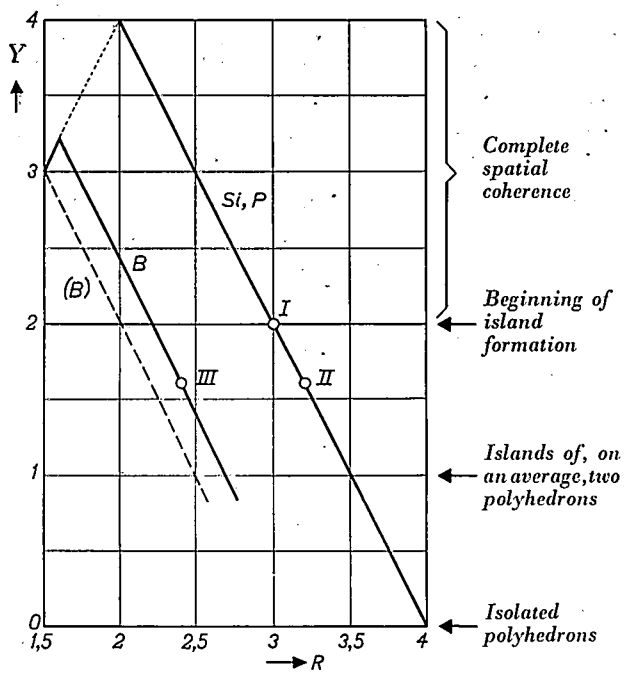


Fig. 4. Relation between Y, the average number of bridging oxygen ions per polyhedron, and R, the total number of oxygen ions divided by the total number of network-forming ions. The more metallic oxides are added to the glass-forming oxide, the greater is the value of R. The line *Si,P* applies for a network of tetrahedrons (eq. 1), the broken line (*B*) for a network of triangles (eq. 2).

From $Y = 2$ onwards island-formation occurs. In the case of silicate glasses this indicates the limit to which R can be raised without devitrification occurring (point *I*). In phosphate glasses the vitreous state is still tenable up to $Y = 1.6$, i.e. $R = 3.2$ (point *II*).

In the case of borate glasses the addition of metallic oxide is at first accompanied by a transition of B^{3+} ions from the coordination number 3 to the number 4; then $Y = 2R$ (beginning of the sharply bent, fully-drawn line *B*). When the coordination number has reached the average value $Z=3.22$ it remains constant and Y drops with increasing R according to eq. (3). Island formation ($Y=2$) then begins at $R = 2.22$, but for most borate glasses the vitreous state appears to be still possible up to $R = 2.4$ (point *III*).

of their composition. This curve shows a minimum just at that composition where, according to the foregoing considerations, the network has the strongest structure.

Something similar is found in the case of borosilicate glasses. On the boundary line between the accumulation region and the destruction region *(fig. 6)* are the glasses which, compared with other

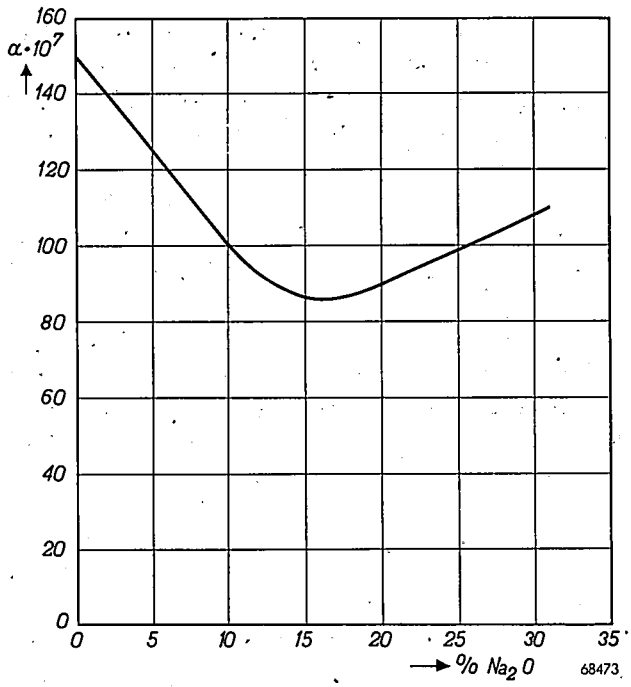


Fig. 5. The expansion coefficient of pure sodium-borate glasses as a function of the Na_2O content, expressed in weight %.

borosilicate glasses, are characterized by a maximum strength of structure (small expansion coefficient and high softening point), for example the "Pyrex" glasses.

The changing of the coordination number of the B^{3+} ion is therefore responsible for the fact that, under suitably chosen conditions, the incorporation in glass of B_2O_3, which itself has a very high expansion coefficient and very readily melts, results in a reduction of the expansion coeffiicent and a raising of the softening point. Various other

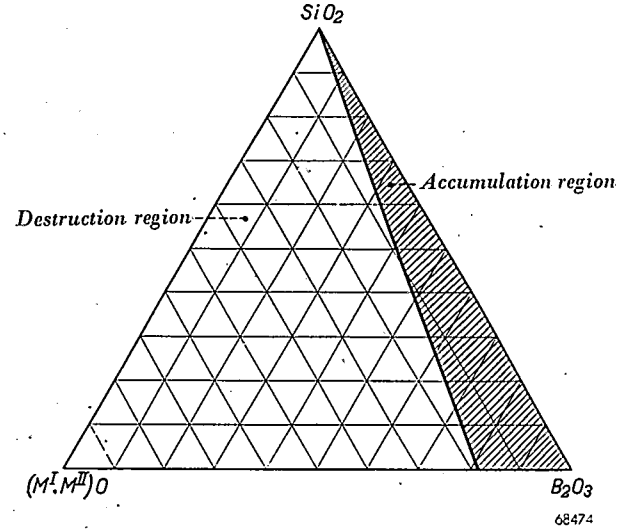


Fig. 6. Phase diagram of the system $(M^I\ M^{II})O$-B_2O_3-SiO_2, where M^I and M^{II} represent respectively a monovalent and a bivalent metal, which may occur in any proportions. The diagram can be divided into an accumulation region and a destruction region.

properties of the glass are likewise influenced by the structural change described, as for instance the dielectric losses. This will be discussed in greater detail in another article to be published shortly in this journal.

The conceptions of network-forming and network-modifying ions

Zachariasen thought that the positive ions could be divided into two groups, which in the foregoing have been denoted as network formers and network modifiers, but it has meanwhile become apparent that this division does not hold in all cases. It is now known that there are a large number of ions which may occur in glass in both forms, and often simultaneously. This means that in glass there are some ions of a certain type which have 4 oxygen ions surrounding them, thus being situated in the centre of the tetrahedrons forming the network typical for glass, while other ions of that type are surrounded by a larger number of oxygen ions. In some cases the positon of the "equilibrium" between the two forms may be roughly determined by the colour of the glass, as is the case, for instance, with the nickel ion. Ni^{2+}ions may occur in a glass mainly as network formers, in which case the glass has a purple colour; under other circumstances these Ni^{2+}ions mainly occur as network modifiers, the glass then being yellow. Such differences in colour in cases where the coordination number is changed have been found, for example, with Cu^{2+}, Co^{2+}, Fe^{2+}, Fe^{3+}, Mn^{2+}, Mn^{3+} and U^{6+}. But it is also possible to detect these changes in the coordination number among non-colouring ions, by a method which has recently been indicated [6]) and which will be discussed in our next article.

The "equilibrium" between the forms of coordination depends not only upon a number of external factors (temperature and furnace atmosphere when making the glass, and the rate of cooling) but also to a large extent upon the quantities of the different ions contained in the glass. The simple theory of Zachariasen could not furnish any explanation for this, but it has been made comprehensible by means of a "competition principle". In general, the small, highly charged ions which are capable of attracting the oxygen ions close to them, and strongly binding them, will easily surround themselves with only four oxygen ions (or three in the case of B^{3+}). The larger and less charged cations are then surrounded by a larger number of oxygen ions, to which they are less strongly bound. In principle, however, also these latter cations may be surrounded by only four oxygen ions, thus entering into competition with the small, highly charged ions, in which they may be successful, for instance, when relatively few oxygen ions are present.

Various investigators have attempted to express this "competition capacity" of the ions numerically. A suitable measure has been found to be the "field strength" of the ion at the centre of an adjacent oxygen ion, z/a^2, where z is the charge of the ion and a the distance between the centres of the two adjacent ions [7]). When the ions are arranged in the order of their calculated field strength (see table I) their sequence gives an idea of the preference that the various ions have for occurring either as network formers or as network modifiers. When different cations occur in a glass simultaneously, those with the highest z/a^2 value will preferably occupy the network-forming positions (smallest coordination numbers), while those with the lowest z/a^2 value show a preference for the network-modifying positions (largest coordination numbers).

Table I. The "field strength" z/a^2 for a number of cations, according to Dietzel [7]); z = charge of the ion, expressed in elementary charges; a = distance (in Ångstrom) from the centre of the ion to the centre of an adjacent oxygen ion. (If the cation is rather large it then no longer fits in the interstice of a tetrahedron formed by four oxygen ions, and for this reason alone the natural coordination number is greater than four).

Ion \ Coordination number	3	4	6	8
P^{5+}		2.08		
B^{3+}	1.65	1.45		
Si^{4+}		1.57		
Al^{3+}		.97	0.84	
Zr^{4+}				0.78
Be^{2+}		0.87		
Ni^{2+}		0.61	0.55	
Zn^{2+}		0.59	0.53	
Mg^{2+}		0.51	0.45	
Ca^{2+}			0.35	0.33
Ba^{2+}				0.24
Li^{+}			0.23	
Na^{+}			0.19	0.17
K^{+}				0.13

Some striking examples can be given to illustrate this.

When silicon ions and phosphor ions enter into competition with each other, as in the glass of the composition SiP_2O_7 (i.e. $SiO_2.P_2O_5$), from the table

[6]) J. M. Stevels, Verres et Réfractaires 2, 2, 1948.

[7]) A. Dietzel, Glastechn. Ber. 22, 41, 1948.

it is to be predicted that the phosphor ions will "win". For this glass it is difficult to determine directly with what coordination numbers the ions Si^{4+} and P^{5+} occur in it. The crystalline SiP_2O_7 has never been examined in this respect either, but crystalline ZrP_2O_7, which is isomorphoush, as. By X-ray diffraction analysis it was found that each P^{5+}ion is indeed surrounded by four bridging oxygen ions and each Zr^{4+}ion by six bridging oxygen ions. Since it may be assumed that the coordination numbers will be the same in the vitreous state and in the crystalline state, we have here a case where Si^{4+} occurs in a coordination of 6, i.e. the Si^{4+}ion, the network former *par excellence,* occupies only a network-modifying position.

The opposite effect is seen in the following example. Of the compound $2CaO.SiO_2$ only a crystalline form is known. This is in conformity with the aforementioned considerations about Y and R: for four oxygen ions there is only one network-forming ion, thus $R = 4$ and therefore, according to eq. (1), $Y = 0$. Though this substance is built up from SiO_4 tetrahedrons, these are all isolated and therefore arrange themselves into a well-ordered structure, with Ca^{2+}ions as "binder". The remarkable fact is that the entirely analogous compound $2ZnO.SiO_2$ may indeed occur in the vitreous state. An explanation for this may find support in the fact that Zn^{2+} stands some steps higher in table I than Ca^{2+}. It is therefore feasible that in the system in question Ca^{2+} occurs exclusively with a coordination number six or higher, i.e. as a network modifier, whereas under suitably chosen conditions Zn^{2+} may occur also with coordination number four, thus possibly playing the part of a network-forming ion. If in the compound $2ZnO.SiO_2$ only 16.7% of the Zn^{2+}ions occur as network formers — instead of writing Zn_2SiO_4 the composition can then be better written as $Zn_{5/3}\ (Zn_{1/3}Si)O_4$ — then the quotient R is already reduced to exactly 3 and thus vitrification is possible ($Y = 2$). Actually slightly more than 16.7% of the Zn^{2+}ions will be present as network formers. Vitrification in the system ZnO-SiO_2 continues until roughly the composition 2.2 $ZnO.SiO_2$ is reached. The condition $R \leqq 3$ can then only be satisfied if at least 18.2% of the Zn^{2+}ions occupy a network-forming position [8]).

[8]) Here it is tacitly assumed that the limit found in practice for the occurrence of the vitreous silicates ($R=3$) would likewise apply for the glasses containing Si^{4+} and few Zn^{2+} ions as network formers. *A priori* this need not strictly be the case, it being possible for the limit to lie at a somewhat greater value of R ($Y<2$) (cf. the phosphate and borate glasses). Then smaller percentages of network-forming Zn^{2+}ions would be sufficient to account for vitrification in the case of the compositions mentioned.

Certain barium-containing silicate glasses (the heavy barium crown glasses) have been shown to contain a number of barium ions forming part of a typically tetrahedral formation. Thus even Ba^{2+}ions may sometime play the part of network formers.

Resuming, the following advances have thus been made upon the old theory. According to Zachariasen's criteria Zn^{2+} and Ba^{2+}ions should always behave as network modifiers. The new conception (the competition principle) makes it feasible that such ions may also accur as network formers, and that such is more likely to be the case the greater the concentrations in which these ions are present. This is of great practical importance, since the addition of one and the same ion to a glass may influence the properties to a different extent (or even in the opposite direction), depending upon the position it takes in the network: the Mg^{2+}ion, for instance, yields, as network modifier, a contribution towards the dielectric losses of a glass at 10^6 c/s, whereas as a network former it is harmless in this frequency range.

How the nature of the bond influences vitrification

The theory of Zachariasen has undergone a considerable evolution in recent years in yet another respect. Whereas Zachariasen never discussed the nature of the bond in vitreous systems and, moreover, regarded the elements (the polyhedrons) as being invariable, in recent years it has been pointed out, especially by German investigators, that the nature of the chemical bond in the elements is of great importance in determining whether a certain substance may or may not occur in the vitreous state. Smekal, to whose work this theory is mainly to be ascribed, has indicated that it is feasible [9]) that a condition for vitrification is that the system must contain "mixed" bonds. By this it is meant that, in addition to directed bonding forces (homopolar bonds, which are confined to certain mutual "valence angles", such as, e.g., in CO_2 or NH_3 or $SiCl_4$), also non-directed bonding forces (heteropolar bond or Van der Waals bond) must be in action. The two kinds of bonding forces may be united in one bonding direction, such as, e.g., in the Si-O bond, which is to be described as a combination of a force of the type of a homopolar bond and one of the type of a heteropolar bond, or they may be present in different bonding directions. The latter is found to be the case, for instance, in the vitreous selenium and in

[9]) A. Smekal, Nova Acta Leopoldina **11**, 511, 1942; worked out in more detail in Glastechn. Ber. **22**, 278, 1949.

the chain macromolecules, where in the direction of the chains a homopolar bond prevails while the chains are mutually kept together mainly by a Van der Waals bond [10]).

These considerations are of great importance because they relate to all kinds of glasses, whereas Zachariasen's theories apply exclusively to kinds of glass formed by inorganic oxides.

The difference between Smekal's theory and that of Zachariasen may be formulated by saying that the latter theory ascribes the irregular structure of the network exclusively to the irregular packing of the oxygen polyhedrons. According to Smekal's theory, applied to the oxide glasses, the type of bond within the polyhedrons may change, in that the bond may bear a more homopolar or a more heteropolar character. Consequently, therefore, interatomic distances in those polyhedrons may change, and this may also contribute towards the irregular structure of the network in vitreous systems.

Inorganic glasses in which Zachariasen's rules do not hold

Having accepted this line of thought as being correct, one will not be surprised to find that inorganic systems exist which do not at all obey the rules given by Zachariasen and nevertheless occur in the vitreous state.

Excluding those resembling selenium (such as vitreous sulphur and tellurium), the earliest known example is the system of carbonates K_2CO_3-$MgCO_3$. According to Zachariasen's rules, as applied to the oxide CO_2, there would be no question of vitrification here; one would rather expect a regular crystal lattice formed by the positive metal ions and negatively charged CO_3^{2-}groups. It has now been found, however, that the heating of a mixture of K_2CO_3 and $MgCO_3$ in equimolecular quantities does indeed lead to vitrification. The glass certainly has a strong devitrifying tendency: at 300 °C it completely devitrifies in an hour, but at temperatures below 150 °C the vitreous state is maintained. Quite analogous systems, such as $Na_2Ca(CO_3)_2$, $K_2Ca(CO_3)_2$, $Na_2Mg(CO_3)_2$, $NaLiCO_3$ and $KLiCO_3$, on the other hand, show no trace of vitrification.

Looking back at table I, it is seen that of all the cations in question here K^+ and Mg^{2+} differ most in "field strength". The fact that of the systems mentioned only $K_2Mg(CO_3)_2$ forms a glass is an indication that the condition favourable for vitrification is created by the combination of a very weak and a much stronger cation. It may, therefore, be so interpreted that the greatly varying electric fields prevailing in the system deform the CO_3^{2-}groups in a different way, thereby, according to Smekal, promoting vitrification.

Something similar has been found with the nitrates. A mixture of equimolecular quantities of $Ca(NO_3)_2$ and KNO_3 becomes vitreous when cooled from the molten state, whereas such is not the case with the system Ca (NO_3)2-$NaNO_3$. Apparently the difference in strength of the Ca^{2+}ion and the Na^+ ion is not large enough for a sufficient deformation of the NO_3^-groups. On the other hand the system $Mg(NO_3)_2$-KNO_3 would, indeed, be expected to occur in the vitreous state, and this has in fact been found to be the case. This is not so easily proved because it is difficult to obtain anhydrous $Mg(NO_3)_2$ in a chemically pure state. In the procedure followed by us a mixture of $Mg(NO_3)_2.4H_2O$ was melted in a platinum crucible (the melting point is 92 °C) and then carefully heated further, thereby removing all the water contained in it. As soon as the melt begins to give off nitrous vapour (at about 290 °C) an equimolecular quantity of KNO_3 is added. The homogeneous melt of $KMg(NO_3)_2$ thereby obtained is quenched by decanting in droplets in liquid air. X-ray diffraction photographs showed that the product was vitreous.

With the sulphates there is a still more striking example. The system K_2SO_4-$MgSO_4$ does not, in any composition, yield a vitreous product, notwithstanding the great difference in strength between the K^+ and the Mg^{2+}ions. When, however, the Mg^{2+}ion is replaced by an ion with a still stronger field, viz. the H^+ion, vitrification is possible: by heating $KHSO_4$ to just above the melting point and then quenching by decanting onto an iron plate, a glass is obtained. At a temperature of 70 °C this glass can be drawn into threads.

Finally, a very striking confirmation of Smekal's theory has been found with normal oxide glasses themselves [11]). According to what has been set forth above, in a system such as, for instance, Na_2O-B_2O_3 vitrification occurs as long as the amount of Na_2O does not exceed a certain percentage corresponding to a value of $R \approx 2.4$. The same limit is found for the system K_2O-B_2O_3. When examining, however, the mixed alkali borates, such as the system Na_2O-K_2O-B_2O_3, it is found that there are kinds of glass where R is much greater

[10]) In the light of these considerations it will be realized that theoretically it is difficult to find justification for the classification of the ions according to Dietzel's criterion, which is based entirely upon a heteropolar bond picture. Nevertheless it has been found very useful as a heuristic and didactic principle.

[11]) A. W. Bastress, Glass Science Bull. 4, 133, 1946; 6, 9, 1947; 6, 12, 1947.

than the limit mentioned. The same phenomenon is found with the mixed alkali silicates and alkali phosphates, such as illustrated for some systems by the phase diagrams in *figs 7* and *8*. As far as the value of R is concerned, the area of vitrification would be expected to be bounded in each diagram by the broken line: as a matter of fact vitrification

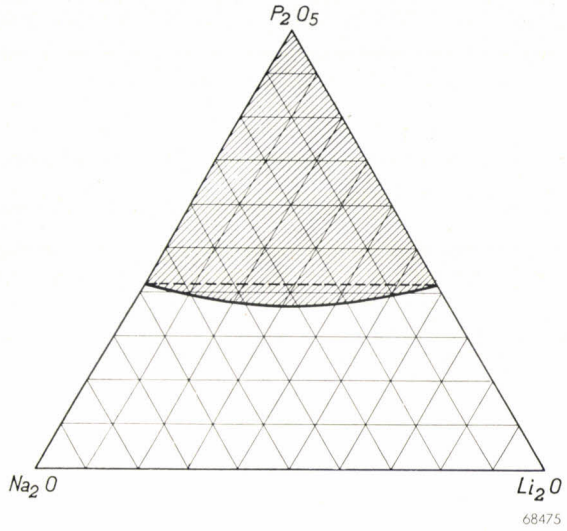


Fig. 7. Phase diagram of the system Na_2O-Li_2O-P_2O_5. The vitreous state may be obtained with any composition in the hatched area. (Figs 7 and 8 have been taken from the publications by Bastress quoted in footnote [11].)

occurs within the whole of the hatched area. The most striking fact is that the relative increase of the vitrification area is on the whole larger the more the respective cations differ in „field strength";

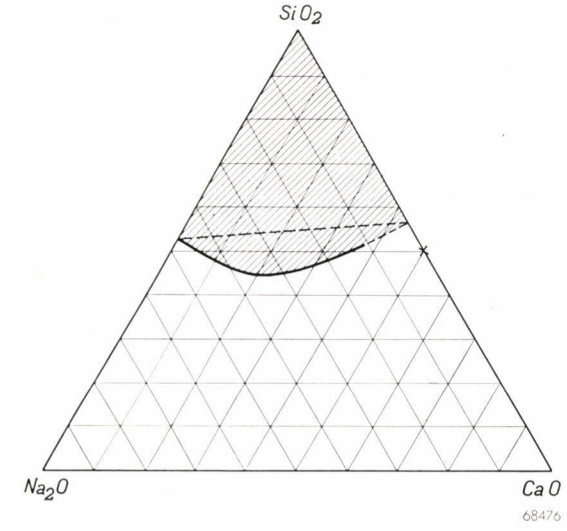


Fig. 8. Phase diagram of the system Na_2O-CaO-SiO_2. The vitreous state may be obtained with any composition in the hatched area. Contrary to this diagram, according to J. B. Ferguson and H. E. Merwin (Amer. J. Sci. **48**, 81, 1919) a substance of the composition $CaO.SiO_2$ (denoted by a cross in the diagram) can occur in the vitreous state; thus here $R = 3$.

for an example see *table II*. Herein lies the explanation for the whole phenomenon. The threshold value of R is related to the fact that with increasing number of oxygen ions the polyhedrons, originally considered as being entirely rigid, are ultimately no longer able to retain their mutual coherence in a network that is becoming more and more open. Owing to the great variations in the electric fields, however, the oxygen polyhedrons are themselves somewhat deformed. This deformation makes it possible for a coherent vitreous network to be formed also when the value of R is somewhat greater (smaller Y).

Table II. Relative expansion of the vitrification area in the phase diagram of mixed alkali borates, compared with the area in which the condition for R is satisfied [11]).

System	Increase in %
Na_2O-K_2O-B_2O_3	15
Na_2O-Li_2O-B_2O_3	50
Li_2O-K_2O-B_2O_3	100

In conclusion it may be said that due to recent refinements of Zachariasen's theory many details in the phenomenon of vitrification can now be better understood. In particular the existence of all sorts of glasses not fitting in the old scheme has become understandable, while the new conceptions may serve as a guide — and in many cases have already so served — when seeking new glasses possessing entirely new combinations of properties.

Summary. Zachariasen's theory indicates under what conditions oxides may occur in the vitreous state. These conditions are based upon a conception of the structure of glass as an irregular network of oxygen tetrahedrons (or triangles) in whose centres are situated the small, highly charged ions B^{3+}, Si^{4+} or P^{5+} (network formers), while in the interstices of the network there may be taken up, as network modifiers, all sorts of large, less charged, cations surrounded by a larger number of oxygen ions. According to new conceptions this picture has to be refined in some respects. The coordination numbers are not fixed to the extent originally supposed. The number of oxygen ions surrounding B^{3+} may change from 3 to 4, thereby explaining various phenomena encountered with borate glasses. Further, under certain conditions typical network modifiers (with normal surroundings of, say, 6 or 8 oxygen ions) may act as network formers (with coordination number four), and vice versa. The behaviour of the cations in this respect is governed by a competition principle, to which expression is given by a classification according to the order of their "field strength". According to Smekal the occurrence of bonding forces of different types between the particles in the network is essential for vitrification. Many facts not covered by Zachariasen's theory, or in contradiction therewith, thus find a plausible explanation. A number of examples are given by way of illustration.

VOL. 13 No. 11 pp. 301-336 MAY 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

SOME APPLICATIONS OF FERROXCUBE

by W. SIX.

621.318.1:538.246.1

It is not possible to indicate any general rules governing the uses of Ferroxcube materials; in fact, it is necessary to investigate each individual case on its merits to determine whether and in what from Ferroxcube can be employed. At the same time, a few examples in which Ferroxcube has amply proved the value of its properties will serve to furnish some idea of the field of possibilities which this material has opened, and of methods that will ensure the best practical results.

Ferroxcube is a ceramic magnetic material which may be described chemically as consisting of mixed crystals of simple cubic ferrites such as Mn-Zn-ferrite (Ferroxcube III) and Ni-Zn-ferrite (Ferroxcube IV). These materials were originally developed as material for the cores of inductors such as are used in telephone work, where the eddy current losses inherent in the highly conductive metallic ferromagnetic materials at the higher frequencies render the use of these materials extremely difficult. It may be assumed to be sufficiently well-known that, so far, we have open to us a choice of two methods by means of which iron (i.e. highly conductive) cores can be made suitable for use at high frequencies, namely by laminating the core or by using the so-called powder-core. In either case the eddy current losses are certainly reduced but, as we shall presently show, these expedients do not provide an ideal solution to the problem. On the other hand, owing to the generally high volume resistivity of Ferroxcube the eddy current losses in this material are relatively small, even in the case of the solid material.

Beyond a certain frequency, which is dependent on the composition of this material, the "residual" losses in Ferroxcube may be quite considerable; it appears that this frequency (described as the "ferromagnetic resonant frequency" [1])) is proportionately higher according as the initial permeability of the material is lower. It is therefore possible, at the cost of the initial permeability, to extend the range of frequencies within which the residual losses will remain sufficiently low, this being but one example of the much more general variability of the properties of Ferroxcube. The facility that exists for varying the composition of the material as desired enables us so to vary the various characteristics that the requirements imposed by a particular problem will be met in the best possible manner. At the same time, in order to secure the fullest advantages of these new ferromagnetic materials, such problems should be approached as far as possible with a certain open-mindedness, in not only endeavouring to suit the choice of material to the problem, but, if necessary, to present the problem in such a light that it will fit in with the possibilities offered by the unusual properties of the new material.

Before proceeding to a discussion of the actual examples of the uses of Ferroxcube, we should say something about the general characteristics of magnetic circuits in which an air-gap is incorporated.

The effective permeability

As long as the flux density is kept low, as explained in I, the losses in Ferroxcube may almost be regarded as falling entirely under the heading of residual losses. These can best be described by introducing a complex permeability value $\mu = \mu' - j\mu''$. That the occurrence of losses is thus actually implied will be seen from the following.

The inductance L of an ideal toroidal coil forming a closed magnetic circuit (that is to say

[1]) J. J. Went and E. W. Gorter, The magnetic and electrical properties of Ferroxcube materials, Philips Techn. Rev. 13, 181-193, 1952 (No. 7). Hereinafter denoted by I.

with all the lines of force passing wholly through the ferromagnetic core) is represented by:

$$L = \mu_r L_0 ,$$

where μ_r is the relative permeability ($\mu = \mu_r \mu_0$, $\mu_0 = 4\pi \cdot 10^{-7}$ H/m = permeability of a vacuum) and L_0 is the inductance of the same coil without core [2]). For an alternating voltage of angular frequency ω the impedance of the coil is:

$$Z = j\omega L = j\omega \mu L_0 .$$

If we write μ as a complex quantity in the manner suggested above, it will be seen that the impedance Z is not purely imaginary, but that it has a real (resistive) component, which means that losses are involved. The magnitude of this component is $\omega \mu' L_0 \cdot \tan \delta$, where $\tan \delta = \mu''/\mu'$ defines the tangent of the loss angle.

The quality factor of such a coil is then:

$$Q = \frac{1}{\tan \delta} .$$

For an open magnetic circuit derived from the kind of circuit just described by introducing an air-gap in the core (*fig. 1*), we can now define the relative effective permeability μ_1 by putting:

$$L = \mu_1 L_0 .$$

If the air gap is so placed as to be perpendicular to the lines of force and is at the same time so small that the field outside the coil will be negligible (no leakage), μ_1 may be represented by the equation:

$$\frac{1}{\mu_1} = \frac{\lambda^k}{\mu} + \lambda^l , \quad \ldots\ldots\ldots \quad (1)$$

where λ^k and λ^l are the fractions of the total magnetic path occupied by the core material and air respectively ($\lambda^k + \lambda^l = 1$) (see fig. 1). As an approximation the losses can now be calculated from:

$$\frac{\tan \delta_1}{\mu_1} = \frac{\tan \delta}{\mu} \quad \ldots\ldots\ldots \quad (2)$$

where $\tan \delta_1 = \mu_1''/\mu_1'$.

The conditions under which equation (2) holds are revealed by the following derivations of equations (1) and (2).

Take once more the above-mentioned case of the ideal toroidal coil. H^k and B^k will denote the magnetising force and flux density in the core respectively, whilst H^l and $B^l = \mu_0 H^l$ are the magnetising force and flux density in the air-gap.

Now the following relation holds for a closed line of force S (see fig. 1):

$$\oint_S H_s ds = iN , \quad \ldots\ldots\ldots \quad (3)$$

in which i is the current flowing in the coil and N is the number of turns. Further, $B^k = B^l$. Let a denote the total length of the magnetic path; then, in view of equation (3), we may write:

$$H^l \lambda^l a + H^k \lambda^k a = iN = H^s a, \quad \ldots\ldots \quad (4)$$

where H^s is the magnetising force produced by the coil *in vacuo*. By definition, moreover, $\mu_0 H^l = B^l = B^k$ and $\mu_0 H^k = B^k/\mu$. Multiplication of equation (4) by μ_0/a then gives:

$$B^k \lambda^l + \frac{B^k}{\mu} \lambda^k = \mu_0 H^s . \quad \ldots\ldots \quad (5)$$

Now the inductance of the ideal toroidal coil in question is equal to N times the total magnetic flux in a cross-section of the coil, when a direct current of 1 A is passing through it. Since the flux density of such a coil may be regarded as being homogeneous, it may be said that $L = NB^k O$ and $L_0 = N\mu H^s O$, where O is the cross-sectional area of the coil. In view of the fact that $L = \mu' L_0$, it thus follows that:

$$B^k = \mu' \mu_0 H^s ,$$

and this, in conjunction with equation (5), gives us equation (1).

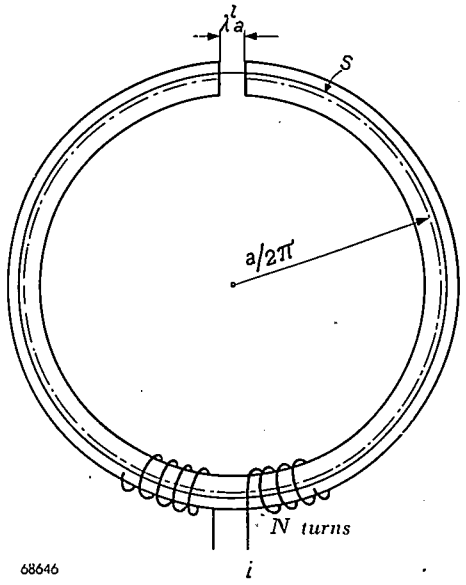


Fig. 1. Diagram illustrating the deduction of equation (1).

Equation (2) can be derived from equation (1), the simplest method being by means of the (relative) parallel permeability, which complex quantity is defined as follows (see also *fig. 2*):

$$\frac{1}{\mu_p'} + \frac{j}{\mu_p''} = \frac{1}{\mu} = \frac{1}{\mu' - j\mu''} .$$

Explicitly then:

$$\left.\begin{aligned} \mu_p' &= \mu' (1 + \tan^2 \delta) \\ \mu_p'' &= \mu'' \frac{1 + \tan^2 \delta}{\tan^2 \delta} \end{aligned}\right\} \quad \ldots\ldots \quad (6)$$

[2]) In the further text only the relative permeability is referred to, whilst, for the sake of simplicity, the index r will be omitted.

If we now express formula (1) in terms of μ'_p and μ''_p, at the same time making use of the relationship $\lambda^l = 1 - \lambda^k$, we find that:

$$\frac{1}{\mu'_{1p}} - 1 + j\frac{1}{\mu''_{1p}} = \lambda^k\left(\frac{1}{\mu'_p} - 1\right) + j\lambda^k\frac{1}{\mu''_p}.$$

This virtually constitutes two equations, one of which refers to the real and one to the imaginary part. By eliminating λ^k from these two equations and using the relationship expressed in (6), viz:

$$\tan\delta = \frac{\mu''}{\mu'} = \frac{\mu'_p}{\mu''_p},$$

it is found that:

$$\frac{\tan\delta_1}{\mu'_{1p} - 1} = \frac{\tan\delta}{\mu'_p - 1}. \quad \ldots\ldots\ldots \quad (7)$$

If we assume that $\tan\delta$ and $\tan\delta_1$ are small with respect to unity, it may be said, disregarding the second powers of $\tan\delta$ and $\tan\delta_1$, that:

$$\mu'_{1p} = \mu'_1 \quad \text{and} \quad \mu'_p = \mu'.$$

Assuming, further, that μ_1 and μ' are large compared with unity — this being generally justifiable — equation (7) reduces to equation (2).

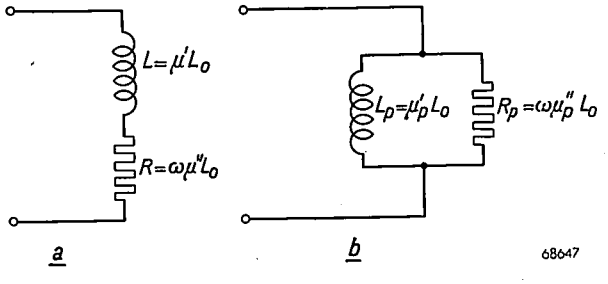


Fig. 2. Equivalent circuits of a ferromagnetic system (with losses), in the form of a core (closed magnetic circuit).
The core has a complex (series) permeability $\mu = \mu' - j\mu''$, the impedance being represented by $Z = j\omega\mu L_0 = j\omega\mu' L_0 + \omega\mu'' L_0$. This is the equivalent of an inductance $\mu' L_0$ with resistance $\omega\mu'' L_0$ in series (see *a*).
The (complex) parallel permeability is defined as $1/\mu = 1/\mu'_p + j/\mu''_p$ from which it follows that $1/Z = 1/j\omega\mu L_0 = 1/j\omega\mu'_p L_0 + 1/\omega\mu''_p L_0$. This is the equivalent of an inductance $L_p = \mu'_p L_0$ with resistance $R_p = \omega\mu''_p L_0$ in parallel (see *b*).

From equations (1) and (2) it will be noted that the introduction of an air-gap does certainly reduce the permeability, but that the tangent of the loss angle $\tan\delta$ is then also smaller. It will be found, as applied to Ferroxcube materials, that this possibility of reducing the losses may be of great advantage.

Let us now proceed to a review of some practical examples of the uses of Ferroxcube, dealing in turn with filter coils, loading coils, wide-band H.F. transformers as used in telephony; an I.F. transformer for radio receivers, a high tension generator for television receivers, and cores such as are employed for H.F. heating purposes. In conclusion a few remarks will be devoted to certain applications in which the otherwise undesirable high losses above a certain frequency can be put to good use.

This selection of examples makes no pretence at completeness; it is intended rather to demonstrate through the individual examples themselves the new possibilities which Ferroxcube has to offer to designers of network components.

Uses in telephony

Filter coils

For an appreciation of the reasons why Ferroxcube is so particularly adapted for use as core material for filter coils of the type used in carrier telephony, let us first consider the more important requirements that such filters have to satisfy.

In the first place, losses must be small, notably in order to ensure sharp separation between the pass and attenuation bands. Further, the flux density of the core should always be low; otherwise distortion and intermodulation due to non-linearity of the magnetisation curve will give rise to difficulties.

Again, in view of the fact that a large number of filters must be provided in every carrier-telephony station for the expenditure of the least possible amount of space, it is important for the filter coils to be not only of the required quality, but of the smallest practicable dimensions.

Another important requisite is that the inductance of the filter coil shall be as insensitive as possible to variations in temperature. Since the temperature coefficient of the whole assembly is largely dependent on the temperature coefficient of the permeability $(1/\mu)\cdot(d\mu dT)$, this should be as low as possible.

The field outside the coil should be so small that the various coils comprising a complete filter and mounted close together will not interfere with each other; in other words the magnetic shielding of the coil must be effective.

From the point of view of economy in space and cost it is preferable for the inductance of the coils to be adjustable, as this dispenses with the need for trimming capacitors for aligning the various circuits; such trimmers are relatively costly and take up extra space.

It will be seen from equation (2) that, owing to the residual magnetic losses, the loss angle δ depends on the size of any air-gap that may be incorporated in the circuit, this loss angle being smaller according as the effective permeability is lower. We shall now show that the values of other forms of losses occurring in the coil are also changed when an air-gap is provided, and that the total coil losses can be mini-

mized by suitable dimensioning of the air-gap.

The total losses in an inductor comprise the following:

1) Magnetic losses in the core.
2) Copper loss in the turns, which can be subdivided into:
   a) losses due to D.C. resistance,
   b) losses due to eddy currents in the copper.
3) Dielectric losses in the insulating materials.

The losses mentioned under (3) can be reduced to negligible proportions by taking a suitable value for the inductance.

D.C. losses in the copper are represented by loss angle δ_0. This loss angle is increased by introducing an air-gap in the circuit. It can easily be seen that:

$$\frac{\tan \delta_{1_0}}{\tan \delta_0} = \frac{\mu'}{\mu_1'}. \quad \ldots\ldots\ldots \quad (8)$$

A similar relationship appears to exist in respect of the eddy current losses in the copper, represented by a loss angle δ_w, viz;

$$\frac{\tan \delta_{1w}}{\tan \delta_w} = \frac{\mu'}{\mu_1'} \ldots\ldots\ldots \quad (9)$$

Derivation of equations (8) and (9):

Let the current flowing in the coil be the same with and without air-gap. The equivalent resistance ΔR_0, in which the D.C. resistance is discounted, is therefore also the same. The real component of the inductance of the coil, however, changes from $\mu' L_0$ to $\mu_1' L_0$, so that, before and after providing the air-gap:

$$\tan \delta_0 = \frac{\Delta R_0}{\mu' L_0} \quad \text{and} \quad \tan \delta_{1_0} = \frac{\Delta R_0}{\mu_1' L_0},$$

from which equation (8) immediately follows.

Equation (9) is obtained in exactly the same manner by assuming that the amount of power dissipated by the eddy currents in the copper is independent of the presence of the air-gap. It may be concluded that this is indeed the case, seeing that equation (9) is of actual practical significance; in other words when an air-gap is introduced, the variation in the stray field around the turns of wire is small compared with the variation in the permeability.

From the foregoing it becomes clear that a part of the total loss is proportional to μ_1', namely the residual losses in the core, whereas another part, the copper losses, is inversely proportional to μ_1'. If the loss angle of each of the two forms of loss is small compared with unity, it may be said that, for the tangent of the total loss angle $\delta_{tot} = \delta + \delta_w + \delta_0$:

$$\tan \delta_{tot} = \tan \delta + \tan \delta_w + \tan \delta_0$$

as an approximation. The quality factor of the coil under consideration is given by $Q = 1/\tan \delta_{tot}$. Owing to the dependence of the losses upon the value of μ_1', it is possible to make the value of μ_1' such that Q will be at a maximum, and this is the case when:

$$\mu_1' = \mu' \sqrt{\frac{\tan \delta_0 + \tan \delta_w}{\tan \delta}}.$$

Now Ferroxcube lends itself much more readily to this facility for reducing the losses than metallic ferromagnetic materials, for the following reasons.

Metallic ferromagnetic materials are suitable for high frequencies only when employed in the powdered form, and in this form the effective permeability is low because the insulating layers between the ferromagnetic particles produce the same effect as a very large number of air-gaps. This objection is met as far as possible by making the coil toroidal, for the magnetic field produced by this kind of coil is almost wholly concentrated in the space within the coil; in effect then, there is no stray field, and no further steps are necessary to avoid magnetic leakage. At the same time, if this leakage is not to become too large, a macroscopic air-gap cannot be employed; hence μ_1 can be varied only by modifying the composition of the material used for the core.

Apart from this limitation, a toroidal coil has the disadvantage that it is a complicated and costly process to wind such coils; moreover, powder cores have the inherent disadvantage that the packing of the magnetic material in the core is not homogeneous and that the flux density is therefore not uniform, the lines of force tending to pass as far as possible through the ferromagnetic particles. At those points where the particles touch each other the flux density is very high. This involves additional losses and distortion in consequence of hysteresis.

This objection could naturally be eliminated by making the toroidal coil with a Ferroxcube core, but the advantages of this new material would not be fully manifested by substituting Ferroxcube for a powder core in a toroidal coil. These advantages come into their own only when the coil is given an entirely different form (since the toroidal form is now no longer necessary) and notably one that will best be suited to the characteristics of the new core material. We will now describe the design for a filter coil for use in carrier telephony adopted by the Philips Laboratories, Eindhoven.

The filter is constructed on the lines of a "pot" core (see *fig. 3*). The core comprises a central piece (*C*), on which the coil is mounted, two plates (*B*) and (*O*), and a ring (*R*) which encloses the coil as if it were in a pot. The cylindrical centre piece is cemented to the bottom plate with an intermediate

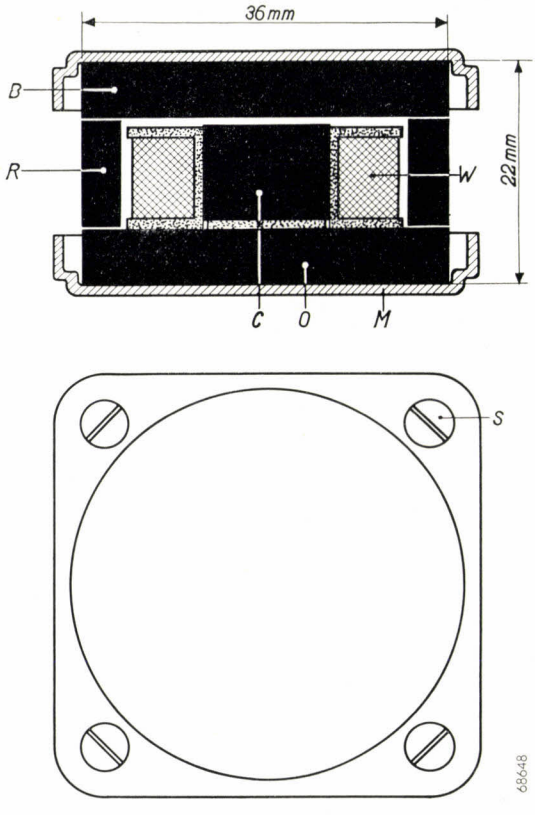


Fig. 3. Contruction of a filter coil for carrier telephony. W = coil. O = bottom plate to which the centre portion of the core C is cemented with an intermediate layer of non-ferromagnetic material. R = annular core section. C, B, R and O are of Ferroxcube. M = brass plates which, with bolts S, clamp the assemblage together.

layer of a non-ferromagnetic material, and the height of this piece is such that an air-gap is formed between the top plate and the top face of the cylinder, of approximately the same size as the layer of non-ferromagnetic material between the other end of the cylinder and the bottom plate. The use of an air-gap of this kind is made possible by the relatively high permeability of the core and enables an almost perfect screening. It is, moreover, possible to adjust the size of the gap by removing a small quantity of material either from the top of the cylindrical piece or from the ring. This may be done by using emery cloth, as Ferroxcube is a hard ceramic material and as such can only be processed by some form of grinding operation. Owing to this facility of adjustment of the air-gap, a ready control on the effective permeability μ_1 is obtained and it is accordingly possible to determine the inductance of the coil to within $^1/_2\,\%$, thus giving μ_1 the particular value that will ensure maximum quality of the coil (see above).

When assembled, the coil with the ring and Ferroxcube plates are clamped between two brass plates (M) (see fig. 3) and the whole is impregnated to protect the coil from the effects of humidity.

Should it be necessary to make adjustments to the self-inductance after the filter has been impregnated, the following procedure is adopted. Before the coil is immersed in the impregnating medium, a thin metal strip is inserted between the top plate and the upper edge of the ring (slots are provided in the latter for this purpose), so that it passes through the air-gap. Then, when the assembly has been impregnated, the metal strip is replaced by a strip of thermo-plastic material to which a wedge-shaped layer of Ferroxcube powder has been applied. The amount of Ferroxcube powder within the air-gap can then be varied by pushing in or pulling out the wafer as required. In this way the inductance can be varied by amounts which may lie between 0.1 $^0/_{00}$ and 1 or 2 $^0/_0$.

The method of construction in question also offers facilities for varying the temperature coefficient of the permeability, which is reduced by the presence of the air-gap proportionately with μ_1/μ, as will be seen at once on differentiating equation (1), viz: $-(1/\mu^2{}_1)(\mathrm{d}\mu_1/\mathrm{d}T) = -\lambda^k(1/\mu^2)(\mathrm{d}\mu/\mathrm{d}T)$. If the air-gap is sufficiently small, $\lambda^k \approx 1$; hence:

$$\frac{1}{\mu_1{}^2}\frac{\mathrm{d}\mu_1}{\mathrm{d}T} \approx \frac{1}{\mu^2}\frac{\mathrm{d}\mu}{\mathrm{d}T}. \quad \ldots \quad (10)\ ^3)$$

and multiplication of equation (10) by μ_1 then gives the ratio in question.

Should it be desirable still further to reduce the temperature coefficient, the method of providing a polarising field described in I must be resorted to. For this purpose a piece of permanent magnetic material is incorporated in the circuit. In order not to introduce eddy-current losses thereby, the volume resistivity of this material must be high. For this reason it is of great advantage to employ Ferroxdure, the new ceramic material for permanent magnets discussed at length in a recent issue of this review [4]).

The method of construction outlined above thus quite simply meets our requirements of effective screening, easy correction of the inductance and a low temperature coefficient. The extent to which it has proved possible to combine a high quality factor Q with small dimensions is amply demonstrated in *fig. 4* which depicts two filter coils designed for the same purpose, one with Ferroxcube and

[3]) Seeing that according to this formula the quantity $(1/\mu^2)(\mathrm{d}\mu/\mathrm{d}T)$ is hardly changed by the presence of an air-gap, this quantity is often stated as temperature coefficient (see I, table III, column 5).

[4]) J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Ferroxdure, a class of new permanent magnet materials, Philips Techn. Review **13**, 194, 1952 (No. 7).

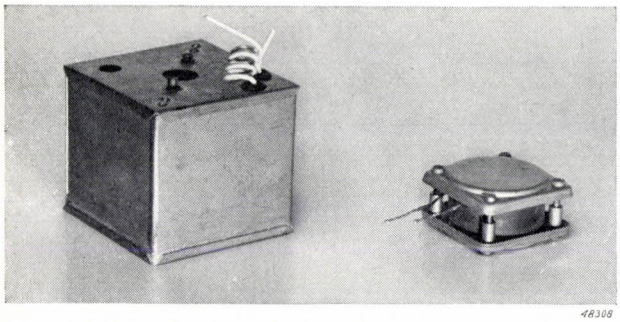


Fig. 4. Filter coil for carrier telephony, housed in a screening can, as formerly used (at left) and a similar coil made with Ferroxcube core. The volume of the old type of coil is 210 cm³ and the quality factor Q is 220 at 60 kc/s. The new coil requires no screening, the volume is only 44 cm³ and the Q-factor is 600 at 60 kc/s.

the other without. The new coil, although having only $^1/_5$th of the volume of the other, has a quality factor Q which is three times higher (44 cm³ volume, $Q = 600$ at 60 kc/s, as against 210 cm³, $Q = 220$ at 60 kc/s).

Loading coils

In general, the requirements to be met by loading coils are not essentially different from those imposed on filter coils. We have just explained in some detail the manner in which Ferroxcube may be made quite simply to satisfy all these requirements, namely by adopting the "pot" type of construction, and this arrangement is even more suitable for loading coils. In fact these coils are very much more sensitive than filter coils to mutual disturbance because of inadequate screening, but it has been proved that, by reason of the excellent screening provided by Ferroxcube, the resultant cross-talk between different circuits can be reduced to negligible proportions, even when two loading coils are placed one on the other.

In a certain class of laoding coils, namely those intended for phantom circuits, another demand is imposed on the performance of the coil. As will be seen from *fig. 5*, these coils must be composed of two parts, the inductances of which must be eaxctly equal.

The pot construction has been adopted also for this type of coil (see *fig. 6*); the necessary accurate matching of the inductance values is effected by means of external screws which provide a small amount of adjustment of the coil up and down, over the central core. This has the effect of varying the positions of the two parts of the coil with respect to the stray field in the vicinity of the air-gap and thus producing small variations in the inductance. It is accordingly a simple matter so to adjust the coils that their inductances will be the same.

Loading coils which are used in carrier telephony systems or in telephone lines which are also employed for telegraphic communications have to conform to an additional requirement: they must produce as little distortion as possible, so as to avoid intermodulation; in fact, very much less distortion than is permissible with filter coils.

For this reason it is usual to characterise ferromagnetic materials for such coils by a value indicating their distortional tendencies, and, among the metallic materials, the hysteresis constant C_h defined in I has been found quite suitable for this purpose. Let us now see why this quantity cannot be used in the case of Ferroxcube.

It is well known that distortion is caused by the non-linearity of the magnetisation curve; in other words it is dependent on the shape of this curve. Now, with the metallic ferromagnetic materials hitherto employed, the shape of the magnetisation curve is almost independent of frequency; for low

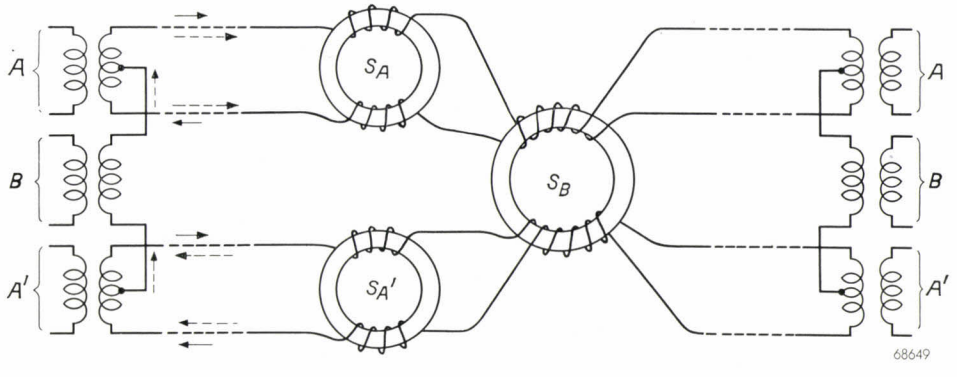


Fig. 5. Method of arranging three telephone circuits using two double pairs, viz. two side circuits A and A' (full-line arrows showing current flow) and a phantom circuit B (dotted arrows). S_A and $S_{A'}$ are the loading coils for the side circuits, S_B is that for the phantom circuit. The winding directions shown in the diagram ensure that each coil will possess inductance only with respect to its own circuit.

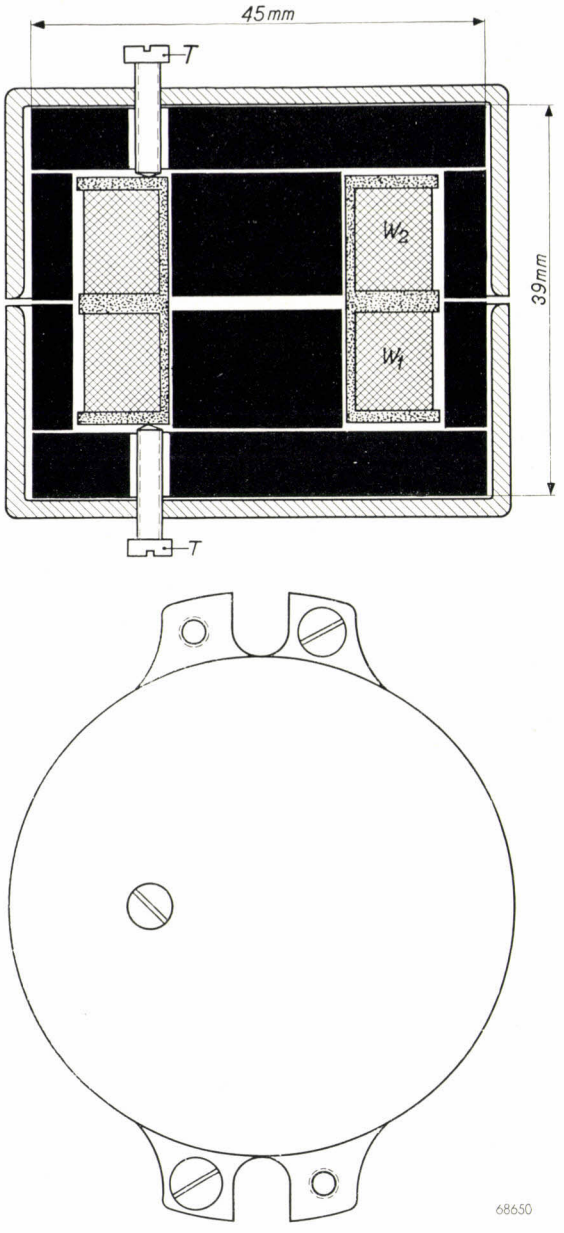


Fig. 6. Details of a loading coil for a phantom circuit. By means of two screws T, the coil, consisting of two sections W_1 and W_2, can be moved up and down slightly in order to equalize the inductance of the sections.

values of the magnetising force it may be described mathematically in accordance with Rayleigh's law as:

$$B = \mu_i \mu_0 H + \nu \mu_0 H^2,$$

where μ_i is the relative initial permeability and ν the "Rayleigh constant" for the particular kind of magnetic material used. In this instance there is a direct relationship between the distortion and the hysteresis constant C_h. When C_h has been ascertained from the losses, the degree of distortion is thus known.

On the other hand, as explained in detail in I, the shape of the magnetisation curve of most materials of the Ferroxcube group is in high degree dependent on frequency; hence there is now no object in classifying the losses due to hysteresis by means of a single C_h figure.

With Ferroxcube, then, a direct measurement of the distortion must be taken, viz. at the lowest frequency at which the coil in question is to be operated. For, as will be seen from figs 7 and 8 in I, the magnetisation curves of Ferroxcube materials below the ferromagnetic resonant frequency approximate more and more to a straight line as the frequency increases, which means that the distortion decreases accordingly.

Wide-band high-frequency transformers

An important field of application for Ferroxcube is to be found in the type of transformer often employed for wide-band amplifiers in carrier telephone systems. The fact that Ferroxcube III is also useful for this kind of transformer — which in some instances (coaxial systems) may have to operate at frequencies up to 4 Mc/s — is not very generally appreciated. This may be ascribed to the belief that ferromagnetic material for transformer cores should have a high permeability at all frequencies, and this certainly is not the case with Ferroxcube III at the frequencies concerned (see I, fig. 5).

In the following paragraphs we shall analyse the requirements usually imposed on this kind of transformer; it will be shown that Ferroxcube is well able to meet these requirements, despite the fact that the permeability drops sharply in the region above 1 Mc/s.

The more important requirements imposed on transformers suitable for carrier telephone amplifiers are as follows :

1) Flat response curve to within 0.1 dB throughout the whole, very wide, band of frequencies employed in the system.
2) The output impedance W, which is mainly determined by the output impedance of the transformer itself, should differ as little as possible from the characteristic impedance Z of the cable (see *fig.* 7). This ensures that reflections in the cable will be avoided or, more precisely, that the ratio in decibels of the reflected signal to the incoming signal, amounting to

$$20 \log \left| \frac{Z - W}{Z + W} \right|,$$

will remain sufficiently low.

Two reasons can be given why reflections should be kept within certain limits, viz:

a) Difficulties would otherwise be encountered in balancing the conductors in the cable for the purpose of preventing cross-talk. Such balancing would then be different for the direct and reflected signals.

b) Owing to irregularities in the construction of cables, the attenuation characteristic exhibits small discrepancies which are magnified by reflection and give rise to difficulty in the equalizing of the response characteristics.

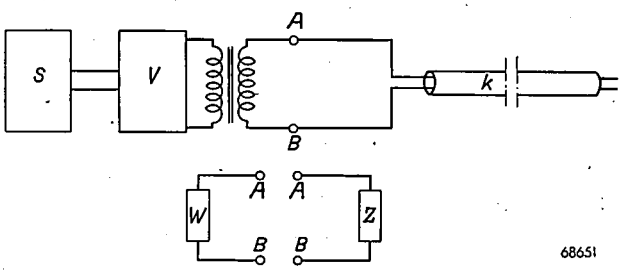


Fig. 7. Diagram showing the point in a carrier telephone system where the wide-band H.F. transformer is located. S = transmitting station. V= amplifier. The cable k from points A and B to the receiver has a characteristic impedance Z, whilst on the left of these points A and B there is the output impedance W.

3) The total loss in the transformer must be less than 2%.

The significance of these demands can best be appreciated by investigating the equivalent circuit of the transformer in *fig. 8*. For the sake of convenience let us assume that the transformer ratio is 1 : 1. In the diagram, E is an alternating voltage source, of internal resistance R_i. C represents the primary and secondary capacitances, these being made equal. R_l is the load resistance, L_k the leakage inductance, L_p the mutual inductance and R_p the parallel loss resistance (The dissipative resistance of the windings is disregarded.)

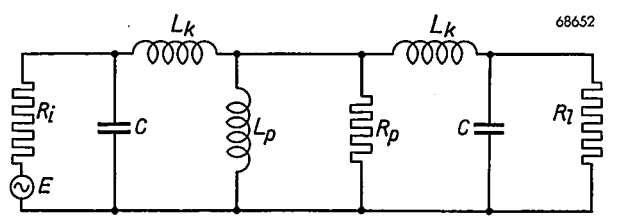


Fig. 8. Equivalent circuit of wide-band H.F. transformer. Transformer ratio 1:1. R_l = load resistance, equal to the characteristic impedance Z of the line. C = primary and secondary capacitances. L_k = leakage inductance. L_p = mutual inductance. R_p = parallel loss-resistance. E = equivalent source of alternating voltage of amplifier, with internal resistance R_i.

It will be noted that this circuit diagram of the transformer is the equivalent of a low-pass filter. Now, the requirements mentioned under (1) and (2) above mean that ωL_p must be high compared with R_l, whilst (3) is another way of saying that R_p should be high with respect to R_l. Calculation shows that the following should apply:

$$\left.\begin{aligned} R_l &\leqq {}^1/_5\,\omega L_p \\ R_l &\leqq {}^1/_{50}\,R_p \end{aligned}\right\} \quad \ldots\ldots\ (11)$$

L_p and R_p are both dependent on the (complex) parallel permeability μ_p of the material used for the transformer core, which relationship follows at once from the equations given in fig. 2, viz:

$$\omega L_p = \omega \mu_p' L_0 = 2\pi f \mu_p' L_0$$

and

$$R_p = \omega \mu_p'' L_0 = 2\pi f \mu_p'' L_0.$$

If the parallel permeability were not dependent on the frequency it would only be necessary to satisfy equation (11) for the lowest frequency concerned for it to be automatically satisfied at all higher frequencies as well. Actually, however, μ_p

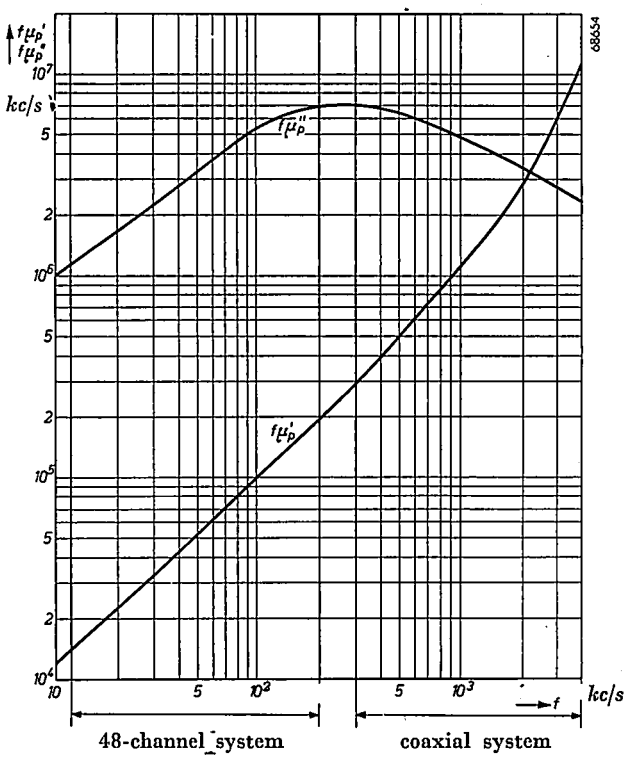
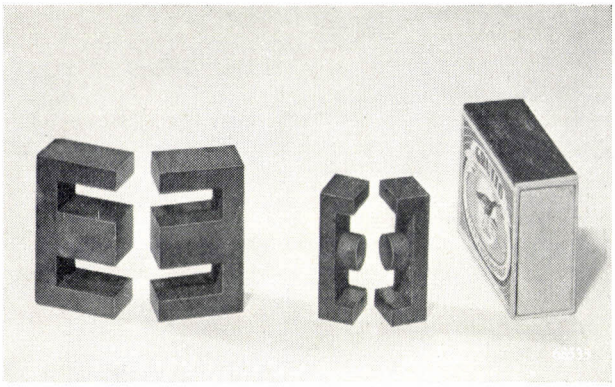


Fig. 9. The quantities $f\mu_p'$ and $f\mu_p''$ relating to the transformer core of Ferroxcube III^A illustrated in fig. 10*b* as functions of the frequency.

is a function of the frequency which, according to the definition given in page (302) is dependent on the behaviour of both μ' and μ'' as a function of frequency and which accordingly cannot at once be evaluated. In principle, therefore, it is possible that the right-hand terms of (11) assume smaller values at higher frequencies than at the lowest operating frequency (12 kc/s).

Fig. 9 depicts the curves for $f\mu_p'$ and $f\mu_p''$ as functions of the frequency for the transformer core



a *b*

Fig. 10. Ferroxcube cores for wide-band H.F. transformers. That on the left (*a*) is employed in a system comprising 48 channels (12-200 kc/s) and that on the right (*b*) for a coaxial system (0.3-4 Mc/s).

of Ferroxcube III[A] illustrated in *fig. 10b*. A glance at these curves will reveal the fact that $f\mu'_p$ increases throughout the whole band of frequencies in question (up to 4 Mc/s), that is, far beyond that frequency at which μ' drops sharply (0.5 Mc/s) [5]). It is true that $f\mu''_p$ fluctuates to a certain extent, but it does not drop below the value at 12 kc/s.

Summarising, it may be said that it is important that $f\mu'_p$ and $f\mu''_p$, i.e. $(f/\mu')(\mu'^2 + \mu''^2)$ and $(f/\mu'')(\mu'^2 + \mu''^2)$ should not fall below certain values at any frequency within the given band; hence the marked drop in the real component μ' of the permeability, above the ferromagnetic resonant frequency, is no deterrent to the use of Ferroxcube for the cores of the type of transformer under review.

Figure 10 illustrates two Ferroxcube cores of the kind used for these transformers, operating in a (48-channel system (12-200 kc/s) and in a coaxial system (0.3-4 Mc/s). The homogeneity of Ferroxcube enables the core to be assembled in two sections and the coil-winding process is thereby greatly simplified; by grinding the mating faces of the two sections of the core these can be made to fit together leaving the smallest possible air-gap, this being in effect so small that the effective permeability is only 10% less than the permeability of the material itself.

Uses of Ferroxcube in radio

Again owing to its high permeability and low losses, Ferroxcube has become invaluable in the manufacture of radio components. The material is used in the form of small rods and tubes and, especially where the working frequencies are not too high, it has many advantages over materials in powder form.

An example of the use of Ferroxcube in an I.F. transformer is illustrated in *fig. 11*. The inductance of the coils (*S*) is adjustable, this being effected by means of Ferroxcube rods which are moved one way or the other with respect to the coils with the aid of external screws. Each rod is held in position by a spring (*V*) and two short glass rods.

In order to reduce the losses in the aluminium can, each of the coils is flanked by three Ferroxcube rods (*F*), an arrangement known as "palisade" screening. Owing to the high permeability of the Ferroxcube, a large part of the field is concentrated in the rods, in consequence of which the magnetic flux passing through the aluminium, and accordingly the loss, is reduced.

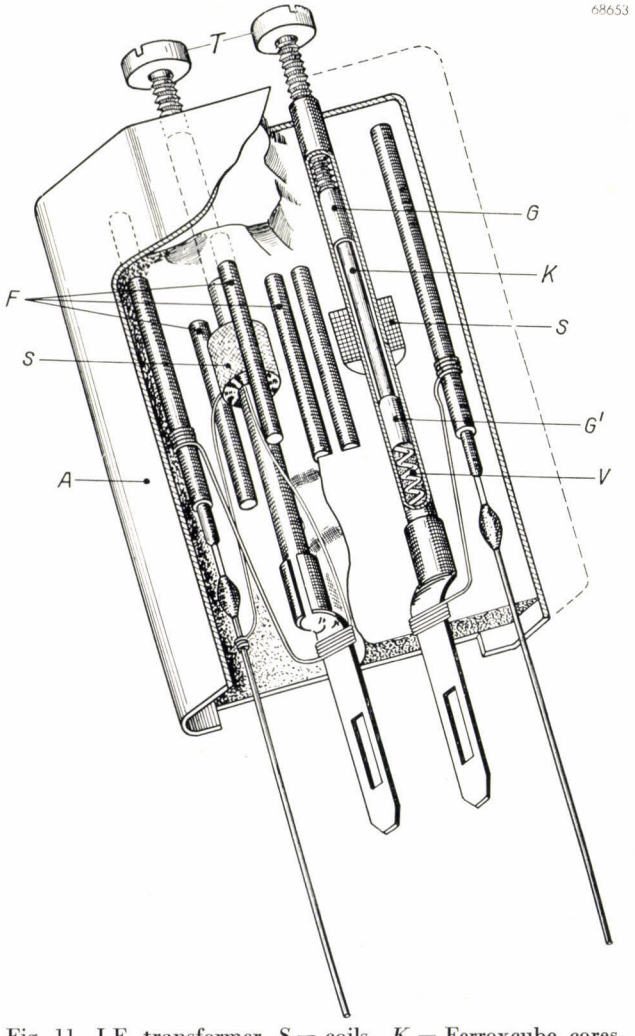


Fig. 11. I.F. transformer. *S* = coils. *K* = Ferroxcube cores, adjustable by means of screws *T* and glass rods *G*. The cores are held in position by glass rods *G'* and springs *V*. Ferroxcube rods *F* constitute a so-called "palisade" screen for reducing losses in the aluminium can *A*. The components on the extreme right and left-hand sides are "drawn" wire capacitors (see Philips Techn. Rev. **13**, 145-156, 1951 (No. 6).

[5]) The quantities μ'_p and μ''_p as read from these curves are not material values but, at any rate at frequencies above 1 Mc/s, refer only to the transformer cores under review by reason of the dimensional resonance effects mentioned in I.

The results obtained in this manner are best seen from the following details (see also *fig. 12*). The volume of a conventional I.F. transformer of cylindrical form and containing no Ferroxcube is 64 cc, the Q-factor being 206 at 452 kc/s, whereas details of I.F. transformers equipped with Ferroxcube are as follows:

type 5730: cylindrical type; volume 34.5 cm^3; $Q = 226$ at 452 kc/s.

type AP 1000: prismatic type; volume 8.75 cm^3; $Q = 174$ at 452 kc/s.

Since the losses in Ferroxcube cores are small even in the kilocycle band it has been found possible to develop the required E.H.T. by utilising the voltage peaks which occur across the circuit of an inductor with its own inherent capacitance, when a current flowing in it is suddenly interrupted.

The E.H.T. needed for direct-vision equipment lies somewhere between 8 and 14 kV, and as the power dissipated is small compared with that required for the deflection, the potential in question can be derived from the peaks occurring during the flyback in the horizontal deflection. The E.H.T. of the picture tube thus becomes, as it were, a by-product of the horizontal deflection voltage.

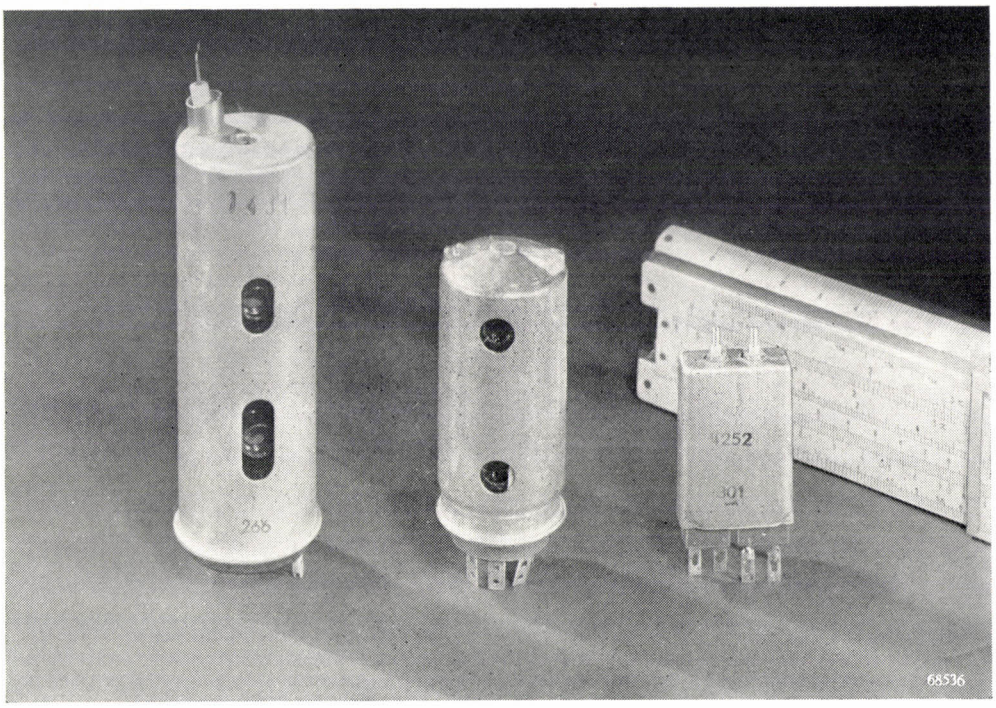


Fig. 12. I.F. transformers for radio receivers.
Left: An old type, without Ferroxcube. Volume 64 cm^3; $Q = 206$ at 452 kc/s.
Centre: type 5730. Volume 34.5 cm^3; $Q = 226$ at 452 kc/s.
Right: type AP 1000. Volume 8.75 cm; $Q = 174$ at 452 kc/s.

Uses in television

Another important use of Ferroxcube is to be found in the equipment which serves to generate the extra high tension required for television picture tubes. The most obvious method, of transforming and rectifying the mains voltage, is unsuitable seeing that transformers capable of producing such high voltages would be very expensive on account of the large quantity of very fine copper wire that would be needed for them. The rectifier, too, would be far too large and heavy. As the actual amount of power required is very small indeed, another solution was sought, as described in a previous issue of this Review in connection with projection television [6]).

The E.H.T. required for projection-television receivers is 25 kV and, although the power employed is certainly low, it is comparable with that needed for deflection purposes. In this case, therefore, separate equipment is used for the anode voltage, as described in the article referred to in reference [6]).

To this may be added that, here too, the special properties of Ferroxcube have enabled designs to be adopted which would otherwise involve much space, material and money.

Other uses of Ferroxcube

Advantage is taken of the small losses in Ferroxcube also at the higher frequencies in concentrating

[6]) H. J. Siezen and F. Kerkhof. Philips Techn. Rev. **10**, 157, 1948.

high-frequency alternating fields (of the order of 0.5 Mc/s) at given points, as for example, in H.F. heating. If cores of suitable shape are prepared, certain parts of the work can be heated as required, leaving contiguous areas at a relatively low temperature.

A number of other uses of this new material are based not on the smallness of the losses below the ferromagnetic resonant frequency, but actually on the fact that the losses are quite high beyond that frequency. In such applications the high volume resistivity of Ferroxcube is also utilised. These possibilities are barely out of the research stage, however, and are not yet being utilised on a large scale.

For example, Ferroxcube can be used for modulation purposes at very high frequencies [7]); when a piece of Ferroxcube is placed in a cavity resonator the Q-factor of the resonator is reduced considerably, owing to ferromagnetic resonance absorption. If a field is applied so as to polarise the ferromagnetic material, the Q-factor, in the case of specially prepared Ferroxcube, may be greatly increased again; apart from the grade of material and to a certain extent also the frequency, this increase in quality is dependent on the strength of the polarising field so that, if the strength be varied in accordance with a low-frequency signal, the variation in Q will result in amplitude modulation of the high-frequency field by that signal.

The high dielectric and magnetic losses in Ferroxcube at frequencies above the ferromagnetic resonance frequency can also be utilised to introduce a D.C. voltage or A.F. voltage into cavity resonators from which no leakage of H.F. field may occur (e.g. in a standard signal generator for frequencies of 1000 Mc/s or more). To this end a piece of coaxial cable filled with Ferroxcube can be used; in this case enamelled wire is used for the central conductor which is to carry the direct current.

As mentioned in the introduction, this review of the applications of Ferroxcube is by no means complete. In the first place, no mention has been made of a great many other instances in which Ferroxcube is at the moment being employed with every success. Secondly, it must be remembered that Ferroxcube is a comparatively new material whose possibilities have not yet by any means been investigated from every aspect.

Summary. The name Ferroxcube covers a range of ceramic ferromagnetic materials of high volume resistivity. It has proved to be particularly suitable for the cores of filters as used in carrier telephony; by adopting a so-called "pot" construction it becomes a simple matter to satisfy all the requirements imposed on such filters. Moreover, in contrast with the older powder cores, it is possible to make use of the fact that the magnetic losses in a magnetic circuit are decreased by incorporating an air-gap of suitable size in the circuit. This facility is also extended to the manufacture of loading coils, in which an original method has further been adopted to meet the conditions relating to symmetry usually imposed on such coils in phantom circuits. Notwithstanding the fact that the real component of the permeability of Ferroxcube drops sharply above a certain frequency, this material is also suitable for the cores of the wide-band high-frequency transformers met with in carrier telephony. It is merely necessary for the absolute value of the (complex) permeability at all operating frequencies to be sufficiently high.

In radio, Ferroxcube is used *inter alia* for I.F. transformers. On account of the relatively high permeability of the material (compared with metallic materials in the form of powder), a small number of Ferroxcube rods placed round a coil will provide adequate screening ("palisade" screening).

Ferroxcube is important in television by reason of the light and compact E.H.T. generator — now universally employed — which the introduction of this material has made possible.

In H.F. heating, Ferroxcube can be employed to advantage to produce locally concentrated alternating fields of high frequency.

Even the fact that the losses in Ferroxcube above certain frequencies may be very considerable will in certain cases be advantageous, viz. where a material is required that will ensure heavy losses at high frequencies, whilst having a high volume resistivity, for example, in the amplitude modulation of an H.F. field, or as isolating material for D.C. leading-in wires impermeable to U.H.F. energy.

Summarising, it may be said that Ferroxcube ensures a smaller and cheaper form of construction than is possible with metallic core material; hence designs are now possible which were formerly discarded as being too cumbersome.

[7]) H. G. Beljers, W. J. van de Lindt and J. J. Went, Journal of applied Physics (now printing).

FLYWHEEL SYNCHRONIZATION OF SAW-TOOTH GENERATORS IN TELEVISION RECEIVERS

by P. A. NEETESON. 621.397.335:621.396.615.17:621.397.62

As compared with an ordinary radio set, a television receiver contains a whole series of new components and circuits required for performing just as many new functions. Of particular interest are the saw-tooth generators, present in any television receiver, which so direct the beam of electrons from the cathode-ray tube that the spot on the screen neatly traces the succession of parallel lines of the picture. The synchronizing signals from the transmitter ensure that this tracing is synchronized with the scanning in the pick-up (camera) tube at the transmitter. It is of the utmost importance, however, that this synchronization should not be disturbed by interference, and it is with this end in view that new circuits have recently been developed.

The horizontal and vertical deflection of the electron beam in the cathode-ray tube, by means of which the picture frame is traced in a television receiver, is brought about by saw-tooth voltages or currents produced in the receiver by two generators [1] [2]). In recent years new types of saw-tooth generators have been designed which are often referred to under the collective name of "flywheel time bases". The word "flywheel" typifies excellently the new principle according to which these saw-tooth generators are synchronized with the synchronizing signals sent out by the transmitter together with the video signals.

In mechanical constructions the purpose of the flywheel is to steady the speed of a driving spindle to which it is fixed, or, in other words, to minimize the influence of variations in the driving torque upon the speed of the engine of which that spindle forms a part. The inertia of the flywheel provides for a certain measure of smoothing of disturbances of short duration. When the value of the driving torque changes permanently to a different level the number of revolutions will likewise assume a different final value after a delay which depends upon the momentum of the flywheel. In the case of flywheel synchronization of saw-tooth generators the influence of disturbances accompanying the synchronizing signals upon the frequency of the saw-tooth generator is similarly reduced or smoothed out by electrical means which will be discussed in this article.

These disturbances may arise from various local sources, such as electrical-ignition engines (e.g. in motor vehicles), commutator motors (e.g. in vacuum cleaners), diathermic apparatus, etc. Much depends upon the position of the receiving aerial with respect to the source of the interference. In the vicinity of a busy thoroughfare, for instance, much trouble may be experienced from motorcar interference. Still more serious, however, is the situation in fringe areas at such a distance from the transmitter that the signal is apt to be drowned in the interference noise. Such an interference is of a permanent nature and may make any reasonable television reception impossible, mainly on account of the synchronization being entirely upset by the continuous noise, with the result that there is no coherence in the picture. A striking improvement can be obtained by means of saw-tooth generators with flywheel synchronization. It is true that the effect of the noise upon the video signal is still seen in the form of a whirling of light and dark spots over the whole picture, like a snow shower, but there is nevertheless a considerable improvement in that the coherence of the picture is maintained. This is the main purpose and effect of flywheel synchronization.

The effect of interference on the synchronization of a saw-tooth generator

Most saw-tooth generators used in television engineering are based upon a very simple principle, namely the charging of a capacitor to a high potential via a resistance, the voltage across the capacitor increasing exponentially with time and asymptotically approaching the value of the voltage applied. Long before the voltage across the capacitor reaches this final value however, the capacitor is rapidly discharged by means of a switching system, which, for the sake of simplicity, will be termed the "switch". The condition that is to

[1]) See, e.g., Philips Techn. Rev. **1**, 20, 1936; **2**, 37, 1937; **4**, 346, 1939; **10**, 308 and 364, 1949.

[2]) From now on, for the sake of brevity, only saw-tooth voltages will be mentioned, though often saw-tooth currents may also be intended.

be imposed upon this switch is that its internal resistance should be as small as possible. After the capacitor has been discharged as far as possible the switch is opened again and the process of charging the capacitor begins anew. Thus there are two distinct phases in the operation of a saw-tooth generator, the first phase — the gradual charging of the capacitor — being called the stroke, or scan, while the second phase — the rapid discharge of the capacitor — is called the flyback. It is not intended to deal here with such details as the measures that can be taken for improving the linearity of the scan or for preventing undesired oscillations (ringing) during the flyback [3]). For the time being only the synchronization of the saw-tooth signals and the various means of achieving this will be considered.

The synchronizing signals are sent out by the transmitter in the form of pulses. *Fig. 1* shows schematically the shape of the video signal as now commonly obtaining, or to be adopted, in various countries; the form of the signal is explained in the subscript.

In the transmitter a synchronizing pulse is generated and transmitted every time a line has been scanned. The same applies when a frame has been scanned from top to bottom, though these frame synchronizing signals are of a more complicated nature, as explained in fig. 1. In the receiver the line and frame synchronizing pulses are separated from the video signal and from each other. The line synchronizing pulses (for the time being only these will be considered, and not the frame synchronizing pulses) have to ensure that the aforementioned switch in the line saw-tooth generator is closed just at the right moment, and this can be achieved in various ways.

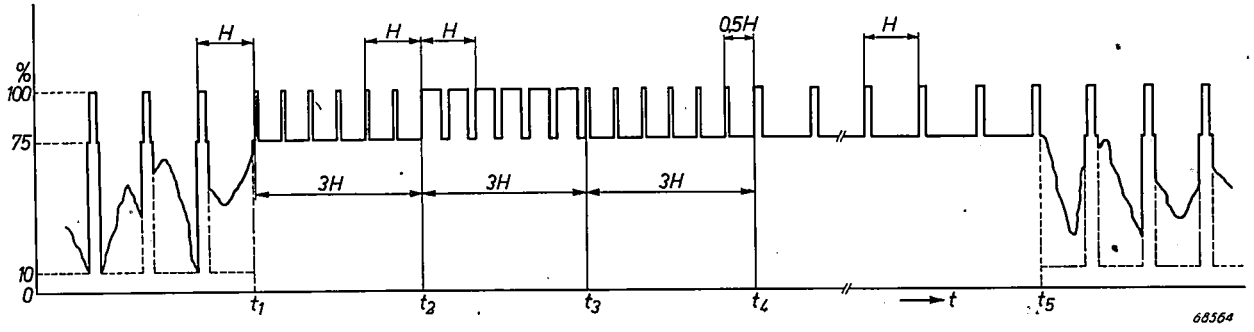


Fig. 1. Form of the video signal as a function of time t according to a proposal made by an international study committee of the Comité Consultatif International des Radiocommunications [4]). To the left of the instant t_1 and to the right of the instant t_5 are some lines of the picture signal alternated by a line synchronizing pulse. The modulation is negative: 75% of the maximum carrier amplitude corresponds to the black level, 10% corresponding to the level of the maximum brightness in the picture (peak white). While the synchronizing pulses are present the carrier is at full strength. The cycle H amounts to $1/(25 \times 625) = 1/15\,625$ sec.

The interval t_1-t_5, the duration of which should lie between 6% and 10% of $1/50$ sec, occurs only at the transition from the frame of the odd lines to that of the even lines and vice versa, thus 50 times per second. The frame synchronizing pulses occur between t_2 and t_3. The pulses between t_1 and t_2 and those between t_3 and t_4 are the equalizing pulses; these cancel out the difference which, without these pulses, would arise as a result of the instant t_1 occurring alternately at the end of the preceding line (as shown here) and in the middle of that line.

The horizontal dimensions are not drawn to scale (pulses too wide in proportion to H).

The first possibility is to employ a "switch" which, while remaining open in the absence of a synchronizing pulse, immediately reacts to a synchronizing pulse by closing and remaining closed as long as that pulse continues. Such a switch might, for instance, consist of a triode so adjusted that during normal operation the anode current is just cut off, but which becomes conductive and thus discharges the capacitor of the saw-tooth generator as soon as a positive voltage pulse is applied to its control grid. There are, however, two obvious objections to this type of switch. In the first place, if for some reason or other the synchronizing pulses should fail to come through then no saw-tooth signal will be supplied and consequently the electron beam will no longer sweep over the screen of the picture

[3]) See Philips Techn. Rev. **10**, 309, 1949.

[4]) Sub-group Gerber, Standards for the international 625-line black and white television system, C.C.I.R., Geneva 10th October 1950. In this committee there were delegates from Belgium, Denmark, Italy, the Netherlands, Sweden and Switzerland. For the somewhat different form of the video signal according to the British and the (then) French standards, see Philips Techn. Rev. **10**, 365 (fig. 1), 1949.

tube, the spot remaining stationary on one point and tending to cause serious damage to the luminescent screen. The second objection is that such a switch is highly sensitive to interference, any interfering pulse of sufficient amplitude being capable of closing the switch and thus making the tube conductive, so that the capacitor will be discharged.

Another method of synchronization, whereby the first of these drawbacks is eliminated and the other, the high sensitivity to interference, is considerably reduced, consists in the use of a self-blocking saw-tooth generator, which might also be called self-switching. Even when no synchronizing signal is present, such a circuit continues to supply a periodic saw-tooth signal, the frequency of which is mainly determined by an adjustable time constant, i.e. an RC product. As an example of such a self-blocking generator may be mentioned the circuit employing a triode with very heavy positive feedback, often denoted as a blocking oscillator, as represented in *fig. 2* [5]). At the beginning of an oscillation in this circuit the grid current suddenly assumes such a high value that a capacitor is rapidly charged in such a way that the control grid becomes highly negative with respect to the cathode. Thus both the grid and anode currents of the valve are limited to short pulses. The current flowing from the voltage source V_0 via the resistor R_1 to the capacitor C_1 causes the capacitor potential to change gradually in the inverse sense, as a result of which the grid becomes less negative with respect to the cathode, the voltage decreasing according to an exponential law with a time constant R_1C_1. As soon as the negative grid voltage becomes less than the cut-off voltage of the triode, anode current again begins to flow and, as a result of the heavy feedback between the anode and grid circuits, both the anode and grid currents suddenly increase. The grid current recharges the capacitor C_1 and the valve blocks itself, thus completing the cycle. *Fig. 3* represents the variations of the anode current i_a and of the grid voltage v_g as functions of time.

The frequency of this self-blocking saw-tooth generator may be synchronized, for instance, by applying the synchronizing pulses to the grid of the triode (see fig. 2). The amplitude of these pulses and at the same time that of any interfering signals is fixed by some limiting circuit to a value a. In fig. 3 a broken line has been drawn at the height a above the exponential part of the v_g curve. The instant at which this line crosses the level of the cut-off voltage V_c of the triode is denoted by t_2 and the instant at which this level is crossed by the v_g curve is denoted by t_3. From the course of the broken line it is seen that only between the instants t_2 and t_3 will a pulse of amplitude a advance

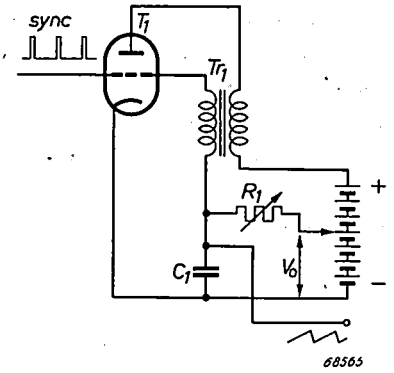


Fig. 2. Blocking oscillator for generating a saw-tooth voltage. T_1 triode given a blocking action by the tight coupling between the coils of the transformer Tr_1. The voltage across the capacitor C_1 has a saw-tooth shape. The frequency of the free-running oscillator can be adjusted by varying the resistance R_1 or the direct voltage V_0. Pulses (*sync*) are applied to the grid of T_1 for synchronization.

the flyback of the saw-tooth signal. It will be clear that synchronization will be possible as long as the cycle of the synchronizing signals lies between t_3—t_1 (i.e. the cycle of the free-running oscillator) and t_2—t_1. This illustrates the well-known fact that for proper synchronization the frequency of the free-running generator has to be adjusted to a value slightly below that of the synchronizing signals. Moreover, fig. 3 shows the reduced sensitivity to interfering signals, since these pulses (likewise limited to the amplitude a) can only initiate the flyback during the interval from t_2 to t_3, and not during the interval from t_1 to t_2.

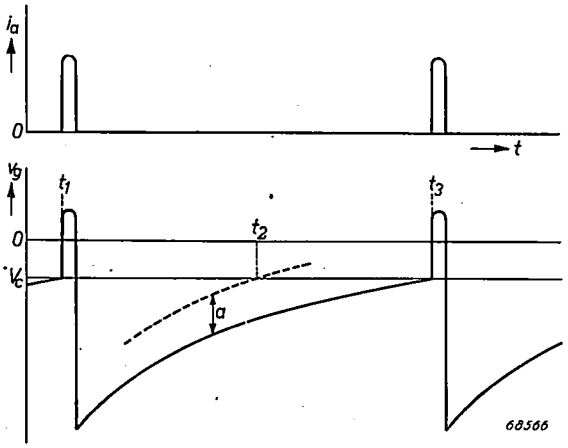


Fig. 3. Anode current i_a and grid voltage v_g in the circuit of fig. 2 as functions of the time t. The grid voltage below which the triode is cut off is V_c. A limiter limits the synchronizing pulses and the interference to the amplitude a.

[5]) See also Philips Techn. Rev. **10**, 317, 1949.

The two systems mentioned above have one feature in common, namely that each synchronizing pulse always determines the instant of the flyback (disregarding, of course, the case of an interfering signal initiating the flyback). Thus the generator never has a chance to produce an oscillation at its natural frequency. It is in this feature that the great difference lies between these two systems and the third method of synchronization to be dealt with here, viz. flywheel synchronization. With this method the saw-tooth generator is allowed to operate at its natural frequency and the influence of the synchronizing pulses is restricted to a readjustment of that frequency as soon as it begins to deviate from the synchronizing frequency. How this is brought about will be explained in the following pages.

The principle of flywheel synchronization

There is also a method intermediate between directly synchronized saw-tooth generators, where each individual synchronizing pulse determines the instant of the flyback, and that form of generator where its natural frequency is maintained at the correct value by comparison with the frequency of the synchronizing signal. What is referred to here is the possibility of initiating the flyback directly by pulses derived from the transmitted synchronizing pulses by first passing the latter through a flywheel circuit, thereby introducing a certain inertia effect. As electrical flywheel circuit use may be made of a resonant circuit. When a series of pulses are allowed to act upon the resonant circuit at a frequency equal to the natural frequency of that circuit the first pulse will cause only a slight oscillation in the circuit, while the following pulses reaching the circuit at exactly the right instants will boost the oscillations more and more until they approach a certain final amplitude. The interval of time τ that has to elapse before the final stationary state is almost reached [6]) is a measure for the inertia of the circuit. This time constant is determined by the circuit quality Q and the cycle T of the natural oscillations of the circuit, such that

$$\tau = QT/\pi .$$

The number of pulses occurring in the time τ is τ/T, thus equal to Q/π, so that it may be said that owing to its inertia the circuit does not undergo the full effect of the succession of pulses until a series of Q/π pulses has acted upon it; and the same applies if, for instance, the amplitude of the pulses were to be suddenly increased. When, therefore, only the periodical synchronizing pulses are acting upon the circuit it will oscillate sinusoidally at exactly the same frame frequency as that of the synchronizing signals, and these sinusoidal oscillations can easily be converted into pulses again, for instance by allowing only the positive peak of each sine to pass and thus obtaining a new series of pulses with the same frequency as the synchronizing pulses. This new series of pulses is then employed in the manner described above for synchronizing the saw-tooth generator.

When, therefore, the synchronizing pulses are accompanied by an interfering signal this will not have any noticeable effect upon the oscillations of the circuit unless it occurs at regular intervals corresponding approximately to the natural frequency of the circuit and, moreover, continues long enough to form a regular series of at least Q/π pulses. The greater the value of Q, the less chance there is of such an interference occurring. There are, however, various factors setting a limit to the extent to which Q can be increased. If, for instance, as is often the case, the frequency of the synchronizing signals in the transmitter is governed by the mains frequency, any fluctuations in the latter cause variations in the frequency of the synchronizing pulses, with the result that in a circuit with a very high Q value the phase variations will soon assume intolerable proportions [7]).

This form of synchronization, the medium between direct pulse synchronization and the more usual flywheel synchronization with automatic phase control now to be discussed, has purposely been dealt with in detail because in the case considered the "flywheel", i.e. the resonant circuit, is so clearly distinguishable.

Flywheel synchronization with automatic phase control

Before proceeding to discuss how automatic phase control can be used for synchronizing the saw-tooth voltage with the synchronizing signal received, the case will be considered where two sinusoidal voltages are mutually synchronized by means of this automatic phase control. Let the one voltage be that of an oscillator O_s with a given, fixed, frequency f_s, e.g. a crystal-controlled oscilla-

[6]) By this is to be understood, for instance, a value amounting to $1-1/e \approx 63\%$ of the asymptotically approached final value.

[7]) For further particulars see K. Schlesinger, Locked oscillator for television synchronization, Electronics 22-1, 112-117, 1949.

tor. In order to synchronize the second oscillator, O_1, with O_s the two voltages are applied to a phase discriminator *PD* (see *fig. 4*) producing a signal s_d. By means of a device F this signal is converted into a direct voltage which is a measure for the phase difference between the two alternating voltages. This direct voltage serves as control voltage, being fed back to the oscillator O_1 and influencing its frequency f_1.

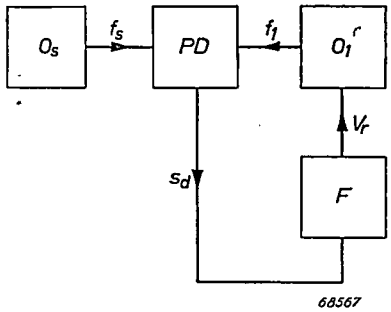


Fig. 4. O_1 is an oscillator (frequency f_1) that has to be synchronized with the standard oscillator O_s (with fixed frequency f_s). *PD* is a phase discriminator with output signal s_d, from which is derived via F a direct voltage V_r controlling the frequency of O_1.

Let us suppose that O_1 and O_s are indeed synchronized, so that $f_1 = f_s$. There is then a certain constant phase shift and the control voltage has likewise a certain constant value. If, for instance owing to a variation in temperature, the resonant frequency of the tuned circuit of O_1 drifts, then the phase shift will not retain its constant value but will increase, or decrease, as the case may be. It is then the function of the phase discriminator to vary the control voltage accordingly, and when this is done the frequency deviation $|f_1 - f_s|$ is reduced, even exactly to zero: the slightest frequency difference remaining would increase the phase shift and the control voltage variation would continue. The fact that the deviation is reduced to zero is due to the response of the discriminator to the phase shift, which is the time integral of the frequency deviation. That is why in regulating technique with controllers of this type one speaks of regulators with an integrating action [8]).

Let us now turn our attention to the synchronization by automatic phase control as applied in television. Here the analogue of the oscillator O_s with the given fixed frequency is formed by the series of synchronizing pulses with fixed recurrence frequency. The analogue of the oscillator O_1 to be synchronized is the saw-tooth generator; it is not, however, always the saw-tooth voltage itself that is applied to the phase discriminator, this sometimes being first converted into pulses (e.g. by means of a differentiating network), or else use is made of the pulses directly supplied by the saw-tooth generator (see fig. 3); in both cases the pulses have, of course, the same frequency as that of the saw-tooth voltage. In contrast to the case of fig. 4, where each of the oscillators produces a sinusoidal voltage, with the phase discriminator in a circuit for flywheel synchronization there are mostly two series of pulses, one having to be synchronized with the other by means of automatic phase control.

Fig. 5 represents such a circuit for this purpose. The valve T_1 with its accessories forms the saw-tooth oscillator (cf. fig. 2); it is schematically indicated how the saw-tooth voltage is converted into pulses. The valve T_2 acts as phase discriminator; the current flowing through this discriminator, i.e. the anode current of the valve T_2, passes through a resistor R_a, where it influences the value of the voltage V_r used for controlling the frequency of the saw-tooth oscillator.

We shall now examine the principal elements of this circuit more closely. As an example of a phase discriminator the valve EQ 80 is indicated; for a description of this "φ-detector" see an article published earlier in this journal [9]). The synchronizing pulses are applied to one of the control grids of this valve, while the series of pulses derived from

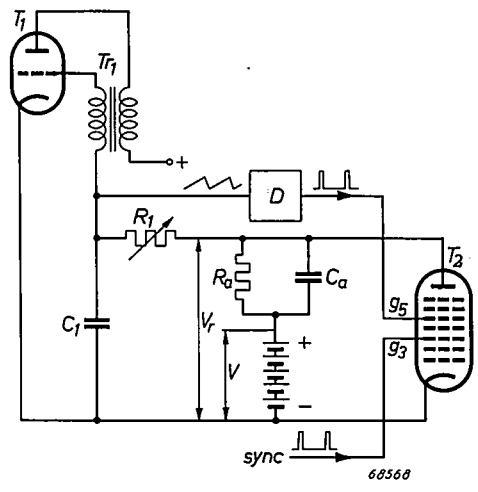


Fig. 5. Example of a circuit for flywheel synchronization with automatic phase control. T_1, Tr_1, C_1 and R_1 form a blocking oscillator (cf. fig. 2). D differentiating network forming pulses from the saw-tooth voltage. T_2 is an EQ 80 φ-detector with control grids g_3 and g_5. *Sync* synchronizing pulses. R_a load resistor and C_a smoothing capacitor, together acting as a flywheel. V supply voltage, V_r control voltage.

[8]) See, e.g., H. J. Roosdorp, On the regulation of industrial processes, Philips Techn. Rev. **12**, 221-227, 1951 (No. 8), in particular p. 226, and also Philips Techn. Rev. **12**, 263 (fig. 14), 1951 (No. 9).

[9]) Philips Techn. Rev. **11**, 1-11, 1949.

the saw-tooth voltage are applied to the second control grid. Now the EQ 80 has the property of allowing anode current to pass only in those intervals when both control grids are positive with respect to the cathode. When the pulses of the two series partly coincide, this gives rise to an anode current pulse, the width of which depends upon the extent to which the grid pulses coincide, thus upon their mutual phase difference. As is at once evident from *fig. 6*, the anode current pulse is widest when the phase difference is zero (fig. 6*a*). If the pulses in the two series are of the same width — as is presumed in fig. 6 for the sake of simplicity — then the anode current pulse becomes narrower as soon as those pulses begin to differ in phase (fig. 6*b*). When the phase shift becomes so great that the pulses of the two series no longer coincide (fig. 6*c*), there is no longer any anode current pulse. The variation in the width of the current pulses is shown as a function of the phase shift in *fig. 7a*. If the pulses of the two series are unequal in width, the variation in the width of the current pulses will be as represented in fig. 7*b*, as will be readily understood.

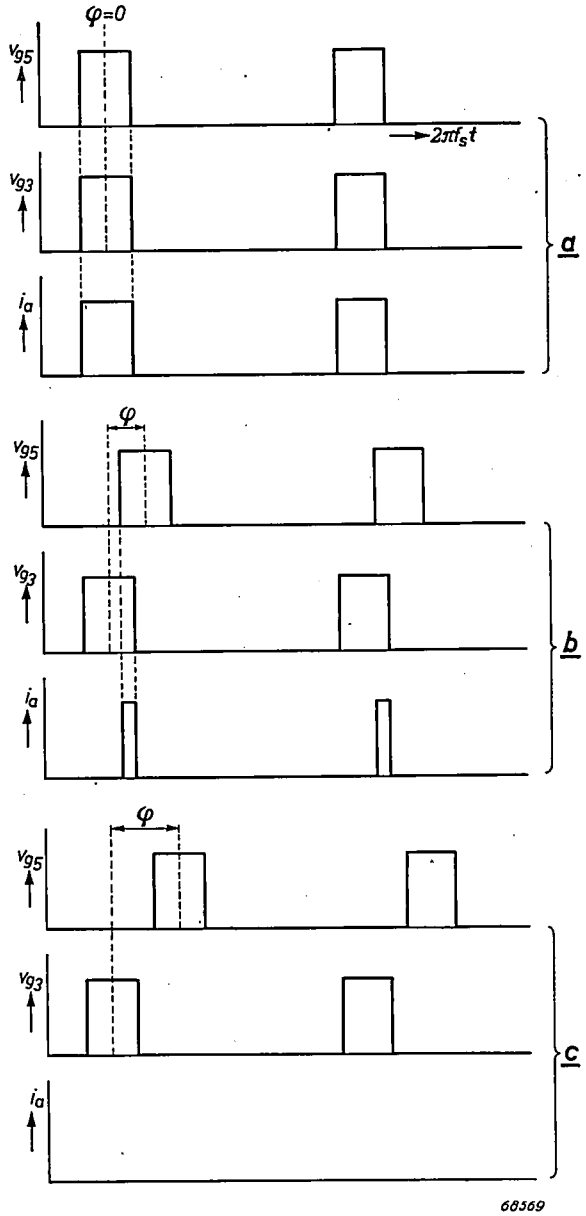


Fig. 6. Represented as functions of the phase $2\pi f_s t$: the pulses v_{g5} (derived from the saw-tooth voltage) on the grid g_5, the synchronizing pulses v_{g3} on the grid g_3, and the anode current pulses i_a, of the valve T_2 in fig. 5. Here the voltage pulses have been taken to be equal in width.
a) The two series of voltage pulses are in phase; the width of the anode current pulses is at its maximum.
b) Small phase difference φ between the voltage pulses. The current pulses have become narrower.
c) Large phase difference φ between the voltage pulses. There is no longer any anode current.

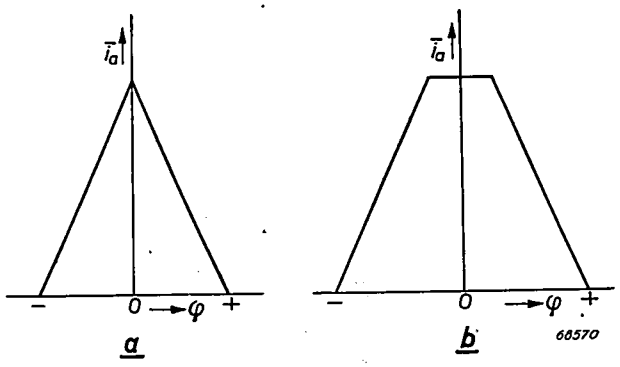


Fig. 7. Average anode current $\bar{i}_a$ in the valve T_2 of fig. 5 as a function of the phase difference φ between the two series of voltage pulses, for pulses of equal width (*a*) and of unequal width (*b*).

The anode current pulses form the output signal s_d of the phase discriminator (cf. fig. 4). They pass through the resistor R_a shunted by the capacitor C_a, so that across this combination (denoted in fig. 4 by F) a voltage is produced with an average value governed by the phase difference between the two series of pulses, in one of the two ways indicated in fig. 7. Here the sum of the voltage across the combination R_a-C_a and the fixed voltage V acts as the control voltage V_r serving to control the frequency of the saw-tooth oscillator.

Now how is such a control of the frequency of a saw-tooth generator by a control voltage V_r (see figs 2 and 5) brought about? To explain this, the variation of the exponential part of the grid voltage curve drawn in fig. 3 for one particular value of V_r (denoted in fig. 2 by V_0) is represented in *fig. 8* for different values of V_r, namely for the values V_1, V_2 and V_3. Since the flyback of the saw-tooth generator is initiated at the instant at which the grid voltage of the triode reaches the cut-off value, it is at once evident from fig. 8 that the cycle of the saw-tooth depends upon the value of the control voltage; the higher the latter, the shorter is the cycle.

It having now been explained how the synchronization is brought about by automatic phase control, the question may be asked where the "flywheel" is localised in such a circuit. In point of fact the flywheel action is due to the capacitor C_a connected in parallel to the resistor R_a, in that the voltage

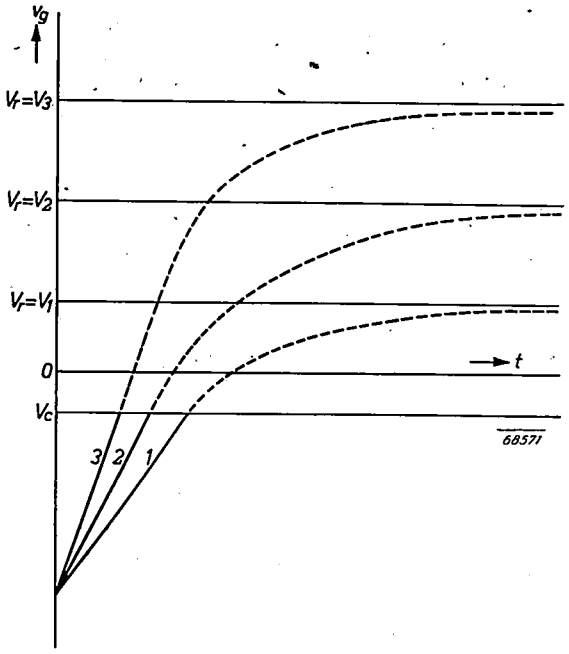


Fig. 8. Grid voltage v_g of the triode T_1 in fig. 5 as a function of t for different values (V_1, V_2, V_3) of the control voltage V_r. The higher the value of V_r the quicker v_g reaches the cut-off level V_c and thus the higher is the frequency.

appearing across this combination equals the product of R_a and the average value of a series of anode current pulses formed by the combined action of the two series of voltage pulses applied to the phase discriminator. When, in addition to the synchronizing pulses, interfering pulses also reach the control grid of the phase discriminator, they will have to occur in a rather large number and also at regular intervals before they can have any noticeable effect upon the average value of a series of anode current pulses. As in the case considered in the preceding section, here one may rightly speak of a flywheel effect and the corresponding reduced sensitivity for interfering signals.

In the circuit of fig. 5 the EQ 80 "φ-detector" may be replaced by an ordinary mixing valve (e.g. a hexode, or possibly a pentode) for deriving from two pulsatory voltages current pulses whose width is directly proportional to the phase difference between the voltage pulses. In fact, in the last section of this article a circuit will be described in which the EQ 80 is replaced by a pentode performing another function at the same time.

Another way of achieving the purpose in view is the following: the phase of the saw-tooth voltage is compared directly with that of the synchronizing pulses, without that voltage first being converted into pulses. When these voltages are combined in an ordinary mixing valve and any variation takes place in the phase difference, then the pulse slides, as it were, over the slope of the saw-tooth. This results in a series of anode current pulses varying in amplitude but not in width, contrary to the case in fig. 5. An example of this method has been described in an article in this journal dealing with synchronization in facsimile transmission [10]).

Instead of employing the saw-tooth voltage itself this can first be converted into a sinusoidal voltage with the aid of a resonant circuit, the pulses then running along the sloping sides of the sine curve. Or else, proceeding just the other way round, the sinusoidal voltage can be derived from the synchronizing signals and the pulses taken from the saw-tooth generator.

Owing to the flywheel action of the resonant circuit these methods are to be preferred to the direct use of the saw-tooth voltage. An example of the last-mentioned method will also be given later.

Further comments on the synchronization by automatic phase control

From the foregoing it has been seen that the frequency f_1 of the saw-tooth generator depends upon the control voltage V_r. To a certain approximation this may be said to be a linear dependency. But this frequency also depends upon the value of the time constant R_1C_1, where C_1 is the value of the capacitor in the grid circuit of the blocking oscillator (fig. 2) and R_1 is the value of the resistor across which C_1 is discharged. Use is made of this latter dependency in a television receiver by making the resistor R_1 variable, so that the saw-tooth frequency can be adjusted by hand to bring it close to the frequency of the synchronizing signals. Via the control voltage the automatic phase control then ensures that the saw-tooth frequency is maintained at its correct value. It may be said that the saw-tooth frequency is approximately inversely proportional to the product R_1C_1.

Taking these two dependencies together, the frequency of the saw-tooth generator may therefore be formulated as:

$$f_1 = \frac{a}{R_1C_1}\left(\frac{V_r}{V} + b\right); \quad \ldots\ldots \quad (1)$$

[10]) Philips Techn. Rev. **10**, 329 et seq., 1949.

where V is the fixed part of the control voltage (see fig. 5), while a and b are constant quantities depending upon the characteristics of the blocking triode.

In addition to this eq. (1), for a full description of the synchronization by automatic phase control it is necessary to indicate the relation between the control voltage V_r and the phase difference φ between the two series of pulses. According to the foregoing, the control voltage is given by $V_r = V - \bar{i}_a R_a$, where $\bar{i}_a$ is the mean value of the anode current pulses. This equation can be written in the form:

$$\frac{V_r}{V} = 1 - \frac{\bar{i}_a R_a}{V} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now $\bar{i}_a$ varies as a function of the phase difference φ, for instance, in the manner represented in fig. 7*b*. From eq. (2) and fig. 7*b* the relation between V_r/V and φ can be represented graphically as shown on the right in *fig. 9*. Farther to the left in the same

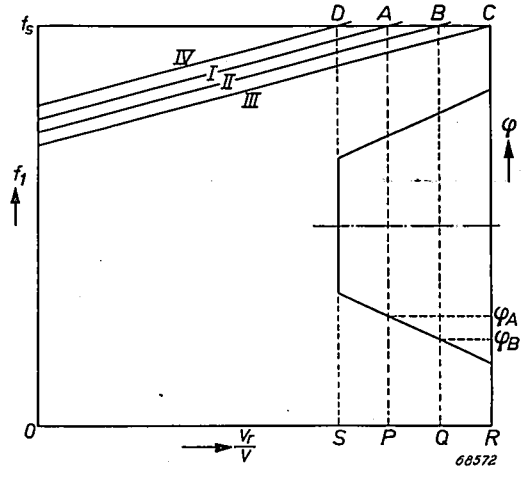


Fig. 9. The straight lines *I*...*IV* represent graphically eq. (1) for different values of the time constant R_1C_1: the frequency f_1 as a function of V_r/V. The trapezium on the right represents V_r/V as a function of the phase difference φ (cf. fig. 7*b*). f_s is the frequency of the synchronizing signals.

diagram are some straight lines representing the relation of eq. (1) for different values of the variable resistor R_1. Supposing that R_1 has such a value that eq. (1) is represented by the line *I*, then the working point of the saw-tooth generator adjusted to the fixed frequency f_s of the synchronizing signals will be the point *A*, i.e. the control voltage (in relation to V) has the value OP, while the phase shift is φ_A.

When the product R_1C_1 changes, either through R_1 being adjusted by hand or owing to a temperature variation changing R_1 and/or C_1, then eq. (1) would be represented, for instance, by the line *II*. Then the saw-tooth generator operates in the point *B*, the control voltage then having the value OQ and the phase shift amounting to φ_B. From this it is clearly seen how the sawtooth frequency is kept equal to the synchronizing frequency f_s at the cost of a greater phase shift between the two series of pulses. This change in the phase difference manifests itself on the screen of the picture tube as a displacement of the picture. This is a typical feature of a saw-tooth generator with automatic phase control: when the frequency of the line saw-tooth generator in a receiver employing phase control is adjusted, the picture is seen to move across the screen in the horizontal direction, often within wide limits. If, however, the knob is turned too far (one way or the other) synchronization is lost and the picture becomes incoherent; the fact that this should be so can be seen from fig. 9: if the value of R_1 (eq. (1)) is changed so far as to follow the lines *III* or *IV* then the limits of the "retaining zone" of the automatic phase control are reached, since control voltages smaller than OS or greater than OR are impossible.

Some practical circuits for flywheel synchronization with automatic phase control

For the general arrangement of a synchronizing circuit with automatic phase control, reference may be made to fig. 4, where in the place of O_s we now have the source of the synchronizing signals and for O_1 the saw-tooth generator. The working of the circuit may be summarized as follows:

The output signal s_d of the phase discriminator *PD* is dependent in width, or in amplitude, or in both respects — in general thus in surface area (time integral) — upon the phase difference φ between the saw-tooth voltage and the synchronizing pulses. The filter F produces from s_d a direct (control) voltage V_r influencing the frequency f_1 of the saw-tooth generator. The latter is so adjusted that in the absence of synchronizing signals f_1 is lower than the recurrence frequency f_s of the synchronizing pulses. If, in the presence of synchronizing pulses, for some reason or other f_1 and f_s should no longer be equal, e.g. $f_1 < f_s$, then φ changes, V_r increases and f_1 is raised. Since the slightest difference between f_1 and f_s causes the phase difference φ to increase infinitely with time, f_1 is brought exactly to the value of f_s (regulating system with integral action).

In principle this method could be employed equally well for synchronizing the frame saw-tooth generator as the line saw-tooth generator. It is, however,

applied less for the former purpose than for the latter on account of the low value of the frame frequency (50 frames per second), which would require impracticably large capacitances. In the following, therefore, only the applications of this method to the line saw-tooth generator will be considered.

For the phase discriminator one of the known circuits can be used, e.g. with two or with four diodes, details of which may be found in the literature on the subject [11]).

screen grid g_2, forms, together with a triode T_4, part of a multivibrator so adjusted that at the screen grid a saw-tooth voltage is obtained with a frequency that is slightly too low when synchronizing pulses are absent. The frequency can be adjusted with the aid of the variable resistor R_1.

During each flyback of the saw-tooth signal the control grid of T_3 is momentarily at about cathode potential and a current pulse then reaches g_2. It depends upon the potential of the third grid,

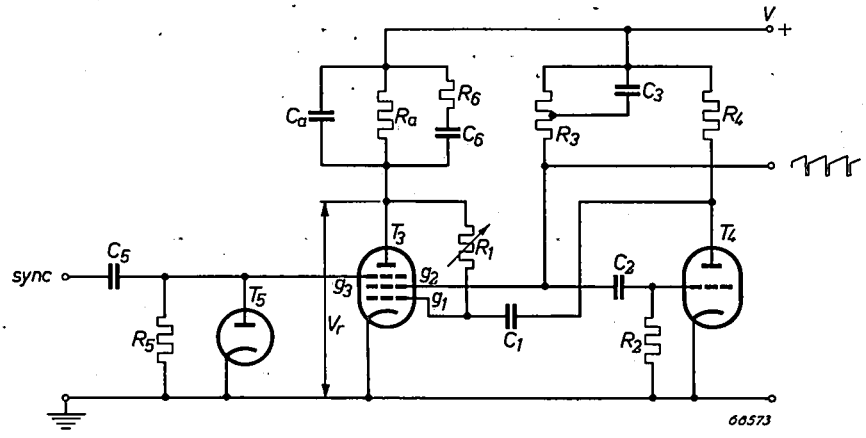


Fig. 10. Circuit for flywheel synchronization of the line saw-tooth generator with automatic phase control. The saw-tooth generator is a multivibrator consisting of the "triode" formed by the cathode and the grids g_1 and g_2 of the pentode T_3, the triode T_4 and the capacitors C_1 and C_2, the grid resistors R_1 and R_2 and the anode resistors R_3 and R_4. To the grid g_3 are applied the synchronizing pulses plus an automatically produced negative grid bias (obtained with the aid of the grid capacitor C_5, the resistor R_5 and the diode T_5). The pentode serves not only as saw-tooth generator but also as phase discriminator. The correcting network R_6—C_6 gives the control an aperiodic character. Other symbols have the same meaning as in fig. 5.

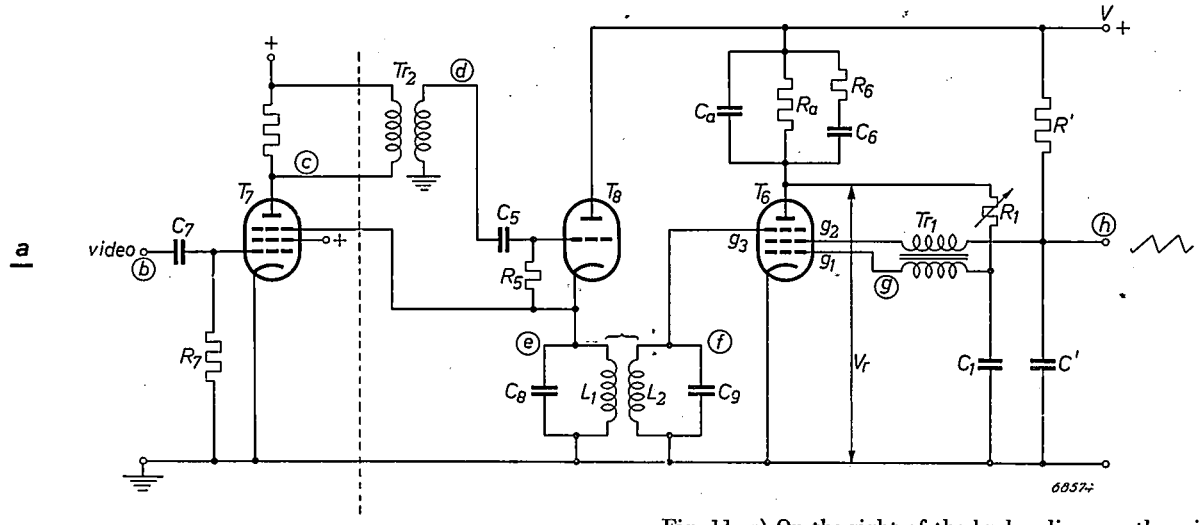
Two circuits will now be described which have been greatly simplified by *combining in one normal amplifying valve the functions of phase discriminator and saw-tooth voltage generator*. This valve must be at least a pentode. The cathode, the control grid and the screen grid act together as a triode forming part of the saw-tooth generator. The synchronizing pulses (or a signal derived therefrom) are applied to the third grid, while the anode circuit supplies the control voltage for restoring the saw-tooth frequency to its right value in the event of any deviation.

Fig. 10 shows how this idea can be worked out. T_3 is the pentode referred to. The "triode", constituted by the cathode, the control grid g_1 and the g_3, whether anode current also flows or not: if that potential is zero an anode current pulse will be produced, but if g_3 is at a sufficiently negative potential there will not be any anode current.

In the circuit of fig. 10 the synchronizing pulses are applied to g_3 via a coupling capacitor with a grid leak shunted by a diode. Consequently the third grid adjusts itself to a negative potential in the order of the amplitude of the pulses, assuming the potential zero only while a pulse is present. Further the operation is the same as that of the EQ 80 valve in fig. 5: anode current flows through the pentode only when both g_1 and g_3 allow it (that is to say, in this case, when g_1 and g_3 are both at about cathode potential). The width of the anode current pulses thus depends upon the phase difference φ between the saw-tooth voltage and the synchronizing pulses. Since the anode current has to pass a resistor R_a (shunted by a capacitor C_a), the phase difference φ influences the anode voltage V_r acting as control voltage.

[11]) See, e.g., K. R. Wendt and G. L. Fredendall, Automatic frequency and phase control of synchronization in television receivers, Proc. Inst. Rad. Engrs **31**, 7-15, 1943; E. L. Clark, Automatic frequency phase control of television sweep circuits, Proc. Inst. Rad. Engrs **37**, 497-500, 1949; Milton S. Kiver, Modern television receivers, Radio and Television News **43**, 50-52 and 110-111, March 1950.

Actually it is not the synchronizing signals themselves that are fed to the input terminals on the left in fig. 10, but the pulses obtained by differentiating those signals. Otherwise the series of wide frame-synchronizing pulses occurring 50 times per second (see fig. 1) would disturb the working of the circuit. The differentiation causes the leading edge of each square-wave pulse to yield a positive pulse, while the trailing edge yields a negative pulse. As can readily be seen from fig. 1, the positive pulses resulting from the differentiation are equidistant — distance $H = {}^1/_{15\,625}$ sec = 1 line cycle = $= 1/f_s$ — except while the equalizing pulses are present, when the distance becomes $\frac{1}{2}H$; the positive pulses are then, however, acting alternately, because the negative voltage on the control grid of the pentode prevents any anode current flowing. Naturally the (non-equidistant) negative pulses obtained from the differentiation are also inactive.



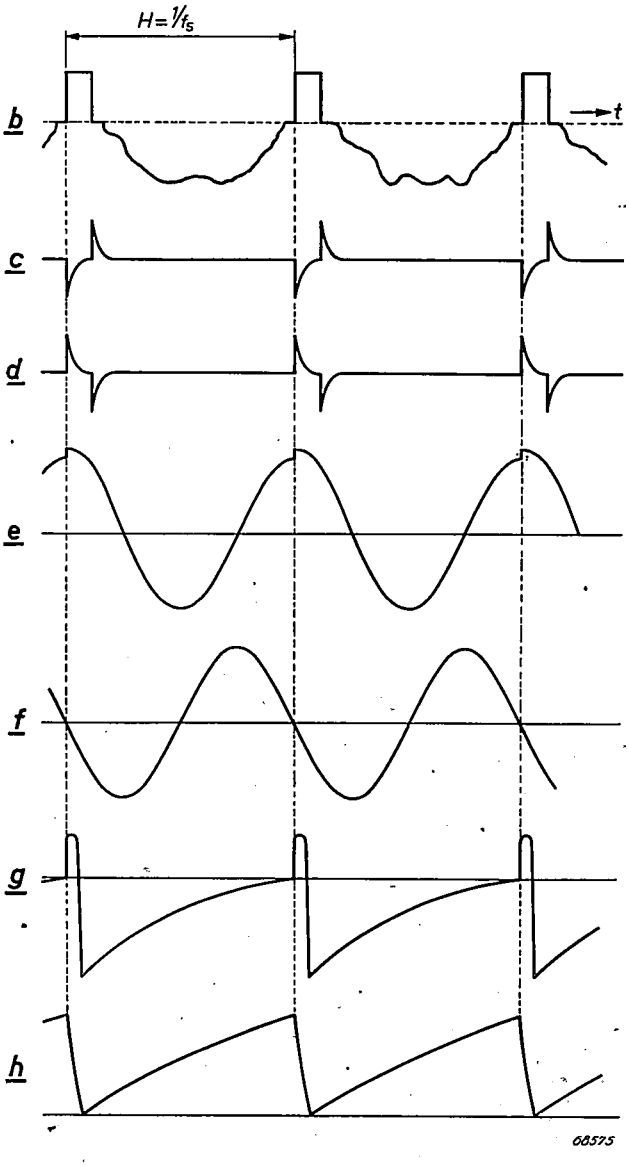


Fig. 11. *a*) On the right of the broken line: another circuit for flywheel synchronization of the line saw-tooth generator with automatic phase control. On the left: pentode T_7, with grid capacitor C_7 and resistor R_7, cutting off the picture signals from the video signal, so that only the synchronizing pulses remain.

Tr_2 differentiating transformer. T_8 triode, whose cathode current pulses excite the resonant circuit L_1-C_8. L_2-C_9 circuit critically coupled to L_1-C_8. T_6 pentode acting both as blocking oscillator (grids g_1 and g_2, transformer Tr_1, combinations R_1-C_1 and R'-C') and as phase discriminator (grids g_1 and g_3 and the combination R_a-C_a-R_6-C_6; cf. fig. 10).

The figures (*b*) ... (*h*) represent as functions of t the voltages at the points of the circuit denoted in (*a*) by the corresponding letters.

Attention is also to be drawn to an advantage of the automatic negative voltage produced at the third grid by the synchronizing signals themselves with the aid of the coupling capacitor, the resistor and the diode. When the synchronizing signals fail (e.g. owing to a defect in the transmitter) the potential of g_3 becomes zero and the anode current pulses which are then produced are only a little wider than the pulses normally produced during synchronization. Therefore the control voltage V_r and thus the frequency f_1 of the multivibrator then running free do not differ much from their normal values, and when the synchronizing signals come through again the synchronism is automatically restored. The situation would be different if the negative grid voltage were not obtained automatically but derived, for instance, from a battery: if in that case the synchronizing signals drop out then the negative grid voltage is maintained, the anode current becomes zero, V_r rises to the value V of the supply voltage, and f_1 rises to a value so much higher than the value f_s that, upon the return of the synchronizing signals, the resistor R_1 has to be readjusted in order to restore the synchronism.

Another small detail to be considered in the circuit of fig. 10 is the series connection of the resistor R_6 and the capacitor C_6 shunted by the combination R_a-C_a. If R_6 and C_6 were omitted then after a disturbance the control would die away according to a damped oscillation, manifesting itself in waving sides of the picture. With a suitable choice of R_6 and C_6 the control can be given a damped aperiodic character, the sides of the picture then remaining practically straight.

Another form of circuit is represented in *fig. 11a*, where the functions of phase discrimination and saw-tooth generating are combined in one pentode. Here again the cathode, the control grid and the screen grid of this pentode (T_6) make a triode forming part of the saw-tooth generator (in this case a self-blocking oscillator; cf. fig. 2). The difference between this and the other circuit is that g_3 in this case receives a sinusoidal voltage derived from the synchronizing signals (in a manner to be shown later) and having the frequency f_s. The amplitude is so chosen that the cut-off voltage of the third grid is not reached, so that this alternating voltage never entirely prevents anode current from flowing. Here the anode current pulses are constant in width (equal to the time during which the control-grid voltage is approximately zero) but variable in amplitude, according to the instantaneous value of the alternating voltage on the third grid. Here we have, therefore, an example of the method of phase control, already mentioned, where the amplitude of the signal supplied by the discriminator varies with the phase difference between a sinusoidal voltage of the right frequency and the saw-tooth voltage to be synchronized.

A filter R_a-C_a (fig. 11*a*) ensures, in the known manner, that a direct voltage V_r, serving as control voltage for the saw-tooth generator, corresponds to the "surface area" of the anode current pulses.

It remains to be explained how the sinusoidal voltage for the third grid of T_6 is derived from the synchronizing signals.

T_7 in fig. 11*a* is the valve, present in any television receiver, serving to separate the square-wave synchronizing pulses from the video signal (fig. 11*b*); this valve does not, therefore, belong to the special circuit with which we are dealing here. This separation is brought about simply by cutting off the picture signals. Thus the anode current of T_7 has the form of square-wave pulses. Owing to the differentiating action of the inductance of the primary of the transformer Tr_2, the anode voltage of T_7 assumes the form of pulses as represented in fig. 11*c*. The secondary of Tr_2 supplies identical pulses of opposed polarity (fig. 11*d*) to the grid of the triode T_8. The grid and cathode of T_8 form a diode, which, together with the grid capacitor and the grid leak (C_5-R_5), ensures that the valve conducts only at the positive peaks of the grid voltage. These current pulses excite a resonant circuit L_1-C_8, tuned to the frequency f_s of the synchronizing signals, as is also the case with a second circuit L_2-C_9. The voltage across the latter circuit is applied to the third grid of T_6. These two circuits are critically coupled, with the result that the sinusoidal voltage across L_2-C_9 (fig. 11*f*) is in quadrature with the voltage across L_1-C_8 (fig. 11*e*).

When the saw-tooth generator is synchronized, the discharge of the capacitor C_1 (thus the flyback of the saw-tooth, see fig. 11*g*) begins at the instant that the alternating voltage on g_3 passes through zero, thus at the instant that the voltage across L_1-C_8 is at its maximum, i.e. on the leading edge of the square-wave synchronizing pulses. Any deviation in the synchronism causes the pulse on g_1 to slide, as it were, along the flank of the sine curve in fig. 11*f*, thereby initiating the control process.

A refinement commonly employed, which, for the sake of simplicity, was not mentioned when dealing with the blocking oscillator of fig. 2, consists in the introduction of a resistor and a capacitor as represented in fig. 11*a* by R' and C'. The saw-tooth obtained across C' (fig. 11*h*) is more linear than that across C_1.

With this circuit, too, synchronism is restored without readjustment after an interruption of the synchronizing signals.

The combination R_6-C_6 performs the same function as in the circuit of fig. 10.

As regards the connection of the cathode of T_8 to the third grid of T_7 (fig. 11*a*), this prevents any trouble being experienced from the frequency $2f_s$ of the equalizing pulses of the frame-synchronizing signals (see fig. 1). The negative peaks of the alternating voltage across the circuit L_1-C_8 (fig. 11*e*) make the third grid so highly negative that the extra pulses cannot cause any anode current to flow in T_7 and thus are inactive as far as the line synchronization is concerned.

Summary. In a television receiver the saw-tooth signals serving for the horizontal and vertical deflections of the electron beam in the picture tube are synchronized by means of the synchronizing signals contained in the video signal received from the transmitter. This synchronization may be seriously disturbed by interference. After a survey of the sensitivity to interference of different systems of synchronization, a detailed description is given of a method of line synchronization, known as flywheel synchronization, by means of which the influence of interference can be minimized, a system which is becoming more and more popular. The principles of automatic phase control, which are particularly important with this system, are dealt with at length. Finally two circuits are described which are simpler than those hitherto known, by reason of the fact that a normal pentode serving for generating the saw-tooth voltage functions at the same time as a phase discriminator.

HIGH-PRESSURE MERCURY VAPOUR LAMPS FOR PHOTO-COPYING PROCESSES

by W. ELENBAAS and K. R. LABBERTÉ. 621.327.312:778.1

In view of the importance which photo-copying processes have assumed in the reproduction of technical documents, the efficiency of the necessary light-source has become a significant factor. The high-pressure mercury vapour lamp has proved to be eminently suitable for such purposes. The efficiency of this kind of lamp is greatly improved by making the envelope from hard ultra-violet transmitting glass and can be increased still further by employing quartz glass, seeing that the power rating of the lamp can then be increased and that the transmissibility of this kind of glass to the radiations in question is higher.

Lamps for photo-copying purposes

Photo-copying is in effect a process for the duplication of documents based on a photo-chemical principle. The textual matter or drawing to be reproduced is applied in opaque ink to a transparent paper and this "transparency" is placed in contact with the "dye-line" paper which has a specially prepared surface. These are then exposed to a powerful source of light and, the colour and intensity of the light being appropriate, a "print" is produced in a very short space of time. This print has to be subsequently developed. The advantages of this system of contact-printing are obvious in that an unlimited number of prints of the same size as the transparency can be made very quickly. Photo-copying machines have been developed on the endless-belt principle which greatly facilitates the production of a large number of prints in quick succession.

The conventional grades of dye-line paper are mainly sensitive to violet and ultra-violet radiations (the great advantage of this is that the undeveloped paper can be more or less safely exposed to daylight). *Fig. 1* shows the spectral sensitivity curve of a popular grade of paper; it will be seen that maximum sensitivity occurs at a wavelength of about 4000 Å.

The main requirement to be imposed on an efficient photo-copying lamp is that a large part of the radiated energy shall lie within the range of wavelengths to which the paper is sensitive. Now, as pointed out in an earlier number of the Review [1]), the lamps most widely employed for this purpose are the carbon arc and high-pressure mercury vapour lamps. Of these the first-mentioned, which were formerly used very extensively, suffer from several drawbacks, such as the maintenance which such lamps always demand, the wide variations in luminous flux which are liable to occur and the high current needed to operate them. Another no less important objection to the arc lamp is the fact that

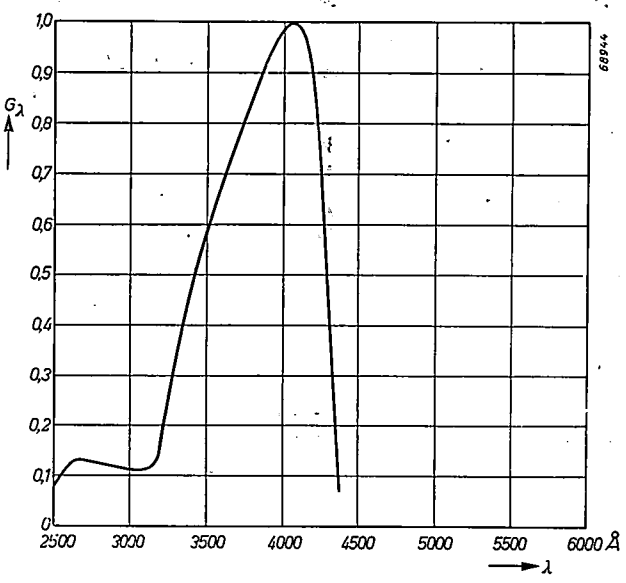


Fig. 1. Spectral sensitivity of a popular grade of dye-line paper (on relative scale). (Vide R. J. H. Alink, Ned T. Natuurkunde **9**, 135-147, 1942 (Fig. 2)).

the light-source is practically a point-source, which means that the illumination varies considerably between the centre and the edges of the transparency. For the whole transparency to be exposed uniformly, either the lamp must be placed at some distance from it, necessitating long exposure times and resulting in uneconomical consumption of power, or the lamp (or paper) must be shifted continuously.

The distribution of spectral energy of the high-pressure mercury vapour lamp is illustrated in *fig. 2*, from which it will be seen that, since a number

[1]) A. A. Padmos and J. Voogd, Mercury lamps for use in making heliographic prints, Philips Techn. Rev. **6**, 250-252, 1941.

of strong spectral lines occurs in the region of $\lambda = 4000$ Å, the mercury vapour lamp must be very suitable for this class of work. Furthermore, the discharge lamp has none of the disadvantages of the carbon arc already mentioned, whilst the lamp itself is of elongated form and the length of the actual discharge can be easily matched with the width of the paper. By guiding the paper cylindrically over the lamp, uniform exposure is then ensured without the necessity of moving the paper laterally.

If the high-pressure mercury vapour lamp is to operate efficiently, the glass of which the bulb is made must be capable of transmitting practically without loss those rays to which the paper is the most sensitive.

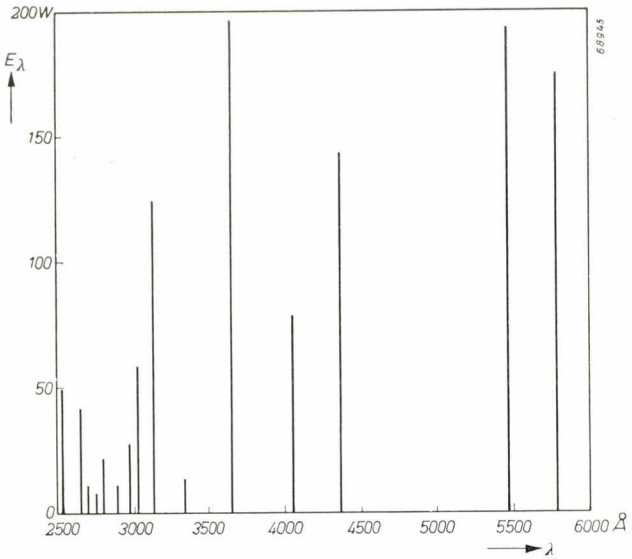


Fig. 2. Spectrum of the high-pressure mercury vapour lamp (HOK 3000 W). The lengths of the lines indicate the spectral energy emitted, expressed in watts. There is little or no continuum.

Lamps made of hard, ultra-violet transmitting glass satisfy the requirements imposed in many respects but, besides these, Philips manufacture tubular lamps made of fused silica. This kind of glass is certainly more expensive, but it has the advantage that it admits of higher loading of the lamp per unit length, and this, together with the higher transmissivity for the appropriate radiations, promotes increased printing speeds. In some instances, then, the more costly lamp of fused silica will be an economical proposition.

Some details of the photo-copying lamps marketed by Philips are given in *table I*. The HOG lamps are made with a hard-glass tube, whereas the HOK lamps have a fused silica tube. The types are distinguished according to the power consumption, corresponding to different luminous lengths of the tube. (The best results are obtained when the light-emitting length of the lamp is equal to, or a little greater than, the width of the paper.) By way of illustration *fig. 3* depicts part of an opened photo-copying machine.

Table I. Details of Philips photo-copying lamps.

Type	luminous length cm	arc voltage V	maxim. ignition voltage V	open-circuit voltage V	current A
HOG 700 W	43	190	340	390	3.9
HOG 2000 W	122	550	680	850	4.2
HOK 1200 W	40	550	600	850	2.5
HOK 2000 W	55	550	600	850	4.2
HOK 3000 W	130	1250	1500	2000	2.9

Manufacturers can therefore choose between two hard-glass and three lamps of fused silica.

Design and characteristics of the lamp

As will be seen from table I, the arc voltage is fairly high, especially in the longer tubes, for which reason the open-circuit potential of the voltage source (leak transformer, which dispenses with the need for a separate choke) must also be high to ensure stable operation. The higher arc voltage is employed, because, on lower voltages, either the vapour pressure of the mercury must be reduced, or the diameter of the tube must be made very much greater. A larger tube, especially of fused silica, would increase the cost of the lamp, whilst the warming-up time (the time from the moment of ignition until the arc voltage assumes a steady value) would then be very long.

The first alternative, i.e. a reduction in the vapour pressure, results in a discharge column which is not so concentrated, with consequent rapid discoloration of the glass wall. This effect is largely

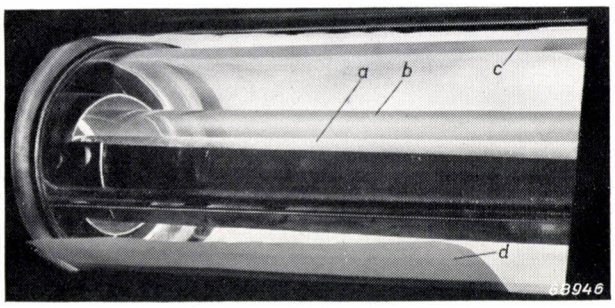


Fig. 3. Part of a photo-copying machine with cover removed (make: N. V. van der Grinten, Venlo). *a* fused silica lamp, type HOK 2000 W. *b* glass jacket to restrict the evolution of ozone to a confined area around the lamp. *c* rotating glass cylinder against which the transparency with sensitive paper are pressed.

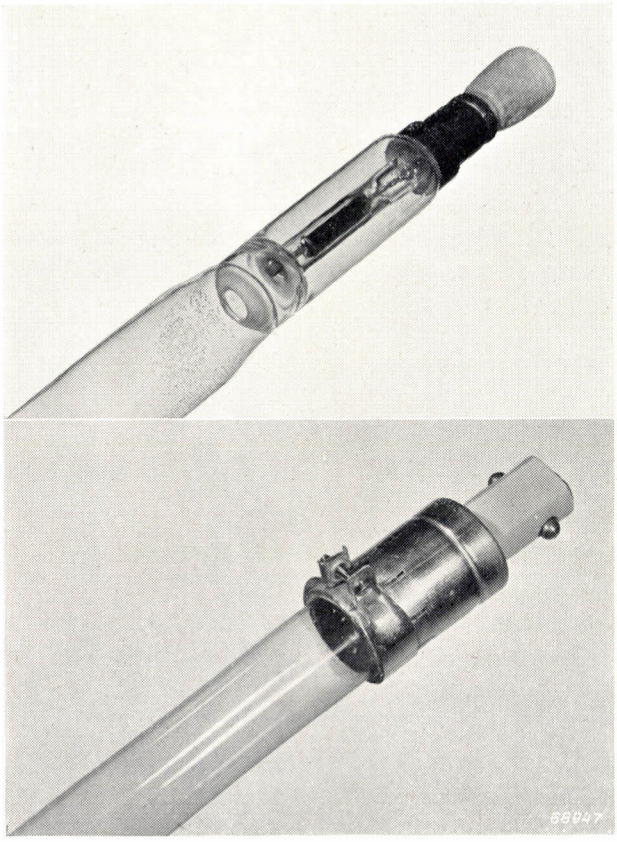


Fig. 4. *a*) Extremity of lamp type "HOK" (that part which is normally sand-blasted has been left clear for better visibility in the illustration). *b*) The same, type "HOG". In the HOK lamp will be seen, from left to right: drops of mercury on the inner wall of the discharge tube, the diaphragm of fused silica which serves, inter alia, to concentrate the discharge, the cathode, and the seal, consisting of molybdenum strip, the stranded wire and, just visible near the cap, an extra seal made from glass as an intermediate material between the fused silica and the leading-in wire. The quartz tube is extended, and carries the cap with insulated terminal nut. The HOG lamp has a universal base, mounted on the tube by means of a metal strip.

due to adsorption of positive mercury ions in the wall. These ions are neutralised in the wall and, owing to the large diameter of their atom, have difficulty in leaving the wall again. (Efforts to demonstrate the presence of the mercury chemically in coloured fused silica have proved successful.) The more the discharge is concentrated, the fewer the ions reaching the wall and consequently the less the discoloration.

The electrodes of the lamp may also be a source of blackening of the tube; in order to maintain the discharge along the axis of the tube and also to limit as much as possible any blackening of walls near the electrodes, the HOK lamp is fitted with diaphragms of fused silica 7 mm in diameter, roughly 1 cm in front of the electrodes.

A lamp with constant luminance along the whole length of the tube is thus ensured, and this is an important condition of a good photo-copying lamp.

The electrodes of the photo-copying lamps are oxide-coated and, in the lamp made of fused silica, the current is applied through molybdenum strips 4 mm wide and 12 to 13 microns in thickness. These strips attain a very high temperature and that part which lies outside the tube tends to oxidize very rapidly; this may result in cracking of the seal. For this reason and also from the point of view of the mounting of the diaphragm, a special seal has been devised, as shown in *fig. 4a*.

It will be noted that the tube of fused silica is extended beyond the special seal; this has been so arranged in order to avoid over-heating of the cap due to the high temperature in the region of the seal. In the type of construction depicted, in which the cap is some distance from the seal, the temperature at that point is therefore sufficiently low to admit of attaching the cap quite simply by means of heat-resistant cement.

The seal of the HOG lamp is of the conventional kind for ordinary glass. Since the loading of this lamp is lower than that of the lamp of fused silica, the temperature in the region of the seal is accordingly lower, so that the caps can be attached to the tube by means of a metal strip with a lining of asbestos (see fig. 4*b*). These caps are universal, i.e. suitable for side-contact, end-contact or screw-contact. The various lamps discussed are depicted in *fig. 5*.

Fig. 5. Philips photo-copying lamps. From below: HOK 1200 W, HOK 2000 W, HOK 3000 W, HOG 2000 W and HOG 700 W. (The HOG 3000 W is roughly 5 feet in length.)

The power balance of the various lamps may be seen from *table II*. It will be noted that, notwithstanding the increased difficulty of manufacture entailed by the use of fused silica and the higher cost of the lamp, the use of this material is amply

justified; the percentage of applied power emitted by the tube in the form of radiation of the discharge is almost three times as much as that of the HOG lamp.

Table II. Power balance of photo-copying lamps.

Type	Average wall temp.	Radiant power passing through tube wall		Heat radiated by tube wall		Heat dissipated by convection and conduction	
	°C	in watts	in %	in watts	in %	in watts	in %
HOG 700 W	425	120	17	380	55	200	28
HOG 2000 W	425	350	18	1100	55	540	27
HOK 1200 W	600	575	48	400	33	225	19
HOK 2000 W	620	1000	50	625	21	375	19
HOK 3000 W	580	1250	42	1025	35	700	23

This great difference is in large part due to the higher loading per unit length of the HOK lamp (by reason of which the total radiation per watt of applied electrical power is also considerably greater) and, for the rest, to the lower absorption in the tube wall. Absorption in the hard-glass tube of the HOG lamp in the near ultra-violet, at the high working temperature (about 425 °C), is much more pronounced than at room temperature, at which level it is not very different from that of fused silica. If the HOG lamps were fitted with a tube of fused silica instead of hard glass, the amount of power transmitted through the tube wall would be 175 W in the case of the HOG 700 W, instead of 120 W; that in the HOG 2000 W would be 500 W instead of 350 W. In both instances, therefore, roughly 25% of the power consumed by the lamp would be emitted as radiation, as against 17% with the hard-glass tube.

The effect of absorption in the tube wall upon the spectral distribution of the radiations is illustrated in *table III*. In the visible range of wavelengths the HOK 1200 W and HOG 2000 W lamps are roughly of equal intensity; at the wavelengths of $\lambda =$ $= 3655$ Å and 3341 Å, which are still important for the exposure of dye-line paper, the radiant power of the HOG lamp is 55% and 18% respectively of the radiation of the HOK lamp. In the last column of table III the efficiency η of the lamps is shown (lamp as used with the sensitive paper referred to in fig. 1). This is obtained from the equation:

$$\eta = \frac{\int E_\lambda G_\lambda \, d\lambda}{W},$$

where E_λ denotes the power in watts emitted at the various wavelengths (as given in table III), G_λ is the sensitivity of the dye-line paper as in fig. 1, maximum sensitivity being taken to be unity, and W the power consumption of the lamp in watts.

Radiations at wavelengths below $\lambda = 3000$ Å are not taken into account as these are absorbed by the glass between the lamp and the paper over which the paper is transported, whilst radiations of $\lambda = 3022$ Å and 3130 Å are included to the extent of one half for the same reason. As will be observed from table III, the HOK lamps attain an efficiency of nearly 10%.

By reason of the high transmission of the far ultra-violet radiations through the tube wall of the HOK lamp, this lamp, apart from photo-copying purposes, is particularly useful for other photo-chemical processes, more especially in cases where a linear source is an advantage. In this connection we may mention the manufacture of vitamin D [2]) and ozone, and investigations into the processes of discoloration.

Table III. Power in watts radiated by the complete lamp at the more important spectral lines, and efficiency η.

Type \ λ(Å)=	2537	2653	2699	2753	2804	2894	2967	3022	3130	3341	3655	4047	4358	5461	5780	η
HOG 700 W	—	—	—	—	—	—	—	—	0.5	0.3	7.5	8.5	13	19	12	2½%
HOG 2000 W	—	—	—	—	—	—	—	<0.5	1.5	1	36	35	58	77	58	3½%
HOK 1200 W	18	14.5	3.5	3	7.5	4.5	11	19	41	5.5	64	30	55	75	58	8 %
HOK 2000 W	37	34	7.5	6	17.5	8.5	21	43	85	10	128	52	96	122	128	9 %
HOK 3000 W	49	42	10.5	7.5	21.5	11	29	59	125	13.5	190	79	144	187	174	8½%

In photo-copying machines the formation of ozone can be a drawback (odour, oxidation). Ozone is produced by the action of wavelengths of roughly 1850 Å on the atmosphere. Large quantities of air circulate in the space where the photo-copying

[2]) See amongst others A. van Wijk, Lamp manufacture and vitamine research, Philips Techn. Rev. 3, 33-39, 1938.

machine is situated, to cool the glass cylinder over which the transparency and sensitive paper are guided (the temperature of the paper must not exceed about 70 °C) and, in order to check the formation of ozone from so much air, a glass jacket which is capable of absorbing the radiations in question is always used with the HOK lamps. This jacket is sometimes also included with the HOG lamp if the photo-copying machine is relatively compact, or if the particular dye-line paper is not so well able to withstand the usual high temperature. If the jacket were not used, the cooling would have to be increased to such an extent that there would be a risk of condensation of the mercury in the lamp with consequent drop in the arc voltage. Although the softening point of the glass of which both the cylinder on the machine and the jacket are made may be lower than that of the tube wall of the lamp, it is strictly essential that the radiations required to effect the exposure ($\lambda > 3000$ Å) shall not be absorbed to any appreciable extent, if the advantages of the HOK lamp over its HOG counterpart are not to be lost.

Summary. In photo-copying a light source is required that will emit a high proportion of its radiations in the violet and near ultra-violet and which at the same time demands but little maintenance. The carbon arc originally used for this purpose is being completely ousted by the high-pressure mercury vapour lamp which not only has the qualities referred to, but which constitutes a linear light source, this being of considerable importance in photo-copying work. The material of which the tube is made must be capable of transmitting ultra-violet rays even at the high temperatures involved; it must also be able to withstand the effects of such temperatures for a considerable time without deterioration. Hard, ultra-violet transmitting glass and fused silica fulfil these conditions. A number of Philips photo-copying lamps, type HOG (hard glass) and type HOK (fused silica) are discussed and details are given in tabular form. The photo-copying efficiency of the HOK lamps made of fused silica is more than twice that of the hard-glass HOG lamps, owing inter alia to the higher loading per unit length. Apart from photo-coyping, HOK lamps can be employed for diverse photo-chemical processes by reason of their high emission among the far ultra-violet spectral lines.

ELECTRIC FENCING OF GRAZING LAND

by F. H. de JONG and L. W. ROOSENDAAL.

621.315.1:631.27

Many people may think it rather peculiar that electrotechnical, even electronic means, should be employed for confining cattle within certain bounds, an object that can be achieved in a much simpler way, for instance, with barbed wire. This matter appears in an entirely different light, however, when it is understood that thanks to electrical fencing grazing land can be made to yield far more grass and thus the production of dairy produce can be considerably increased. For a country like the Netherlands this is a matter of such great economic importance that in certain cases a government grant is given towards meeting the expense of having electric fencing installed.

Economical cattle grazing

In countries where the acreage under grass is limited and cannot be extended, either owing to the soil being unsuitable or on account of the density of the population, economically planned grazing is a matter of great importance not only for the cattle raisers but for the country at large. A typical example of such a country is the Netherlands, which, with its enormous density of population, is obliged to increase its export trade as far as possible, a trade which consists for a large part of dairy produce. The same applies more or less to Belgium, Denmark, France and Switzerland, where somewhat similar conditions prevail. This explains why in recent years so much attention has been paid in these countries to economical grazing.

Grazing is not such a simple matter as the layman imagines it to be. When for instance, a number of cows are put out to graze on a certain pasture for the whole of the season the amount of grass they consume is not by any means so much as what that land could produce if there were no cattle on it. The cattle are continually trampling down quite a considerable amount of grass which never has the time to develop into a luxuriant crop.

For economical grazing the cattle have to be confined to one particular plot of pasture land at a time (e.g. 1 acre per 8 cows), being moved on every

two or three days to another plot, but not back to the first plot until the grass there has had time to recover from the trampling and grazing, which may take anything between one and two weeks according to the conditions prevailing. In this way the "loss of grass" can be reduced from about 50% to 20 or even 10%.

But this involves a suitable fencing off of the plot being grazed. A fencing of ordinary, smooth wire is quite inadequate because cattle are in the habit of rubbing against the wire and the posts, which would have to be very strong, and thus expensive, to withstand this. That is why barbed wire is usually employed, though it costs more. Putting up a barbed wire fence, however, is no pleasant job and this makes it impracticable to move the fencing every time the grazing plot is changed, so that really all the pasture land would have to be fenced off in equal plots, which would be rather costly. Moreover, what constitutes a more serious objection is the fact that the cattle often injure themselves on barbed wire, sometimes so seriously that they have to be attended to by a veternary surgeon, and the scars left in the hides render them inferior for the leather trade.

Electrical fencing

These objections against the use of barbed wire have now been overcome by setting up a fencing of smooth wire attached to insulators (screwed into posts) and electrically energized at such a voltage that when touched it gives shocks sufficient to cure the cattle of their rubbing habit, without being dangerous either to animals or to human beings.

This has proved to be quite successful. The first part of the problem — how to give the wire sufficient voltage to act as electric fencing — is of course quite easy to solve, but the difficulty lies in meeting the requirement that the wire must not be dangerous. Neither with an uninterrupted direct voltage nor with an ordinary alternating voltage could a suitable compromise be found. Even an alternating voltage of 60 V, with a frequency of 50 c/s, is not without danger to human beings, and yet it would by no means give sufficient shock effect for cattle.

However a very good solution was found, based upon the fact that both human beings and animals can bear much higher voltages when these are pulsatory, with sufficiently long intervals, than when they are constant or sinusoidal. Anticipating the safety regulations — which will be discussed later —, it might be mentioned here that under certain conditions voltage pulses with a peak value of 5000 V are harmless. It is obvious that such a voltage has sufficient shock effect for the purpose.

It might be feared that cattle would then suffer from burns, instead of the wounds inflicted by the barbed wire, but as a matter of fact, notwithstanding the high voltage, the energy of the pulses has been made so small that there is no question of burns. The absence of any risk of injury at all is therefore considered in cattle-rearing circles to be a very great advantage. This has made electric fencing popular not only for cattle grazing but also for horse breeders.

This electric pasture fencing thus consists of ordinary, smooth, wire attached to insulators on posts and connected to a fence controller supplying the electric pulses. Mostly only one single wire suffices for the fencing. Since the animals will refrain from rubbing against them, the posts need not be particularly strong and it is not necessary to set them up at intervals of less than about 30 feet (10 m). Compared with barbed wire such a fencing is very cheap, so that all plots of pasture land can be fenced off in this way at little cost. Furthermore, since the fencing consists of only one or two smooth wires and a small number of light posts, it can quite easily be shifted from one plot to another when the cattle are moved on to fresh grazing ground. Both these methods are being followed.

Though it may not be essential to lay the fence wire along the side of a plot that is bounded by a ditch, this is advisable if it is desired to prevent cattle from getting too close to the ditch and breaking down the banks.

Incidentally it is to be noted that electric fencing is also economical for poultry farms, for remarkably enough the birds do not attempt to fly over the fence wire once they have been in contact with it. For this reason the wire has to be laid rather close to the ground, at a height of about 8 inches (20 cm). The advantage of electric fencing in this case has, of course, nothing to do with the cropping of the grass but lies in the saving in cost compared with a fencing of 6-ft (2 m) wire netting and the convenience of being able to step over the wire anywhere instead of having to walk round to a gate.

Safety regulations

Most countries have their specific safety regulations for electrical material and appliances. In the Netherlands, for instance, the regulations are formulated and inspections are carried out by a government-authorized body briefly referred to as the

"KEMA". Apart from a number of regulations of general application for all electrical appliances handled by the public there are others of particular importance for the controllers used for this electric fencing. The reason why these controllers have been designed in the manner described below will be more readily understood when the following abbreviated extracts from the regulations concerning the voltage and current in the fence wire circuit [1]) are considered:

a) No continuous current of more than 0.7 mA (peak value) may flow through the fence wire circuit [2]).
b) The duration of a pulse must not exceed 0.1 second (a definition of duration is given).
c) The interval between two pulses must not be shorter than 0.75 sec.
d) The electric charge per pulse must not exceed 2.5 milli coulomb (absolute value).
e) The peak value of the voltage must not exceed 5000 V (measured with the fence-wire terminals of the apparatus connected by the most unfavourable combination of a resistance of 500 ohms or more and a capacitance of 0 to 2 μF).
f) The peak value of the current in the fence-wire circuit [2]) must not exceed 300 mA.
g) Mains-operated fence controllers must comply with the requirements (a) ... (f) when they are supplied with a voltage 1.15 times the rated voltage or less.
h) Mains-operated fence controllers must function when they are supplied with a voltage between 0.85 and 1.15 times the rated voltage. (From (g) and (h) it follows that with mains voltages below 85% of the nominal rating the controller must either still function — and at the same time comply with the conditions (a)... (f)! — or not function at all.)
j) In the event of a component part becoming defective, an inadequately secured internal connection working loose, an insulation breakdown or air gap of less than 10 mm or an air path [3]) carrying emission current being shorted, the apparatus being supplied with the wrong kind of current, or the supply source being connected in the wrong manner, such must not result in the conditions (a) ... (g) no longer being complied with.

The regulations existing in other countries, in so far as they have been imposed at all, are of the same purport, though in some cases differing somewhat in detail.

[1]) Taken from section 7 of the "Conditions imposed for the inspection of electric fence controllers" issued by "N.V. tot Keuring van Electrotechnische Materialen, Arnhem", December 1948.
[2]) According to a preceding clause (2g), unless otherwise expressly stated, the voltage and the current in the fence circuit have to be measured while the terminals for that circuit are connected by the most unfavourable combination of a resistance between 500 ohms and 2 megohms and a capacitance of at most 0.2 μF — in accordance with the widely divergent impedances which may be present between the fence wire and earth when it is touched.
[3]) By "air gap" is to be understood here any discharge path in gas or in vacuo.

The fence controller

The fence controller designed by Philips (see fig. 1) is intended for use on A. C. mains. The first requirement was that no moving contacts should be used, because these involve maintenance and periodical readjustment.

Fig. 1. The Philips electric fence controller, type 7937, designed for use on A. C. mains.

The basic principle taken was the well-known circuit for generating relaxation oscillations (*fig. 2*). A capacitor C_1 is charged via a resistor R_1 from a direct-voltage source. This capacitor is shunted by a gas-discharge tube with an ignition voltage

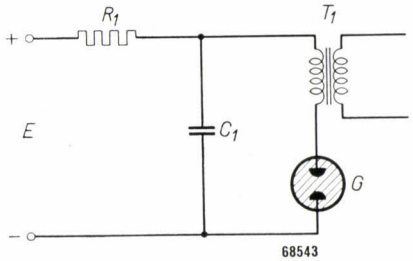


Fig. 2. Basic circuit for generating relaxation oscillations. G glowlamp igniting as soon as the exponentially increasing voltage across the capacitor C_1, charged by the direct-voltage source E via a resistor R_1, reaches the ignition level. The capacitor then discharges rapidly, the glow discharge is interrupted and the capacitor is recharged. Each ignition produces a voltage pulse at the transformer T_1.

higher than its extinction voltage. When the capacitor reaches a potential equal to the ignition voltage of the tube the latter is triggered and the capacitor rapidly discharges with a powerful current surge. When the discharge is completed and the tube is again non-conducting the capacitor is recharged. The current surges can be passed through the primary of a transformer, which then, like a spark inductor, produces high-voltage pulses on the secondary (high-voltage) side.

The repetition frequency of the pulses depends upon the relaxation time R_1C_1 and also upon the magnitude of the direct voltage E and the ignition voltage of the tube. It is the dependency on these voltages that renders the simple circuit of fig. 2 inadequate for the purpose. In practice the voltage E will be obtained by rectifying an alternating voltage taken from the mains, and unless it is stabilized in some way or other it will be subject to the same fluctuations as occur in the mains, which particularly in rural districts are apt to be considerable. But even if E were kept constant there is still the difficulty of the ignition voltage of a gas-discharge tube varying in the course of time. The consequence of this voltage dependency would be that under adverse conditions the recurrence frequency of the pulses might either drop so low that no shock would be felt at all when momentary contact is made with the wire, or it might rise so high that the aforementioned requirement of at least 0.75 sec between two successive pulses is no longer complied with.

In order to overcome this difficulty the ignition of the gas-discharge tube is synchronized by means of small auxiliary pulses separately generated with a practically constant recurrence frequency. To this end the gas-discharge tube is fitted with an ignition electrode (thus making it a relay tube, or thyratron) and so dimensioned that in the absence of any voltage on the ignition electrode the ignition voltage lies far above the direct voltage E.

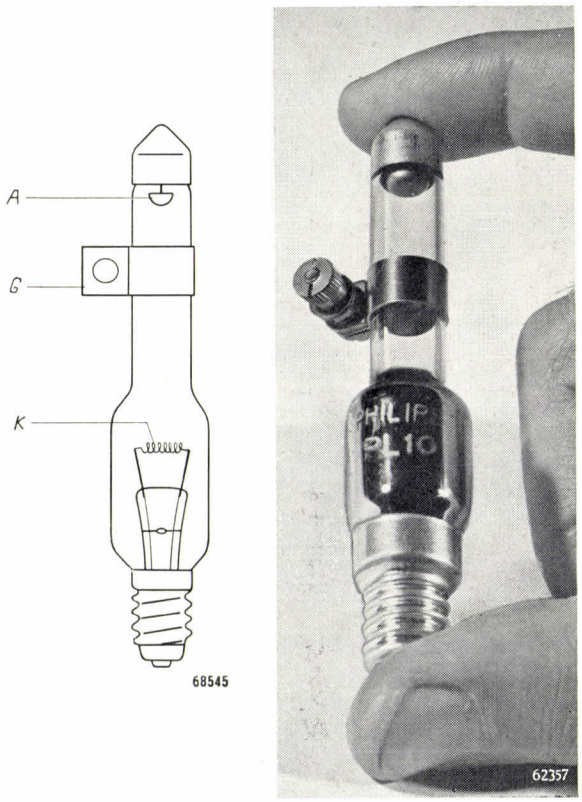


Fig. 4. Cross section and photograph of the type PL 10 thyratron (*Th* in fig. 3). *K* directly heated cathode (2 V, 2,8 A), *A* anode, *G* external ignition electrode in the form of a metal strip.

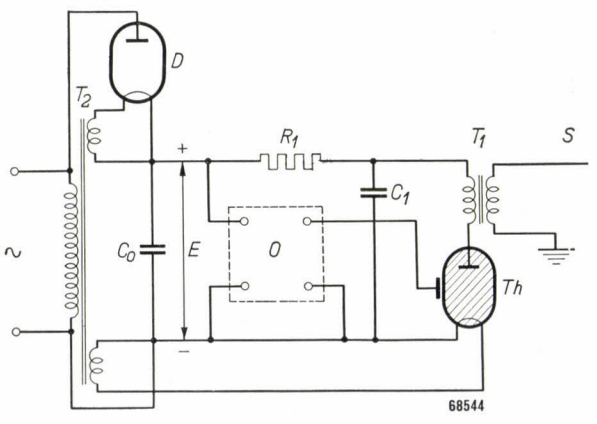


Fig. 3. Extension of the circuit of fig. 2. The glowlamp is replaced by a thyratron *Th* with hot cathode and external ignition electrode, pulses originating from the oscillator *O* being fed to this electrode. The direct voltage E is obtained by rectifying (with diode *D*) and smoothing (capacitor C_0) the alternating mains voltage. T_2 heater-current transformer, *S* fence wire. R_1, C_1 and T_1 as in fig. 2.

This system is represented in *fig. 3*. The thyratron, illustrated in *fig. 4*, has a hot cathode because this is able to withstand, better than a cold cathode, the powerful current surges, which in this apparatus reach a peak value of about 4 A (the discharge is therefore an arc and not a glow discharge). The high value of the ignition voltage is obtained by placing the anode in a narrow neck of the glass envelope at a relatively large distance from the cathode. It was found unnecessary to use an internal grid for the ignition electrode, a metal strip clamped round the glass neck answering the purpose very well (see fig. 4).

The auxiliary pulses triggering the thyratron are generated by an oscillator denoted by *O* in fig. 3 and shown in detail in *fig. 5*. This oscillator consists of a triode *P* (actually an EL 42 pentode connected as a triode is used), a coil T_3, a grid capacitor C_2

and a grid leak R_2. It is fed with the direct voltage E.

The circuit may be regarded as a Hartley oscillator with the coupling between the two parts of the coil T_3 serving as feedback. This feedback is so heavy as to cause periodical self-blocking. When the control-grid voltage approaches zero, the tube passes anode current and the circuit begins to oscillate. The oscillation is so strong, however, that before one cycle is completed such a large negative grid voltage is produced across the capacitor C_2 as to block the tube, so that the oscillation ceases. The discharge of this grid capacitor via the resistor R_2 takes place slowly. When the grid potential has again risen sufficiently close to zero the cycle begins anew. Thus in the coil T_3 voltage pulses are produced which trigger the thyratron by means of its ignition electrode.

The *table* below gives some electrical values of the Philips fence controller as taken from a test report of the "KEMA", in respect to the points (d), (e) and (f) in the foregoing extract of the Netherlands regulations.

The duration of the voltage pulse on the fence wire (see *fig. 6*) is shorter than 0.04 sec (mostly no longer than 0.01 sec, depending upon the load), thus well within the maximum limit of 0.1 sec given under (b) in the extract from the said regulations.

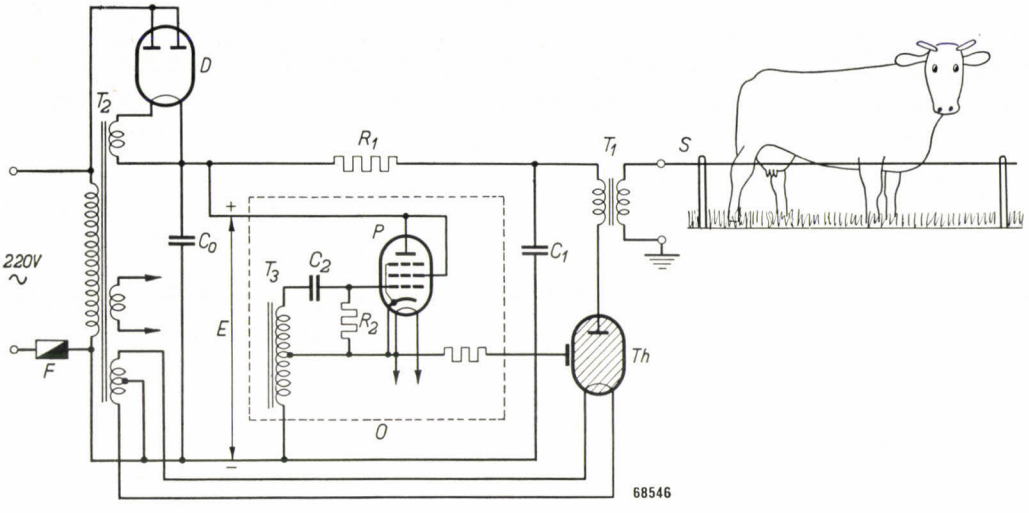
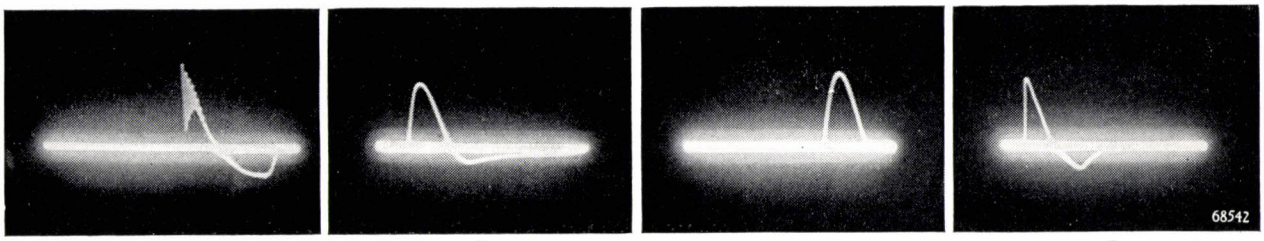


Fig. 5. Circuit of the Philips electric fence controller (somewhat simplified). P pentode EL 42 connected as triode, with grid leak R_2, grid capacitor C_2 and coil T_3. D diode AZ 41, F fuse; other symbols as in fig. 3.

Some values of the Philips electric fence controller as measured by the KEMA.

	Measured	Permitted	Mains voltage	Load
Charge per pulse	1.7 mC	2.5 mC	115 %	500 Ω [4]
Peak voltage on fence wire	100 V	150 V	115 %	500 Ω
	4700 V	5000 V	115 %	no load
Peak current in the fence wire circuit	200 mA	300 mA	115 %	500 Ω

[4]) 500 ohms is the most unfavourable load referred to in footnote [2]).



a *b* *c* *d*

Fig. 6. Oscillograms of the secondary voltage of the fence controller loaded with (*a*) a resistance of 30 MΩ, (*b*) 10.000 Ω, (*c*) 500 Ω and (*d*) a resistance of 1.0 MΩ in parallel with a capacitance of 3000 pF. In all four oscillograms the length of the thick white line corresponds to 1/50 sec. It is seen that clause (*b*) of the KEMA regulations, stipulating that the duration of the pulse shall not exceed. 0.1 sec, is amply complied with. The vertical scale differs in the four oscillograms.

Reckoning with a duration of 0.04 sec, according to (c) the maximum recurrence frequency allowed by these regulations is $1/(0.75+0.04) = 1.27$ per sec = 76 pulses per minute. The recurrence frequency is inversely proportional to the relaxation time R_2C_2 (fig. 5), which is so chosen that under normal conditions about 60 pulses are given per minute. Small variations in the characteristics of the EL 42 tube (P in fig. 5), as may occur in the course of time or when the tube has been replaced, prove to have little influence upon the recurrence frequency. The frequency is almost independent of the mains voltage within wide ranges (see *fig. 7*).

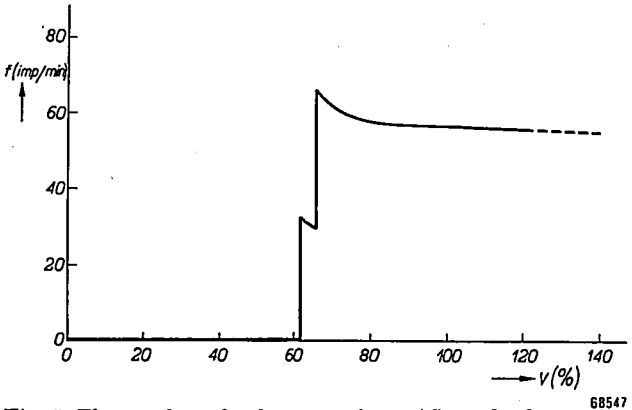


Fig. 7. The number of pulses per minute (f) on the fence wire as a function of the mains voltage V (in % of the nominal rating). f always remains below the maximum permissible value of 76 pulses per minute.

There is a slight tendency for the recurrence frequency to be reduced when the mains voltage rises considerably above the nominal rating, but even when the level reaches 200% of the nominal there is still very little change in the frequency [5]). Too low a mains voltage would tend to increase the frequency. According to condition (h) in the said regulations the frequency limit determined from (c) and (b) — which in our case is the 76 pulses per minute mentioned above — may not be exceeded at any mains voltages below 115% of the nominal rating. This condition has been met by choosing the conditions such that just before this limit is reached the thyratron is not triggered by every pulse on the ignition electrode but only by every second one: the construction of the thyratron is such that, as the direct voltage E is lowered proportionally with the mains voltage, the capacitor C_1 (fig. 5), after the last discharge, is not sufficiently charged for triggering the tube when the next pulse comes through on the ignition electrode, though it will be so for the second pulse. Thus at a certain level of the mains voltage (60 to 70% of the nominal rating) frequency division takes place in the ratio 2:1 (fig. 7), the pulse recurrence frequency in the fence circuit then being well below the critical value. As the mains voltage drops still further the frequency rises a little at first but very soon the system ceases to work. In this way condition (h) of the regulations is complied with, but it is to be noted that in practice it will hardly ever occur that the mains voltage drops to the level where frequency division takes place.

[5]) This excessive voltage was, of course, applied only by way of a test; the fuse in one of the leads (see later) was shorted for this test.

Safety precautions against defects or misuse of the controller

Something remains to be said about clause (j) of the regulations referred to, concerning safeguards against defects in the controller or its misuse.

From the diagram in fig. 5 it will readily be seen that any defects in the form of an interruption of an electrical connection will simply put the controller out of operation and thus not give rise to any danger.

Neither is there any risk if the primary of the heater-current transformer were connected to a voltage lower than that for which it is designed. As to the risk of the apparatus being connected to too high a mains voltage, if, for instance, a controller designed for 127 V were connected to 220 V it would be put out of action by the blowing of a fuse (F in fig. 5).

As to the stipulation that no danger may arise if "an air gap (read: discharge tube) carrying emission current is shorted", it is to be borne in mind that the controller contains three discharge tubes, which will now be considered in turn.

Short-circuiting of the diode *(D)* would result in the entire A. C. mains voltage of 220 V being applied to the smoothing capacitor *(C_0)*. In that case the apparatus would cease to function. Moreover the high alternating current then flowing through this capacitor would blow the fuse and thus prevent damage to the electrolytic capacitors C_0 and C_1.

A short-circuit between two or more electrodes of the pentode *(P)* only results in the oscillator being put out of action, so that the whole apparatus then ceases to function, because without ignition pulses the thyratron cannot work.

Although, considering the shape of the thyratron *(Th)*, any spontaneous short-circuit between the

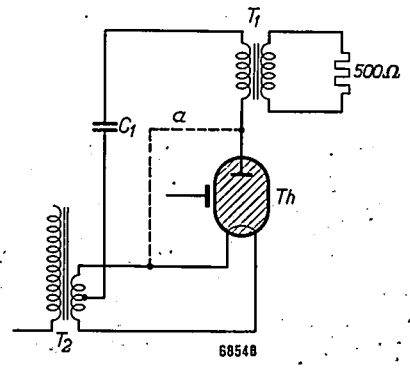


Fig. 8 If the thyratron *Th* is shorted by the connection *a* then half (1 V) of the heater voltage is across the series circuit consisting of the capacitor C_1 and the primary coil of the transformer T_1. If the secondary winding is loaded with the most unfavourable impedance (500 ohms) the continuous current in this circuit still remains below the maximum permissible value (0.7 mA peak).

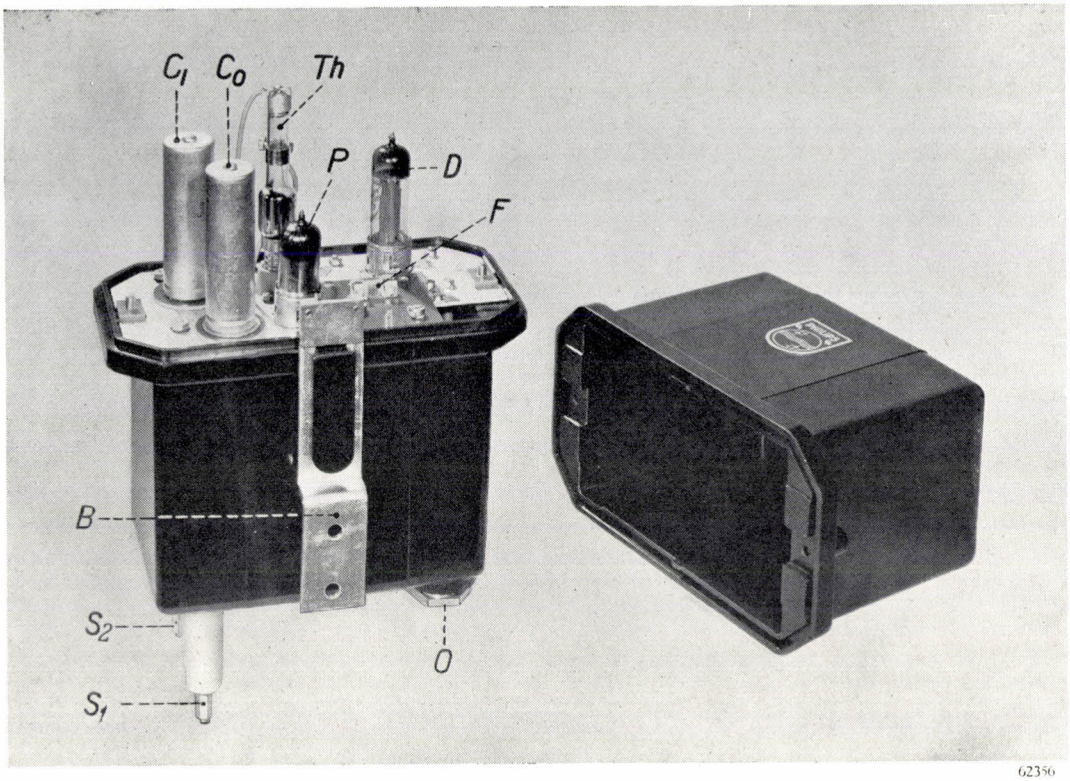


Fig. 9. The controller type 7937 opened and viewed from the back. *B* mounting bracket, *O* port for leading in the mains connections, S_1 and S_2 terminals for the fence wire and earth. C_0, C_1, *D*, *F*, *P* and *Th* as in fig. 5.

anode and the cathode of this tube is quite impossible, to comply with the regulations the consequences of such a thing ever happening have to be considered. Any shorting of the thyratron would give rise to a situation as shown in *fig. 8*, half the heater voltage being across the series connection between the capacitor C_1 and the primary of the transformer T_1. In that event, if the fence-wire circuit is loaded with an impedance, a continuous alternating current would flow through that circuit, and, according to clause (a) of the regulations, the peak value of this current must not exceed 0.7 mA, even if the impedance mentioned has the most unfavourable value (in this case 500 ohms; see footnote [2]). Since the heater voltage is low (2 V) and in the eventuality in question only half of it is effective (C_1 is connected to the centre tap of the heater current winding), it has been possible to comply with this requirement by suitably dimensioning the transformer T_1 and limiting the capacitance of C_1 to 15 μF.

All the constituent elements of the controller are contained in a drip-proof casing of a special kind of "Philite" made with an asbestos filler and thus very tough (fig. 1 and *fig. 9*); account has been taken of the possibility of the controller having to be mounted out of doors.

An audible signal indicates whether the installation is working properly. This signal is produced by two small strips of transformer sheet, spring-mounted on the core of the high-voltage transformer (T_1 in fig. 5), clapping against each other or against the core as each pulse comes through. If, for instance, the fence wire should make contact with the ground anywhere, this signal will not be heard, because the voltage at the transformer would be too low.

Summary. When grazing cattle are confined to one plot of pasture land and regularly moved on to another plot, instead of being allowed to roam over the whole of the field, the grass is given an opportunity to recover from the damage it suffers from the cattle trampling it down. As a result the land yields a much larger crop of grass than when it is grazed in the ordinary way. This, however, requires a fencing, either a permanent one round all the grazing plots individually, or a movable one placed round the plot being grazed. For both methods a smooth wire connected to an electric fence controller offers considerable advantages over barbed wire, because it is inexpensive and easily movable and the cattle cannot injure themselves on the wire. Provided the controller complies with the safety regulations there is no danger for human beings or for animals. A description is given of a controller operating with relaxation oscillations synchronized by means of a self-blocking oscillator. The controller is designed for A. C. mains supply and complies with the conditions imposed, which is explained for some less obvious cases. The proper functioning of the installation is indicated by an audible signal.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon applicatiou to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherland.

1977: A. J. W. M. van Overbeek: Voltage-controlled secondary-emission multipliers (Wireless Engineer **28**, 114-125 1951, April),.

Difficulties with life in secondary-emission amplifier valves have been overcome by employing a layer of caesium oxide kept at a temperature below 180 °C. Several constructions of experimental valves are shown and a description is given of their characteristic properties. A variable-mu valve and a very-high-slope valve employing four stages of multiplication are shown. The use of grid-shaped dynodes (secondary-emission electrodes) is discussed. Some circuits in which secondary-emission valves offer typical advantages are described, in particular generators of sinusoidal and non-sinusoidal oscillations and monostable and bistable trigger circuits.

1978*: W. Elenbaas: The high-pressure mercury-vapour discharge, Noordholl. Uitg. Mij., Amsterdam 1951 (XII + 173 pp., 80 figs.).

This book contains a survey of about 30 papers by the author and some hundred papers by other writers on the high-pressure mercury discharge. After a short historical introduction and a description of the mercury lamps now in use, the suppositions on which the theory of this discharge is founded are exposed, viz. that in every point of the discharge there is thermodynamical equilibrium as regards excitation and ionisation and that the power developed is dissipated by radiation and conduction. Different methods of measuring the temperature of the discharge are discussed. From the differential equation governing the radial temperature distribution similarity laws are derived and it is shown how the equation may be solved graphically. Subsequently the influence of a certain number of phenomena, such as convection, external magnetic fields, addition of rare gasses and of metals with an ionisation potential lower than that of mercury is discussed. The above considerations mostly pertain to the discharge in mercury vapour of about 1 atm. A special chapter is devoted to small mercury discharges with higher pressure and high luminance, which may be artificially cooled. Elaborate discussions are devoted to the spectrum and the gradual transition from a pure line spectrum to a spectrum with broadened lines and continuous background at higher vapour densities. A number of miscellaneous questions are discussed, such as the relation between the pressure and the potential gradient, the part played by positive ions, the conditions near the electrodes, the influence of self-absorption on the intensity of spectral lines; the question of the dimensioning of mercury lamps. Finally, attempts are made to give account of the quasi-thermodynamical equilibrium as found experimentally, by means of kinetic considerations, using known data about probabilities of excitation and ionisation. The result is rather convincing, considering the uncertainty regarding the functions governing the elementary processes.

1979: K. F. Niessen: On the condition determining the transition temperature of a superconductor (Physica **17**, 43-43, 1951, No. 1).

The analogy between superconductivity and ferromagnetism suggests an analogous dependence upon temperature for the part of the Fermi sphere covered by Heisenberg's superconducting layer and the relative magnetisation I/I_0 in ferromagnetic materials below the Curie temperature, I_0 being the maximum magnetisation (at $T = 0$). This leads to an assumption about what occurs at the transition temperature.

1980: H. Bremmer: On the theory of optical images affected by artificial influences in the focal plane (Physica **17**, 63-70, 1951, No. 7).

Optical images may be influenced by an artificial modification of the diffraction field in the focal plane in the image space. The effects of this modification are discussed in this paper, starting from an arbitrary distribution of the corresponding attenuation and phase retardation imposed at each point of the focal plane. These distributions may be chosen such as to accentuate very special features of the object in its paraxial image. The variety of possibilities in this respect is illustrated by a number of examples. For instance, it appears to be possible to reproduce theoretically the modulus of the gradient of the transparency for absorbing objects, and the modulus of the gradient of the phase-rotational distribution for transparent objects.

R 156: J. L. H. Jonker: The internal resistance of a pentode (Philips Res. Rep. **6**, 1-13, 1951, No. 1).

The phenomena determining the internal resistance of a pentode are calculated. For output pentodes the electrostatic influence of the anode voltage on the cathode current appears to be the main cause of this resistance. For high-frequency pentodes the two main causes are (1) the primary electrons that are repelled in the neighbourhood of the suppressor-grid wires and absorbed by the screen grid, and (2) the reflected electrons from the anode that can pass the suppressor grid and reach the screen grid. As both phenomena depend on the electrode voltages, the current distribution between screen grid and anode is affected by anode-voltage variation.

R 157: P. J. H. A. Kleijnen: The penetration factor and the potential field of a planar triode (Philips Res. Rep. **6**, 15-33, 1951, No. 1).

The potential field of a planar triode is discussed by making use of the penetration factor defined by means of the field of one plate and one grid. A calculation of this penetration factor is given, valid for all possible values of the wire diameter and the distance between the grid and the plate. This calculation being too complicated for immediate application, the numerical evaluation is given for a large number of configurations. A formula is deduced for the effective potential in the grid plane of a planar triode.

R 158: J. M. Stevels: Some experiments and theories on the power factor of glasses as a function of their composition, II (Philips Res. Rep. **6**, 34-53, 1951, No. 1).

It is shown that the measurement of the dielectric constant and the power factor for room temperature at a frequency $f = 1.5$ Mc/s of series of glasses in which the composition is varied by substituting ions on a molar basis, the rest of the glass being kept constant, gives valuable information about the structure of glasses in general. The replacement of alkali-metal ions by other alkali-metal ions gives rise to power-factor curves with a very deep minimum. The reasons for this minimum are discussed and explained. The substitution of alkali-metal ions by Mg^{++} and Ni^{++} ions sometimes causes the curves to assume a different shape. This shows that for these series part of the Mg^{++} and Ni^{++} ions take a network-former position in the glass, this giving rise to the deviating curves, especially for glasses with higher concentrations of these ions. Finally the replacement of alkali-metal ions by Zn^{++} and Pb^{++} ions is studied and the results are discussed.

R 159: W. Ch. van Geel and J. W. A. Scholte: Capacité et pertes diélectriques d'une couche d'oxyde déposée par oxydation anodique sur l'aluminium (Philips Res. Rep. **6**, 54-74, 1951, No. 1). (Capacitance and dielectric losses of an oxide layer deposited on aluminium by anodic oxydation; in French).

In this paper a model of the structure of the oxide layer on Al is given, based on measurements of the capacitance and dielectric losses of this layer. It is found that the dielectric losses decrease with increasing thickness of the oxide layer. The capacitance as a function of the forming voltage is measured by means of an A.C. bridge and also by the ballistic method. The equivalent circuit for the capacitance of the oxide layer is found to be a mounting in series of capacitors with resistors in parallel. The values of these resistors increase steadily from low to very high values. In consequence of the measurements the authors consider the layer to be built up in the following way: At the boundary aluminium - aluminium oxide there exists a layer containing an excess of Al atoms. These atoms produce a good conductivity of electronic character. In the direction of the electrolyte the concentration of the excess of atoms decreases. First comes an intermediate layer with higher resistance, and next a layer of very poor conductivity. An applied electric field seems to displace the atoms in excess and to change the mutual ratio of thicknesses of the different layers. The well-conducting layer is the cause of the dielectric losses.

VOL. 13 No. 12 pp. 337-372 JUNE 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THERMIONIC EMITTERS UNDER PULSED OPERATION

by R. LOOSJES, H. J. VINK and C. G. J. JANSEN.

537.583:621.385.032.216

The sphere of electronics has become so extensive that those who are employed in the design of circuits for specific purposes may easily lose sight of the physical principles of the medium in which they work. Although this need not be a deterrent to devising new, and even more ingenious, circuits, it should, nevertheless, be realized that real progress in the sphere of electronics can only be made through painstaking and patient study of the basic principles. These are, in the first place, the physical processes in an electronic tube, and the emission of electrons from the cathode is one of the most important. An investigation of this emission in the special case of pulsed operation — so important for radar — has shown that some very remarkable and entirely unexpected physical phenomena occur, of which the consequences cannot yet be fully grasped.

The use of thermionic emitters in valves for pulsed operation has become of the utmost importance during the last ten years. By pulsed operation is understood causing a cathode to emit intensively for a short period (for example 1-100 μsec) followed by a comparatively long interval (for example 1000 μsec).

The application of pulsed operation is of great importance in radar, apart from many other things. The use of pulses creates, in the first place, the possibility for a clear echo effect. Moreover, it is possible to obtain, with a comparatively small valve, a very large temporary output. This was already discovered in 1940, thus enabling the rapid development of radar. The extent of the difference in behaviour at pulsed and D.C. operation is made clear by the fact that a normal oxide-coated cathode ((BaSr)O) can supply current densities of tens of amps per cm^2 for one microsecond, whereas at D.C. operation the emission cannot be more than a few hundreds of milliamps per cm^2. It is therefore not surprising that this curious behaviour of the oxide-coated cathode should have been studied by a number of investigators [1]).

[1]) See, for example, E. A. Coomes, J. appl. Phys. **17**, 646, 1946. M. H. Pomerantz, Proc. I. R. E. **34**, 903, 1946. A. S. Eisenstein, Advances in Electronics **1**, 1, 1948.

These investigations were divided mainly into research into two different problems. The more practical one was designed to find the limits of thermionic emission for pulsed operation. The other, more theoretical, problem was designed to investigate what exactly happened in the oxide-coated cathode during a pulse and why there should be such a considerable difference between D.C. and pulsed emission.

Despite most intensive investigations, there is still no conclusive reply to the latter questions. In this article, an idea is given of the investigations which have been conducted in this direction in the Philips Laboratories at Eindhoven and a report of the remarkable phenomena which were found. Particularly remarkable is that the gradient of the potential in the oxide coating differs considerably from that in a normal resistor, and that the velocities of the electrons which are emitted by the oxide-coated cathode are not distributed according to Maxwell's rule (this would mean that the greatest differences in velocities that occur would amount to a few tenths of eV), but that the emitted electrons can be divided into several groups having different velocities. The velocities of the groups appear to differ by from some tens to some hundreds of eV.

Before the apparatus used to measure the velocity spectra is described and before discussing the results obtained in this manner, a short review will be given of the published results of other investigators of the behaviour of thermionic emitters for pulsed operation. This will be followed by a description of our investigations concerning the potential gradient in the oxide coating of the cathode. It were the results of these investigations which led us to investigate the velocity distribution. Finally, we shall sum up the consequences of the spectra found, as applied to the use of cathodes for pulsed operation.

Published results of other investigators

The investigations mentioned in the footnote [1]) resulted in a number of findings of great practical and theoretical importance. It was established, for example, that the maximum pulsed emission that can be obtained from an oxide-coated cathode is just over 100 A/cm². Moreover, it was found that with a new cathode this current density and the concomitant anode voltage corresponded to a point on that part of the current-voltage characteristic which is determined by the presence of space charge (see *fig. 1*, point *P*). The fact that the emission cannot be further increased by increasing the anode voltage is due to the occurrence of sparking. It is easy to understand that this sparking is not a normal breakdown between cathode and anode. When a normal breakdown occurs, the field strength at the surface of the cathode must be large, and positive in the direction of the anode. This is not possible as long as a negative space charge is present near the cathode. All this is true only for cathodes which are comparatively new. With older cathodes the current-voltage characteristic is slightly lower, and the current is saturated at a sufficiently high voltage, causing sparking to occur in the saturation area (see fig. 1, point *Q*). In that case there is a strong electrical field at the surface of the cathode which draws the electrons from the cathode.

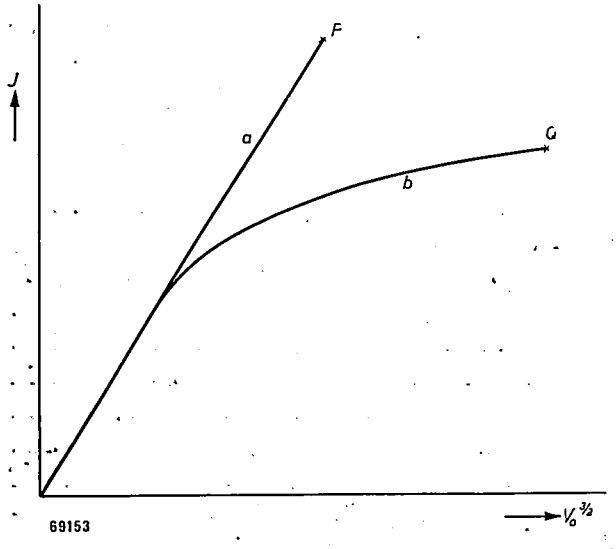


Fig. 1. Current-density-voltage-characteristic of a thermionic emitter under pulsed conditions. A potential gradient is caused by the negative space-charge present in front of the cathode provided the temperature is not too low and the voltage not too high, so that not all the electrons emitted can reach the anode. In that case $J \sim V_a^{3/2}$ (Langmuir-Child formula) is true. When increasing the anode voltage V_a in excess of a certain value, the space charge no longer influences the emission and so-called saturation occurs. In the characteristic, in which J has been plotted against $V_a^{3/2}$, this is shown by a deviation from a straight line. With a new cathode (curve *a*) sparking occurs in that part of the characteristic determined by the space charge (*P*); with an old cathode (curve *b*) sparking occurs in the saturation area (*Q*).

Another very important property, in addition to the maximum density of the current, is the electrical resistance of the oxide coating. This determines the additional heating of the cathode during the period of emission. At pulsed operation this additional heating (which is proportional to the resistance R and the square of the current density i) may be considerable at average current strengths which would have only a slight heating effect at D.C. operation. An increase in cathode temperature to the extent of 50 °C, as a result of the passage of a pulse current, is not exceptional.

Apart from the normal electrical resistance of the entire oxide coating, there may be also other resistances in the interfacial area between the metal base and the oxide coating. These resistances may be greater than those of the oxide coating itself. resulting in a large potential fall in the interface layer which, in turn, can lead to the greatest part of the heat due to the Joule effect being developed in that layer. It is assumed that the combination of high temperature and great field strength may be the cause of sparking.

The high resistance of the interface layer is caused by coumpounds resulting from reactions between the oxide coating on the one hand and the reducing components (for example Si, Ti), which are present as impurities in the nickel cathode base on the other hand. In other words; this high-resistance layer consists of bariumsilicate, barium titanate and similar substances. As such, reducing reactions are indispensable to obtaining good emission from an oxide-coated cathode, although a high-resistance layer is undesired because it influences sparking, an attempt has been made to reduce the influence of the resistance layer and a search has also been made for reducing materials which will not form such a layer. Our investigations

have thrown a little light on the last-mentioned point, i.e. the avoidance of a layer with a high resistance. A description of this will be given later. Some reduction of the influence of the resistance layer is achieved by constructions in which a skeleton of nickel is incorporated in the oxide coating [2]). This reduces the interface resistance as a result of the larger contact surface between nickel and oxide.

A description of our own studies and of the phenomena which we observed, some of which have not hitherto been published, follows. These notes may be valuable in explaining the occurrence of heating in the oxide coating and sparking phenomena.

The potential gradient in the oxide coating

Method of measurement

Two methods were used to investigate the potential gradient in the oxide coating during a current pulse. The first method [3]) was intended to give an overall picture of the total potential drop through the total thickness of the oxide coating. Diodes, with a flat cathode and adjustable anode, were used for this purpose (see *fig. 2*). The voltage between cathode and anode (V_a) is determined at a certain current pulse and the anode is brought closer to the cathode, step by step. The relation

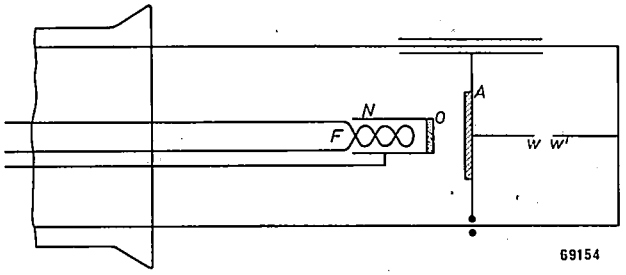


Fig. 2. Diode for measuring the potential gradient across the oxide coating. F filament, N nickel cathode base, O oxide coating, A adjustable anode. The distance between anode and cathode is determined by the distance between the points W and W'.

found between V_a and the distance between anode and cathode was extrapolated to the zero distance in the assumption that the difference in potential established in this way would be the voltage through the oxide coating [4]).

It would have been simpler and more direct to measure the potential difference whilst the anode was pressed against the cathode. This was indeed done; but when comparing these results with the result obtained by extrapolating to zero distance, it appeared that the two potential differences obtained varied considerably. This led to the conclusion that for pulsed operation the measurements were seriously disturbed by the contact between cathode and anode. This conclusion was supported by the fact that, when the cathode and anode were pressed together, sparking was observed.

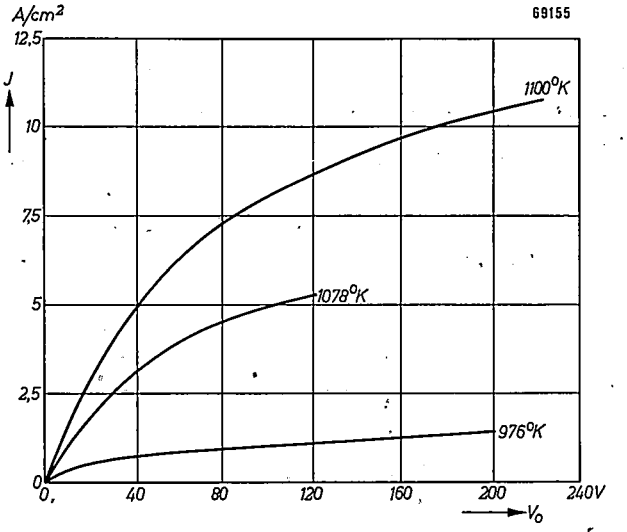


Fig. 3. Current-density-voltage-characteristics of an oxide-coated cathode under pulsed conditions at different temperatures. The voltage V_0 is plotted for the whole of the oxide coating as a function of the current density J. Total thickness of the oxide coating 25 μ.

The J-V characteristics at different temperatures, illustrated in *fig. 3*, were found by measuring, in addition to the potential gradient V through the oxide coating, also the current per cm² (J) emitted by the cathode. The graph shows that very great differences in potential (a few hundreds of volts) may occur in the oxide coating.

The next problem was to investigate the distribution of the potential gradient through the thickness of the coating: whether it was evenly distributed through the total thickness or whether it was localized in a certain part of the coating (for example, in the interface layer between nickel and oxide).

In order to investigate this potential gradient, one or more metal probes were placed in the oxide coating. A drawback of this method is, of course, that, in spite of the fineness of the wires used (thickness 8 μ), these might cause a considerable disturbance in the situation within the layer. We think, however, that the results found in this way, are, nevertheless, broadly correct for the potential gradient in the cathode.

[2]) Cathodes of this type are made in the first place to improve the bond between the oxide and the nickel. It was feared that strong electrical fields would pull the oxide away.
[3]) R. Loosjes and H. J. Vink, Philips Res. Rep. 2, 190, 1947.
[4]) The justification, proving the result obtained by means of extrapolation, is given in the article referred to in footnote [3]).

Fig. 4 gives an idea of the potential differences which were found, at different current densities, between the metal base of a cathode and three probes placed at distances of resp. 35 μ, 85 μ and 180 μ from the base. (The tests were made with a cathode covered with an oxide coating having a thickness of 270 μ). The fact that the potential difference between the nickel cathode base and the first probe and that between the second probe and the first were of the same order, was conspicuous. It meant that there was no extraordinarily large potential difference between the metal base and the first probe in this case; this is contrary to the experience of most other investigators.

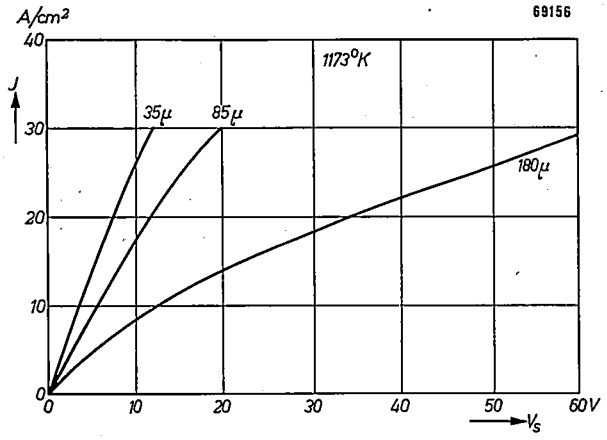


Fig. 4. Potential difference V_s between metal base and probes in the oxide coating as a function of the current density J. As is indicated, the various lines belong to probes which were placed at distances of 35 μ, 85 μ and 180 μ, respectively.

As the total potential difference across the oxide coating is much greater, with a coating of only 25 μ thick, than the above-mentioned potential differences between probes and nickel base (see fig. 3), we can only conclude that the potential gradient found by us for the total layer must be localized mainly somewhere near the surface.

The potential gradient as a function of the distance to the metal base, in a diode during a pulse, must therefore be as sketched in *fig. 5a*. The way in which this potential gradient was generally thought to be, has also been illustrated for purposes of comparison (fig. 5*b*). The absence of an area with large resistance in the boundary layer between metal and oxide is probably to be ascribed to the cathode nickel used by us. The contamination consisted principally of 0.1% to 0.2% magnesium and some silicon. The boundary layers with large resistance were found mostly when nickel had been used which contained principally silicon or chromium. Reverting to a remark made in the introduction, it is now clear that there are kinds of nickel in which there will not be an interfacial layer with large resistance.

This conclusion concerning a layer with large resistance close to the surface was tested by investigating an effect which might be expected from a potential distribution as illustrated in fig. 5*a*. The very porous structure of the oxide coating (50 to 80% pores; see also fig. 11) must lead to the assumption that the electrons are emitted not only from the surface but also from places deeper

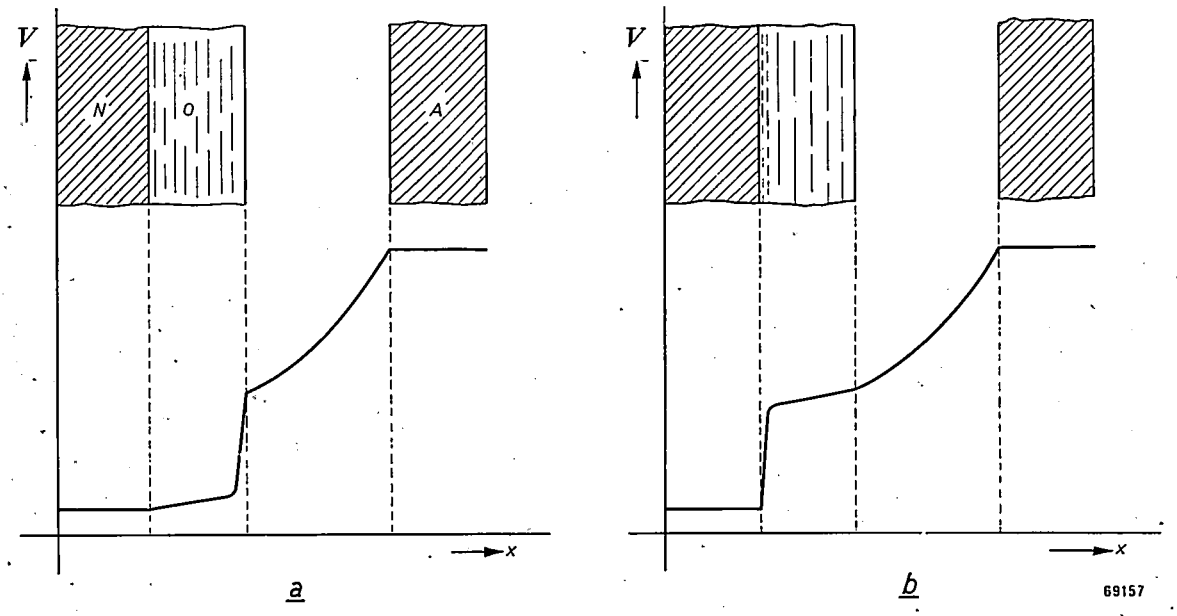


Fig. 5. Potential gradient in a diode during a pulse, according to our measurements (*a*) and according to Eisenstein (*b*) (see the article mentioned in footnote [1]). *N* metal base, *O* oxide coating, *A* anode.

in the coating. As the latter pass through a fairly large potential difference before arriving at the surface, it was to be expected that the emitted electrons would have velocities which differ by some tens of electron volts from each other. The velocity distribution of the electrons must, therefore, deviate from the Maxwell distribution which implies, as we have already remarked, that the differences are only a few tenths of electronvolts at a working temperature of 800 °C. This led us to investigate the velocity distribution of the emitted electrons.

The velocity distribution of emitted electrons

Description of the measuring tube

A special tube was constructed (*fig. 6*) to measure, in the best possible way, the velocity of emitted electrons. The principal parts of this tube were: the cathode (*K*), the anode (*A*) with a round aperture with a diameter of 200 μ, a screen (*B*) with a round aperture with a diameter of 4 mm, and a collector (*C*), with a rectangular slit measuring 24 mm × 0.3 mm. Deflection plates (*D*) were constructed behind the collector and, finally, a fluorescent screen (*Fl*).

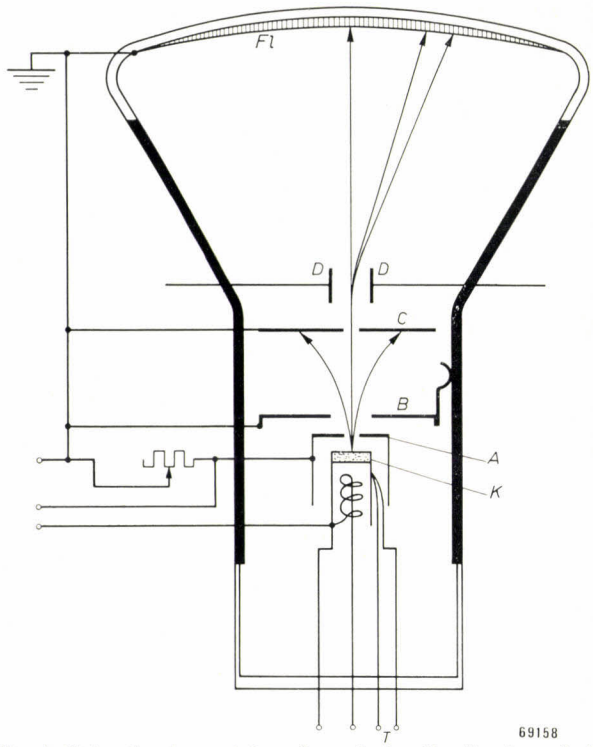


Fig. 6. Tube for determining the velocity distribution of the emitted electrons. *K* cathode, *A* anode, *B* screen, *C* collector, *D* deflection plates, *Fl* fluorescent screen, *T* thermocouple.

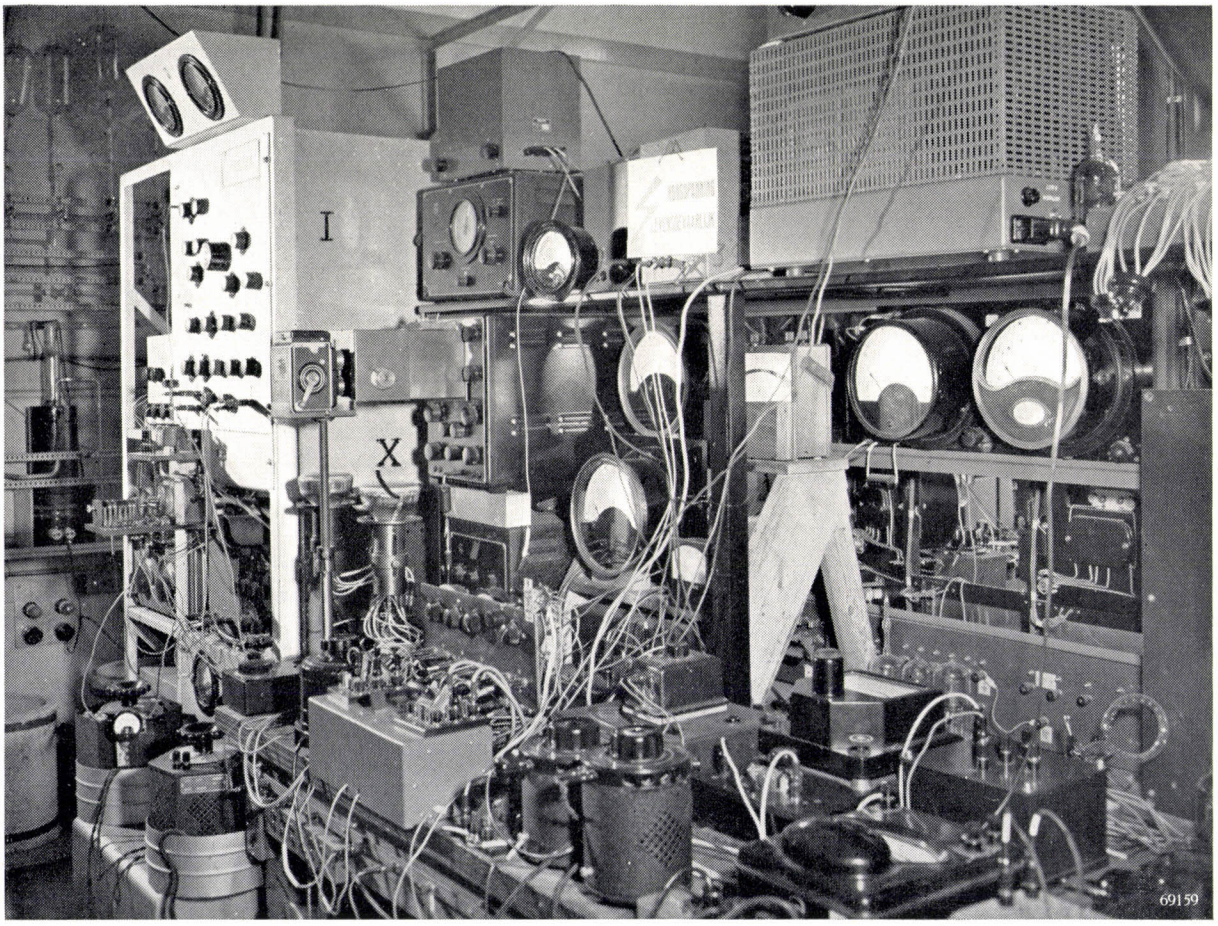


Fig. 7. Set-up for measuring electron velocity spectra. *X* measuring tube. *I* pulse generator.

Fig. 7 gives a general idea of the complete apparatus used for these measurements.

Operation of the measuring tube

If the cathode is at a negative potential with respect to the anode, part of the electrons flowing to the anode will pass through the aperture in the anode and enter the equipotential space behind the anode. The lens effect of the anode diaphragm and electrostatic repulsion between the electrons themselves, will cause the beam to spread. Because of this, only part of the beam will pass through the slit in the collector *C*. If both the deflection plates are at anode potential, the electron beam will form a sharp picture of the slit upon the fluorescent screen. If a potential difference is applied across the deflection plates, the place where an electron strikes the screen *Fl* will depend upon its velocity: the image of the slit will change into a spectrum. From the spectra obtained in this manner it was indeed shown that normal oxide-coated cathodes ((BaSr)O) will emit electrons, during a square-wave pulse, at velocities which, in some cases, varied by as much as two hundred electron volts.

A more detailed description will now be given of the velocity spectra found.

Resultant spectra

A number of characteristic electron velocity spectra are shown in *fig. 8*. Besides investigating the emission from (BaSr)O-cathodes, we also examined the spectra of a number of other kinds of cathodes. It must be admitted that the spectra of different cathodes, although coated with the same oxide (for example (BaSr)O) at a temperature of, for example, 925 °C [5]), are not identical with the one illustrated in fig. 8*b*. This spectrum is, however, representative of the behaviour of the majority of different (BaSr)O-cathodes at that temperature.

In the spectra of a tantalum cathode or an L-cathode [6]) (both metallic cathodes without an oxide coating) we can observe only one sharp line, thus electrons of one velocity, and the velocity corresponds to that which the electrons possess as a result of the acceleration in the field between the cathode and the anode. This is in accordance with expectations, as we would not be able to observe a Maxwell distribution and these metallic cathodes have no noticeable resistance; this signifies that all electrons are emitted from places having practically the same potential. These spectra also indicate that the tubes we used do not cause "false" lines, for example, due to secondary emission of the anode.

The spectrum of the normal oxide-coated cathode ((BaSr)O) shows the curious phenomenon referred to in the introduction; that there is not only a considerable range of electron velocities (to be expected in view of the potential gradient in the oxide coating), but that there are discrete lines, which mean that the electrons are divided into groups having certain velocities. It may further be observed that the velocity of even the fastest

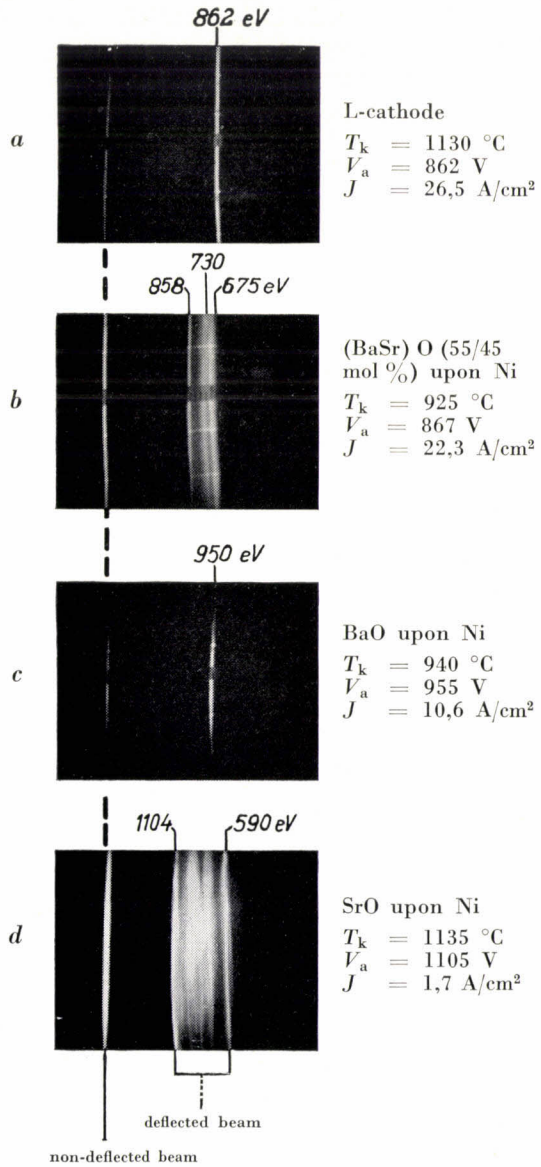


Fig. 8. Electron velocity spectra for different kinds of cathodes under square-pulsed operation. Pulse width $\approx 2.10^{-6}$ sec. The pulse recurrence frequency in *a*, *b*, *c* and *d* was 1000, 500, 1000 and 6000 c/s, respectively. The thickness of the oxide coating of *b*, *c* and *d* was about 60 μ.

[5]) All the temperatures communicated have been measured by means of thermo couples (see *T* in fig. 6): these are, therefore, so-called true temperatures.

[6]) See H. J. Lemmens, M. J. Jansen and R. Loosjes, A new thermionic cathode for heavy loads, Philips Techn. Rev. **11**, 341-350, 1949/1950.

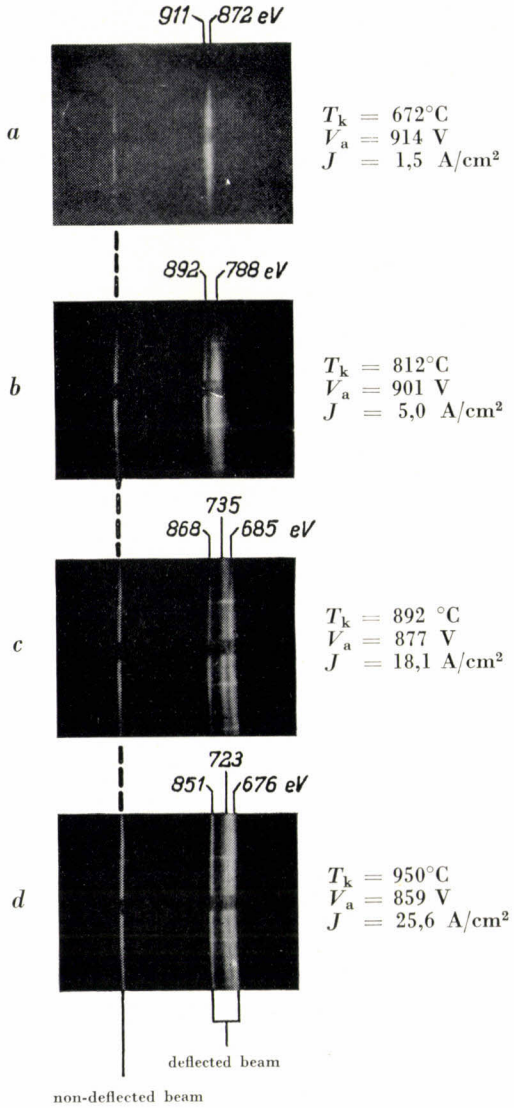


Fig. 9. Electron velocity spectra of a (BaSr)O-cathode (55/45 mol %) at different temperatures; square-pulsed operation. Pulse width 1.9×10^{-6} sec. Pulse recurrence frequency 1000 c/s.

electrons is smaller than the maximum attainable (V_a). The placing of the lines, as shown here, — two close together on the "slow" side of the spectrum and one fairly sharp one at the "rapid" side — is found in most cases. Occasionally, however, spectra are found which have one, two or more lines. A continuous background is always present.

Compared with the spectrum of a (BaSr)O-cathode, that of a BaO-cathode is very simple. As with the metallic cathodes, there is only one line and the velocities of the electrons differ only a few eV from the maximum that can be attained. In contrast to this, a SrO-cathode shows a very broad and confused line-spectrum. However, the spectrum of a ThO_2-cathode always consists of one sharp line only.

The influence of the temperature upon the character of the spectra, if any, has not, so far, been considered. *Fig. 9* illustrates the spectra of a (BaSr)O-cathode at different temperatures. It is remarkable that the variation as a function of the temperature shows a maximum. Below 640 °C no velocity spread could be observed; the strongest spread occurred at about 850 °C.

Finally, we have also investigated the relation between the velocity spectrum of the (BaSr)O-cathode and the current density. The results of these measurements have been summarized in *fig. 10*. The potential difference traversed by each group of electrons has been plotted as a function of the total current density. The dotted line indicates the connection between the anode voltage and the total current density.

These curves show that, in general, the velocity spread increases with increasing current density. It is also noticeable that spread occurred with current densities which belong to the extreme left-hand side of the current-density-voltage-characteristic of the diode, in other words, that part where the current is determined by the space charge (also see fig. 1). Conspicuous in fig. 10 is that the (dotted) diode characteristic, once past the saturation area, again curves upwards. This is a phenomenon which has been observed before with oxide-coated cathodes for pulsed operation (abnormal Schottky-effect). In this part of the characteristic the cathode is heated most.

This summary of so many uncorrelated facts is perhaps confusing. Unfortunately it is not possible

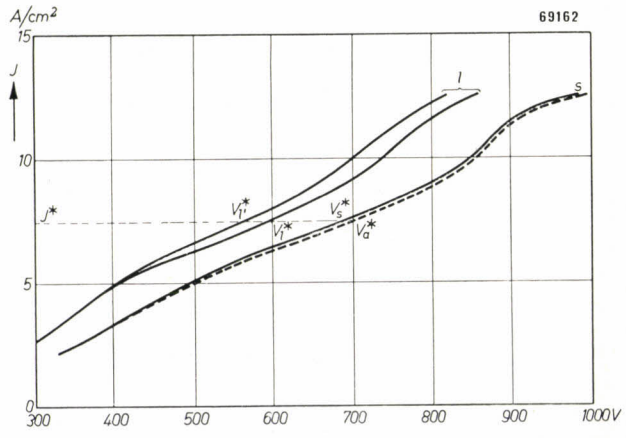


Fig. 10. Current-density-voltage-characteristics for the different electron groups of a (BaSr)O-cathode at a temperature of 840 °C. The voltage difference traversed by each group of electrons is plotted as a function of the total current density. *l* "slow lines", *s* "rapid line". The broken line represents the anode voltage as a function of the total current density. The line is, therefore, the normal diode characteristic. The graph must be read as follows (note the points indicated in the figure): at a current density J^* and an anode voltage V_a^* the spectrum shows groups of electrons with velocities (in eV): V_l^*, V_l^*, V_s^*.

to form a theoretical picture which will coordinate all the phenomena discovered into a form which is easy to understand. It is possible, however, to explain some of the phenomena. With regard to other phenomena it will be possible only to make a few observations which serve merely to indicate the direction in which the possible explanation lies.

Possible explanation of the phenomena discovered

In the first place we must try to understand what causes a (BaSr)O-cathode to have a resistance layer at its surface during a current pulse. There is evidence enough to assume that the occurrence of a resistance layer is closely related to the mechanism that is responsible for the conduction of the electrons in the oxide coating [7]). In another article [8]), the two processes which play a part in electron conduction have been extensively discussed. These are:

1. the conduction through the grains of the oxide material;
2. the conduction through the pores between the grains.

The potential which will be found at the outside layer of the oxide coating depends upon the relative contributions of the two mechanisms to the conductivity. This should be seen as follows: First of all assume that there is no conduction through the grains at all and that the electrons in a pore adjacent to the outside layer originate exclusively from the surrounding grains. By applying a difference in potential between the cathode and anode, all the electrons will then be drawn out of this pore. Thus, at that moment, there is no longer pore conduction. (It can be calculated that the time necessary to empty a pore is very much smaller than the duration of a voltage pulse.) Electrons will continue to flow as a result of emission from surface grains until the grains under consideration reach anode potential, thus there will exist a large potential drop in the surface layer. There is, however, conduction through the grains as well as residual conduction through the pores, so that the potential of the outside grains which actually occurs will be between that of the anode and that of the cathode. In cases in which the pore conduction plays no great part, the potential gradient in the oxide coating will be equal to that in a normal resistance. No great potential drop will take place, in that case, at the surface of the oxide coating.

This argument, which holds that the relative importance of the pore conduction determines the magnitude of the potential drop and hence also the extent of the velocity spread, is supported by the following facts, It is known that (BaSr)O, which shows a great velocity spread, has a relatively high pore conduction. In the case of ThO_2, for which only one "line" was found, it can be shown that pore conduction is negligible.

This explains that in (BaSr)O no abnormal velocity spread was observed (see fig. 9*a*) at low temperatures (for example 650 °C). For, at this temperature, the pore conduction is negligible also for this oxide. When the temperature increases, the pore conduction becomes increasingly important, up to a certain limit, because the emission increases more with increasing temperature than the conductivity of the grains. This also actuates the mechanism which originates the potential drop. A further increase in temperature causes, at constant anode potential V_a, a space charge in front of the cathode which limits an increase of the emission at increasing temperature. As the conductivity of the solid material continues to increase, however, the velocity spread of the electrons will diminish in the end.

Although this has given us an idea of the mechanism which causes the large velocity spreads, it is much more difficult to understand why the velocity spectrum is discrete instead of continuous. A definite explanation cannot be given. One of the ways in which the lines of the discrete spectrum might have been caused is as follows:

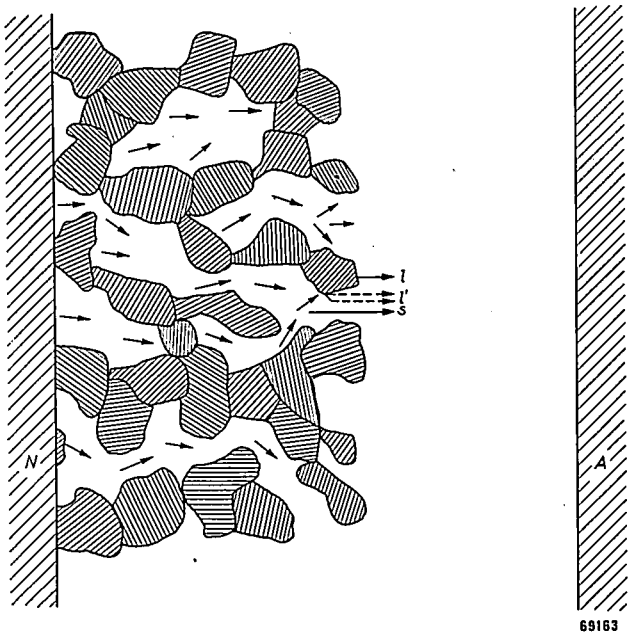
The electrons with velocities corresponding to the "fastest" line might be emitted from grains which are situated so deeply inside the coating that the conduction is no longer disturbed by the emission, as is the case with the grains in the outer layer (see *fig. 11*). The potential difference of these grains in relation to the metal base is determined only by the total conduction of the oxide and by the current density and one spectral line is obtained.

The fact that the slow electrons are also grouped in "lines" can only be explained by assuming that the ratio between residual conduction and emission is approximately equal for all grains situated in such a way that they can charge themselves.

The occurrence of two "slow" lines indicates that two kinds of electrons are emitted from the charged grains. Superficially one might suppose that these are thermal and secondary electrons. It would be easy for secondary electrons to be formed because the rapid electrons between the grains (emitted from the more deeply situated grains) would liberate secondary electrons through collision.

The explanation given is not, however, conclusive, because the initial velocity of thermal and secondary

[7]) R. Loosjes and C. G. J. Jansen, Le Vide, 7, 1131, 1952.
[8]) R. Loosjes and H. J. Vink, Philips Techn. Rev. 11, 271, 1949, 1950.



69163

Fig. 11. Cross-section through an oxide coating. Because of the porous structure of the oxide coating the emitted electrons may be divided into: l slow thermal electrons, s rapid thermal electrons, l' slow secondary (?) electrons.

electrons does not differ so much as to be solely responsible for the difference in energy of the two "slow' lines (sometimes some tens of eV):

Consequences of the results found

The existence of a layer with large resistance at the surface of an oxide-coated cathode has all kinds of consequences in practice.

In the first place, the heat development due to the passage of current through the oxide coating, will, for the greater part, occur in the outer layer. As this outer layer has a comparatively bad thermal contact with the vicinity, the temperature increase must be considerable in this region.

The large potential gradient in the outer layer of the oxide coating may also directly influence the behaviour of the cathode. If we assume that the potential changes 200 V within a trajectory of 10 μ (i.e. within a layer having a thickness of only a few grains), a field of 2×10^5 V/cm will be present in this area. It is conceivable that slightly larger field strengths might cause a breakdown in the outer layer (which moreover also has a high temperature). Such a breakdown might start sparking between cathode and anode. This fits in with the fact that ThO_2, which, as we saw, gives no indication of a surface charge, is known to be an emitting oxide in which no sparking phenomena occur. As can be seen from fig. 8c, there is no charging of the outer layer with BaO. It is, indeed, a fact that the voltage necessary to produce a spark between cathode and anode is 2 to 3 times as large as that necessary for (BaSr)O and SrO [9]).

The occurrence of a velocity spread of the electrons which is caused by the existence of the resistance layer, also affects the behaviour of the cathode. For, if there is a space charge in front of the cathode, this will influence electrons with different velocities in different ways. The fast electrons will be able to penetrate practically untrammelled through the space-charge electron sheath. Thus, the current caused by the rapid electrons will be saturated and become larger only when the temperature is raised. Of the slow electrons, however, only a part will be able to penetrate the space-charge electron sheath. The current caused by these electrons is not saturated and is determined by the space charge. The current contribution of the different lines was measured by means of a tube in which a collector had been incorporated in front of the fluorescent screen. The above-mentioned saturation and space-charge character of the two electron groups was confirmed by the behaviour of the currents determined in this manner, when the anode voltage was increased. As a result of the existence of these groups, the J-V characteristic of a diode with an oxide-coated cathode ((BaSr)O) must deviate from the characteristic ($J \sim V^{3/2}$), according to Langmuir-Child's formula.

Finally, it should be stated that when our system of measuring was applied to D.C. operation, no abnormal velocity spread [10]) was found.

Summary. Following a short summary of the results of former investigations in the sphere of thermionic emitters under pulsed conditions, a more detailed description is given of investigations in this sphere made by the authors. The potential gradient in the oxide coating of a cathode for pulsed operation was first investigated. This indicated that the layer with large resistance in the interface between nickel and oxide, often encountered by other investigators, was absent in the cathodes investigated by us. It appears, however, that in the case of some oxide-coated cathodes, such as (BaSr)O and SrO, a layer with large resistance occurs at the side nearest the anode, which results in a considerable increase in potential in the outer layer of the oxide coating. This, in turn, causes a large velocity spread of the emitted electrons which often amounts to hundreds of electron volts. By means of a specially constructed tube, the velocity spectrum was examined for different kinds of cathodes under different conditions. The spectrum, curiously enough, appeared to be discrete. The occurrence of the layer with large resistance, which results in the considerable velocity spread, is probably due to two mechanisms causing the conduction through the oxide coating: the pore conduction and the conduction of the grains of the oxide. The relative importance of the two conduction mechanisms will have a decisive influence on whether the layer of large resistance occurs or not. No satisfactory explanation has been given so far for the occurrence of discrete lines in a velocity spectrum. The considerable potential gradient in the outer layer of the oxide may well be one of the causes of the sparking observed with oxide cathodes under pulsed conditions.

[9]) E. A. Coomes, J. appl. Phys. **17**, 446, 1946.

[10]) During D.C. operation abnormal spread was found only in the case of SrO cathodes. For further investigations in this field the reader is referred to the work by G.C. Dalman, Report R 224-49, Polyt. Inst. Brooklyn 169, New York 1949.

NEW PHOSPHORS FOR FLUORESCENT LAMPS

by J. L. OUWELTJES.

621.327.43:546.411.85:535.625

The demand for suitable phosphors for fluorescent lamps has led to widespread research during the last few years. Until recently, it was necessary to mix two or more phosphors in order to obtain white fluorescent light, but the discovery of halophosphate phosphors has made it possible to obtain white light with only one phosphor. This is a great asset in the manufacture of fluorescent lamps, and, moreover, halophosphates have a higher efficiency and are better able to withstand the effects of mercury vapour. If a more faithful rendering of red is desired, magnesium-arsenate, which emits red light, can be mixed with the halophosphate. The spectral distribution of this combination very closely approximates to that of a black body.

Introduction

Tubular fluorescent lamps contain a filling of mercury vapour at low pressure. Under the influence of an electric discharge, this vapour emits radiation which is for the greater part concentrated in several spectral lines in the ultra-violet part of the spectrum, with about 80% in the resonance line at $\lambda = 253.7\ m\mu$. In order to convert this radiation into visible light, the interior of the tube is coated with a fluorescent substance (a phosphor) which emits visible light rays when subjected to the ultra-violet irradiation. The spectral distribution of this light as well as the efficiency with which the electric power is converted into light are wholly dependent on the composition and properties of the phosphor. To obtain maximum efficiency, the phosphor must absorb all the ultra-violet rays, and emit a quantum of visible light for every quantum of ultra-violet absorbed (quantum efficiency 100%).

In order that the lamp shall function as a useful light source, the spectral distribution should as nearly as possible meet a number of requirements, the most important of which are as follows:

1) The light emitted should be white. This means that an object with a coefficient of reflection which, in the visible spectrum, is independent of wavelenght, should appear to the eye as white when illuminated by the lamp, i.e. the visual impression should correspond to that obtained in daylight.
2) Good colour rendering. When an object having a coefficient of reflection that is dependent on the wavelength is illuminated by the lamp, the object must appear to the eye to be of the same colour as when it is irradiated by a black body of the same colour temperature as that of the lamp.

In order to satisfy the requirements concerning efficiency and spectral distribution, it has until recently been common practice to coat the tube with a mixture of phosphors, viz. zinc-beryllium silicate and magnesium tungstate. The spectral distribution of neither of these phosphors even approaches the ideal, but when the two are mixed, the resultant spectral distribution, being the sum of the two single distributions, is well within the limits imposed. The efficiency of the mixture is also satisfactory.

For the lamp manufacturer, however, the necessity of having to use a mixture of two phosphors is in some respects a drawback. It is often difficult to produce the exact colour required, since differences in the grain size and compositon of the constituents, and consequent variations in the efficiency of the two phosphors, are practically unavoidable. These variations lead to displacements in the colour value.

The search for a single phosphor which would emit light with a spectral distribution conforming to the above requirements, was instigated in order to obviate these difficulties. Some years ago, a group of such phosphors, namely the halophosphates were discovered [1]). As a result of this discovery, it is now possible to prepare white-fluorescent single phosphors in great variety, and, although the compositions of these phosphors may vary slightly, the basic constituent is always the same. It is possible to produce with these phosphors the colours of all present-day fluorescent lamps, ranging from "cold" daylight (colour temperature 6500 °K) to the "warm" colours (3000 °K). The demand for a single phosphor capable of emitting white light has thus been amply met.

In addition to the advantages of halophosphates already mentioned, the efficiency of these substances is greater than that of the silicate-tungstate mixture. Furthermore, they are better able to withstand the effects of mercury vapour, and partly because they absorb little of the vapour, this being a drawback

[1]) H. G. Jenkins, A. H. Mc Keag and P. W. Ranby J. Electrochem. Soc. **96**, 1-12, 1949.

N. V. PHILIPS' GLOEILAMPENFABRIEKEN

which renders many other phosphors unsuitable for fluorescent lamps. As a consequence a lamp coated with halophosphate phosphor retains more of its good properties toward the end of its life than does a silicate-tungstate lamp.

It is a remarkable fact that the development which followed the discovery of the halophosphates ultimately produced an entirely new range of lamps coated with phosphor mixtures even more complex than those previously used. This development was founded on the second requirement imposed on the spectral distribution of the light source, viz. good colour-rendering. In those cases where good colour rendering is not of primary importance, e.g. in street- and workshop lighting, the halophosphates are excellent, but if true colour rendering is desired, particularly in the red region, a phosphor with red fluorescence must be added. In this case, too, the new phosphors are, however, still to be recommended, since they give a better colour rendering than silicate-tungstate mixtures.

In this article, a discussion of the properties of the halophosphates is followed by a description of the method employed to improve the colour rendering.

Fluorescence with ultra-violet irradiation

As already mentioned, a mixture of zinc-beryllium silicate (activated with manganese) and magnesium-tungstate has hitherto been used in most fluorescent lamps. The colour of the fluorescence of silicate

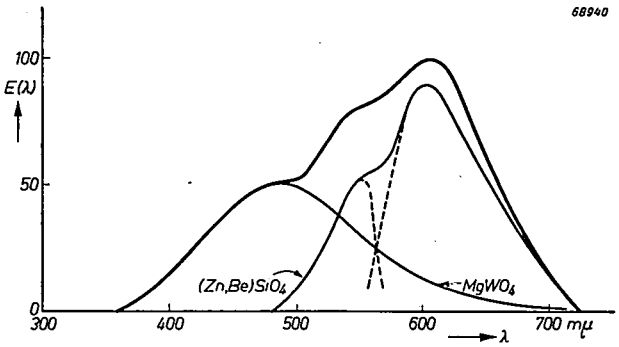


Fig. 1. Spectral energy distribution of $(Zn,Be)_2SiO_4$, activated with Mn (yellow fluorescence), and that of $MgWO_4$ (blue fluorescence) (thin lines). The curve relating to $MgWO_4$ is flat-topped at $\lambda = 480$ mμ (in the blue-green region), whereas the other curve has its major peak at $\lambda = 610$ mμ (orange) and a smaller peak at $\lambda = 560$ mμ. The emission of the silicate actually consists of two emission bands (dotted lines). The heavy line represents the total energy distribution of the mixture generally used in practice, comprising two individual components. The energy is plotted on a relative scale.

irradiated with ultra-violet light (mercury line $\lambda = 253.7$ mμ) is yellow, that of the tungstate blue. *Fig. 1* shows the spectral energy distribution of the individual phosphors, together with that of the mixture normally used.

Since the mechanism of fluorescence has already been fully discussed in this journal [2]), a brief summary will now suffice. When a substance such as zinc-beryllium silicate containing a small quantity of manganese is irradiated by ultra-violet rays of the wavelength $\lambda = 253.7$ mμ, much of this radiation is absorbed: due to the incorporation of manganese atoms, the crystal lattice absorbs the energy of many radiation quanta. Part of this energy is converted into lattice vibration energy, that is, into heat; in the instance under discussion, however, by far the greater part of the energy is again transferred from the crystal lattice to the manganese atoms, which are thus excited. This state of excitation lasts only a very short time. After this period the atom emits rays of visible wavelength, mainly in the region of $\lambda = 600$ mμ, i.e. in the orange region. This fluorescence radiation is therefore brought about entirely by the manganese atoms in the crystal lattice, for which reason these atoms are known as activating atoms, or activators.

The fluorescence of magnesium tungstate takes place without the presence of any activators; the WO_4 ions function as selective absorption centres which are excited when rays of a given wavelength reach the crystal. When these ions return to their original level, which they do in stages, a radiation quantum of greater wavelength, mainly in the blue-green region, is emitted.

We shall confine ourselves to this simple description of the mechanism of fluorescence, without touching upon the complications which frequently occur.

Generally speaking, the fluorescence of phosphors, apart from that of the comparatively rare substances which require no activator, is governed mainly by the nature and concentration of the activating atoms, and by the manner in which these atoms are incorporated in the crystal lattice.

Let us consider the fluorescence of zinc-beryllium silicate more closely: it appears that there are actually two "emission bands" (see dotted lines in fig. 1) which together form a single, less sharply defined band with a hump on one side. It is therefore assumed that there are two categories of activating atoms which differ, for example, in the manner in which they are surrounded by the other atoms.

With this in view, an attempt might be made to evolve a phosphor which would emit white light without the addition of any other phosphors, by introducing activating atoms into a substance in a

[2]) See, for example, F. A. Kröger and W. de Groot, The influence of temperature on the fluorescence of solids, Philips Techn. Rev. **12**, 6-14, 1950/1951 (No. 1).

way such that the various emission bands will combine to produce white light. This is not actually practicable, however, but it is possible to obtain the same effect by introducing two (or more) different activators into a single parent crystal, one activator giving, for example, yellow fluorescence and the other blue.

The halophosphates

Halophosphates constitute a group of substances which permit of the interposition of two activators. The crystal lattice and composition of halophosphates are closely related to those of the natural mineral apatite. The chemical formula of pure fluor-apatite is $3Ca_3(PO_4)_2.CaF_2$. From a stoechiometrical point of view, the molecule is, therefore, built up of 3 calcium phosphate molecules and 1 calcium fluoride molecule. The fluorine is sometimes completely or partly replaced by another halogen, say chlorine, whilst the calcium atoms can be replaced, also wholly or in part, by other bivalent metals (e.g. cadmium, iron, magnesium). It is even possible to incorporate an OH^--group, a CO_3^{2-}-group or an SO_4^{2-}-group instead of halogen ions, whilst trivalent rare earths may be substituted for the metal atoms.

In halophosphate phosphors producing white light, a small proportion of the calcium ions has been replaced by trivalent antimony ions and bivalent manganese ions, both of which are activators, the antimony ions giving fluorescence in the blue emission band and the manganese producing a yellow band. Some of the fluorine ions are replaced by chlorine ions.

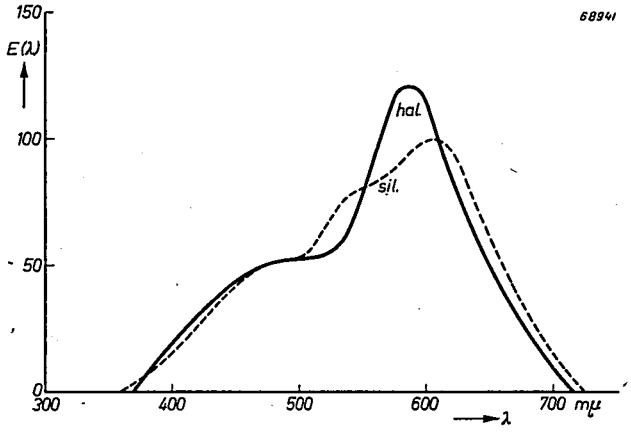


Fig. 2. Spectral energy distribution of a halophosphate phosphor activated with antimony and manganese. The blue antimony band with a peak at $\lambda = 475$ mμ is readily distinguished from the yellow, manganese band with a maximum at $\lambda = 590$ mμ. For comparison, the spectral energy distribution of a silicate-tungstate mixture (fig. 1) is shown (broken line). The halophosphate produces more yellow light, but less green and red. Energies are again plotted on a relative scale; the areas enclosed by the two curves are equal.

Fig. 2 shows the spectral energy distribution of a certain halophosphate phosphor (heavy line). The blue emission band, with a peak at $\lambda = 475$ mμ, is clearly distinguishable from the yellow emission band which peaks at $\lambda = 590$ mμ. The actual location of each band and the height of the peaks appear to be dependent on the composition of the phosphor. The blue (antimony) band shows a maximum at $\lambda = 475$ mμ, the location of this maximum being almost independent of the composition, whereas that of the yellow (manganese) band is governed mainly by the ratio of chlorine-to-fluorine in the crystal: the greater the quantity of fluorine replaced by the chlorine, the longer the wavelength at which the maximum occurs. Moreover, the quantity of manganese introduced to obtain emission also affects the position of this band.

It is assumed that both antimony and manganese ions replace those Ca^{2+} ions which are in the vicinity of halogen ions. When an Sb^{3+} ion is incorporated, the adjacent halogen ion is replaced by an O^{2-} ion in order to maintain electric neutrality. The Sb ions are not, therefore, in direct contact with the halogen ions, and the substitution of chlorine for fluorine will have little effect on the excitation of the antimony. On the other hand, bivalent manganese ions are in the vicinity of halogen ions, and the emissivity of the manganese is definitely affected by the substitution of fluorine by chlorine.

The intensity ratio of the emission is determined mainly by the quantity of manganese present in the phosphor. It is possibly surprising to note that the quantity of antimony introduced has no effect on this ratio (although it does, naturally, affect the absolute heights of the maxima). To explain this, it must be assumed that the excitation of manganese ions is not the direct result of ultra-violet irradiation, but the indirect result of the transfer of part of the excitation energy from antimony ions to manganese ions. This is usually expressed by stating that the antimony ions are not only activators, but also sensitizers for emission by the manganese. The fraction of excitation energy transferred from Sb ions to Mn ions is dependent only on the number of Mn ions available, i.e. on the manganese concentration. This, then determines the intensity ratio of the emission bands.

The hypothesis relating to the transfer of energy is confirmed by the following: A halophosphate to which only antimony is added gives blue antimony fluorescence when subject to radiation of wavelength 253.7 mμ, but a halophosphate containing only manganese emits no light at all when thus irradiated. The manner in which energy is transferred from Sb to Mn cannot be dealt with here.

Colour and colour rendering of halophosphate phosphors

According to the foregoing, the white fluorescence of a halophosphate activated with antimony and manganese is a result of the superimposition of the blue fluorescence of antimony on the yellow of the manganese. The colour of the resultant light can be controlled to some degree by varying the quantity of manganese added (since this determines the position of the yellow emission band, and the intensity ratio of this band to the other) and, to a lesser degree, by varying the chlorine-to-fluorine ratio: by increasing the chlorine content the wavelength of the yellow emission band is increased and in consequence the total spectral distribution is displaced towards the red zone. The result of the two effects can best be illustrated with reference to the colour triangle [3]).

Any change in the composition of the phosphor will produce a change in the colour of the light emitted, which means that the relevant colour point in the colour triangle is displaced. This is illustrated in *fig. 3*, where the curved line *BGR* represents the spectral colours, the straight line *BR* being the so-called purple line connecting the two colour points extreme red and extreme blue. Curve *b* is the "black-body" line, indicating the position of the colour point of the black body as a function of the absolute temperature (a few of the values are shown). The colour point of daylight corresponds roughly to that of a black body at 6500 °K.

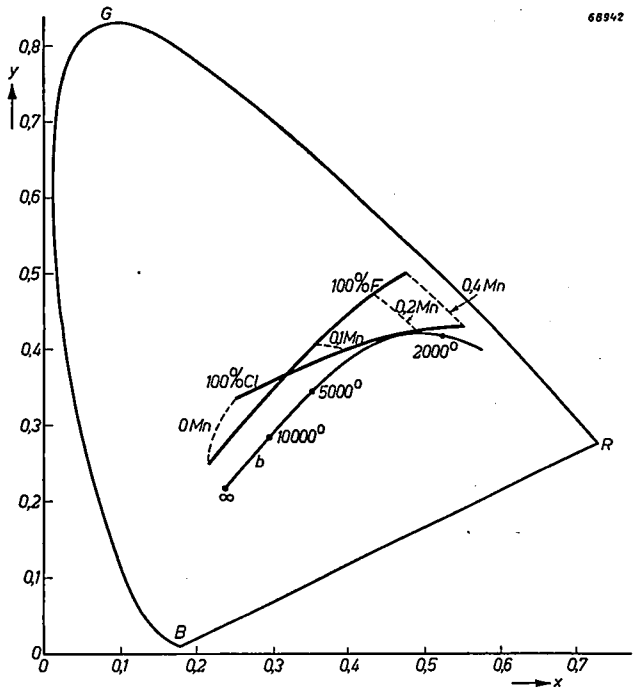


Fig. 3. Colour triangle on *x-y* co-ordinates. The curve *BGR* represents the spectral colours; *BR* is the purple line, whilst line *b* connects the colour points of black-body radiation at various absolute temperatures. A few of the temperatures are indicated (daylight corresponds roughly to black-body radiation of 6500 °K). The thick lines show the displacement of the colour point of a halophosphate phosphor when the Mn content of the phosphor, which contains a given quantity of fluorine-chlorine, is varied. The manganese content is indicated by the number of Mn atoms per calcium-halophosphate molecule specified on the broken lines. The lines representing constant halogen ratios are almost parallel to the black-body line. (Derived, in part, from K. H. Butler and C. W. Jerome, J. Electrochem. Soc. **97**, 265-270, 1950 (No. 9).

Any change in the Mn-content of the phosphor has the effect of displacing the colour point of the fluorescent light along a line almost parallel to the black-body line (see thick lines in fig. 3). At the same time, the actual "path" of the colour point is also slightly dependent on the fluorine-chlorine content. With a given Mn-content, the colour point almost describes a straight line when the fluorine-chlorine content is varied, the angle of the path being governed once again by the Mn-content (see broken lines).

Although the position of the colour point is sufficient indication of the colour of the light, the colour rendering cannot be assessed without a more or less comprehensive survey of the spectral distribution of the emission. Let us compare the colour rendering of a new halophosphate lamp with that of the old type silicate-tungstate lamp. The *table* on the next page contains details of the spectral distribution of a "white" silicate-tungstate lamp the colour temperature of which is 4000 °K (line 1) and that of a "white" halophosphate lamp with a colour temperature of 4200 °K (line 2). The halophosphate lamp is designed to replace the silicate-tungstate lamp; the difference in colour temperature is the outcome of international standardization. Also included in the table are data relating to black-body radiation (line 4) with a colour temperature of 4000 °K, and particulars of another new lamp which will be discussed later. Those spectral regions have been chosen [4]) which will give a representative impression of the colour rendering; the components of the luminous flux are shown as relative quantities (percentages), the total quantity of light emitted by each lamp over the entire spectrum being taken as 100.

According to the data in the table, the agreement between halophosphate and the black body is not so good as that between silicate-tungstate and the black body. This in understood on comparing the

[3]) See W. de Groot and A. A. Kruithof, The colour triangle, Philips Techn. Rev. **12**, 137-144, 1950/1951 (No. 5).

[4]) See P. J. Bouma, The colour reproduction in the use of different sources of "white" light, Philips Techn. Rev. **2**, 1-7, 1937.

Percentage of luminous flux contributions in 8 spectral regions of a lamp with (1) white-fluorescent silicate-tungstate coating: white silicate lamp, colour temperature 4000 °K, (2) white-fluorescent halophosphate coating: "standard" type, colour temperature 4200 °K, (3) white-fluorescent mixture comprising halophosphate and magnesium-arsenate (plus a small quantity of willemite): "de Luxe" lamp, colour temperature 4200 °K. The values on line 4 represent the spectral distribution of a black-body at a colour temperature of 4000 °K. The colour matching of any two lamps is practically identical when the luminous flux of the one lamp does not differ from that of the other by more than 25% in any one of the eight columns.

Wavelength range (mμ)	380-420	420-440	440-460	460-510	510-560	560-610	610-660	660-760
Light source	violet	blue-violet	blue	green-blue	green	yellow	orange	red
1) White silicate lamp	0.010	0.35	0.30	4.80	38.0	43.0	13.0	0.43
2) "Standard" lamp	0.012	0.335	0.405	6.00	35.5	49.0	8.4	0.20
3) White "de Luxe" lamp	0.010	0.32	0.29	4.93	40.1	41.4	12.3	0.65
4) Black-body 4000 °K	0.014	0.13	0.47	7.89	38.0	39.5	12.9	1.05

spectral energy distributions of the two phosphors. To facilitate this comparison, the spectral energy distribution curve of the silicate-tungstate phosphor in lamp 1 is shown in Fig. 2 (broken line). In the yellow region, the halophosphate curve has a higher peak than the curve of the mixture, but, on the other hand, the phosphate emission contains less green and, more particularly, less red. On considering lines 2 and 4 more closely, we see that the difference in the spectral distribution of the halophosphate and the black body is manifested mainly in the last two columns, i.e. in the red. The differences in the first four columns (blue) are due chiefly to the previously mentioned difference in colour temperature, and are therefore unimportant.

It follows from the above, then, that the halophosphates give insufficient red light, the most noticeable result of this being that red and pink objects (the colour of the skin) appear in an unnatural brownish, deathly tinge.

This can be remedied to some extent by reducing the fluorine : chlorine ratio (see fig. 3), as the Mn content exceeds one-tenth per molecule. The Mn-content cannot be increased, seeing that this would affect the colour temperature, The increase in the chlorine content results in a displacement of the colour point towards the purple line. Even so, red objects would not appear entirely natural. This is not objectionable for many practical applications.

This procedure was adopted to produce the "standard" type of lamp intended for industrial purposes, street lighting, etc. Compared with a silicate-tungstate lamp, the new lamp has three distinct advantages: namely only one phosphor is required in manufacture, the efficiency is higher, and the characteristics of the lamp towards the end of its effective life show a distinct improvement.

Improvement of colour rendering: the "de Luxe" lamp

In all cases where good colour rendering is of the first importance, the "standard" lamp is unsuitable. This difficulty could be overcome by reverting to silicate lamps, but in view of the lower efficiency and poorer maintenance of these lamps, it is worth while to seek to improve the colour matching properties of the halophosphate lamp. Theoretically, this can always be achieved by adding a red fluorescent substance to the white, although the principal advantage of the halophosphate lamp is then of course lost. The efficiency, too, decreases slightly when a red phosphor is added to a halophosphate: this is unavoidable since the eye is less sensitive to red than to yellow or green. On the other hand, if a red phosphor with a quantum efficiency approaching 100% and suitable spectral distribution were available, this phosphor could be mixed with halophosphate to produce a lamp having much better matching properties than a silicate-tungstate lamp with almost the same efficiency. Such a phosphor, in the form of magnesium arsenate, was actually discovered; this is a fluorescent substance having a deep red emission and the highest quantum efficiency of all known red fluorescent substances (85%). Magnesium-arsenate has already been mentioned in a previous article [5]) (in this article the quantum efficiency is given as 75%, but recent samples showed an efficiency 10% higher).

Apart from the high quantum efficiency of magnesium-arsenate, there are two other factors to which the high efficiency of lamps using this phosphor is attributable, viz:

[5]) J. L. Ouweltjes, W. Elenbaas and K. R. Labberté, A new high-pressure mercury lamp with fluorescent bulb, Philips Techn. Rev. **13**, 109-118, 1951/1952 (No. 5).

a) This phosphor responds to the entire ultraviolet spectrum (and part of the blue spectrum) of the low-pressure mercury discharge. This spectrum consists not only of the resonance line at $\lambda = 253.7$ mμ, but also lines at $\lambda = 313$ mμ, $\lambda = 365$ mμ, etc., with a weak, continuous background. Halophosphates, as well as silicates and tungstates, are excited by the resonance line only. When magnesium arsenate is added, some of the previously unused rays are utilized and, moreover, part of the visible blue light emitted by the mercury vapour, mainly in the spectral line $\lambda = 436$ mμ, is absorbed by the arsenate, resulting in a slightly better rendering of blues and yellows.

b) Magnesium-arsenate has a very sharply defined emission band (see fig. 5 in the article quoted in footnote [5])); in contrast with other phosphors emitting red light, magnesium-arsenate emits no infrared.

Halophosphate fluorescence is lacking not only in red light, but also, to a lesser extent, in green. For this reason a small quantity of willemite (zinc silicate) emitting green radiation is added. In this way, the colour matching is improved without displacement of the colour point.

This development has led to an entirely new type of lamp, namely the "de Luxe" lamp. Although the advantage of a single phosphor is no longer applicable (a mixture comprising three phosphors is used), the colour matching properties compare very favourably with those of the silicate-tungstate lamp. The spectral distribution of the white "de Luxe" lamp with a colour temperature of 4200 °K ("warm-white" and "daylight" lamps are also available) is given in the table (line 3); the values correspond very closely to those of a black body with radiations of roughly the same colour temperature (line 4). This is also apparent in fig. 4, in which the spectral energy distribution of the lamp and that of the black body are reproduced.

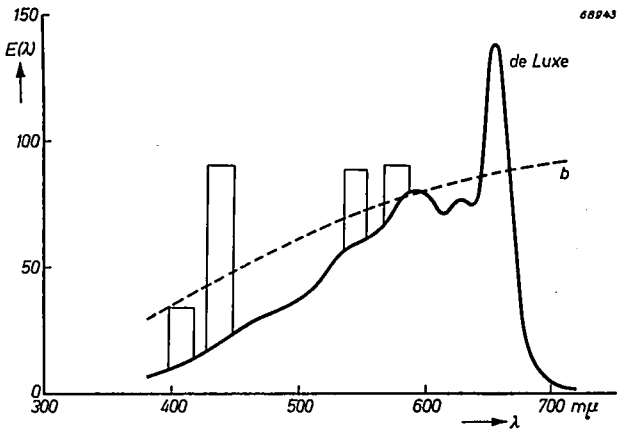


Fig. 4. Relative spectral energy distribution of the "white" "de Luxe" lamp with a colour temperature of 4200 °K (fully drawn line) compared with that of a black body with a colour temperature of 4000 °K (broken line). Superimposed on the fluorescent light are a few spectral lines emitted by the mercury vapour.

Summary. A light source for general lighting purposes must meet certain requirements as regards colour and spectral distribution of the light emitted (colour rendering). Until recently, low-pressure fluorescent lamps have fulfilled these requirements if coated with a mixture of two phosphors, zinc-beryllium silicate and magnesium-tungstate. To avoid the difficulties entailed by the use of such a mixture in the manufacture of fluorescent lamps, a phosphor was sought which would by itself emit white light. Phosphors having this property were discovered in the halophosphate group; these phosphors contain two kinds of activating atoms; the one, antimony, emits blue light and the other, manganese, yellow. The spectral distribution and the colour triangle are used as basis for determining the colour and matching properties of the phosphors, respectively. It appears that some control can be effected by varying the composition of the phosphor. The efficiency of fluorescent lamps with halophosphate phosphors is greater than that of lamps coated with the silicate tungstate mixture. In those cases where good colour matching is essential, also in the red region, red-luminescing magnesium arsenate and a small quantity of willemite are mixed with the phosphor. Because of the excellent properties of magnesium arsenate, the efficiency is reduced only slightly by the inclusion of this substance, thus producing a lamp (the new "de Luxe" lamp) which ensures high efficiency and excellent colour matching properties.

AN INDENTATION METER FOR PAINT

by J. HOEKSTRA and J. A. W. van LAAR.

667.6:620.178.152:621.317.39

In all industries in which paints are made or used, the need has long been felt for a means of measuring those characteristics of paints which determine their quality, by a method capable of being accurately reproduced. One of the most important properties of a coat of paint is its resistance to the penetration of hard objects. A method hitherto all too frequently employed to determine this resistance consists in scratching the surface with the finger nail — a "test" which, it is true, yields more information to experienced persons than might be supposed, but which cannot possibly provide a basis for scientific research.

The authors of this article have substituted for the finger nail a sapphire which is specially shaped to give a reproducible impression when applied to a coat of paint. The depth of this impression is measured by an electrical zero method.

The greater the resistance of a coat of paint to damage in the form of indentations, scratches, etc., in other words, the "harder" the paint, the longer it is able to fulfil its task of providing a decorative or protective coating. In order objectively to determine the composition, method of drying, etc. that is most suitable for a given purpose, it must be possible to measure the resistance of the paint to the penetration of hard objects, i.e. the "hardness" of the paint. The lack of a reproducible method of determining this quality more often than not leads only to fruitless argument.

To measure the hardness of a metal, the Brinell, Rockwell or Vickers [1]) method is usually employed. In the Brinell and Vickers tests respectively round and tapered rams of great hardness and standardised dimensions are pressed to the metal with a certain force. When the ram is withdrawn, the superficial area of the indentation thus produced is measured, and the applied force divided by the area then represents the hardness of the metal. It is true that this definition of the hardness is not directly related to purely physical constants, but the tests are fairly easily reproducible and are therefore of great importance from the technological point of view; the quantity thus measured may be termed a "technoligical quantity".

When applied to coats of paint the above methods suffer from the objection that the impression is very often not clearly defined, so that the area of the indentation cannot be measured with any accuracy. Furthermore, when the load is removed, the impression contracts, firstly owing to the not inconsiderable recovery of the elastic part of the deformation and, secondly, as a result of elasto-plastic flowing of the material ("retarded elasticity"), by reason of which the impression shrinks still further, according to the time lapse. Moreover, experiments have proved that in contrast with most metals the area of the indentation in a coat of paint is not proportional to the applied load, so that there is no object in specifying a certain "hardness" value without stating the load.

Principle of the instrument

The method employed by the authors has eliminated the first two objections mentioned above. As in the Vickers test, a hard, pointed ram is pressed against the surface to be tested with suitable force. This ram takes the form of a regular square pyramid having an apex angle of 136°. In this case, however, instead of measuring the dimensions of the resultant impression *in* the plane of the surface *after* removal of the load, the *depth* of the impression is measured *while* the force is *still* being applied, thus eliminating the problem of the vague outline. By again measuring the depth of the impression a certain time after the load is removed it is possible to ascertain roughly the relationship between the plastic and elastic parts of the deformation.

In view of the absence of a proportional relationship between the area (or square of the depth) of the indentation in a coat of paint and the applied force, we give the preference not to computing the "hardness" as such, but to state the depth of the impression (together with the applied force).

As with hardness tests on metals, the depth of the impression is here again in some degree dependent on the time during which the force is actually operative. This is not a drawback from the point of view of the reproducibility of the test, provided that the force is always applied during the same

[1]) See, for example, Philips Techn. Rev. **2**, 177-181, 1937.

period of time, or at any rate that this duration is indicated in the result. Thirty seconds has been found suitable for tests on most paints, varnishes, lacquers, etc.

The force applied to the ram should always be such that the depth of the impression will be small compared with the thickness of the coat of paint. If the point of the sapphire is permitted to penetrate too close to the surface to which the paint is applied, it will reach a zone in which this surface noticeably hinders the yielding of the paint; the value measured then represents an "apparent" hardness which is considerably higher than the much more important value related to the outer surface. Since most coats of paint are not more than 25 to 100 μ thick, the penetration depth must be limited to a few microns, and the measurement must be accurate to within 0.1 μ.

The procedure is as follows: The paint is applied to a solid surface (specimen plate); after drying, this plate is clamped to an adjustable platform on the instrument. Before the sapphire is brought into contact with the paint, the control knob of the bridge is set to its centre position and the bridge is roughly balanced by raising or lowering the centre plate of the differential capacitor by means of a set-screw until it is central between the two fixed plates (fig. 1*b*). Exact balance of the bridge is then obtained by carefully acting on the control knob of the bridge, thus compensating slight deviations of the centre plate of the differential capacitor.

The platform holding the specimen plate is now carefully lowered over the sapphire by turning a calibrated micrometer screw until the surface of the paint just touches the tip of the sapphire (fig. 1*c*), this being shown by the indicator which

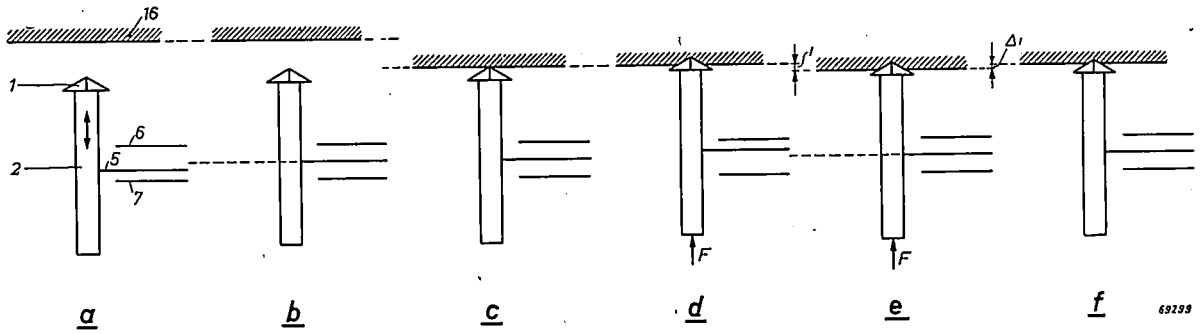


Fig. 1. Diagram showing the principle of the indentation meter: a sapphire *1* is mounted on a ram *2* which is capable of vertical movement and which carries the centre plate *5* of a differential capacitor, *6* and *7* being the fixed plates. The specimen plate with paint *16* can be raised or lowered by means of a micrometer screw (not shown). (The numbers in this diagram correspond to those shown in the other figures.)

a) Starting point (arbitrary position).
b) Plate *5* adjusted centrally between the fixed plates, *6* and *7*.
c) Specimen lowered until it just touches the tip of the sapphire.
d) Force *F* (e.g. 10 g) applied to the ram presses the tip of the sapphire into the paint.
e) Specimen quickly lowered until plate *5* is again central between plates *6* and *7*, cancelling out the upward movement of the ram and restoring the load exactly to 10 g. This downward pressure is maintained for 30 seconds, care being taken that the bridge remains exactly balanced. The difference between the micrometer reading at stage (*c*) and that at the end of the 30-second period then represents the penetration depth *i*.
f) The variation Δi in the penetration depth can be determined when the force *F* is withdrawn and balance is restored by adjusting the micrometer screw.

With the instrument described in this article the penetration depth is measured by an electrical zero method. Every movement of the ram is communicated to the centre plate of a differential capacitor of which the other two electrodes are fixed plates (*fig. 1a*). The two halves of this capacitor, together with two identical resistors, form a bridge; an alternating voltage is applied across one diagonal and the voltage across the other is amplified and applied to an indicator. The bridge is in balance when the centre plate of the capacitor is central between the two fixed plates, i.e. when the tip of the sapphire is at a certain level.

responds immediately as the balance of the bridge is disturbed.

At this stage the reading of the micrometer screw is taken and a certain force, say 10 g, is applied to the ram (fig. 1*d*) — the manner in which this load is applied is described later. The tip of the sapphire is thus forced into the coat of paint. The load is applied for 30 seconds, during which period the bridge is kept in balance by constant adjustment of the micrometer screw, so that the sample is lowered accordingly (fig. 1*e*). During this 30-second period the depth of penetration of the sapphire into the paint will increase to some ex-

tent, but the position of the ram with respect to the fixed parts of the instrument remains unchanged. At the end of this period, a second reading is taken from the scale on the micrometer screw, the difference between this reading and the first being the depth i to which the sapphire has penetrated into the paint.

If so desired, the load can be removed immediately after the second reading has been taken. Then, after a definite time, the bridge is again balanced by readjusting the micrometer screw (fig. 1*f*), and another reading is taken. The difference Δi between the second and third readings gives us the residual, irrevocable penetration depth which, as already mentioned, furnishes an indication of the relationship between the elastic and plastic properties of the specimen.

Construction

The first idea for this penetration meter dates back to 1942, but since then the original design has been improved in many ways. Disregarding the now obsolete models which have been reported else where [2] [3], we shall now describe only the most recent model (*fig. 2*) which is in use in many departments of the Philips factories [4].

Fig. 3 depicts a vertical cross section of the instrument: *1* is the sapphire for producing the indentation. In the Vickers test a diamond is used, but for paints a sapphire ground to the shape of the Vickers diamond (a square pyramid having an apex angle of 136°) is sufficiently hard. This sapphire is mounted in a metal sleeve at the upper end of the ram, which is held in place by two flat springs. Also attached to this ram, but insulated from it, is the movable plate of the capacitor, measuring approximately 4 cm × 5 cm. Since the gap between the movable plate and each of the fixed plates is only about 50 μ, it is essential that the fixed plates be held securely in place and they are therefore supported by sturdy porcelain rods, cemented in place. As the ram is held in position by two flat springs, movement of the centre plate is restricted to parallel displacement.

Each of the three plates of the capacitor is wired to a plug pin for connection to a "Philoscope" bridge, type GM 4144 [5] which contains all the other components required to form a complete bridge circuit, viz. the two resistance arms, an amplifier, an electronic indicator ("magic eye") and a power unit.

A weak spring in the form of a thin, steel wire (*13* in fig. 3) is secured to the ram holding the sapphire, and, by adjusting a set-screw *14*, which controls this spring, the movable plate of the capacitor can be brought roughly midway between the fixed plates (fig. 1*b*). The exact balancing of the bridge is easily done by electrical adjustment. The bridge must be balanced before every measurement, but as a rule readjustment of the bridge resistors suffices.

The sapphire passes through a hole in a platform *15*, to which the specimen is clamped, painted side down. This platform can be raised or lowered by means of a micrometer screw *18*, which, to ensure maximum accuracy, actuates the platform, not direct, but through reducing gear of 1 in 50.

This mechanism is depicted in *fig. 4*. One end of the platform *15* (the right-hand end in figs 3 and 4) is fixed to the framework *20* of the instrument by means of a flat spring *19*. A beam *21* is also mounted on the framework, by means of a flat spring *22*, and the end of an adjustable screw *23*, protruding from the platform, rests on this support. It will be seen from fig. 4 that any displacement of the micrometer screw, reduced by a factor of $(l_2/l_1) \times (l_4/l_3)$, is transmitted to the platform with respect to the sapphire. The lengths l_1, l_2, l_3 and l_4 are such that this factor is exactly 50.

Screw *23* serves to lower the platform until the top is about

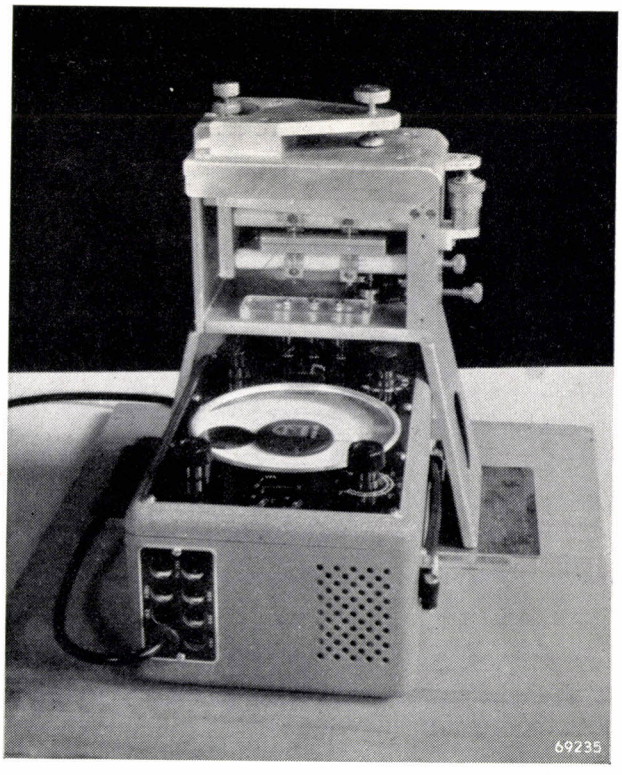


Fig. 2. Indentation meter with "Philoscope" in the foreground.

[2] J. Hoekstra, The hardness of paints (in Dutch), Verfkroniek **19**, 88-90, 1946.

[3] J. Hoekstra and H. A. W. Nijveld, The determination of the hardness of organic films, Rec. Trav. chim. Pays-Bas **67**, 685-689, 1948.

[4] Credit for some of the constructional details is due to the Central Workshop of the Research Lab. Eindhoven, and, more particularly, to Mr. J. M. van Ostade.

[5] An older type (GM 4140) is described in Philips Techn. Rev. **2**, 270-275, 1937.

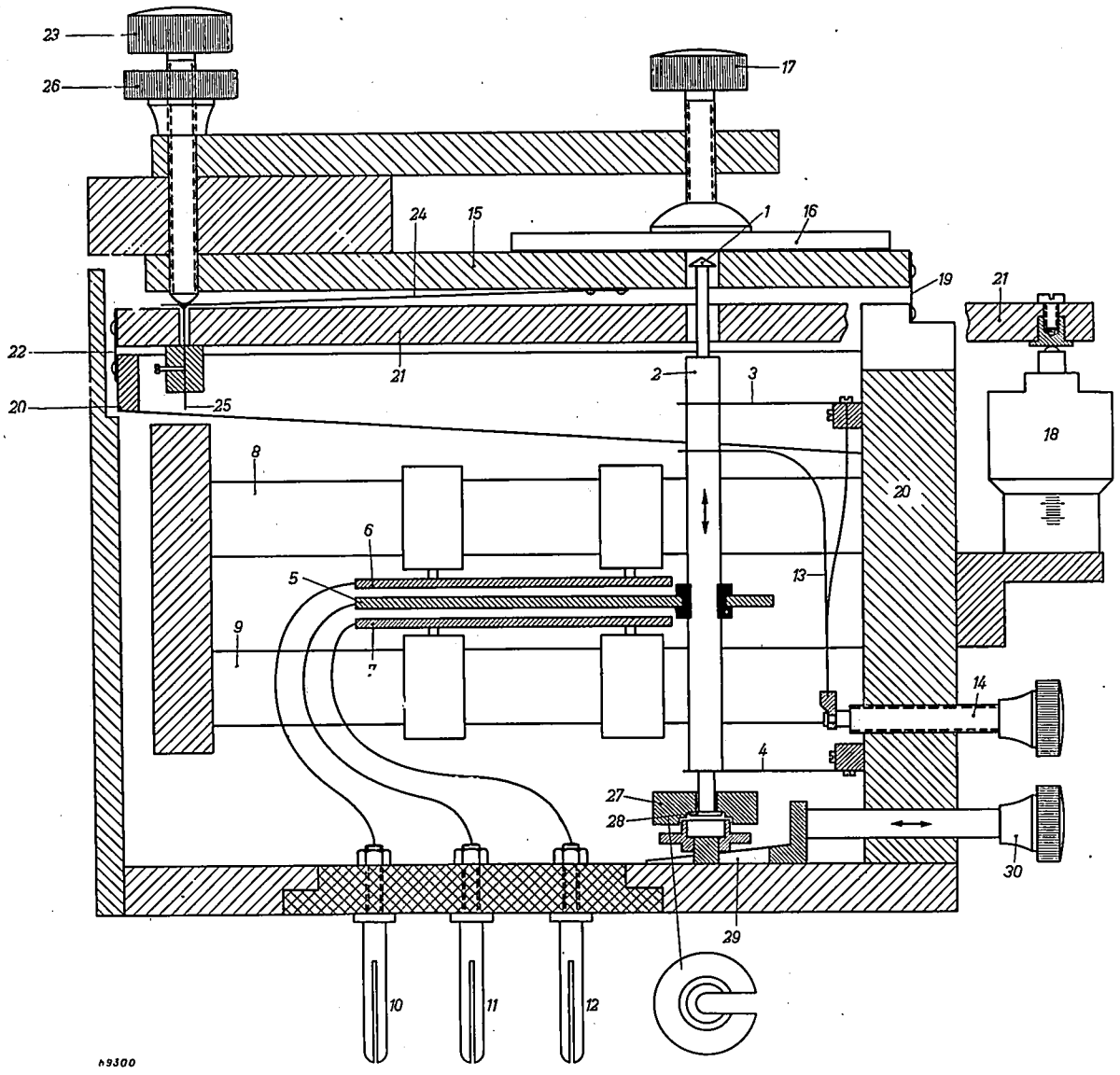


Fig. 3. Vertical cross section of the indentation meter (not to exact scale). The ram *2* with the sapphire *1* is supported by flat springs *3* and *4*; *5* is the movable plate of a differential capacitor, *6* and *7* being the fixed plates, supported by porcelain rods *8* and *9*. Plug pins *10*, *11* and *12* are for connection to the "Philoscope" bridge; *13* is a steel wire spring controlled by screw *14*. The specimen, *16*, is fixed to the platform *15* by means of screw *17*; *18* is the micrometer screw which, in conjunction with the beam *21* and adjustable screw *23*, controls and indicates the height of the platform *15*; *19*, *22* and *24* are flat springs. Spring *24* is attached to beam *21* by a short piece of steel wire *25*; *26* is a lock nut and *27* a slotted ring-shaped weight of 10, 5, $2^1/_2$ or 1 g. This weight rests on a shoulder *38* of the ram *2* and is lifted by a wedge *29* which is operated by means of knob *30*. When this knob is pressed, and the weight lifted, springs *3*, *4* and *13* can force the ram and sapphire upwards.

level with the tip of the sapphire, the micrometer screw *18* being set to a fairly high position. Once this preliminary adjustment has been effected, screw *23* is locked by means of the lock nut *26*. No re-adjustment is then neccesary unless the sapphire is replaced.

A flat spring *24*, situated between screw *23* and the platform *15*, is attached at one end to the platform, the other end being fastened to the beam *21* by the steel wire *25*. This ensures that any movement of the platform with respect to the beam will be free from play and also that the point of the screw *23* does not come in direct contact with the beam (friction between the point of the screw and the beam would cause the platform to move sideways when the screw is turned).

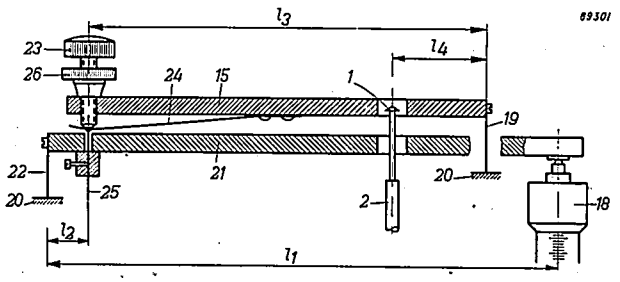


Fig. 4. Transmission of vertical displacement of micrometer screw *18* to platform *15* with a reduction of 50:1 (with respect to the sapphire *1*). Numbers correspond to those in fig. 3. l_1, l_2, l_3 and l_4 are lever dimensions.

The scale on the micrometer screw gives a reading which is accurate to within ½ scale division, or 0.05 μ.

When the specimen is clamped to the platform and the bridge is balanced with the sapphire not yet touching the paint, the next step is to lower the specimen until the paint just touches the sapphire (as in fig. 1*b-c*). This is done by adjusting the micrometer screw, thereby lowering the platform. Contact can be said to be established at the first signs of unbalance of the bridge: sensitivity of the instrument is such that a load of only 3 mg on the sapphire is sufficient to give a visible deflection of the electronic indicator. A load equal to twice this value gives a pronounced indication. The initial load applied to the paint by the sapphire is therefore small compared with the load of 1 to 10 g with which the indentation is effected. Moreover, the initial load is applied only momentarily.

Fig. 3 shows the mechanism through which the initial load is applied. The ram *2* which carries the sapphire and is forced upwards by the flat springs *3* and *4* and, eventually by the steel wire spring *13*, has a ring *27* of which the weight is, say, exactly 10 g. When the knob *30* is pressed, wedge *29* lifts this ring, thus taking its weight off the shoulder of the ram. Instead of the force exerted by this weight of 10 g, an equivalent reactive force is exerted by the film of paint to the sapphire. If the resultant impression in the paint is too deep, the ring weighing 10 g may be replaced by one weighing 5, 2½ or 1 g. At present, all measurements are carried out with a weight of 2.5 g. The indentation is then still small enough to be independent of the thickness of the coat of paint, but sufficiently large to be measured with sufficient accuracy.

As the rings are slotted to fit the ram, they are easily changed. In that case the position of the capacitor plate *5* must be readjusted by means of set-screw *14*.

If the specimen were left in the same position after the sapphire has penetrate the surface of the paint, the deflection of the flat springs (*3* and *4* in fig. 3) would decrease, i.e. they would lose some of their tension, with a corresponding reduction in the force exerted by them. It is desirable, however, to maintain a constant load for 30 seconds, and, accordingly, as soon as the weight *27* is lifted, the platform supporting the specimen is lowered slightly by means of the micrometer screw *18*, the adjustment being continued for 30 seconds so that the bridge remains balanced, i.e. that the spring tension remains constant.

As already mentioned, the minimum force with which the sapphire first comes in contact with the paint is 3 mg. This low pressure is ensured by using very flexible material for the flat springs *3* and *4*. The initial load of 3 mg corresponds (before contact is established) to a displacement of 0.03 μ of the movable plate of the capacitor, this being the minimum displacement to which the electronic indicator will respond when the bridge is balanced. In spite of this high sensitivity, no special steps need normally be taken to protect the instrument from vibration, this being probably due to the damping effect of the air in the confined spaces between the capacitor plates.

The over-all accuracy of the instrument is, however, determined not by the electrical but by the mechanical part of the instrument, i.e. the accuracy with which the adjustment of the micrometer screw can be reproduced, in this case to within 0.1 μ. This very slight amount of play is probably attributable not to the suspension springs, but to the thread on the screw and the point of contact between the screw and the beam.

By using an aluminium alloy for the framework and some of the components, a light but strong construction is secured; owing also to its small dimensions, the unit is easily transportable. Vulnerable components such as the differential capacitor are so housed as to be inaccessible unless the instrument is opened, and they are thus protected against damage and dust under normal working conditions.

The use of the instrument does not require much routine.

Results of some measurements

Measured values of the penetration depth i, taken from some of the most common types of paint, are given in the following *table* for the guidance of the reader. The loads applied by the sapphire were 2.5 g and 10 g. Notwithstanding the objections to the quantitative implication in the word "hardness" as applied to paints, the Vickers hardness H (= ratio of the load F to the area of the impression) is also specified in this table [6]), merely

---

[6]) The lateral area A of a square-based pyramid with height i and apex angle φ is

$$A = i^2 \cdot \frac{4 \tan \frac{1}{2}\varphi}{\cos \frac{1}{2}\varphi}.$$

For $\varphi = 136°$, this is

$$A = 26.4\ i^2.$$

Therefore, the Vickers hardness $H = F/A$ is

$$H = \frac{38\ F}{i^2}\ \text{kg/mm}^2,$$

with F in grammes and i in microns.

to give an impression of the order of magnitude.

For comparison it should be noted that the Vickers hardness of a diamond is 2600, that of a sapphire 1850, of hardened steel 200-950, of non-hardened steel 100-200, of copper 60, of lead and tin 5, all values being in kg/mm².

Penetration depths $i_{2.5}$ and i_{10} and Vickers hardness H of a few types of paint, with a load of 2.5 g, respectively 10 g, applied for 30 seconds at 20 °C.

Type of paint	$i_{2.5}$ μ	i_{10} μ	H kg/mm
Artists' paint	$\gg 20$	$\gg 20$	$\ll 1$
Decorators' paint	>20	>20	<1
Enamel for lighting fittings etc.	4.5-3	9-6	5-10
Synthetic paint for automobiles (synthetic resin, without nitro-cellulose) .	6-3.5	14-7	2- 8
Nitro-synthetic paint for automobiles	5-3	10-6	4-10
Lacquer for radio cabinets	4-3	7-5.5	8-12
Cycle enamel (stoved) . .	3-2.5	6-4.5	10-20

Only those types of paint of which the H-value is under 5 kg/mm² are soft enough to be tested by scratching with the nail. According to the values in the table, many types of paint are too hard to be tested in this manner.

Influence of temperature

Generally speaking, the "hardness" of a paint depends highly on the temperature. In order to enable tests to be carried out at various temperatures, the specimen plate is heated by a flat, brass block having in it channels through which hot water from a thermostatically controlled boiler circulates.

This method was employed to test two types of paint at different temperatures (penetration load 10 g).

Temperature °C	White stoving enamel i_{10} in μ	White nitro-paint i_{10} in μ
20	5.6	6.8
30	7.6	9.4
40	9.5	13.8

This table shows that nitro-paint softens to a greater degree with increasing temperature than stoving enamel; the ratio of the i-value at 40° to that at 20° is 9.5 : 5.6 = 1.7 for enamel, and 13.8 : 6.8 = 2.0 for nitro-paint. This difference is explained by a comparison of the two types of paint: both consist of macro-molecules forming chains of various lengths, but, whereas the stoving enamel structure is strengthened by numerous "bridges" between the chains, no such bridges occur in the structure of the nitro-paint; hence the nitro-paint is more thermo-plastic than the enamel.

In view of the effect of the temperature, it is advisable to state at which temperature the hardness test is carried out whenever accuracy is at all important. It is further recommended that, whenever possible, measurements be effected at not too low a temperature, e.g. 20 $\pm$ 0.5 °C, or that the values measured be referred to this temperature by interpolation or extrapolation.

Effect of layer thickness

The question of the layer thickness is an interesting one. Depicted in *fig. 5* are curves representing i as a function of the thickness d of a layer of a relatively soft, nitro-synthetic paint, each curve being plotted with respect to a constant load F, at a temperature of 25 °C. The thicknesses tested were 190, 120, 50, 20 and 10 μ, and in each case the specimen was thoroughly dried before the test. It appears that similar loads produce deeper impressions according as the thickness of the layer is increased [7]), but that, at thicknesses above a certain limit — the greater the load F, the higher this limit —

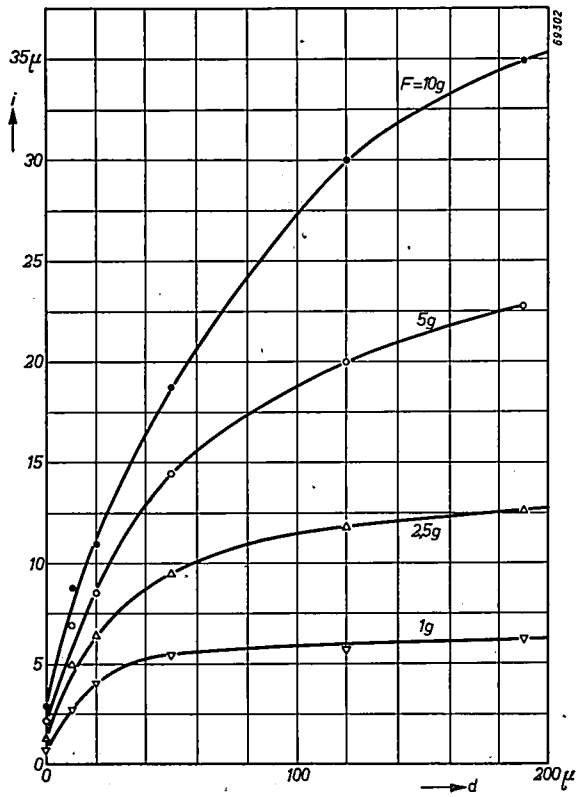


Fig. 5. Penetration depth i as a function of the thickness d of a coat of a relatively soft nitro-synthetic paint, for various values of the load F applied at 25 °C.

[7]) This effect is important not only from the point of view of testing, but also in its application: a thin coat of paint is apparently more resistant to mechanical damage than a thick one, or, for the same resistance to damage, the thin coat may be more plastic. Moreover, a thin film is more economical and is better able to follow deformations of the underlying surface.

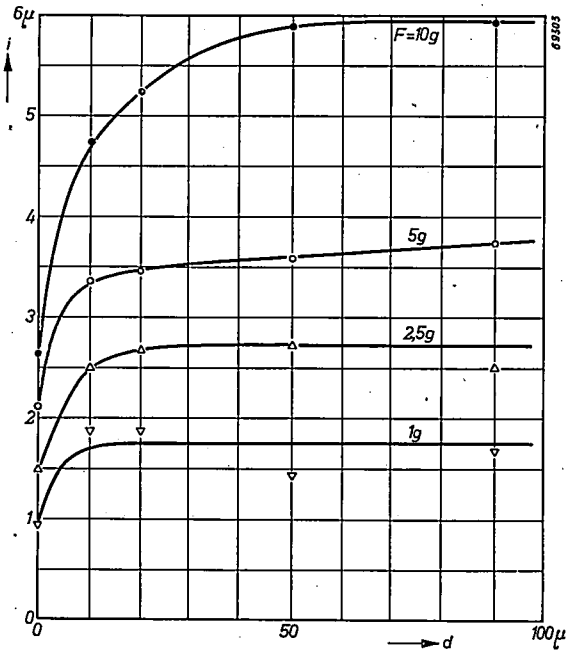


Fig. 6. As fig. 5, but for a slightly harder paint.

any further increase in d produces only a slight increase in i. Numerous tests have proved that, if a measurement is to be taken in the range in which i is only slightly dependent on d, the load F should be so small that i does not exceed one-twelfth of d.

Curves relating to a slightly harder paint and also to a hard stoving enamel are shown in *figs* 6 and 7.

The curves connecting the measured points in figs. 5, 6 and 7 seem by interpolation with respect to $d = 0$ to reach the i-axis at a point which, for the same value of F, gives the depth of penetration into the material to which the paint is applied.

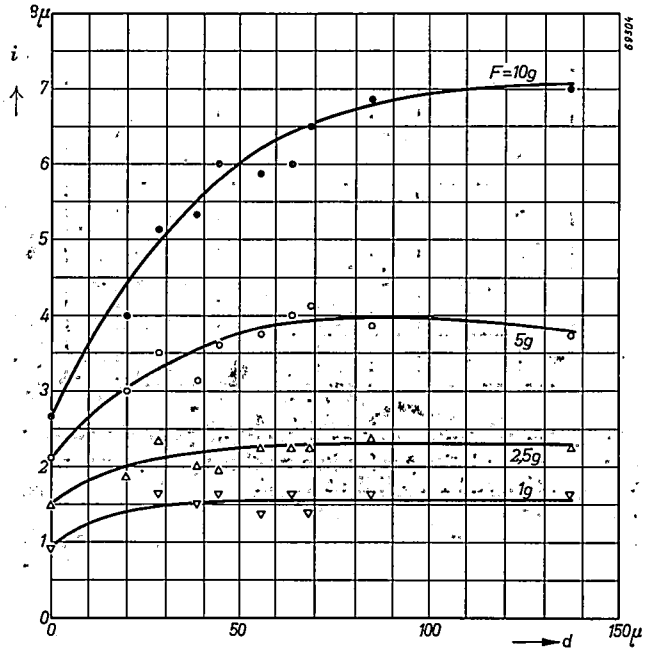


Fig. 7. As fig. 5, but for a hard stoving enamel.

In the tests conducted by the authors this material was aluminium with a hardness value of $H \approx 45$ kg/mm^2. According as the thickness of the coat of paint decreases, the apparent Vickers hardness of the paint, based on the F-value and i-value, more and more approximates the H-value of the underlying material.

Figs. 5, 6 and 7 clearly show that, in the case of $d \rightarrow \infty$, the limit of the penetration depth increases more rapidly than $\sqrt{F}$; hence, with large thicknesses, the apparent hardness drops as F is increased.

Further test results are given in the original paper [8]).

Variation of the "hardness" during the drying process

The ultimate "hardness" in the thoroughly dry condition is not the only important property of a coat of paint: the period involved from the time of application until such time as the painted article can be safely handled is another important factor.

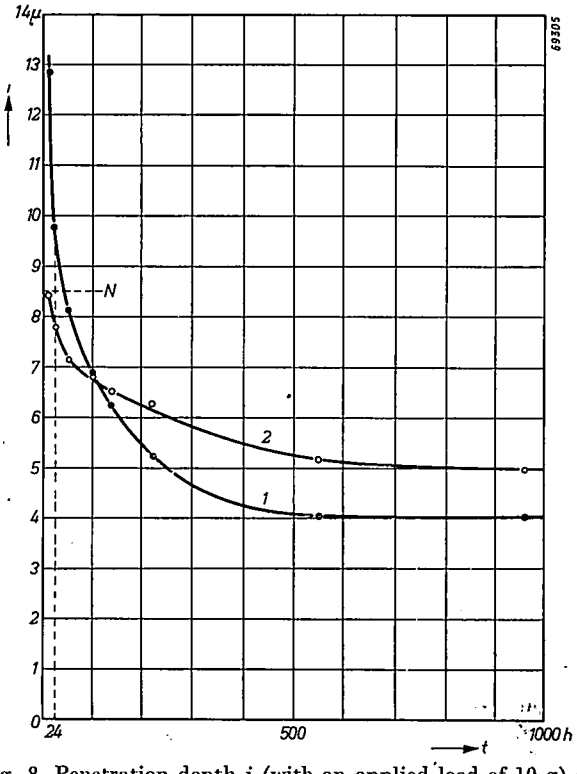


Fig. 8. Penetration depth i (with an applied load of 10 g) as a function of the drying time t, for two types of air-drying paint, both containing TiO_2 as pigment and identical volatiles. Type *2* reaches the value N (resistant to scratching) before type *1*.

The threshold of this condition may be said to be at the point at which the paint is "resistant to scratching" (penetration depth 8.5 μ at $F = 10$ g, or hardness value approx. 5 kg/sq. mm).

[8]) J. Hoekstra (also on behalf of J. A. W. van Laar), The purpose of measuring the hardness of paints (in Dutch), Verfkroniek **25**, 13-17, 1952.

Fig. 8 shows the penetration depth as a function of the drying time of two types of physical drying paint containing identical volatiles and the same pigment (titanium dioxide), but differing slightly in composition (synthetic resin, plasticizer). When i no longer changes noticeably with time the paint in the deeper layers may also be entirely dry.

Curve *1* relates to a paint of a composition such that, although its ultimate hardness is fairly high (fairly low i-value), it reaches the state in which it is "resistant to scratching" much later than the paint represented by curve *2*, which, therefore, is to be preferred. Measurements of this kind are of very great value in the preparation of paints.

Summary. This article describes an instrument for the testing of painted surfaces, more or less along the lines of the Vickers hardness tester for metals. A sapphire of pyramidal form having an apex angle of 136° is pressed against the coating of paint with a force of 10 g or less. When pressure has been maintained for 30 seconds, the depth to which the sapphire has penetrated into the paint is measured by means of a differential capacitor which forms two branches of a bridge, the other two branches (resistors) being incorporated in a "Philoscope" bridge which also includes an amplifier, electronic indicator and power unit. Penetration depths of only a few microns can thus be measured with an accuracy within less than one-tenth of a micron. The instrument is robust, easily transportable and simple in operation.

A Vickers hardness could be deduced from the hardness measured, but this is inadvisable because the latter quantity depends to a large extent on the force with which the sapphire is pressed against the point.

Some examples are given to illustrate the effects of the temperature and the thickness of the coat of paint to be tested. When the thickness is below a certain limit, an "apparent" hardness, which is much higher than the true value, is measured; the penetration depth should therefore never be more than one-twelfth of the thickness of the coating if the effect of the layer thickness on the penetration depth is to be eliminated.

DIELECTRIC LOSSES IN GLASS

by J. M. STEVELS.

666.1:537.226.3

One of the many industries in which glass is used is that of electrical engineering. A problem arising in this industry is the behaviour of the glass in an electric alternating field. The dielectric losses of glass form one of the factors that decide this behaviour. Extensive investigations are still being carried out with a view to producing glasses having small dielectric losses at all frequencies. These investigations have not only led to an improvement of the existing kinds of glass but they have given a better insight into the physical and chemical structure of glass.

Dielectric losses

Glass is a material that is much used in electrical engineering: the bulbs of incandescent lamps and of radio valves (transmitting and receiving valves) are made from glass, that is also indispensable as an insulating material and for very many other purposes. It is therefore of great importance to know how the electrical properties of glass are affected by its composition, for this enables a choice to be made of the best kinds of glass for certain uses. This article will deal with the behaviour of glass in an electric alternating field, attention being mainly focused on the dielectric losses of glass.

When a piece of glass is interposed as a dielectric between the plates of a capacitor and an alternating voltage with a frequency f is applied to that capacitor, if the glass were free of losses, then the current flowing through the circuit would be exactly an angle $\pi/2$ in advance of the voltage across the capacitor. This means that no energy would be dissipated in the capacitor. In practice, however, every dielectric, and thus also glass, dissipates a smaller or larger proportion of the energy applied and converts it into heat. The measure of this dissipation of energy (the dielectric losses) is denoted by the angle δ by which the actual difference in phase between current and voltage deviates from $\pi/2$. The energy dissipation per unit of time for a small δ is approximately equal to $f \tan \delta$; the angle δ is called the loss angle.

In glass, as is the case with other materials, the dielectric losses vary strongly with the frequency of the electric field. The "spectrum" of the losses of glass is broadly represented by the fully drawn curves in *fig. 1*, which have been drawn for two temperatures, viz. 50 °K and room temperature. It is to be noted, however, that these diagrams have no quantitative significance; the losses differ greatly according to the kind of glass considered. The graph gives only a general impression of the losses as a function of frequency. Although the frequency range in these diagrams extends from 1 to 10^{14} c/s, only measurements between the frequencies of 50 c/s and 10^{10} c/s will be dealt with.

Fig. 1 reveals that there are two maxima of the losses, shifting in frequency as the temperature is changed but not in the same direction. Between the maxima is a "valley" of relatively small losses, the depth of this valley decreasing as the temperature drops.

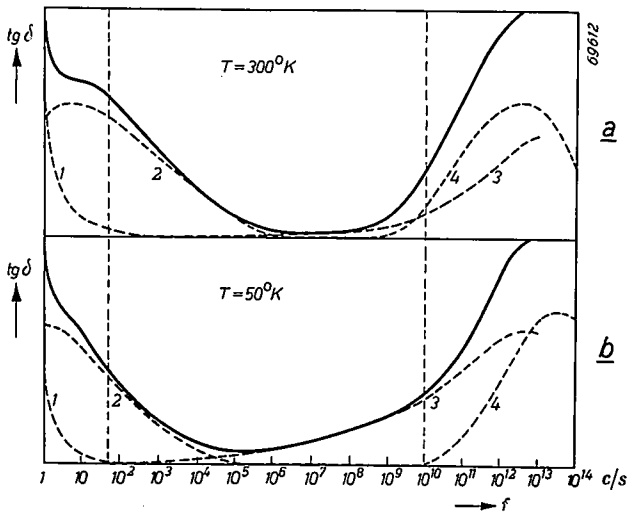


Fig. 1. Diagrammatic representation of the frequency spectrum of the dielectric losses in glass; *a* gives the spectrum at approximately room temperature, *b* shows how the spectrum appears at a very low temperature. The fully-drawn curve gives the total losses. It can be analyzed into four kinds of losses occurring at different frequencies, as represented by the broken-line curves: *1* conduction losses, *2* relaxation losses, *3* deformation losses, *4* vibration losses. It is seen how the vibration losses, occurring at high frequencies, are displaced to still higher frequencies as the temperature falls. The conduction and relaxation losses, on the other hand, are displaced towards lower frequencies, while the deformation losses increase over the whole line. These diagrams are of only theoretical significance. The frequency range here extends from 1 to 10^{14} c/s; this article deals only with the range between 50 and 10^{10} c/s (vertical broken lines).

From what is already known about the behaviour of the dielectric losses in other solids an attempt may be made to analyze the loss spectrum of glass

into various components, taking into account the specific property distinguishing glass from crystalline materials, namely the relatively great disorder in the orientation of the ions from which glass is built up.

It will be seen that there are differences in the manner in which the various kinds of losses depend upon the composition of the glass. Thus from the analysis of the losses some idea can be obtained as to how the composition has to be varied in order to reduce or increase the losses at certain frequencies.

The result of the analysis has been graphically represented in fig. 1 by the broken-line curves. Four kinds of losses are to be distinguished, named respectively conduction losses, relaxation losses, deformation losses and vibration losses, which will now be dealt with in that order.

Conduction losses

Glass is not a perfect insulator. Under the influence of an electric force the network-modifying ions move through the whole of the network [1]), and in doing so they give off part of the energy obtained from this force to the network in the form of heat.

The conduction losses (curve *1* in fig. 1) depend upon the conductivity σ according to the formula

$$\tan \delta = \frac{\sigma}{\omega \varepsilon' \varepsilon_0}, \quad \ldots\ldots \quad (1)$$

where ω $(= 2\pi f)$ represents the angular frequency of the electric alternating field and ε' the dielectric constant of the glass (ε_0 is the dielectric constant of vacuum). Thus $\tan \delta$ is inversely proportional to the frequency (since δ and ε' are not frequency dependent, or scarcely so), and the energy dissipation per time unit does not depend upon the frequency. This is a well-known property of a conductor. When $\tan \delta$ for a lime glass ($\sigma = 10^{-10}\ \Omega^{-1}\mathrm{m}^{-1}$, $\varepsilon' = 6$) is calculated according to the above formula for a frequency of 1000 c/s one finds the small value of 3×10^{-4}, whereas an actual measurement gives $\tan \delta = 100 \times 10^{-4}$, so that at this frequency there must be other causes of losses much more important than conduction. As a general rule, for frequencies higher than 50 c/s the conduction losses are negligible compared with the other losses. In this article, therefore, conduction losses will no longer be considered.

[1]) The structure of glass, and especially the conceptions of network-modifying ions and network-forming ions, have been dealt with in a previous article recently published in this journal (J. M. Stevels, Philips Techn. Rev. **13**, 293-300, 1952, No. 9), which will further be referred to as article I.

Relaxation losses

It is not only the transport of network modifiers through the whole piece of glass considered that causes losses, but also the transport over atomic distances in the glass plays a part. In *fig. 2* the variation of the potential energy of the ions in an arbitrary part of the network is indicated for an arbitrary direction (x) in that network.

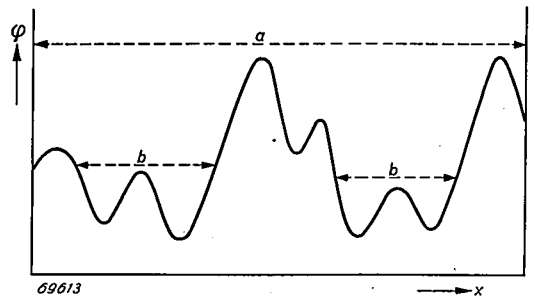


Fig. 2. The undulated curve represents the potential φ of the network-modifying ions along an arbitrary direction in the glass. The movement indicated by a can only be carried out by high-energy ions and corresponds to conduction, while the movement b is carried out by ions with less energy and results in relaxation losses.

The successive potential barriers, with varying heights and distances, form obstacles preventing an ion from travelling further through the glass. The potential barriers arise partly from electrical causes (mutual repulsion of the ions) and partly from purely geometrical causes: the movement of the ion concerned may be obstructed by the other ions. Similar potential barriers are found also in crystals, but there the distances and heights are very regularly distributed.

When an electric field is applied the ions tend to move in the direction of this field. Usually the field alone will not be sufficient to carry the ions over the potential barriers. For that to take place the ions have to take up extra energy, and this is obtained from collisions with other ions due to the thermal movement in the glass. It takes some time, however, before the ions have accumulated sufficient energy to pass over a potential barrier; this time is called the relaxation time τ. If the electric field is an alternating field with angular frequency ω, and ω is not large compared with $1/\tau$, then during each half cycle the ions will have ample time to jump over one or more potential barriers, thereby gaining energy from the field, which is transmitted to the network in the form of heat; one then speaks of losses. At very low frequencies ($\omega \ll 1/\tau$) these losses are small, because then the ions already jump when the instantaneous value of the field differs but little from zero: the energy gain of the ions is then small and the lattice receives little energy.

When, however, ω becomes of the order of $1/\tau$ then, on the average, the ions will not pass over the potential barrier until the field has practically reached the maximum value. The ions then gain much energy and the losses are large. If the frequency of the alternating field is much higher than $1/\tau$ the ions jump independently of the phase of the alternatng field and the losses are again small.

The relation between these relaxation losses and the angular frequency ω of the field is expressed as

$$\tan\delta = \frac{\omega\tau}{1+\omega^2\tau^2}, \quad \ldots\ldots \quad (2)$$

which is entirely in accordance with the foregoing reasoning: the losses occur only at angular frequencies round about $1/\tau$.

As may be understood from fig. 2, in glass there is more than one relaxation time τ, since the local structures differ considerably. This makes the relaxation losses perceptible over a relatively rather wide range of frequencies. The spectrum of these relaxation losses (broken-line curve *2* in fig. 1) shows a wide peak, the limits of the range in which the losses occur lying, at room temperature, at about 10^{-3} and 10^{6} c/s. As already remarked, at frequencies higher than 50 c/s the relaxation losses exceed the conduction losses considerably.

The conduction losses could also be regarded as a kind of relaxation losses. The ions which collect sufficient energy to pass over the highest potential barriers will be able to travel through the whole piece of glass. In order to pass over the highest barriers, a very long relaxation time will obviously be required, and in that case formula (2) becomes

$$\tan\delta \sim \frac{1}{\omega\tau}, \quad \ldots\ldots\ldots \quad (3)$$

$\tan\delta$ being inversely proportional to the frequency, a relation applying, indeed, for conduction losses.

Deformation losses

Relaxation losses are therefore due to an after-effect phenomenon. There is reason to assume that another kind of losses occur in glass which bear the character of after-effect losses. These are denoted by the name of deformation losses to distinguish them from the relaxation losses just dealt with. Deformation losses do not arise from the movement of individual network modifiers, but from the movement of whole sections of the network, in particular a sort of kinking movement of the chains of the network (see I). As is rather obvious, such a phenomenon has a very short relaxation time: it is to be imagined that the chains make movements resembling the jumping of a bent leaf spring. Here again there are a large number of different relaxation times τ. Since on the average τ is very small, for the range of not too high frequencies formula (2) may be written approximately (for the case where there would be only one relaxation time) as:

$$\tan\delta \sim \omega\tau \ldots\ldots \quad (4)$$

From this formula it follows that the deformation losses increase about proportionately with the frequency and in most cases will be very small. The spectrum of the deformation losses is represented in fig. 1 by the broken-line curve *3*.

Vibration losses

The fourth kind of dielectric losses to be expected in glass are due to a resonance phenomenon. The ions in the glass — both the network formers and the network modifiers, as well as the oxygen ions — may vibrate with a certain frequency round about their state of equilibrium, and this they will in fact do as a consequence of the thermal movement. Regarding the ion as a harmonic oscillator, its frequency of vibration may be approximately represented by the formula

$$\omega_{res} = \sqrt{\frac{a}{M}}, \quad \ldots\ldots \quad (5)$$

where a is the proportionality factor denoting the relation between the restoring force and the displacement from the state of equilibrium, and M is the mass of the ion in question. Consequently ions of different mass and ions at different places in the crystal (with different a) will usually vibrate with different frequencies.

When an electric force is applied with a frequency approximately equal to the vibration frequency of an ion, then resonance may occur. Since the vibrations of the ions are always damped, this resonance is accompanied by losses. If all the ions had the same resonant frequency, then the spectrum of these vibration losses would show a maximum round about that frequency. But, as already observed, there are a large number of resonance frequencies. This, and the possibility of the vibrations of the ions being strongly damped, leads to a wide maximum in the spectrum of the vibration losses. This spectrum is represented in fig. 1 by the broken-line curve *4*. Unfortunately, with one single exception, the actual maximum has never yet been determined, since measurements so far carried out have not gone beyond 10^{10} c/s.

Vibration losses in glass are directly comparable to the well-known infra-red absorption in polar crystals, which arises from a similar resonance phenomenon. In the case of a crystal like rock salt this absorption takes place at about 10^{13} to 10^{14} c/s, where the maximum is very narrow. The fact that in the case of glass "spurs" of the vibration losses are to be observed at much lower frequencies (about 10^{10} c/s) may be ascribed, as already explained above, on the one hand to the presence of ions with low resonance frequencies and on the other hand to the possibility of a strong damping of the ion vibrations in glass, which leads to a considerable widening of the loss spectrum. As may be understood theoretically, strongly damped resonance vibrations form, as it were, a transition to the relaxation vibrations previously discussed. The question in how far the vibration losses in glass are to be regarded purely as a resonance phenomenon or as a sort of after-effect phenomenon can to a certain extent be answered, as will presently be shown, when the temperature dependency of the losses is investigated. It is to be added, however, that in order to gain a deeper insight into these phenomena it is necessary to study the losses in glass in the infra-red at the frequencies of 10^{11} to 10^{13} c/s unfortunately it is extremely difficult to carry out absorption measurements in that range.

Having now dissected the total loss spectrum in fig. 1 into four components, the spectrum can be well understood qualitatively. For a quantitative investigation great difficulties have to be overcome, due to the irregular structure of glass. Some confirmation of the picture formed is, however, to be obtained when we come to investigate the temperature dependency of the various kinds of losses. It is then found that their behaviour is roughly as represented in figs 1*a* and *b*, the relaxation and conduction losses being displaced towards lower frequencies as the temperature drops, while the deformation losses increase over the whole line. As to the vibration losses, there are indications that with falling temperature these are displaced towards higher frequencies, and this would be in agreement with what is to be theoretically expected for this kind of loss.

Influence of temperature upon the losses

The foregoing conclusions may be briefly explained as follows.

The temperature dependency of conduction losses is governed entirely by that of the conductivity. As is presumably known, for a fixed frequency the conductivity σ decreases with falling temperature, and thus, according to formula (1), also the conduction losses decrease.

The relaxation times τ occurring with relaxation losses and deformation losses will be greater as the temperature T falls, since the mutual collisions of the ions are then fewer and weaker and it takes longer for the ions to accumulate the energy required to overcome a potential barrier. An increase in τ means a displacement of the losses to lower frequencies, as is indeed observed in the case of relaxation losses when T drops.

As to the deformation losses described by formula (4), an increase in τ means that the losses increase over the whole of the frequency range with which we are concerned. For a quantitative calculation of the temperature dependency of relaxation and deformation losses it is necessary to take into account the great variety of relaxation times occurring in glass, but in their generality the conclusions still hold. To illustrate the relation between the losses at medium frequencies and the temperature, in *fig. 3* a curve has been plotted representing tan δ of a lime glass as a function of temperature at the frequency 1.5×10^6 c/s. It appears that, for temperatures round about 70 °K, tan δ shows a broad maximum, then becoming rather small at about 150 °K and rising again rapidly at temperatures higher than 200 °K. The broad maximum is due to the deformation losses, while the rise above 200 °K is to be ascribed to relaxation losses, as may be understood from the following.

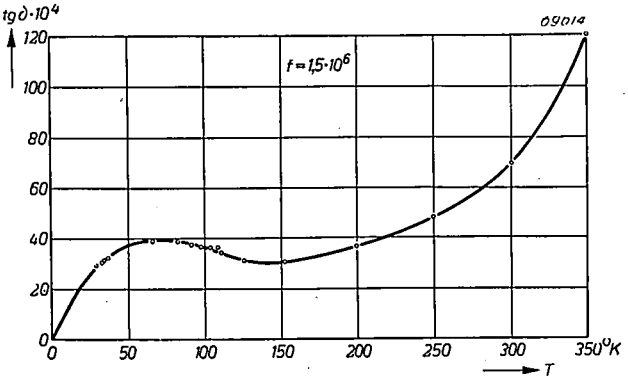


Fig. 3. Dielectric losses in a lime glass at the frequency 1.5×10^6 c/s as a function of temperature. The maximum at low temperatures corresponds to the deformation losses, the rise of the curve on the right is due to relaxation losses.

For a fixed frequency it is possible to determine the temperature at which the deformation losses have reached their maximum. The relaxation times are always much smaller for the deformation losses than for the relaxation losses, so that the frequency mentioned will be situated on the right-hand descending part of the frequency curve for the relaxation losses. When the temperature rises both the spectrum of the deformation losses and that of the relaxation losses will be displaced towards higher frequencies. The maximum of the deformation losses will, therefore, move away from the fixed frequency, whereas the maximum of the relaxation losses moves towards this frequency.

A note is to be added here. The conduction loss is also a kind of relaxation loss, with such a great relaxation time as to occur always on the right-hand descending part of the curve representing the relation between losses and frequency. The conclusion already arrived at in another way, that these conduction losses diminish with falling temperature, is therefore not surprising when considered also from this point of view.

Purely vibrational losses will be displaced towards higher frequencies as the temperature falls. This is because the vibrating ions have on the average higher resonant frequencies at lower temperatures, as may be understood from quantum theory. It has already been pointed out that the vibration losses in glass probably form a transition to relaxation losses. Since the temperature dependency of relaxation losses is just the reverse of that for vibration losses, it is not possible to predict the influence of temperature on these losses at the highest frequencies without carrying out further experiments

in the far infra-red. The indication, already mentioned, that with falling temperature these losses are displaced to higher frequencies makes it highly probable, however, that the character they bear is preponderantly that of resonance.

We now have to consider the influence of the composition of the glass upon the different kinds of losses. First its influence upon the relaxation losses will be dealt with, and then the influence upon the vibration losses; as mentioned above, the conduction losses will be disregarded, since, whatever the composition of the glass may be, for frequencies above 50 c/s they are of no importance; as to the deformation losses too little is yet known about them.

Influence of the chemical composition of glass upon relaxation losses

The manner in which relaxation losses are affected will be considered under five headings.

Nature of the ions

As a rule only those network modifiers which readily migrate through the glass will be capable of yielding a considerable component towards the relaxation losses. The mobility of the ions in a given potential field is determined mainly by their size: it is particularly the relatively small Li^+ and Na^+ ions which play an important part in this respect. Glasses containing a large proportion of these ions show high relaxation losses. The K^+ ion is so much larger that it contributes much less towards the losses. By way of illustration mention may be made of a series of low-melting glasses of the composition 53.3 mol % SiO_2, 10.2 mol % PbO, 4.5 mol % CaF_2 and 32 mol % M_2O, with M representing an alkali metal. For a frequency of 1.5×10^6 c/s and at a temperature of 20 °C, the value of tan δ for M = Li is 132×10^{-4}, for M = Na tan δ is 106×10^{-4} and for M = K it is 54×10^{-4}.

Divalent ions, owing to their larger charge, are bound more strongly to their surroundings and thus are usually less mobile. An exception is the small Mg^{2+} ion which appears to be capable of yielding a considerable contribution towards the losses.

From now on the Li^+, Na^+ and Mg^{2+} ions will be denoted by the name of mobile ions, whilst the divalent network modifiers, such as Ca^{2+}, Ba^{2+}, Pb^{2+}, will be called immobile ions.

It will be clear that the losses depend also largely on the nature of the network. This influence can be reduced to three effects.

Influence of R

As explained in article I, the ratio R of the number of oxygen ions to the number of network-forming ions is a measure for the coherence of the network from which the glass is built up. The larger the ratio R, the less coherent is the network and the less dense is the structure of the glass. This lends more mobility to the network modifiers, and the relaxation losses may be expected to increase.

Influence of the addition of immobile ions

If metal ions showing little mobility — either because they are large or because they are strongly bound electrically — are added to a glass of a certain composition, then there will be less possibility for displacement of the mobile ions. Thus, given equal concentration of mobile ions, the addition of immobile ions to a glass reduces its loss angle.

The effects mentioned here can be quantitatively described by the formula:

$$\tan\delta = A\,C_m\,R\,(1 - BC_i), \quad . \quad . \quad . \quad (6)$$

where C_m represents the concentration of the

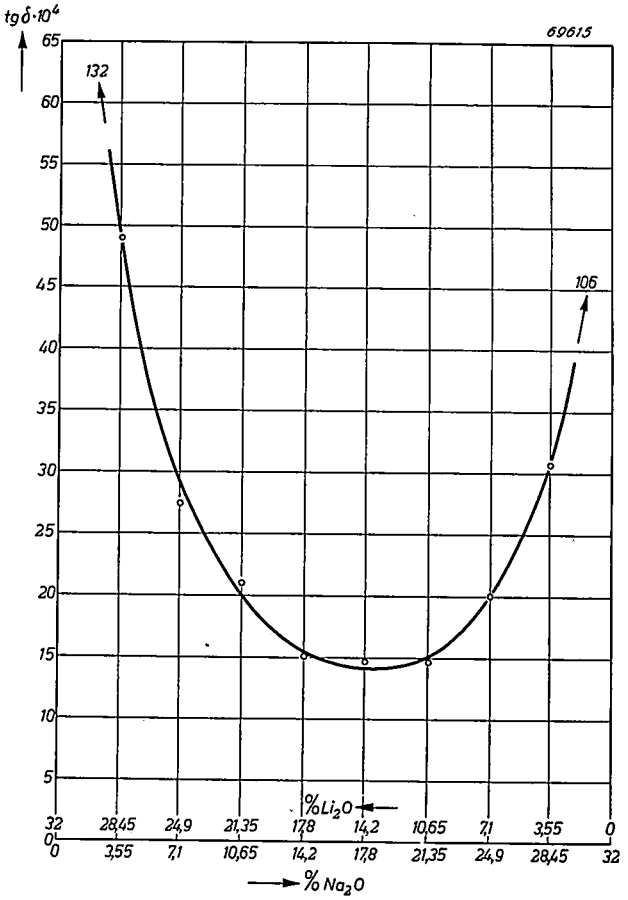


Fig. 4. Replacement of part of the Li+ ions in a lithium-silicate glass by Na+ ions gives a very rigid structure and thus reduces the relaxation losses. If, however, this replacement is carried too far the structure becomes less rigid again. The curve, for tan δ therefore shows a minimum at about equal concentrations of Na^+ and Li^+ ions (measured at the frequency of 1.5×10^6 c/s).

mobile ions and C_i that of the immobile ones; the coefficient A depends on the frequency and temperature; B is a geometrical factor [2]).

Difference in packing, with equal R

If in a glass containing only one kind of network modifiers, some of these ions are replaced by other network modifiers having a different radius, then a more compact network can be obtained. As the glass cools from the melt, the mutually coherent chains of network-forming ions will then preferentially so arrange themselves that the largest interstices in the network are occupied by large network modifiers and the smallest interstices by small network modifiers. The resulting denser structure leads to a reduction of the losses. This is illustrated in *fig. 4*, relating to the glass of the composition mentioned in the beginning of this section, with M representing partly lithium and partly sodium. The value of tan δ (measured for a frequency of 1.5×10^6 c/s) varies with the variation of the proportions of Na^+ and Li^+; the losses are the smallest with about equal molecular concentrations of Na^+ ions and Li^+ ions.

Relaxation losses in borate glasses

In article I it has already been discussed how in certain respects the structure of borate glasses differs from that of other glasses. In a glass of the composition x Na_2O-y B_2O_3, if $x < 18$ mol % (the

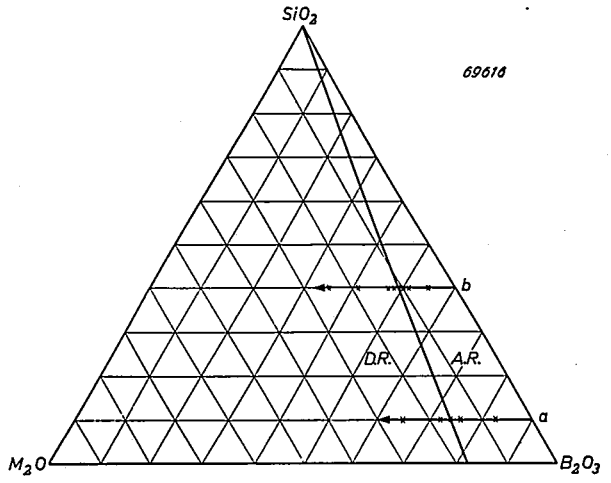


Fig. 5. Phase diagram of the ternary system M_2O-B_2O_3-SiO_2, the borosilicate glasses. M represents an alkali metal. Two regions can be distinguished: the accumulation region *A.R.* and the destruction region *D.R.*

[2]) See, e.g., J. M. Stevels, Progress in the theory of the physical properties of glass, Elsevier Publ. Co., Amsterdam 1948, p. 76.

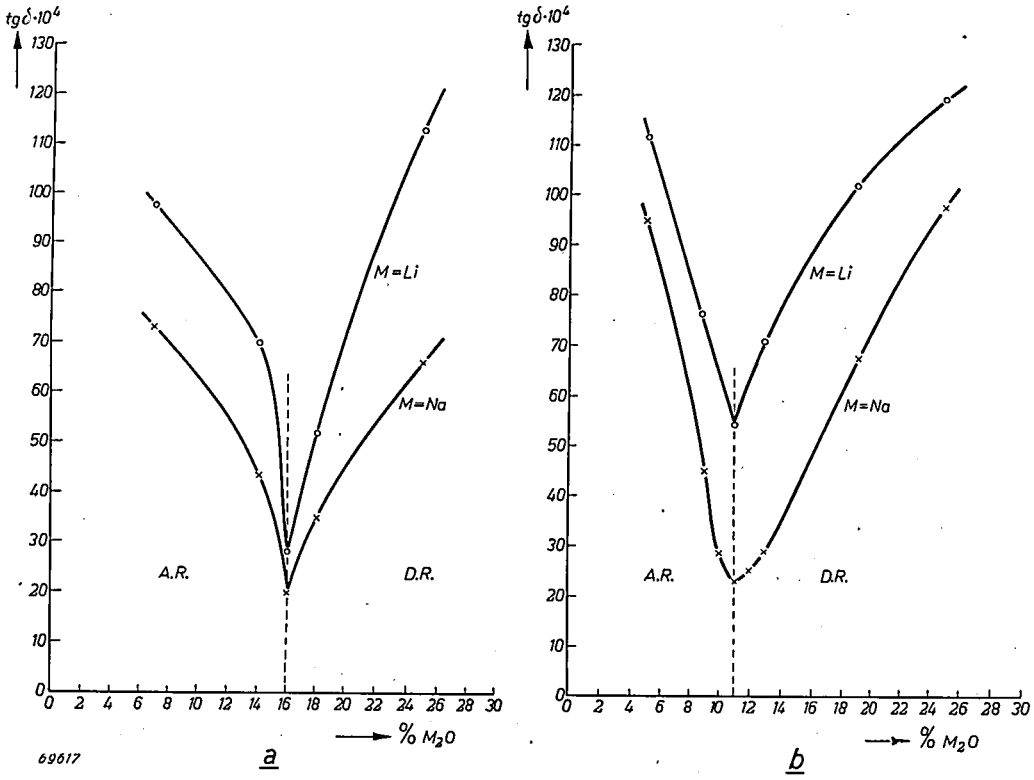


Fig. 6. Variation of tan δ with the Na_2O or Li_2O content of borocilicate glasses; *a*) corresponds to the compositions *a* in fig. 5, *b*) corresponds to the compositions *b*. The curve for tan δ shows a sharp minimum on the *A.R.-D.R.* separating line. It is also seen that Li ions cause a larger loss angle than Na-ions, because the former are more mobile. Measurements were taken at the frequency of 1.5×10^6 c/s.

accumulation region), the network-modifying Na^+ ions are incorporated in the network without any breaking of oxygen bridges, and thus there are no non-bridging oxygen ions. This is due to the fact that the B^{3+} ions may occur both in the centre of an oxygen triangle and in that of an oxygen tetrahedron. With larger concentrations of Na_2O, however, oxygen bridges tend to be broken, and this range of concentrations is called the destruction region.

With increasing concentration of Na_2O in the accumulation region, the network becomes more and more rigid until the maximum rigidity is reached just at the transition concentration of 18 mol % Na_2O, above which it begins to loose its rigidity again. Such has been illustrated in article I with reference to the behaviour of the expansion coefficient of the glass as a function of its composition.

A similar phenomenon arises in the case of borosilicate glasses. In *fig. 5* a phase diagram is given for the system M_2O-B_2O_3-SiO_2, where M stands for a monovalent network modifier. This diagram can be divided into an accumulation region *A.R.* and a destruction region *D.R.* The compositions found along the line separating these two regions are characterized by a maximum rigidity of the structure.

Obviously the dielectric losses (conduction and relaxation losses) will be influenced by this change in structure. As was to be expected, the losses show a decided minimum just on the dividing line, as is illustrated in *fig. 6*. Fig. 6*a* corresponds to the compositions indicated by the arrow points *a* in fig. 5. Fig. 6*b* applies for the compositions *b* in fig. 5. Both Li^+ and Na^+ have been chosen as monovalent network-modifying ions, thus giving two perfectly analogous curves.

Influence of the chemical composition of the glass upon the vibration losses

It is possible to shift the frequency range of the vibration losses by changing the composition of the glass. The influence of the chemical composition will again be considered from a number of points of view.

Nature of the ions

One of the most important factors affecting the frequency of the vibration losses is the mass of the ion (cf. formula (5)). The heaviest ions will give the lowest resonant frequency; thus the addition of heavy ions causes the vibration losses to be displaced to lower frequencies. The charge of the ons plays a less important part (this occurs in the restoring force constant a, which expresses the interaction of the ion with the surroundings).

Influence of R

The larger the ratio R, the less dense is the network. The ions become more loosely bound to their surroundings and the resonant frequencies are shifted to lower frequencies. Phosphate glasses are of a kind in which many non-bridging oxygen ions occur and thus the ratio R is high. *Fig. 7* shows that in a certain phosphate glass the resonant frequencies are shifted so far that the maximum of the

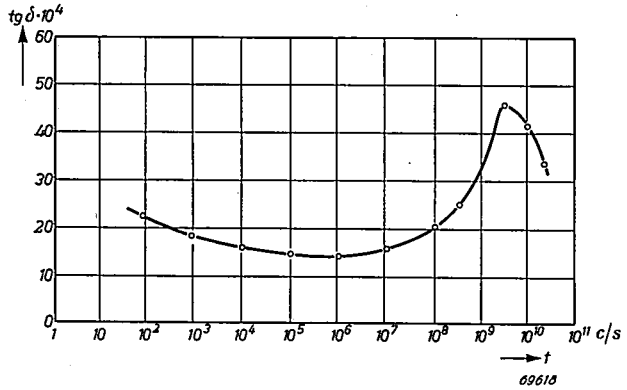


Fig. 7. Example of a glass (phosphate glass) in which the resonant frequencies lie very low, within the frequency range accessible for measurements.

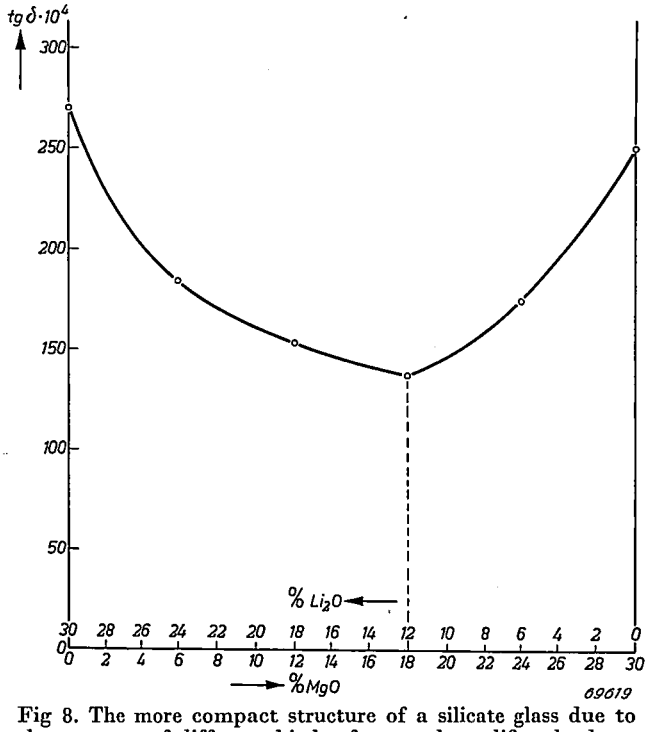


Fig 8. The more compact structure of a silicate glass due to the presence of different kinds of network modifiers leads to a reduction of the vibration losses, measured at a frequency of 2.4×10^{10} c/s. This diagram is fully comparable with fig. 4. It is to be noted that the value of R in this diagram, going to the right, drops somewhat, the reason for this being that Mg^{2+} ions, if present in large concentrations, have the tendency to act as network formers. This diagram relates to a glass of the composition 70% SiO_2 and 30% (Li_2O+ MgO).

vibration losses is situated below 10^{10} c/s (this is so far the only case where it has been possible to measure such a maximum).

The addition of some metal oxides, such as Al_2O_3 and ZnO, which reduce R, will lower the losses at higher frequencies, as follows from the foregoing.

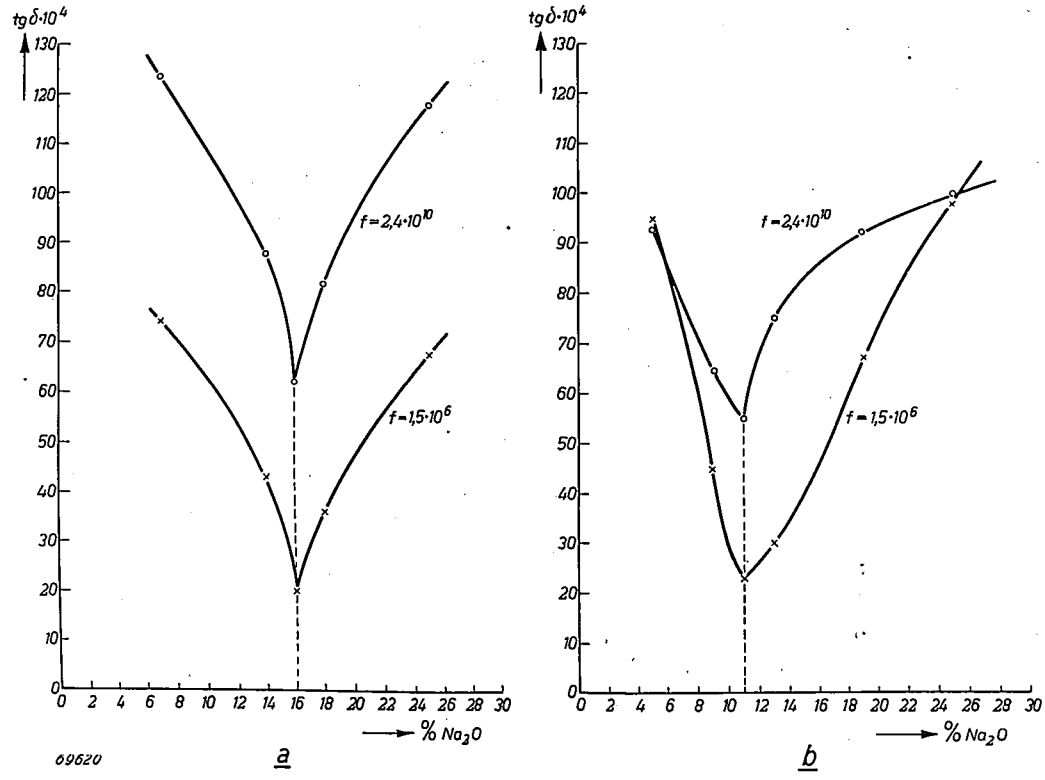


Fig. 9. The vibration losses in borate glass, just as the relaxation losses, are affected by changes in composition going from the *A.R.* to *D.R.* region. Fig. 9*a* relates to the compositions *a* in fig. 5, fig. 9*b* to the compositions *b*, with Na^+ as network modifier. The vibration losses have been measured at 2.4×10^{10} c/s; for comparison also the relaxation losses measured at 1.5×10^6 c/s are reproduced from fig. 6.

Difference in packing, with equal R

As already seen when discussing the relaxation losses, the structure of the network becomes more compact the more the network-modifying ions occurring in the glass differ from each other in their dimensions, even when R remains constant. The more compact the network, the higher are the resonant frequencies of the vibration losses and thus the less are these losses noticed in the range of the relatively low frequencies where the measurements are taken. An illustration of this is given in *fig. 8*, which is directly comparable with fig. 4.

Vibration losses in borate glasses

The effect of the transition from *A.R.* to *D.R.* upon the vibration losses is exactly comparable to the effect upon the relaxation losses. In the neighbourhood of the transition the network is very rigid, and thus the resonant frequencies are very high. The "spur" of the vibration zone that can be measured is then small. This is illustrated in *fig. 9*.

Some examples of the spectrum of dielectric losses in glass

Fig. 10 gives the tan δ as a function of frequency for a number of glasses. The region of the relaxation (and conduction) losses on the left and that of the "spurs" of the vibration losses on the right are clearly distinguished. The glass *A* contains a large proportion of mobile ions and thus shows very large losses. The fact that the curve *D*, for lead glass, rises relatively more steeply at high frequencies than the other curves can be explained, according to the foregoing, by the presence of a large number of heavy ions, as a consequence of which the resonant frequencies lie lower than those in the other glasses.

Fig. 11 shows the losses of a lead-containing glass at a number of frequencies as a function of temperature. This is to be compared with fig. 3. The influence of the deformation loss at low temperatures is clearly seen. This influence is of

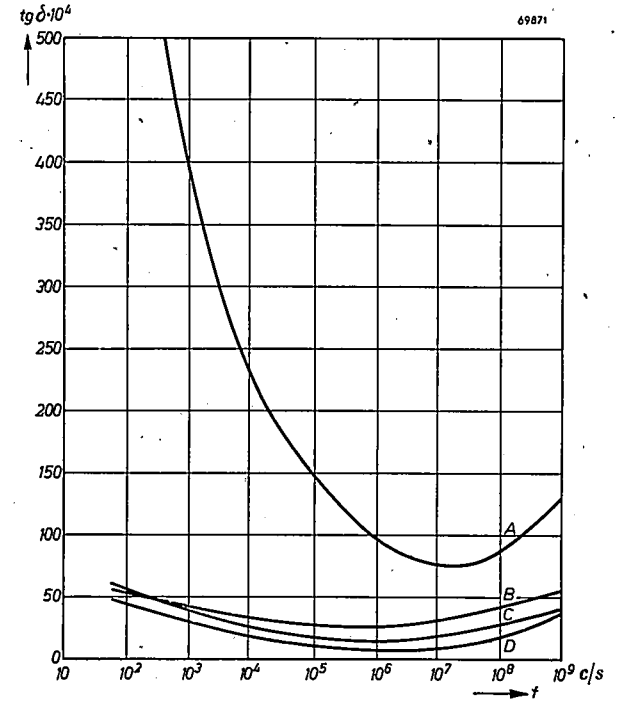


Fig. 10. Loss spectra of a number of glasses. *A* soda-lime glass; the losses, especially the relaxation losses on the left, are very high, owing to the large proportion of mobile Na^+ ions contained in the glass. *B* alkali-free silicate glass, and *C* borosilicate glass: these glasses have low losses. *D* lead glass: the losses for this glass show a relatively steep rise at higher frequencies, because the resonant frequencies in this glass are low.

particular importance for higher frequencies, as follows also from the theory. The increase in the losses at higher temperatures is again due to relaxation losses.

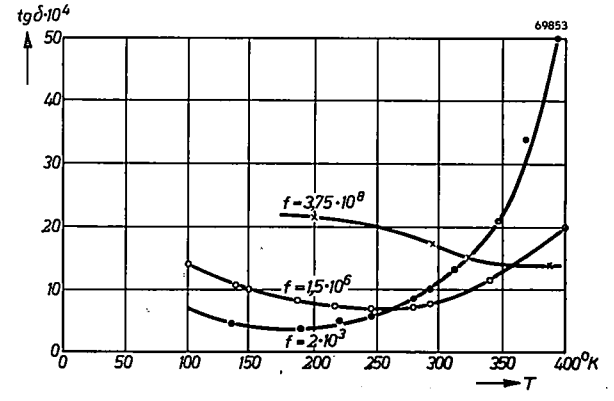


Fig. 11. Loss angle of the glass with the composition 45% SiO_2, 32% PbO, 5% CaF_2, 4% Na_2O and 14% K_2O, as a function of temperature, at different frequencies. The deformation losses (on the left of the diagram) are relatively considerable (cf. fig. 3), particularly at high frequencies.

Practical applications

Glasses with low dielectric losses

Although in practice glasses with high dielectric losses are sometimes desired for special purposes, these will not be discussed here. It has already been made sufficiently clear how such glasses can be produced.

Of most importance are the glasses which show small dielectric losses at normal working temperatures; these are used for the construction of all sorts of electronic tubes and valves. From the foregoing comments it follows that a distinction has to be made between (1) glasses intended for use at intermediate frequencies (radio valves), where the dielectric losses are mainly relaxation losses, and (2) glasses intended for constructions where very high frequencies are applied and mainly vibration losses occur. These are two entirely different fields; a glass that is satisfactory in one field need not at all be so in the other. It is certainly not difficult to find a glass with low dielectric losses at all frequencies — such as fused silica — but the difficulty lies in the fact that for these applications there are nearly always important secondary requirements to be met, e.g. that the glass should be easily workable, that its viscosity should not be influenced much by temperature, and so on. In

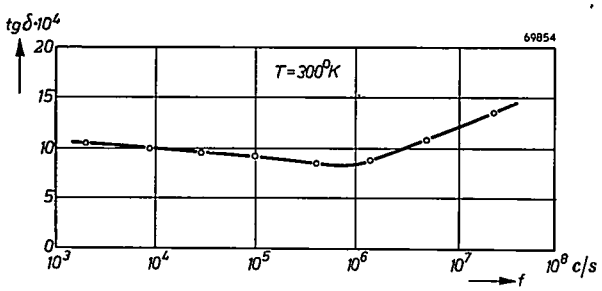


Fig. 12. Frequency spectrum at room temperature of the glass described under fig. 11. The great variety of network modifiers, including, i.a., also immobile Ca^{2+} and Pb^{2+} ions, causes the losses to be small and practically constant over a wide frequency range.

electrical engineering the glass is required to have a certain expansion coefficient to allow of it being fused onto metals such as steel, chromium iron, fernico and tungsten.

Most of the properties just mentioned can be obtained by using a high percentage of Na^+ ions in the glass, but then usually the losses are high. However, by making judicious use of the knowledge now acquired the losses can be appreciably reduced, while still retaining the other properties.

A good example is a glass of the composition 45% SiO_2, 32% PbO, 5% CaF_2, 4% Na_2O, 14% K_2O (weight percentages). Owing to its high contents of Na_2O, K_2O, CaF_2 and PbO, this glass softens at very low temperatures and has such a large expansion coefficient (120×10^{-7}) that it can be fused to iron, while at the same time it shows exceptionally small losses for the intermediate frequencies(see *fig. 12*; fig. 11 applies also for this glass). This is due to the addition of immobile ions

(Pb^{2+}, Ca^{2+}, K^{+}), while the presence of four kinds of network modifiers of different sizes promotes a good packing of the network. It is true that this glass has a relatively large R (viz. 2.64) — which gives it the low softening temperature — but this factor, which is unfavourable for the dielectric losses, is apparently sufficiently compensated by the other favourable factors mentioned above.

Investigation into the structure of glass with the aid of dielectric losses

Since small changes in the structure of glass often result in great changes in the dielectric losses, by applying the reverse process of reasoning it is often possible to draw interesting conclusions as to the structure of a glass from its dielectric losses.

A striking example of this is found in the dielectric properties of the borate glasses already discussed. The position of the separating line between the accumulation region and the destruction region (figs 5, 6 and 9), predicted on theoretical grounds, can be experimentally determined, i.a. from measurements of the dielectric losses.

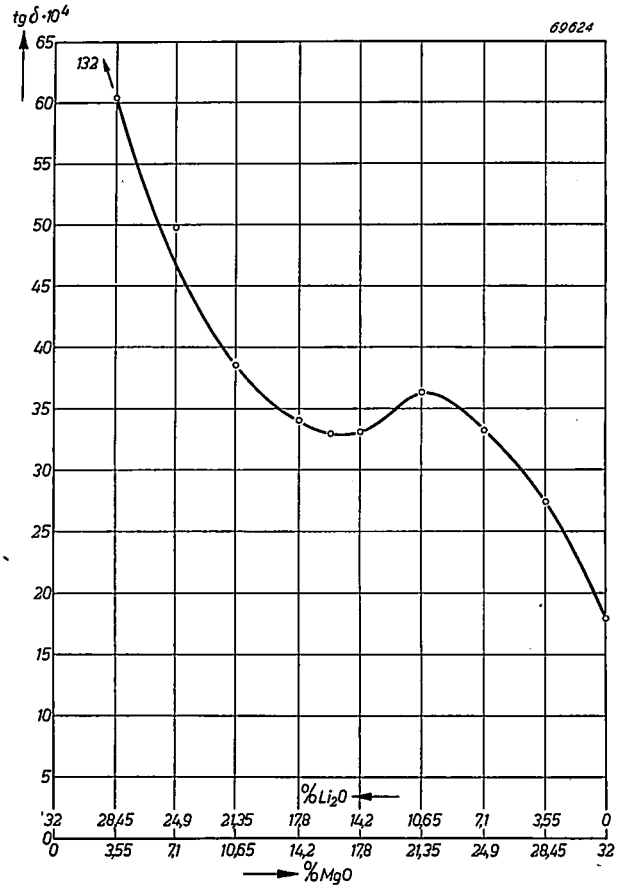


Fig. 13. Variation of tan δ at the frequency 1.5×10^6 c/s and at room temperature for a series of glasses of the composition 53.3% SiO_2, 10.2% PbO, 4.5% CaF_2 and 32% (Li_2O+MgO), as a function of the MgO percentage. As described under fig. 4, when some of the Li^+ ions are replaced by Mg^{2+} ions, tan δ at first decreases, because the structure of the network is then more rigid. A further increase of the Mg^{2+} content reduces the rigidity of the structure and tan δ increases. This does not continue along the same line, in contrast to fig. 4, since the higher the Mg^{2+} concentration, the more Mg^{2+} ions begin to act as network formers. Then they no longer contribute to the relaxation losses, so that ultimately tan δ drops.

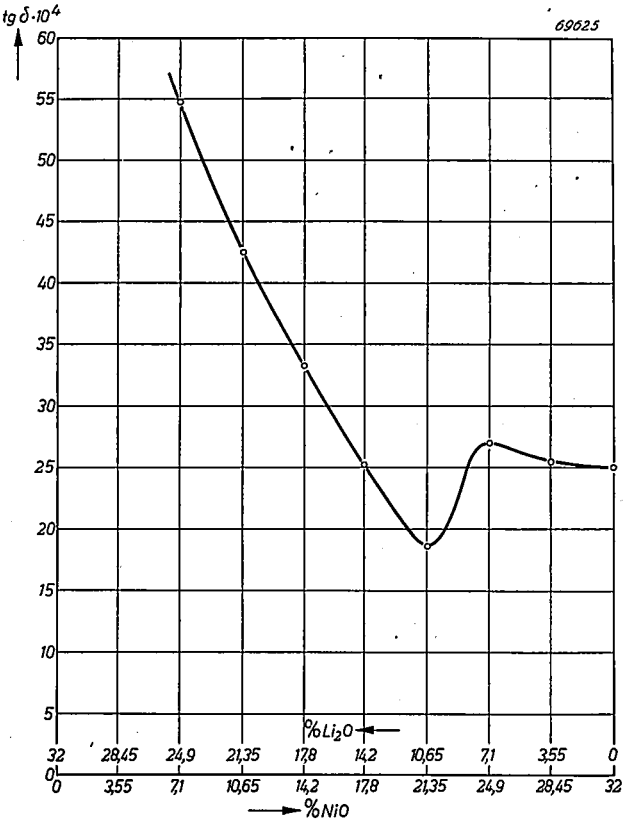


Fig. 14. This diagram relates to a glass analogous to that of fig. 13, but with the Li^+ ions replaced by Ni^{2+} instead of Mg^{2+} ions. The curve has the same trend as that in fig. 13.

There is one other example that may be gone into more deeply. In article I it was stated that there are a number of cations which may occur both in a network-modifying and in a network-forming position. In a number of cases it is possible to see already from the colour of the glass how the ions behave. However, there are some ions (Mg^{2+}, Zn^{2+}) which do not colour the glass, and then the dielectric losses may serve as an indication how these ions are situated in the network. Particularly in the case of the Mg^{2+} ion most valuable information can thereby be obtained, for, as we have seen, this ion yields a rather considerable contribution towards the relaxation losses when occupying a network-modifying position, whereas it obviously contributes hardly anything at all to this kind of losses when it is in a network-forming position. Now it appears that, when the Mg^{2+} ion is present in a large concentration, it shows a tendency to occupy more and more the network-forming positions; such a network is poor in oxygen ions, so that the Mg^{2+} ion can compete successfully against the Si^{4+} ions. This effect is clearly demonstrated in *fig.* 13 (measure-

ments taken at 1.5×10^6 c/s), which is most instructive when compared with fig. 4. Whereas in the latter diagram the dielectric losses show the previously discussed deep minimum at the intermediate concentrations of Na^+ and Li^+ ions, the curve in fig. 13 (where two Li^+ ions have been replaced by one Mg^{2+} ion) shows that with small Mg^{2+} concentrations in the glass the curve for the losses follows an analogous trend, but with higher Mg^{2+} concentrations makes a turn and ends at a very low level, such due to the fact that part of the Mg^{2+} ions no longer contribute towards the losses. As a result of the Mg^{2+} ions changing over from network-modifying to network-forming positions, the value of R, going to the right, drops (in fig. 4 R is constant), and this in itself already leads to a further reduction of the losses.

This picture of the situation is illustrated by *fig. 14*, for the case where two Li^+ ions are replaced by one Ni^{2+} ion. This curve resembles very much that of fig. 13, and in this case it is indeed possible to conclude from the change in colour of the glasses examined that, going to the right, more and more Ni^{2+} ions occupy the network-forming positions.

Finally, it is to be mentioned that, measured at $f = 10^{10}$ c/s, the replacement of two Li^+ ions by one Mg^{2+} ion yields a curve not showing any deflection but retaining the normal shape with a deep minimum. At these frequencies, where the vibration losses play a part, it is of no consequence whether the Mg^{2+} ion occupies a network-forming or a network-modifying position. By way of illustration reference is made to fig. 8.

Summary. The dielectric losses in glass arise from three causes: conduction, after-effect and resonance. The various forms of losses are described, it being shown that the conduction losses are of importance only at very low frequencies and high temperatures. The after-effect losses can be divided into "relaxation losses", occurring at room temperature in the frequency range of 10^{-3} to 10^6 c/s and shifting towards higher frequencies as the temperature rises, and "deformation losses", which are approximately proportional to the frequency and are particularly of importance at low temperatures. The resonance or vibration losses occur mainly at high frequencies (10^{10} to 10^{13} c/s) and as the temperature rises probably shift towards lower frequencies. The influence of the chemical composition and the structure of the glass upon the various losses is investigated, special attention being paid to the borate glasses, in which an accumulation region and a destruction region are to be distinguished according to the composition. It is indicated how the losses can be reduced in certain frequency ranges, some examples being given. Finally, it is shown how the dielectric losses can be taken as indications of the changes taking place in the structure of a glass when its composition is modified.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherland.

1981: P. C. van der Willigen: Grepen uit de ontwikkeling van het booglassen (Electro-techniek **29**, 143-147, 1951, No. 8). (Some remarks on the development of arc welding; in Dutch.)

Two important questions in the field of electric arc-welding by hand are discussed in this article. The first part deals with the choice between direct current and alternating current. Their advantages and disadvantages are described, including the question of the open circuit voltage of the welding transformer. In the second part the different methods for welding of steel are examined from the point of view of shielding of the deposited metal against nitrogen and oxygen from the air. In this connection the use of hydrogen and its influence on the weld are also briefly discussed.

1982: N. Warmoltz: On the application of a Philips ionization gauge type of ion source in a mass spectrometer leak detector (Appl. sci. Res. **B2**, 61-65, 1951, No. 1).

A system of electrodes as in the Philips ionization gauge can function as ion source giving a diverging ion beam in which the ions have different velocities. When placing this source behind a slit, it is possible to focus all the ions of the same mass in a second slit by combining a magnetic field parallel to the slits with an electric field perpendicular to the slits and to the line joining the slits. Between the slits the ions describe a kind of epicycloidal trajectories. In this way it is possible to construct a simple leak detector.

1983: C. J. Bouwkamp and H. Bremmer: A note on Kline's Bessel-function expansion (Proc. Kon. Ned. Akad. Wetensch. Amsterdam A **54**, 130-134, 1951, No. 2).

A certain combination of Bessel functions, viz. $J_\nu(\nu \sec\vartheta)\, Y_\nu(z+\nu\sec\vartheta) - Y_\nu(\nu\sec\vartheta) J_\nu(z+\nu\sec\vartheta)$, in which ν, ϑ, z are complex numbers, is expanded in a power series in the variable $1/\nu$. This series converges if $|\nu| > |z\cos\vartheta|$. The first three coefficients are given explicitly.

1984: J. H. van Santen and G. H. Jonker: Combinaisons ferromagnétiques du manganèse à structure pérovskite (J. Phys. Radium **12**, 202-204, 1951, No. 3). (Ferromagnetic combinations of manganese with perovskite structure; in French.)

From investigations on the magnetic behaviour of ferromagnetic mixed crystals with perovskite structure of the type $(1-x)\,LaMn^{III}O_3 \cdot x MeMn^{IV}O_3$ (Me = Ca, Sr or Ba) it seems probable that the (indirect) exchange interaction between Mn^{IV} and Mn^{IV} is negative and that between Mn^{III} and Mn^{III} is positive. This is the first known example of a positive indirect exchange interaction in oxidic compounds.

1985: J. J. Went: Linear magnetostriction of homogeneous nickel alloys (Physica **17**, 98-116, 1951, No. 2).

The magnetostriction of homogeneous Ni-alloys is investigated. The magnetostriction is measured as a function of the composition of the alloy, of the induction caused by an external magnetic field and of the temperature. A comprehensive table of data is given. Several general relationships for binary alloys have been found. The saturation magnetostriction λ_s at 0 °K can be predicted if the saturation induction I_s is known; λ_s at a higher temperature follows directly from the λ_s value at 0 °K and the relation I_s vs T (except for Ni-Mn-alloys). A single relationship has been found between λ and I for alloys except those containing Co or Mn. From all these data the magnetostriction for ternary alloys may be calculated.

1986: G. W. Rathenau and G. Baas: Grain growth in a texture, studied by means of electron-emission microscopy (Physica **17**, 117-128, 1951, No. 2).

Grain growth and secondary recrystallization in rolled face-entered Ni-Fe-alloys has been studied by electron-optical means, an image of the activated hot metal surface being formed. Grain growth in an imperfect cubic texture proved to be a discontinuous process. Almost all neighbouring grains invade quickly one grain or group of grains. Grain-boundary movement at low-energy boundaries occurs on a small scale. The high-energy parts of the boundary between a cubic crystal and its (near) twin proved to move in a direction parallel to the common (111) plane. The surface tension of the boundary between a secondary crystal and a cubic crystal which is to be absorbed is about twice the value corresponding to the surface between two cubic crystals.

R 160: H. Bremmer: The discharge of a series of equal condensers having arbitrary resistances connected in parallel (Philips Res. Rep. **6**, 81-85, 1951, No.2).

This paper concerns the discharge of a series of condensers through a ballistic galvanometer if a variable resistor is connected in parallel to each condenser, in connection with the model described in **R 159**. The flow of current through the galvanometer is calculated with the aid of the operational calculus.

R 161: B. D. H. Tellegen and E. Klauss: Resonant circuits coupled by a passive four-pole that may violate the reciprocity relation (Philips Res. Rep. **6**, 86-95, 1951, No. 2).

The system of two resonant circuits coupled by a passive four-pole that may violate the reciprocity relation is investigated. To obtain the maximum transfer for a given form of the resonance curve, the circuits must be equally damped and equally tuned. To obtain the maximum transfer with two circuits coupled by a passive four-pole satisfying the reciprocity relation, the circuits must generally be unequally damped and unequally tuned. When the coupled circuits are used as an interstage network in an amplifier, to minimize the influence of changes in valve capacitances on the shape of the resonance curve, the circuits should be equally damped and equally tuned. For symmetrical, flat-topped resonance curves the maximum transfer with non-reciprocal coupling four-poles, the ratios of the maximum transfer with reciprocal coupling four-poles to the transfer with reciprocal transfer with reciprocal coupling four-poles and to equally damped and equally tuned circuits are $(1+\sqrt{2}) : 2 : 1$.

R 162: H. G. Beljers and W. J. van de Lindt: Dielectric measurements with two magic tees on shorted wave guides (Philips Res. Rep. **6**, 96-104, 1951, No. 2).

A description is given of the method and calculations underlying dielectric measurements at microwaves with the aid of wave guides. It is shown that a „magic Tee" (a certain arrangement of four branches of rectangular wave guides at one junction) can be used both as a bridge and as means of establishing any complex waveguide impedance. By using these Tees dielectric measurements can be carried out with greater accuracy than in the conventional method with a standing-wave detector. Some details of the construction of the apparatus and the measurement are given. In conclusion a dielectric measurement of polystyrene is dealt with.

R 163: H. C. Hamaker and Th. Hehenkamp: Minimum-cost transformers and chokes, II (Philips Res. Rep. **6**, 105-134, 1951, No. 2).

In the foregoing paper **R 150** the equations specifying a transformer design with minimum price, P, when the apparent power, VA, and the losses, W, are prescribed, were solved, and the chief characteristics of the resultant designs were discussed. This we shall call the P-W-class of designs. In the present paper two alternative classes of designs are considered, namely, (1) the P-M-class of designs giving minimum price for prescribed values of the power, VA, and the product, $M = BS$, of the peak magnetic flux density and effective electric-current density; and (2) the W-M-class of designs giving minimum losses, W, when VA and M are prescribed. The solutions obtained are casted in such a form that they are directly comparable *inter se* and with the P-W-solution of the foregoing paper. When power and losses are prescribed the P-W-class of designs is always cheapest, but if instead of the losses we prescribe a value of B or of the specific dissipation of heat, v, per cm^2 of external surface, the other classes of designs may be preferable. It is shown that the choice of a design depends mainly on the transformer characteristics which are considered as the limiting factors; a choice between the three classes of designs is of secondary importance. These problems are discussed in detail. In the appendix, a set of tables is provided by means of which the P-M-designs or the W-M-designs can be rapidly obtained. A concluding section deals with the problem of a design such that the sum of the cost of the transformers and the cost of the electric power dissipated as heat during its life, is at a minimum. This problem can also be solved by the methods developed.

R 164: P. Schagen: On the mechanism of high-velocity target stabilization and the mode of operation of television-camera tubes of the image-iconoscope type (Philips Res. Rep. **6**, 135-153, 1951, No. 2).

In this article the mechanism of the high-velocity target stabilization in the iconoscope and the image iconoscope is described. Some hypotheses are formulated for the mechanism of the image iconoscope, and equations are derived therefrom for the potential of target elements as a function of time and for the signal output of these tubes. This theoretical approach indicates that an increase in the secondary-emission coefficient of the target and in the target capacitance will result in better picture quality. Experiments confirm these predictions.