IR.H.A.G.HAZEU

# fifty years of
# ELECTRONIC COMPONENTS

Dr H. W. Horeman

Met mijn hartelijke dank voor Uw waardevolle adviezen en Uw hulp bij de correctie.

v. d. S.

# FIFTY YEARS
# OF ELECTRONIC
# COMPONENTS

U.D.C. 621.38/.39(091)

Translated from the Dutch by J. van Randeraat, Chadderton, England

No representation or warranty is given that the matter treated in this book is free from patent rights, nothing herein should be interpreted as granting, by implication or otherwise, a licence under any patent rights.

Printed in the Netherlands

# FIFTY YEARS OF ELECTRONIC COMPONENTS

## 1921 - 1971

by H.A.G. Hazeu

retired member of the Board of Management of the
N.V. Philips' Gloeilampenfabrieken

Published by Publications Department of
PRODUCT DIVISION ELCOMA of
N.V. PHILIPS' GLOEILAMPENFABRIEKEN.

Printed by the N.V. Eindhovensche Drukkerij

## FOREWORD FROM THE PRODUCT DIVISION ELCOMA

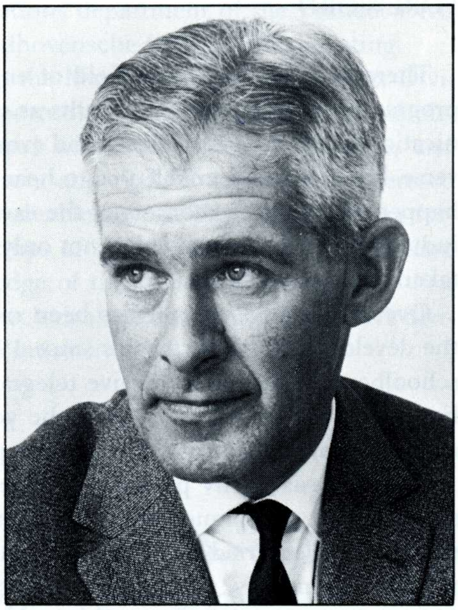
*In 1971 it will be fifty years since the Philips Company started selling radio tubes. A picture of the first catalogue, in which three types were offered, is shown on page 52. To mark this milestone we decided to publish a book that will offer a survey of the technical development of electronics, particularly the components, covering these fifty years.*

*We believe that such a book would be a welcome supplement to the wealth of literature already existing in this field but which, due to the specialized subjects dealt with, is usually limited to a certain part. We were fortunate in finding Mr. H.A.G. Hazeu willing to undertake the difficult task of writing such a book after his retirement. From 1945 to 1957 Mr. Hazeu was the Technical Manager of the Product Division Electron Tubes, and until his retirement in 1968, as a member of the Board of Management, he was responsible for the coordination of the activities concerning electronic components. That Mr. Hazeu has, to our opinion, succeeded in his task is undoubtedly due to his wide experience in this field and his constant vivid interest in engineering.*

*As most people will conclude from reading the book, the developments in electronics followed each other at ever increasing speed across the span of fifty years, and it leaves no doubt that these developments will continue to follow each other at an even higher rate during the period ahead of us. Keeping pace with these developments and, if possible, taking the lead, is the task we have set ourselves in pursuit of an increasingly wide spectrum of fascinating applications.*

*For the management*
*of Product Division Elcoma*
*of N.V. Philips' Gloeilampenfabrieken*

*H.B. Emmerink*



Dr. B.G. Dammers,
† 22nd January, 1968

It is with deep sincerity that I dedicate this book to my friend and cooperator BERT DAMMERS who unfortunately departed this life too soon. He was an electronics engineer, heart and soul, with a profound insight into the fascinating possibilities of electronics. In his function, from 1945 as the leader of the electrical group for electron tube development, later in charge of the coordination of all application laboratories for components of the N.V. Philips' Gloeilampenfabrieken, he contributed largely to the science of electronics.

His endeavours were always aimed at simple and practical solutions which might lead to popularisation of the equipment. His exceptional intelligence and charming personality enabled him to win others for his ideas.

It is with great pleasure that I remember our many discussions during which he informed me of the latest developments in circuit engineering. Indeed the development of our components had to keep pace with the modern circuit development. It is with gratitude that I hereby state that these discussions too, have been a stimulus towards compiling this book.

H.A.G. Hazeu

# FOREWORD

There is hardly any other field of engineering in which science has so rapidly progressed towards practical results as electronics. Not only did it bring communication across land and ocean, and even with the satellites now exploring the universe, but it has also enabled us to hear and see (nowadays even in colour) what is happening in various corners of the earth. Finally electronics has led to a "second industrial revolution" in which not only manual labour but also brain work can be taken over by machines.

Over the last 53 years it has been my privilege to witness from close quarters the development of electronics since 1917, (when as a fourteen-year old grammar schoolboy I managed to receive telegraphed news about the Great War with my home-made radio receiver), until the years 1945-1968 during which I supervised design and production of electronic components within the Philips Company.

On the basis of my personal experience in the industry I have tried to give a survey of the techniques which have made electronics what it is today. In doing so I have also tried to point out the relationship between fundamentals of physics, technology, and production techniques, as well as between the properties of components and their application in circuits. All this I projected against the background of the so rapidly expanding market; for the main objective of industry is, after all, to meet its demands. This is the reason for the four parts of this book: HISTORIC DEVELOPMENT, COMPONENTS, CIRCUITS, APPLICATIONS.

It is not intended as a text book, although one can certainly learn from it. Nor is it an encyclopedia even if it gives a rather complete survey of all electronic components. It should be more regarded as a guide leading the interested reader through the Wonderland of Electronics. He can orient himself about the "how" and "why", about the history, scientific backgrounds, and applications. For a more thorough study it will be necessary to consult the literature of the trade. It may well be that a specialist in one of the many fields discussed here will use this book to get more acquainted with the other fields. It has been my endeavour, however, to keep the text, figures and mathematics (limited to a minimum) intelligible also for non-specialists, be it that a certain basic knowledge of radio engineering will be needed. I hope that the reader will derive as much pleasure from reading the book as I took in writing it!

When collecting data for this book I often resorted to PHILIPS TECHNICAL REVIEW (so excellently edited since 1936) which contained a wealth of information. Furthermore there were a number of books on electronics in the PHILIPS TECHNICAL LIBRARY. And then there were also many editions of the TECHNICAL PUBLICATIONS DEPARTMENT of PRODUCT DIVISION ELCOMA, among which we find "Electronic Applications", "Matronics", "Product Information", "Application Information" and the "Data Handbook System". Lastly, I do want to mention as a valuable source the "Elonco Bulletin" published by Philips Nederland N.V. in the Dutch language.

I hereby express my gratitude to the publishers of the abovementioned publications for putting photographs and figures at my disposal, to Mr. J. Spoelstra for collecting and completing the series of photographs, to Mr. C.J.M. Gladdines for preparing the drawings, and to the publications department of the Product Division ELCOMA for publishing, and the Eindhovensche Drukkerij for printing.

Furthermore acknowledgement is due to the Management of the Product Division ELCOMA who have given me the opportunity to see my "technical memoires" printed. An editorial committee formed within this product division, and consisting of Messrs. J. Jager, P.A. Neeteson, and A.G.W. Uitjens have perused the manuscripts with me, and have given valuable advice. Then there are my former cooperators from the development groups of this product division, who have supported me with their profound knowledge of the trade. I sincerely hope that this book will also reflect the omnilateral technical potential of this product division, in which I have worked with so much satisfaction.

Waalre, 1st October, 1970

H.A.G. Hazeu

# CONTENTS

## PART I. HISTORIC DEVELOPMENT FROM THE MORSE TELEGRAPH TO COLOUR TELEVISION FROM THE MOON

## PART II. ELECTRONIC COMPONENTS

PART III CIRCUITS

## PART IV. ELECTRONIC APPLICATIONS

# PART I. HISTORIC DEVELOPMENT FROM THE MORSE TELEGRAPH TO COLOUR TELEVISION FROM THE MOON

## I.1. THE PREHISTORY OF RADIO

At the beginning of the 19th century attempts were made to convey messages by means of electric current. It was not until 1837, however, that the American artist MORSE introduced the first equipment with which the conveyed message was recorded on paper (RECORDING TELEGRAPH). First he invented a "code", according to which the letters of the alphabet and figures are represented by symbols which are a combination of dots and dashes (MORSE CODE).

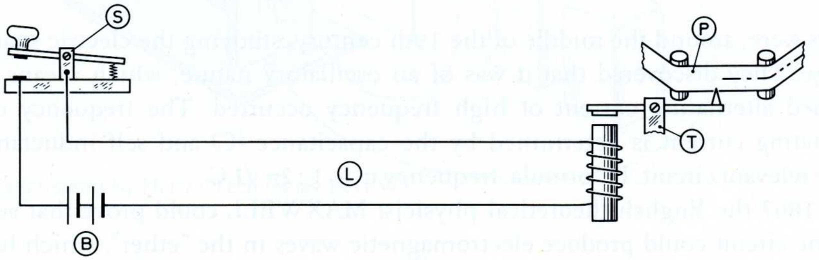


Fig. 1. Morse telegraph. S is the morse key, B the battery, L the wire, T the telegraph apparatus, P the paper tape.

Fig. 1 illustrates how these morse symbols, in the form of long and short current pulses, were fed into an electric wire by pushing a MORSE KEY (S). At the other end of the wire the current pulses energize an electromagnet in the telegraph apparatus so that a stylus is pushed against a travelling paper tape. The telegram in morse code can then be read from this tape.

In 1877 EDISON discovered that the variations in contact resistance between numerous pieces of carbon, which were mounted on a membrane, could be used to convert air vibrations caused by sound (50-2000 Hz) into electric current variations. (MICROPHONE).

As early as 1876 BELL constructed a TELEPHONE in which an iron membrane is placed in the field of a permanent horseshoe magnet whose ends are fitted with coils. If the current variations produced in the microphone are sent through these coils, the telephone membrane will vibrate at the frequency of the sound received by the microphone. A schematic telephone connection is shown in Fig. 2. A transformer is used between microphone and wire to transform the relatively high current at a low voltage, as supplied by the microphone, into a low current at a higher voltage. Thus the electric losses in the wire are minimized.

Towards the end of the 19th century several countries were using telegraph and telephone connections via overhead wires or underground or submarine cables; the costs involved were a multiple of the cost of the electric equipment.

While low-frequency alternating currents were well under way to technical application, the physicists VON HELMHOLTZ, THOMSON (Lord Kelvin) and

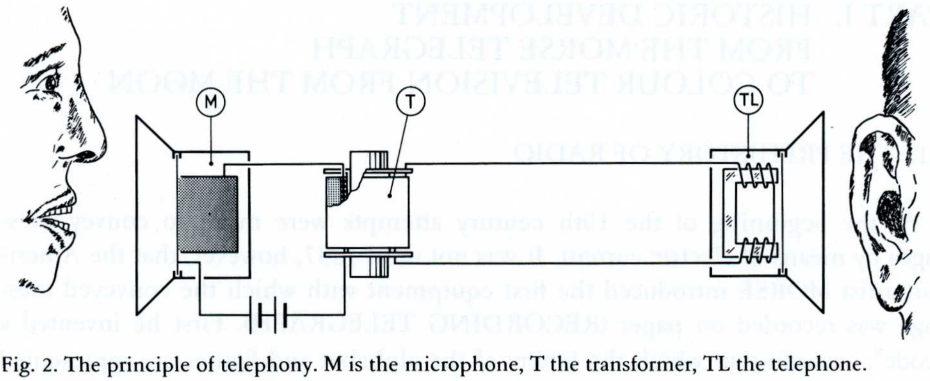


Fig. 2. The principle of telephony. M is the microphone, T the transformer, TL the telephone.

others were, around the middle of the 19th century, studying the electric spark discharge. They discovered that it was of an oscillatory nature, which means that a damped alternating current of high frequency occurred. The frequency of this alternating current is determined by the capacitance (C) and self inductance (L) of the relevant circuit. In formula: frequency $n = 1 : 2\pi\sqrt{LC}$

In 1867 the English theoretical physicist MAXWELL could prove that such an electric circuit could produce electromagnetic waves in the "ether", which had the propagation speed of light (300.000 km per second). (In analogy with the already known sound waves there is the following relationship between the frequency and wavelength of these waves: the number of cycles per second equals 300 000 000 divided by the wavelength in metres).

It was not until 1887 that the German physicist HERTZ succeeded in actually generating these waves with the arrangement as shown in Fig. 3.

A number of capacitors (in those days still in the form of so-called "Leyden jars") were discharged across the primary winding of a transformer. In the secondary winding of the same transformer, which had many more turns, this gave rise to a very high voltage which produced a spark between two metal spheres connected to the terminals of the secondary winding. The space between the metal spheres is known as the "spark gap". An electromagnetic wave was transmitted via two metal plates connected to these spheres, and at a certain distance from these plates this wave proved its existence by producing a spark between the ends of a wire bent to form an almost complete circle. HERTZ could demonstrate that the waves would pass through non-metals, were reflected by metals, and that their speed of propagation indeed amounted to $300 \times 10^6$ m/s.

At about the same time BRANLEY designed an apparatus, the COHERER, consisting of a glass tube filled with small particles of iron, which became conductive under influence of high-frequency currents. POPOFF connected this apparatus to a wire suspended high in the air (AERIAL), in order to register lightning discharges.

Finally it should be mentioned that in 1851, in Paris, RUHMKORFF constructed a piece of equipment (later to be manufactured in his work shop), with which whole series of discharge sparks could be produced. This was a high-voltage trans-

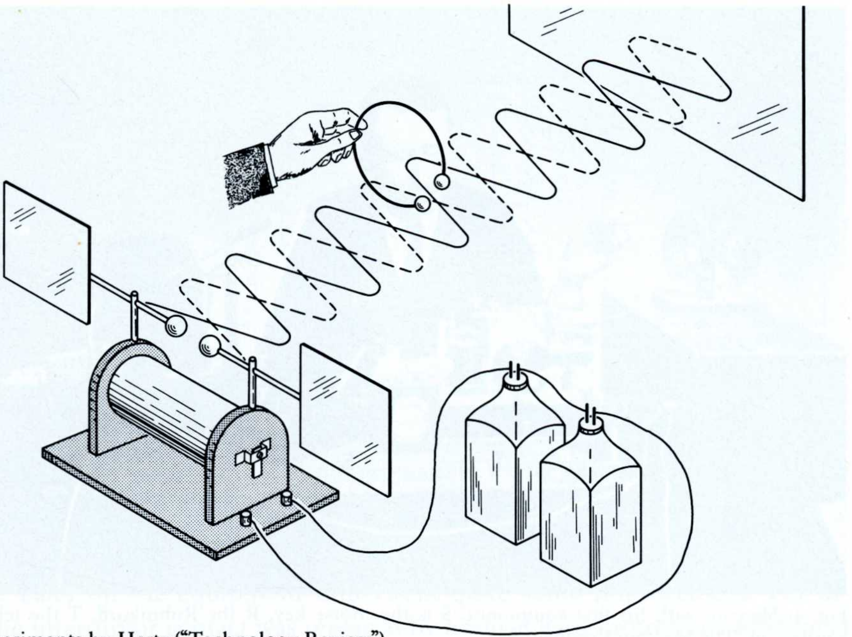


Fig. 3. Experiments by Hertz ("Technology Review").

former with iron core, whose primary current was periodically interrupted by a special contact, the "interruptor" (similar to the system of the electric bell). At every interruption a high voltage was generated in the secondary winding and produced a spark between the electrodes connected to this winding (the principle of the spark "plug" in motor cars).

Towards the end of the 19th century all the elements required for "radio" had been created. However, the electrical engineering companies of those days: Siemens, G.E., Westinghouse, Edison, Bell-Western and others, were too busy developing lighting, electric machines etc., and exploiting electricity, telegraph and telephone systems, to pay sufficient attention to this matter.

## I.2. THE "REALIZER" MARCONI

*"Invention consists in overcoming the practical difficulties of the new advance, not merely talking or writing about the new thing, but in DOING IT"*
(Sir Ambroise Fleming).

It was the twenty-two-year old Italian MARCONI who, on the top floor of his parents' villa, tinkered away at the first radio transmitting and receiving equipment (Fig. 4), and thereby established a wireless connection across a distance of two miles (1896).

Via a morse key Marconi connected a Ruhmkorff inductor to a battery, and the high-voltage winding to an improved spark gap, one sphere of which he connected to the aerial and the other to earth (see Fig. 5). Via this aerial the electromagnetic waves were transmitted high above the surface of the earth from a large area.

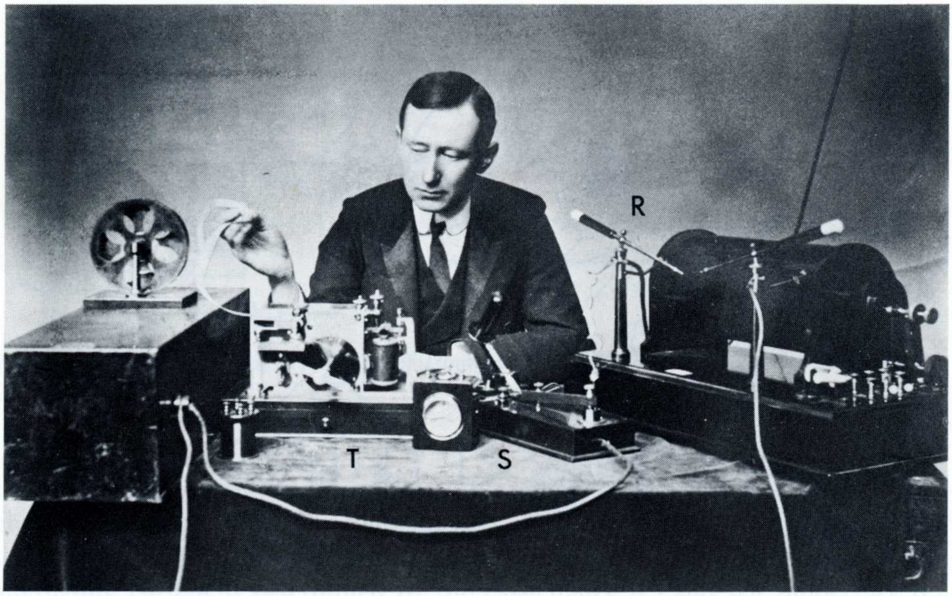


Fig. 4. Marconi with his first equipment. S is the morse key, R the Ruhmkorff, T the telegraph (by courtesy of "The Marconi Company").

At the receiver end a second aerial was connected to the coherer, and via this coherer a battery was connected to the telegraph apparatus. As soon as the morse key in the transmitter was pushed down, a train of discharge sparks occurred; the oscillation generated by these sparks was transmitted via the transmitter aerial. At the receiving end a very small part of the transmitted waves were received by the aerial and fed to the coherer, so that the latter became conductive and the telegraph apparatus drew a line on the paper tape. The only thing further needed was a "tapper", a kind of hammer constantly tapping against the coherer to prevent the iron particles from caking together after the signal had passed through.

Afterwards the system was improved by connecting an "oscillator circuit" between the spark gap and aerial-earth connection. This circuit consisted of a coil and a capacitor which made it possible to transmit a better defined wavelength. A similar circuit built in the receiver was then "tuned" to the particular wavelength of the transmitter (dashed line in Fig. 5).

Marconi's family (he had very influential parents) wasted no time. With his father he travelled to England, where he demonstrated a wireless link covering a distance of six miles. In 1897 "The British Marconi Company" was established (soon to be followed by an affiliation in America), which found the financial means to carry out experiments on a larger scale. In 1899 the first telegram was transmitted across the Channel, and then in 1901 the first telegram went across the Atlantic Ocean from England to New Foundland (the receiving aerial in New Foundland was attached to a kite). On that occasion it was found that radio waves could be propagated far beyond the horizon. At a later stage HEAVISIDE indicated that these waves were reflected by higher layers of air which had become conductive

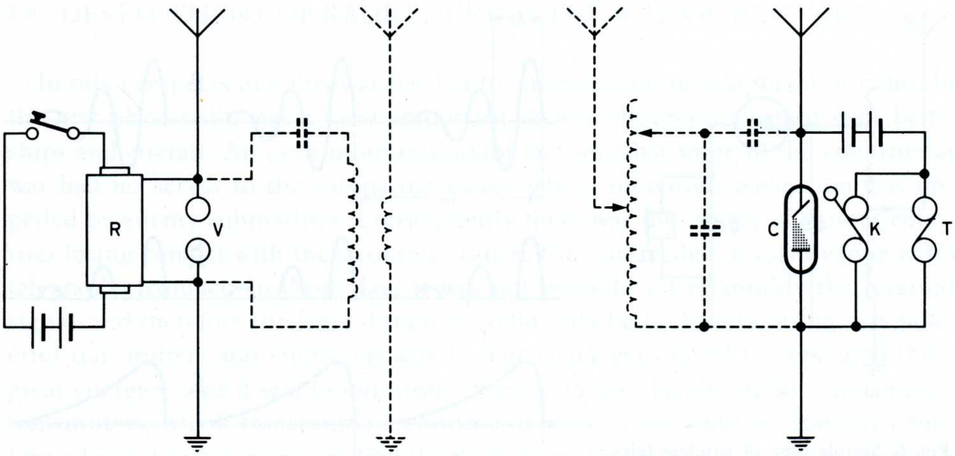


Fig. 5. Transmitter and receiver in Marconi's first experiments. V is the spark gap, C the coherer, K the tapper.

(ionized). Later on the equipment underwent quite some improvement. For instance, the interruptor contact was replaced by a wheel with protrusions rotating at a high speed, so that the spark interrupted itself. A far more powerful current source was found in the form of a dynamo; transformer, coils and capacitors were designed to take stronger currents and higher voltages, the coherer being improved by placing the iron particles in a vacuum.

With the newly designed equipment it was now possible to transmit telegraph signals across rather great distances (even across the ocean) provided sufficient power was fed to the spark gap, and the aerial was large enough. However, wireless telegraphy was only used for places which could not be reached via wire or submarine cable, the main reason being, that governmental bodies and private companies, which had invested large sums of money in telegraph routes and cables, were reluctant to be deprived of the profits their investments yielded. Such competition did not exist for ships, which now were able to communicate from the high seas with land and with each other. Marconi on his part demanded a rather high licence fee, and as a result of this, the application of his invention remained restricted to very large ships until the disaster of the "Titanic" in 1912, when hundreds of passengers perished in the lifeboats, while at a distance of no more than twenty-five miles a ship without radio equipment was passing by. Ever since that time it has been compulsory by law for ships of a certain size to be equipped with a radio installation. Finally small radio stations were established for various information services: weather forecast, stock exchange quotations, general news etc.

## I.3. THE CRYSTAL DETECTOR AND AMATEURS

At the receiving end the coherer, although improved in many respects, remained an unwieldy and unreliable element, and for years endeavours were made to substitute it by something better. There was, for instance, the magnetic detector

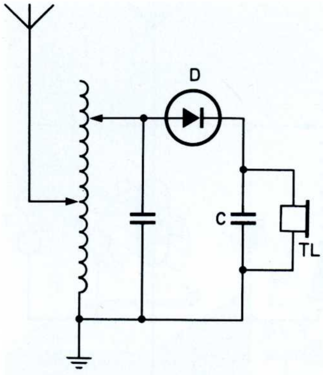


Fig. 6. Simple type of wireless telegraphy and telephony receiver. D is the detector, TL the telephone, C the telephone capacitor.

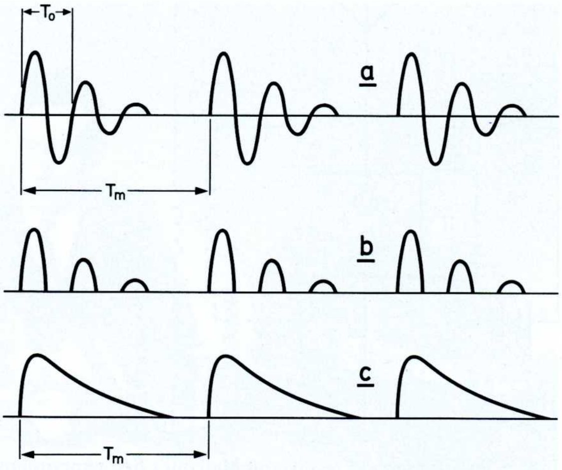


Fig. 7. Damped radio signals and detection.

(predecessor to the magnetophone). A telephone already proved a more sensitive element to detect weak alternating currents than the coherer (at least to make the signals audible). But this telephone failed to reproduce the high radio frequencies. A simple remedy was to connect a rectifier (DETECTOR) in series with the telephone, and in parallel with the telephone a capacitor (Fig. 6).

Fig. 7a gives examples of "damped" oscillations as received from the spark transmitter. $T_0$ is the period between two positive peaks (the reciprocal value of the frequency of the oscillator circuit of the transmitter, to which the receiver must be tuned), $T_m$ is the period of time between two successive sparks travelling across the spark gap. In the receiver (Fig. 6) this damped alternating current is rectified (the negative current pulses are suppressed), so that the current allowed to pass will have a form as shown in Fig. 7b. The capacitor, functioning as a large store of electric charge, will smooth the rapid oscillation so that a train of current pulses (Fig. 7) passes through the telephone, causing its membrane to vibrate at the same frequency as that at which the sparks fly across the gap in the transmitter. It is now of prime importance to choose the frequency matching the greatest sensitivity of the human ear: 1000 Hz ("tone spark transmitters").

In 1906 a particularly simple means of rectifying high-frequency oscillations was found in the form of the so-called CRYSTAL DETECTOR, a point contact made on a certain kind of crystal; for instance: a steel pin on a lead-glance or carborundum crystal (predecessors to germanium and silicon diodes, which were not developed until 1950, and have led to the design of modern transistors).

Any amateur willing to spend a few dollars to purchase a telephone and crystal, and who took the trouble of making his own coil with sliding contact, capacitors and aerial, could be the owner of a simple receiver built according to the diagram shown in Fig. 6. This receiver would enable him to listen in on radio news, an occupation which became quite popular during the war of 1914-1918, when these amateurs could receive news about the war before it appeared in the papers.

## I.4. DEVELOPMENT OF RADIO DURING THE WAR OF 1914-1918

In other respects too, this war has led to an enormous development of radio. In the first place radio was a most important means of communication with battle ships and aircraft. An even more important fact was that some of the countries at war had no access to the submarine cables while, moreover, navigation was impeded by enemy submarines. Consequently there was the danger of various countries losing contact with their colonies unless they succeeded in establishing radio telegraphic contact. In those days it was not possible yet to amplify the received signal, and therefore the large distances could only be bridged by using very powerful transmitters and enormous aerials. The spark gaps failed to cope with these great energies, and it was found more effective to use the energy for "undamped" transmitters, which transmitted a continuous wave. This could be done by coupling a large direct current arc (like the ones in arc lamps) to an oscillator circuit. A reliable transmitter came into existence when the designers of electric machines finally began to occupy themselves with radio, and designed multi-pole alternating current generators with which very high frequences could be obtained. In view of the limited available aerial length, which should preferably be ¼ of the wavelength, a wavelength range of 300-1000 m had been chosen for maritime use. The high-frequency generators could produce frequencies up to about 5000 Hz, and with special transformers these frequencies could be quadrupled to 20 000 Hz. Thus it became possible to use a wavelength of 15 000 metres radiated from gigantic aerials. The German station Nauen could thus maintain contact with Africa. A Dutch engineer of the General Post Office in the East Indies, DE GROOT, was a great pioneer of radio contact between Holland and the then Dutch East Indies. He made use of the reflecting effect of a canyon of the MALABAR, a vulcano near Bandung. He built his own arc transmitter which could radiate 260 kW from its aerial; its signals were received by the Dutch battle ship "De Zeven Provincieën", as far as the Panama Canal.

## I.5. RADIO TELEPHONY

In principle it was possible to use undamped transmitters for radio telephony by connecting a microphone in the aerial circuit. The transmitted telephone signals could be received with a simple receiver as shown in Fig. 6.

Fig. 8a shows the air vibration which the microphone converts into resistance variations. These variations in turn cause the current intensity, fed into the aerial from the high-frequency generator, to vary according to the diagram shown in Fig. 8b. This "modulated" high-frequency oscillation is received, and the detector gives it the form of the curve shown in Fig. 8c. Owing to the "smoothing" of the capacitor in parallel with the telephone, a current of the form shown in Fig. 8d will flow in the telephone, causing the membrane to vibrate at the same frequency as the sound in the microphone. In order to ensure correct sound reproduction, the transmitting frequency had, of course, to be much higher than the sound frequency.

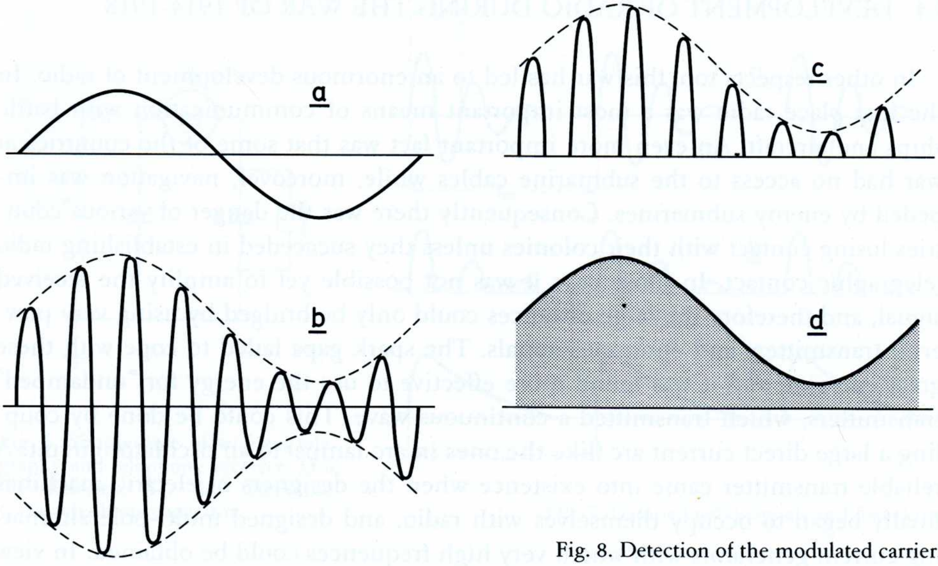


Fig. 8. Detection of the modulated carrier.

The microphone, however, could not handle high aerial powers, so that the communication distances remained rather limited. Means had to be found to amplify the weak microphone currents in the transmitter as well as the signal received. The answer to the problem was found in the form of the "THREE-ELECTRODE TUBE".

## I.6. INTRODUCTION OF THE "ELECTRON TUBE"

Edison had already observed that in the incandescent lamp which he invented, the filament would transmit an electric current if inside the bulb of the lamp a metal plate (ANODE) was fitted to which a positive voltage was applied. FLEMING discussed this phenomenon with Edison as early as 1884, but it was not until 1904 that the idea was conceived to use this phenomenon to rectify high-frequency alternating electric currents. (If the anode was positive with respect to the filament, a current would flow through the lamp, but not so if the anode was negative). This led to the development of the VACUUM RECTIFIER. Yet the crystal detector was so much simpler that the vacuum rectifier found limited use in radio reception.

DE FOREST introduced a metal grid between the filament and the positively charged anode ("THREE-ELECTRODE TUBE"), and could demonstrate that the voltage variations on this grid (which drew hardly any current so that practically no power was needed) caused the anode current to vary. Investigation by the British physicist J.J. THOMSON had already brought to light that the current between filament and anode consists of high-velocity electrons, due to which very rapid grid voltage variations could be reproduced in equally rapid anode current variations. Thus the urgent problem of "AMPLIFICATION" of high-frequency oscillations had been solved (Fig. 9a).

As a point of interest: it took Lee de Forest three weeks to collect the fifteen dollars he needed to pay the application fee for the patent on his invention (29th January 1907)

By connecting a capacitor in series with the grid it proved possible to have the grid negatively charged by the high-frequency signal. From then on the crystal "detector" was no longer needed since the vacuum tube took over its function by also operating as a "DETECTOR" (Fig. 9b).

Furthermore the anode circuit can contain a coil which is magnetically coupled to the grid circuit. The current variations in the anode circuit are transferred to

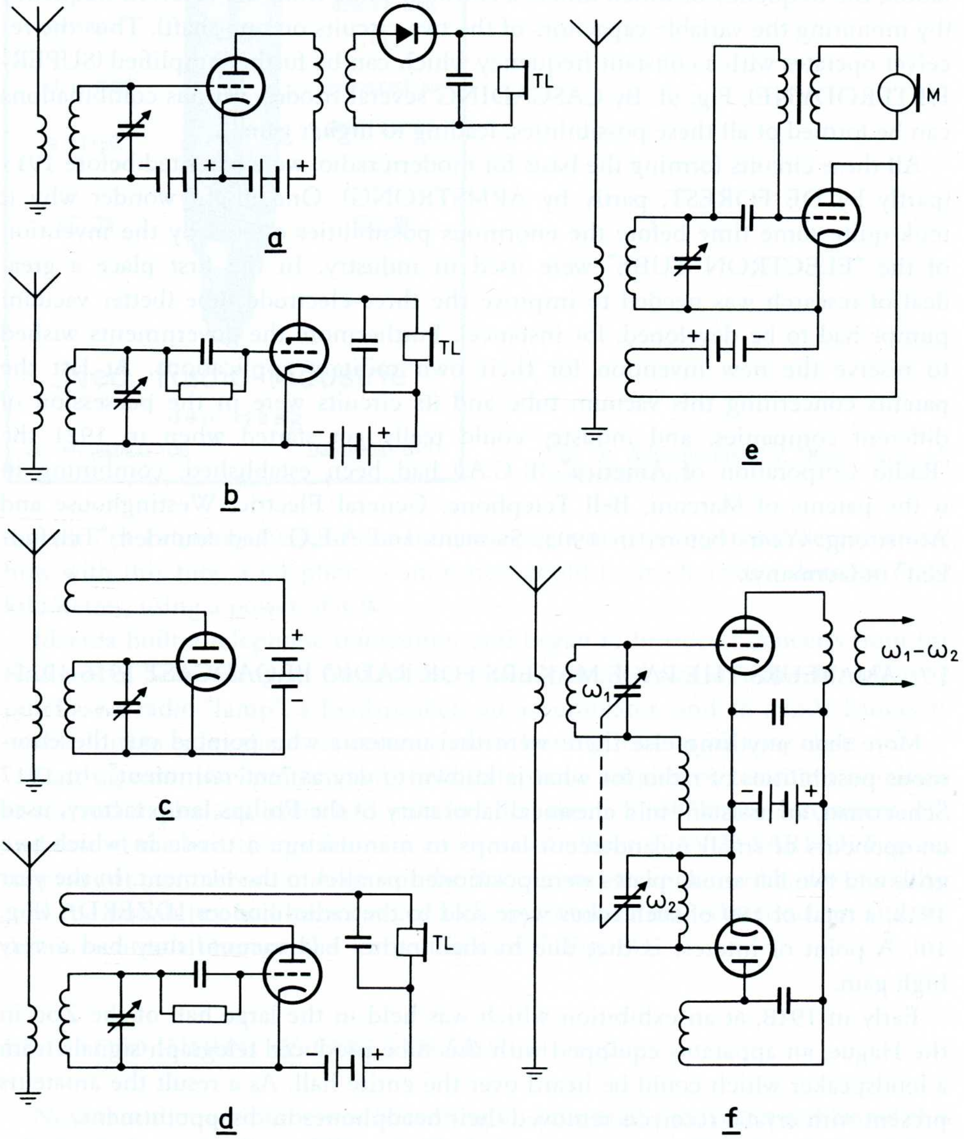


Fig. 9. Basic triode circuits: a. HF amplification, b. detection-LF amplification, c. oscillator, d. regenerative amplification, e. modulation (telephony transmitter), f. super heterodyne.

the grid circuit to be amplified again in the anode circuit. By means of this "feed-back" the system automatically began to GENERATE, thus representing a transmitter of undamped signals (Fig. 9c).

A high degree of amplification is obtained by adjusting a receiver of this design to the verge of generation (REGENERATIVE FEEDBACK, Fig. 9d).

It is also possible to apply the current variation through a microphone to the grid of such a generating tube; in that case the amplitude of the generated high-frequency oscillations will be MODULATED in accordance with the microphone currents (Fig. 9e). This is the principle of a telephone transmitter.

Finally the receiver can be equipped with a second triode to "generate" an oscillation, the frequency of which differs a constant value from the received frequency (by mounting the variable capacitors of the two circuits on one shaft). Thus the receiver operates with a constant frequency which can be further amplified (SUPERHETERODYNE), Fig. 9f. By CASCADING several triodes, various combinations can be formed of all these possibilities, leading to higher gain.

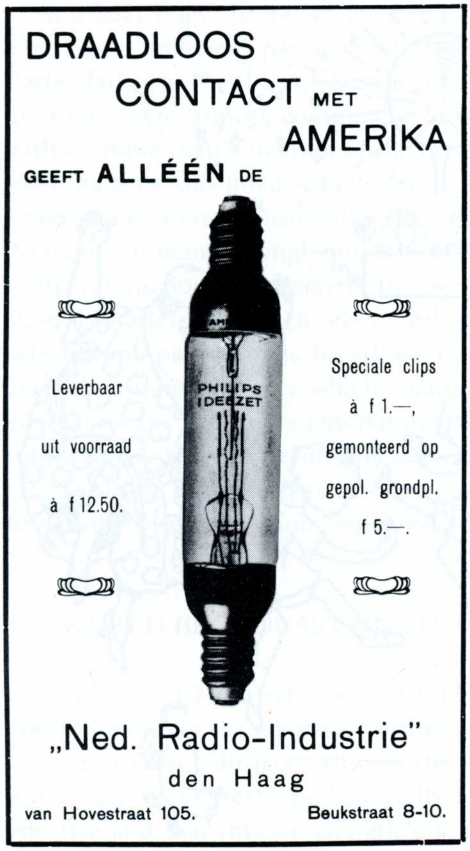
All these circuits forming the basis for modern radio, were patented before 1913 (partly by DE FOREST, partly by ARMSTRONG). One might wonder why it took quite some time before the enormous possibilities offered by the invention of the "ELECTRON TUBE" were used in industry. In the first place a great deal of research was needed to improve the three-electrode tube (better vacuum pumps had to be developed, for instance). Furthermore the governments wished to reserve the new invention for their own military applications. At last the patents concerning this vacuum tube and its circuits were in the possession of different companies, and industry could really get started when in 1921 the "Radio Corporation of America" (R.C.A.) had been established, combining in it the patents of Marconi, Bell Telephone, General Electric, Westinghouse and Armstrong. (Years before, in 1903, Siemens and A.E.G. had founded "Telefunken" in Germany).

## I.7. AMATEURS THE PACE MAKERS FOR RADIO BROADCAST 1916-1921

More than anything else there were the amateurs who pointed out the enormous possibilities of radio for what is known to day as "entertainment". In 1917 Scheerman, an assistant in a chemical laboratory of the Philips lamp factory, used components of small incandescent lamps to manufacture a triode in which two grids and two flat anode plates were positioned parallel to the filament. In the year 1918, a total of 180 of such tubes were sold to the radio-pioneer IDZERDA (Fig. 10). A point of interest is that due to their rather bad vacuum they had a very high gain.

Early in 1918, at an exhibition which was held in the large hall of the Zoo in the Hague, an apparatus equipped with this tube produced telegraph signals from a loudspeaker which could be heard over the entire hall. As a result the amateurs present with crystal receivers removed their headphones in disappointment.

After the war small transmitting tubes developed for military equipment also became available to amateurs. Philips too, could manufacture such a tube and on



DRAADLOOS
CONTACT MET
AMERIKA
GEEFT **ALLÉÉN** DE

Leverbaar
uit voorraad
à f 12.50.

Speciale clips
à f 1.—,
gemonteerd op
gepol. grondpl.
f 5.—.

„Ned. Radio-Industrie"
den Haag
van Hovestraat 105. Beukstraat 8-10.

Fig. 10. Advertisement in "Radio Nieuws", September 1918.

the 24th of February 1919, at the Industries Fair in Utrecht, it was demonstrated how with this tube a telephone connection could be made over a distance of 12 kilometres, using a power of 8 W.

Idzerda built a telephone transmitter and began to broadcast concerts from the Kurhaus in Scheveningen. And all the amateur needed was an extra 40 dollars to purchase a radio "lamp", a loudspeaker, an accumulator, and an anode battery to enable him to enjoy these concerts at home, leaning back in his easy chair.

Starting from the "professional" field (telephone amplifiers, telecommunication at sea and in the air), radio had now invaded the field of "home entertainment", and one of the first things done by an ex-naval officer, DAVID SARNOFF in his function of commercial manager of the R.C.A., was to convince his directors that now the possibility had been created to bring information, education, art and entertainment to all homes: the "ear to the world" (Fig. 11).

## I.8. RADIO INDUSTRY TAKES OVER (1921-1940)

Now that the radio tube was becoming an industrial product, it was found that various improvements could be made: the vacuum was improved and maintained thanks to the so-called "getter" inside the tube; the cylindrical configuration (grid

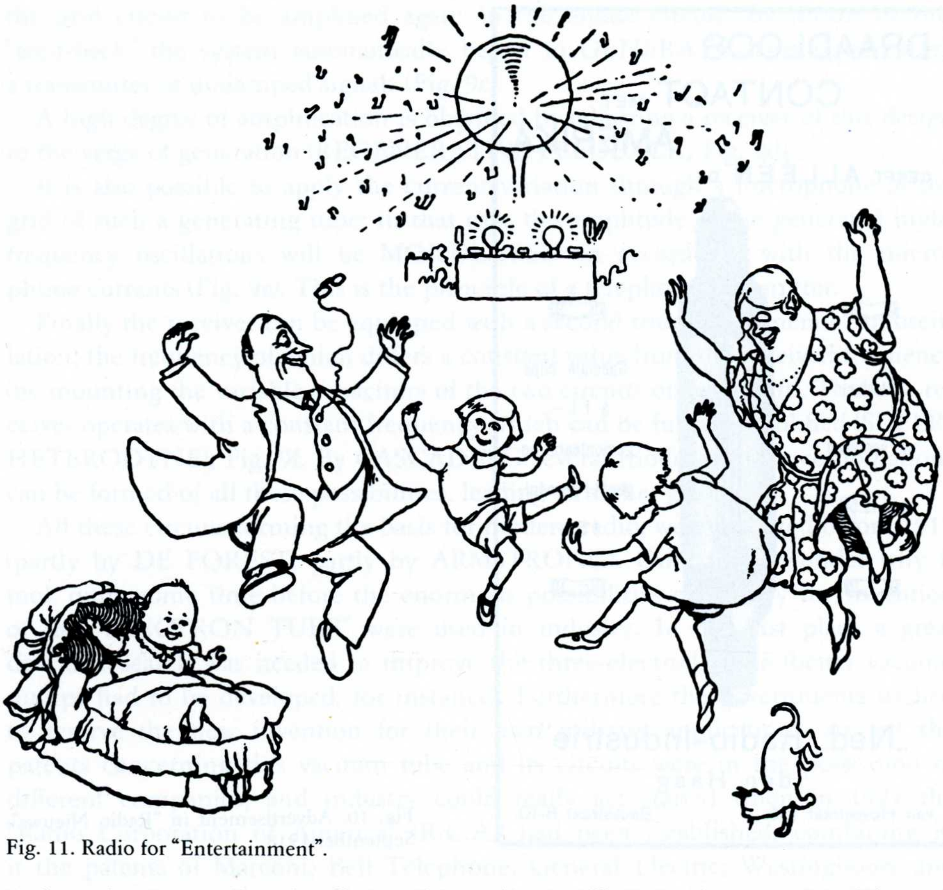


Fig. 11. Radio for "Entertainment".

and anode surrounding the filament) gave better efficiency; by using special materials for the cathode current consumption could be reduced; radio tubes fitted with so-termed indirectly heated cathodes could be fed from the mains, the anode voltage being supplied by a vacuum rectifier; insertion of more grids yielded higher amplification (1925 "PENTODE" etc), and whilst around 1960 the radio industry of the free world was turning out one milliard radio tubes per year, these tubes, five times more effective, were sold at one fifth of the price of the first tubes.

The radio components industry started with loudspeakers, special low-frequency transformers and rectifiers (mains sets), high-frequency coils and variable capacitors, whereas the amateur manufactured his own fixed capacitors (from paper and tin foil) and resistors (pencil stripes on card board); later these components too, were mass produced. In the years 1925-1930 industry began to manufacture complete radio receivers. Transmitters too, entered the industrial atmosphere: transmitting tubes, originally large-sized radio tubes, had to handle greater powers and were made of water-cooled metal vessels sealed to a glass envelope (with their experience in X-ray tubes as a basis, Philips could apply the ferrochrome-glass seal).

Industry began to build transmitters (in 1921 Marconi built an experimental broadcast station in England), and half way 1923 the "Nederlandsche Seintoestellen Fabriek" in Hilversum broadcast their first concert from the "singing towers" on the new factory grounds.

At a later stage the broadcast stations in Europe were governed by the G.P.O. There were sufficient parties interested to reach the growing number of listeners. Particularly in North and South America the radio stations, studios and programmes were almost completely financed by businessmen who were satisfied with no more than a fraction of the broadcasting time for advertising, while the remaining time was filled with concerts, radio plays, sport news etc. In Europe the programmes were sponsored by the Government, the broadcasting system in the Netherlands being strongly influenced by religious societies and political parties.

By international agreement the wavelengths were allocated lest the stations should interfere with each other, and in Europe and America every listener could select from programmes broadcast by about ten stations which were well received, and another fifty which qualified as intelligible. Radio had become the general means for "home entertainment". Radio sets developed into pieces of furniture in the form of beautifully carved wooden cabinets with large station scales etc. Yet there were cheaper versions on the market, so that people who were in a financially weaker position could also afford radio listening.

## I.9. WORLD BROADCAST ON SHORT WAVE

In Chapter I.2 the reflection of higher air layers has already been mentioned. Again there were the amateurs, sitting up all night at their home made equipment to contact their colleagues all over the world, who discovered that it was for short waves (say 30 metres), and especially at night, that this reflection was the most effective and that this wavelength was particularly suitable to bridge long distances. The "Nederlandsche Seintoestellen Fabriek" used a 4 kW watercooled Philips tube to build a 30 metre transmitter in a packing case on the factory grounds (Fig. 12). On the 13th of May 1925 a telegram sent with this "packing case" transmitter was received in Indonesia. This took place only two months before the festive opening of the official radio link between Holland and Indonesia, which had cost about a million dollars spent on machine and arc transmitters, with powers of thousands of kilowatts, set up in large buildings and with gigantic aerials. Another example of how quickly work of previous generations is outdated in electronics!

It was then found also possible to transmit telephone signals across the world by means of a transmitter equipped with a special transmitting tube, which could handle the high frequencies required.

The Philips laboratory at Eindhoven developed a powerful transmitting tube to build an equally powerful telephone transmitter, and on the 1st of June 1927 Queen Wilhelmina was able to speak to the people in the East and West Indies via this transmitter (Fig. 13).

Thus a "world broadcasting system" had been created. From which became known as the "happy station" of the World Broadcasting Company at Huizen (with its rotating aerials which could be pointed to the East or West Indies), life reports of football matches, for instance, could be followed all over the world.

Initially special sets had to be built for this "short wave" reception, but, thanks to the application of the "heterodyne" principle, it became possible to convert both

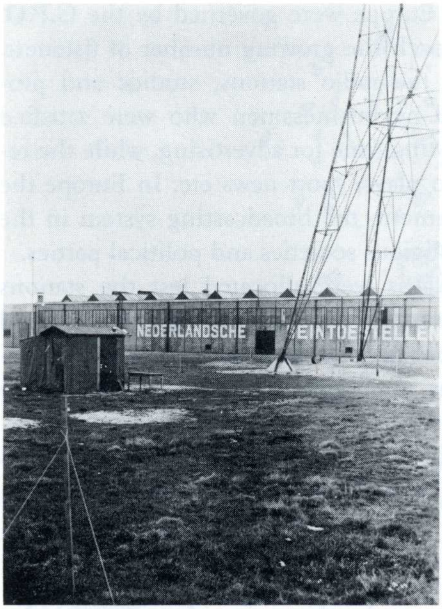


Fig. 12. The "packing-case" transmitter which established a telegraph link with Indonesia.

Fig. 13. Queen Wilhelmina addresses the people of the East- and West-Indies via V.H.F.-radio (Philips Technical Review).

the "long" and the "short" waves to one and the same "intermediate frequency" which is amplified in the receiver. Now it was possible to build receivers for wavelengths ranging from 16 to 2000 metres, the various "bands" being selected by switching the high-frequency part.

Although it was not until 1950 that it was put into practice for broadcasting purposes, the principle of "FREQUENCY MODULATION", demonstrated by ARMSTRONG as early as 1933, must be mentioned as the latest development of radio for entertainment. According to this system it is not the amplitude but the frequency of the carrier that is modulated by the tone frequency.

Armstrong found a circuit which could be used in the receiver to transform these oscillations again into the tone frequency.

The advantage of this system is that there is far less interference from atmospheric discharges (at a later stage also from electric machines and motor-cars and -bikes) and neighbouring stations. Furthermore there is the advantage of a wider frequency range for the transmitted sound oscillations, which the Americans gave the characteristic name of HI(gh) FI(delity). During the years leading on to the second world war, Armstrong had great difficulty in obtaining a wavelength allocation for experimental purposes. This was mainly due to competition from television which was on the march. However, after the war his system proved its great value for a qualitatively better reception (especially of music). After these receivers had in turn been simplified, all industrialized countries went in for F.M. broadcasting. F.M. modulation for sound was introduced for television too.

The aforementioned improvement of radio transmission (telegraphy and telephony) by means of receiving tubes, transmitting tubes, short wave and F.M. has

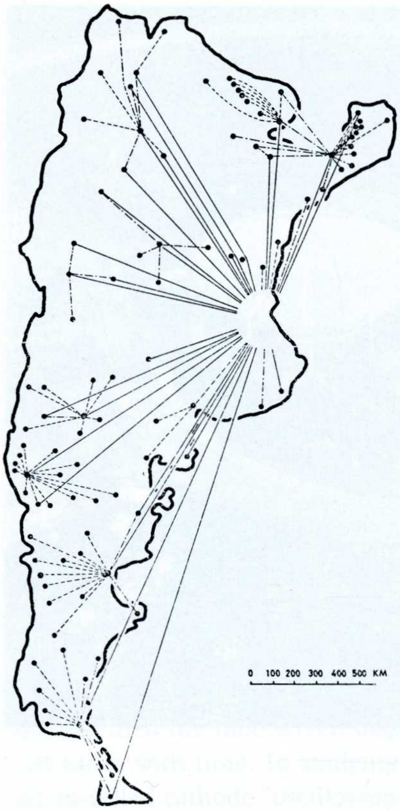


Fig. 14. Radio telecommunication network in Argentine.

apart from "entertainment" also had a great influence on professional telecommunication. Owing to interference caused by thunder and varying propagation through the atmosphere, radio telephony by air has never been as reliable as communication by wire or cable. Nevertheless, an opportunity had presented itself to establish immediate contact between any two or more places on the earth, and that at relatively low cost (the required power is low and the transmission line free). A fascinating example in this respect is a project carried out in Argentina after the war. It concerned the establishment of a network of radio transmitters and receivers, for both telephony and high-speed telegraphy, which ensured continuous contact between the capital Buenos Aires and remote places, some more than 2000 km away and often isolated by mountain chains and extensive swamps (Fig. 14). To prevent the telegrams from being distorted by atmospheric interference, the T(eletype) O(n) R(adio) system came into use, in which a character was automatically repeated if it underwent unacceptable distortion under way.

## I.10. DEVELOPMENTS IN LOW FREQUENCY AMPLIFICATION

It stands to reason that the "amplifying tube" should also find a wide field of application in what is known as low-frequency engineering. Currents having become very weak owing to the long distance they had travelled through telephone wires, could be amplified.

Fig. 15. Eucharistic Congress in Bogotà (1968).

In places where large crowds gathered, microphone currents could be amplified so that music or speech was made audible to thousands from gigantic loudspeakers, PUBLIC ADDRESS, (Fig. 15). Sound quality could be improved by designing better loudspeakers and amplifying tubes, by the so-called "PUSH-PULL" circuits, and the negative feedback system invented by TELLEGEN. Later on HI-FI-sound reproduction could be made still more realistic by stereophony and artificial reverberation.

In electric control and measuring too, the facility of a practically unlimited amplification of currents hardly indicated by means of the existing instuments, was an enormous step forward. Also thanks to the amplifying tube the pick-up element for gramophones could be made much lighter, which resulted in a longer life for the records and improved quality of sound reproduction.

## I.11. TELEVISION IS BORN

The second important item in electronics was the "BRAUN TUBE" or cathode ray tube which existed even before the radio tube (1897), Fig. 16a. In this tube a linear electron beam is thrown onto a fluorescent screen. The beam can be deflected in two directions at right angles with each other by means of coils mounted outside the tube, the deflection being proportional to the current through these two coils. By applying different varying voltages to the mutually perpendicular

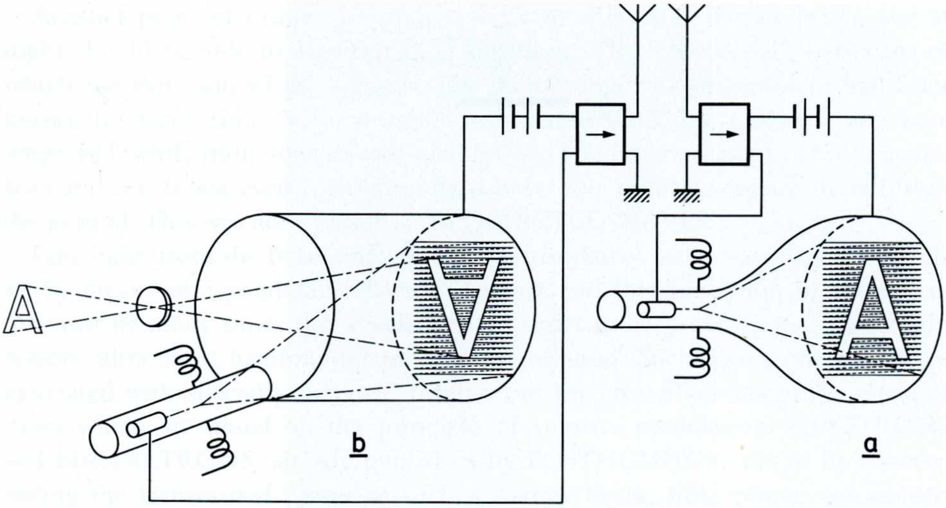


Fig. 16. T.V. transmission at left hand side and T.V. reception to the right.

coils, two electrical phenomena in their mutual relationship can be displayed on the fluorescent screen (the "characteristics" of a radio tube, for instance). Another possibility is to feed a current which increases linearly with time through one set of coils; then the tube screen displays how the electrical phenomena on the other coil varies with time. In studying high-frequency periodic electrical phenomena, the so-called cathode "oscilloscope", fitted with such cathode ray tube, was a most important tool.

If the electron beam is made to describe a zig-zag on the screen by applying currents appropriately varying with time, and if furthermore the intensity of the electron beam is made to vary in accordance with the brightness of each point of the picture, this type of tube can be used for reproducing black and white images in the form of a line pattern (Fig. 16a).

The required variation in current intensity is obtained by having the electron beam of a similar tube in the transmitter describe the same zig-zag (Fig. 16b). On the screen of this tube a light pattern is projected, and the screen is coated with a material which under influence of more or less light produces more or less current. This is basically how a CAMERA TUBE (ICONOSCOPE) operates, which was constructed by ZWORIKIN in 1928.

The high-frequency oscillations radiated by the transmitter can be "modulated" in accordance with the variations in beam current of the camera tube, and thus the wireless transmission of images, TELEVISION, was born! There is one difficulty: the scanning of the picture must take place at a very high speed: to avoid "flickering" the zig-zag pattern on the screen must be repeated 25 times per second. To ensure sufficient picture sharpness, about 500 000 dots on the screen must be scanned and reproduced, which means that all in all 25 × 500 000 = 12 500 000 times per second a piece of information is passed on. The electron beam in the Braun tube is rapid enough, but the transmission frequency must be a multiple of the modulation frequency to ensure satisfactory reproduction of the modulation.

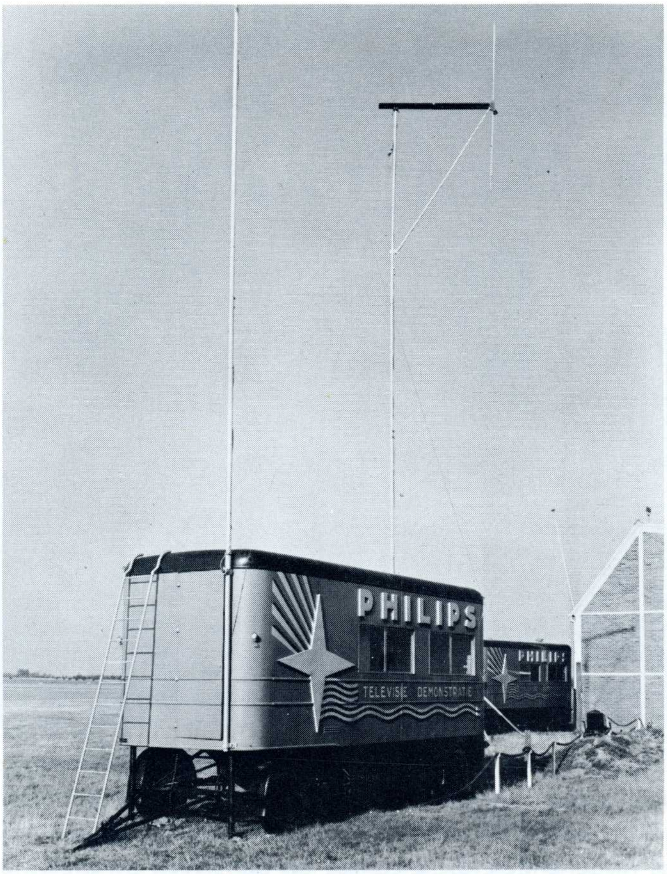


Fig. 17. Philips' T.V. caravan (1937) (Philips Technical Review).

Thus we arrive at a minimum frequency of 50 million Hz (50 mega Hertz), or a maximum wavelength of 6 metres for television. The development of ultra short wave radio systems over preceding years has largely contributed to the development of tubes, components and equipment to "handle" these high frequencies in the transmitting and receiving installations.

Ever since 1929 television transmissions had taken place in Great Britain, initially by means of a mechanical system with rotating mirrors. Round the year 1940 transmissions with 405 lines per second were started. Onwards from 1937 PHILIPS travelled Europe with a television caravan (Fig. 17). In the U.S.A., it took years of battle over the question what system ought to be standardized, before permission was granted in 1941 to transmit according to the standard of 525 lines. The second world war put an end to all these experiments.

## I.12. THE SECOND WORLD WAR BRINGS "RADAR"

In this war, which was being fought at sea, in the air and on land with mobile motorized units, communication between and with these units was a matter of life and death. Obviously, therefore, the mobile transmitters for short wave underwent quite a revolutionary development.

Another point of prime importance was that ships and aircraft in fog and at night should be able to establish their positions. The "loop aerial", by means of which the direction of the transmitting station could be determined, had been known for some time. Various ingenious systems (DECCA, LORAN, etc) were employed, while numerous transmitters formed a radio orientation network across land and sea. It was even more important to be able to locate enemy aircraft from the ground. This was done by means of DIRECTED WAVES.

Like light (from the headlamps of a car, for instance), radio waves can be directed by means of a parabolic reflector. To that end the dimension of the reflector must be many times the wavelength. In order to keep the dimensions within reason, ultra short (centimetre) waves must be used. Such waves can indeed be generated with specially designed triodes, but for greater powers use is made of tubes which are based on the principle of velocity modulation: KLYSTRONS and MAGNETRONS (already published by POSTHUMUS in 1935). By concentrating the transmitted power to such a narrow beam, little power was needed to cover great distances. Then began the era of LINK or BEAM TRANSMITTERS which, using powers of no more than 5-10 W, could transmit television programmes and telephone calls over distances of about 100 km between the various stations.

The phenomenon that a metal object reflects these waves can be used to locate such object. So if an aircraft lands in the beam of the parabola, a reflection is received: the parabola points in the direction of the aircraft. By transmitting a short burst (pulse) of high-frequency power it is possible to determine the distance of the aircraft from the time elapsing between transmission of the pulse and reception of the reflection. This time is 2 × distance in km, divided by 300 000, seconds,

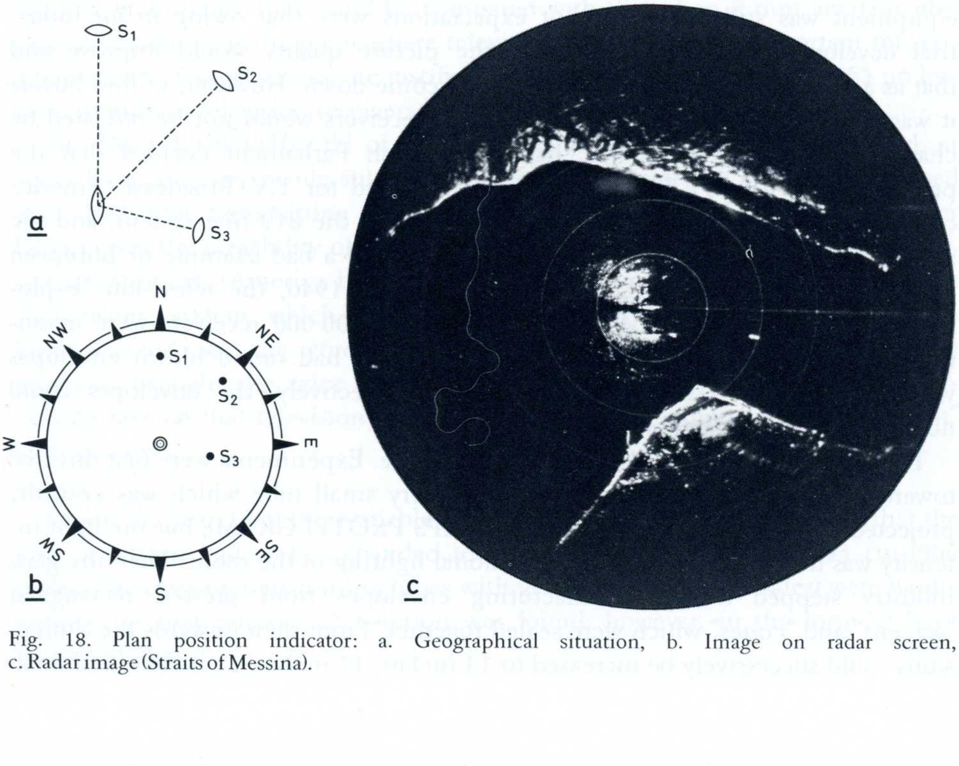


Fig. 18. Plan position indicator: a. Geographical situation, b. Image on radar screen, c. Radar image (Straits of Messina).

and can be measured with an electron beam moving across the screen of a cathode ray tube at a very high speed. Here we have the principle of RADAR.

If the parabolic aerial is rotated round a vertical axis, and the electron beam travels very rapidly from the centre of the screen to the edge like a pointer, and if this pointer is made to rotate at the same speed as the aerial, the screen will show the reflected pulse in a place representing the position of the aircraft on a map drawn on the screen. This is the principle of the P(lan) P(osition) I(ndicator), Fig. 18.

The fact that all the elements required for this complex ultra-short-wave-pulse-transmitting and receiving system were developed in such a short time deserves full admiration. Radar is nowadays used on airports and waterways, on ships, in aircraft and even in tanks, but the principle is still the same as developed by Britain during the second world war under the constant threat from enemy aircraft and V1 missiles.

At a later date HOLLANDSE SIGNAAL developed a system according to which guns were automatically aimed at targets located by radar.

## I.13. DEVELOPMENT OF TELEVISION AFTER SECOND WORLD WAR

When the world war had come to an end, the electronics industry, having played such an important role in this war, had to direct its enormous capacity towards other products. This aim would be achieved if the television receiver could become "the eye on the world" in every home as the radio had already become "the ear to the world". As already stated in Chapter I.11, there had been experimental T.V. transmissions in Britain and the U.S.A. until 1941. Technically the equipment was still imperfect, but expectations were that owing to an industrial development of the equipment, the picture quality would improve and that as a result of mass sales the price could come down. However, to find buyers it was necessary to convince them that their receivers would not be outdated by changes in the system. Hence, in 1946, British Parliament decided that the pre-war system of 405 lines should be maintained for T.V. broadcast. America continued with 525 lines; later France decided for the 819 line system, and the rest of Europe adopted the 625 line system: indeed a bad example of European unity. On the basis of the results obtained until 1940, the television "explosion" started in Britain: in 1948 a total of about 100 000 receivers were manufactured. In those days the cathode ray tubes still had round blown envelopes with diameters of 9 inches and 12 inches respectively; the envelopes could not be blown much larger (Fig. 19).

This picture was too small for "living-room" use. Experiments were first directed towards a very light intensive picture on a very small tube which was optically projected on a screen of frosted glass (PHILIPS PROTELGRAM), but the light intensity was too low in case of some additional lighting in the room. Then the glass industry stepped in by manufacturing envelopes from pressed rectangular "screens" and "cones" which were sealed together. From then onwards the dimensions could successively be increased to 14 inches, 17 inches, 21 inches, and final-

Fig. 19. Television receiver with 9 inch tube.

ly to 25 inches. It may be noted as a point of interest that these glass pieces (Fig. 20) can only be manufactured economically if the annual production totals at least 500 000. In all countries except the U.S.A., Europe, Japan, and Brazil, they still have to be imported.

The larger dimensions, the various measures taken to increase the brightness and contrast of the picture, together with the definition given by the 625 lines, resulted in a picture that could be compared with that of an 8 mm amateur film projection. This was the point where television indeed became important for "entertainment". Since 1950 we are justified to speak of an explosion; in 1955 no less than 10 million television receivers were in use throughout the free world!

The IMAGE-ORTHICON of R.C.A. made it possible to do camera work at lower light intensity inside studios or outdoors. Transmitters could be equipped with short wave transmitting tubes. The beam transmitters discussed in Chapter I.12 opened the possibility of broadcasting one and the same programme via several transmitters. (America built nation wide "Networks", Europe organized the "Eurovision" system which made possible that in 1956 nearly everybody in Europe could see the coronation of Queen Elisabeth). The large number produced brought the price of T.V. receivers down to 150 or 200 dollars (excluding tax), so that television truly could become a medium of "entertainment" for the masses.

Round the year 1960 the available and suitable wave bands were full, so that the wave spectrum had to be expanded towards even shorter waves of 33 cm (900 MHz). The existing transmitting tubes with grids (triodes and tetrodes) were hardly suitable for that purpose. A solution was found, however, in the form of large KLYSTRONS.

Fig. 20. Pressing the glass screens of rectangular picture tubes.

## I.14. COLOUR TELEVISION

Once the problems concerning the picture quality, price and reliability of black-white receivers had been solved, the electronics industry began to take an interest in colour television. In principle the possibility of it had long since been indicated: through coloured filters, red, green, and blue, three camera tubes take pictures in these respective colours and transmit them via three transmitters. The receiver is fitted with three cathode ray tubes whose screens light up in the respective colours red, green and blue; all three pictures thus reproduced are simultaneously projected. Especially at the receiving end, optical composition of the three coloured pictures was a difficult job, but eventually some types of tube were invented which could reproduce the complete coloured picture. The starting point was the phenomenon that dots or lines in the three colours placed close together are seen as a "mixed colour". The R.C.A. were the most successful with their so-called "shadow mask" tube, in which a metal plate containing roughly 500 000 small holes was placed between the three cathodes (for the three coloured pictures) and the fluorescent screen. This screen itself consists of 500 000 red, 500 000 green, and 500 000 blue dots arranged in such a pattern that the electron beam from the cathode producing the "red signal" lands on the "red dot", and so on (Fig. 21).

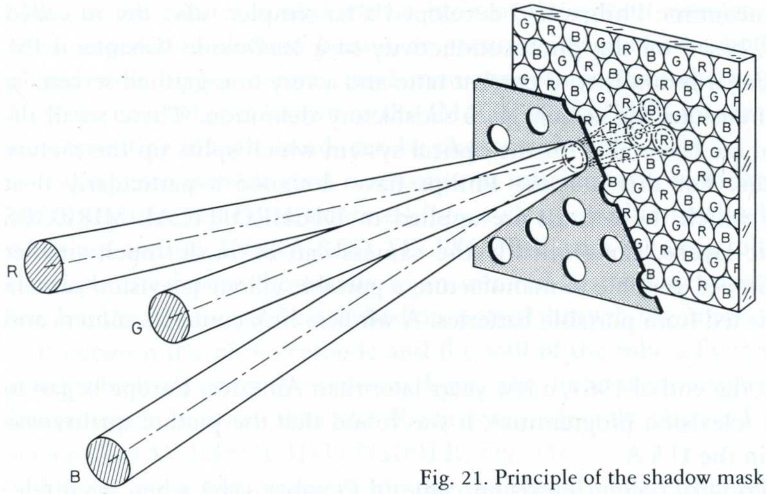


Fig. 21. Principle of the shadow mask colour picture tube.

As early as 1954 the R.C.A. picture tube factory in Lancaster (U.S.A.) started a large-scale trial production of these tubes, and when seven years later colour television broadcast began to develop in the U.S.A., they had gathered so much experience that this tube functioned satisfactorily and reliably. It will probably remain an unanswered question whether the two other developed types of colour picture tube (discussed in Chapter II.C.9.) would have provided an even better solution if they had received the same amount of interest and research as the "shadow mask" tube.

As a camera tube the "Image Orthicon" proved less suitable for this purpose; on the one hand because of its large dimensions, on the other since the luminosity versus current graph is not linear so that in principle it is impossible to mix the

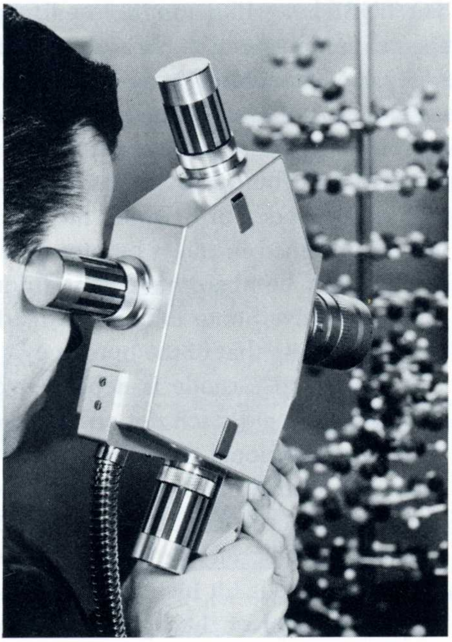


Fig. 22. Philips mini colour camera (1968)
(Philips Technical Review).

colours. In the meantime Philips had developed a far simpler tube, the so-called "PLUMBICON", based on the photoconductivity of a lead-oxide (Chapter I.15). This tube has a linear luminosity to current ratio and a very fine-grained screen, so that at small dimensions the screen gives satisfactory definition. These small dimensions in turn are favourable for the optical system which splits up the picture into three colours. For this purpose Philips have designed a particularly neat prism system to which the colours are applied in DICHROITICAL MIRRORS (vaporized atomic layers). Consequently the camera can be made much smaller and it has even proved possible to manufacture a portable colour television camera (Fig. 22), which is fed from portable batteries. A wireless link connects camera and transmitter.

When towards the end of 1967, a few years later than America, Europe began to broadcast colour television programmes, it was found that the picture quality was better than that in the U.S.A.

The great triumph of colour television came in October 1968 when via a telecommunication satellite (Chapter I.24) the Olympic Games in Mexico could also be received in Europe and Japan.

## I.15. MEASURING, AMPLIFYING, AND TRANSFORMING LIGHT

Already at the beginning of this century it was discovered that thin layers of certain metals (e.g. caesium-antimony) when exposed to light will emit electrons (PHOTO EMISSION). If a metal plate (cathode) in a tube or a glass wall is coated with such a layer, and if the tube contains a second electrode to which a positive voltage is applied, a current will flow between the "photo cathode" and the anode, which is about proportional to the illumination on the photo cathode. Such PHOTO TUBES can therefore be used to measure luminous intensity. By giving it a special composition, such a photo cathode can be made sensitive to white, infrared, or ultraviolet light.

By means of a gas filling it is possible to increase the current ("ion current") and even to obtain an explosive current generation. This is the principle of the PHOTO RELAY which is used, for instance, to control the flame in heating systems with oil burners. Furthermore the electron current emitted by the photo cathode can be amplified in stages by means of a series of anodes which, when exposed to the electron bombardment, emit more electrons (secondary emission). Thus a highly sensitive photo tube is obtained: the PHOTO MULTIPLIER with a sensitivity exceeding that of the human eye.

If the photo cathode is deposited on a flat screen on which a photographic image is projected, each dot on the screen will emit a quantity of electrons which is about proportional to the brightness of each dot. With suitable electrodes it has been found possible to produce an electrostatic field (and/or a magnetic field by placing a coil over the tube) which functions as an optical system.

These fields accelerate and focus the electrons to such an extent that the other end of the tube is hit by electron beams the intensity of which forms a geometrically true image of the light picture projected on the photo cathode. If the end plane of the tube is coated with a layer of fluorescent powder, which lights up when exposed to electron beams, an accurate image of the photo cathode is displayed with higher contrast and more brightness than the original projection. Here we have what is known as the IMAGE AMPLIFIER by means of which, objects travelling at high speeds can be photographed.

If the photo cathode used is sensitive to infrared rays, an infrared image, invisible to the eye, can be transformed into a clearly visible black-white image. Such tubes are used in INFRA-RED TELESCOPES with which a soldier operating in the dark can expose the enemy to an infrared beam, which neither of them can see, but the enemy can be watched or spotted through the image transformer.

If between the photo cathode and the wall of the tube a fluorescent layer is applied which converts X-rays into light, the fluorescent screen will display an image much brighter than the X-ray image directly observed on a fluorescent screen (X-RAY IMAGE INTENSIFIER, Fig. 23).

This is most advantageous for the radiologist; firstly the patient can be exposed to far less radiation, secondly the brighter image can be observed in a light room.

Another form of light-electricity transformation is PHOTO CONDUCTIVITY, in which — under influence of light — electrons are released in certain semiconductors. The SELENIUM CELL, in which a current is generated in a circuit between the two electrodes when the cell is exposed to light, had long been known. It provides the simplest type of PHOTO METER. It was then discovered that the resistance of CADMIUM SULPHIDE plates varied under influence of light. These resistance variations result in current variations when a battery is connected to these cadmium sulphide plates. This is a very simple PHOTO CELL as used in automatic exposure meters in photo cameras. In principle photo conductivity is much slower than photo emission.

Finally it was found that antimonytrisulphide or leadmonoxide could be deposited on a glass wall to form a thin layer whose resistance decreases when exposed to light. By causing an electron beam to describe a zig-zag pattern (Chapter I.11, Fig. 16b) across such a layer, the electron beam current will vary in accordance with the local resistance of the screen, which in turn is governed by the intensity of the incident light. Here we have the principle of the modern VIDICON and

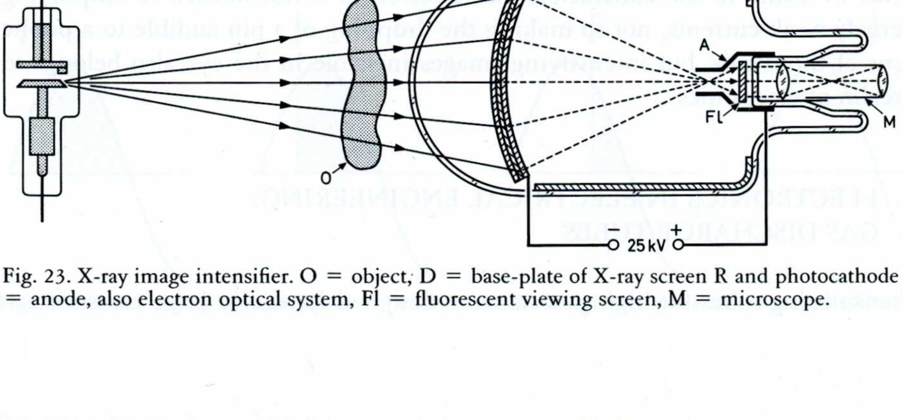


Fig. 23. X-ray image intensifier. O = object, D = base-plate of X-ray screen R and photocathode K, A = anode, also electron optical system, Fl = fluorescent viewing screen, M = microscope.

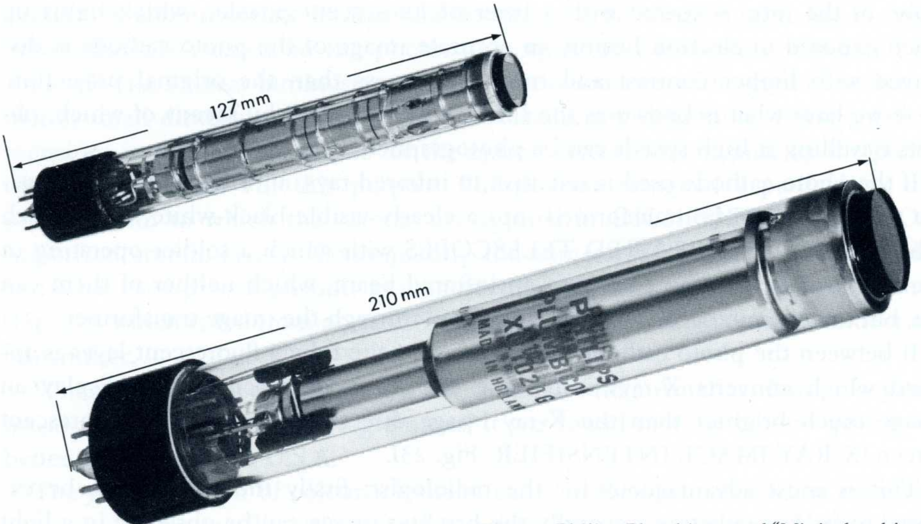


Fig. 24. Philips Plumbicon and "Mini-plumbicon".

PLUMBICON, both T.V. camera tubes (Fig. 24). They are of surprisingly simple design and allow of great picture sharpness. The operating speed of the Plumbicon is sufficient for televizing moving objects.

The light amplification of the aforementioned image intensifiers can be multiplied by using an optical system to project the fluorescent image of one tube on the photo cathode of a next one. Such an optical system requires rather expensive lenses and a great deal of light is lost. A better method of transferring the image is found in FIBRE OPTICS, in which the wall of the envelope is made of a disc consisting of numerous small tubes which conduct light without any diffraction. Telescopes, so sensitive that they provide a clear view of a scene with stars as the only light source, are based on the same principle.

Under conditions of sufficient amplification, the chain from camera tube to television picture tube can also provide important light amplification (X-RAY T.V.). In the X-ray field this system is already widely used by photographing the picture of the image intensifier with a T.V. camera to reproduce it on an ordinary T.V. receiver. Here too, sharpness and contrast can be improved by using fibre optics between image intensifier and camera tube.

Thus we come to the conclusion that electronics is not limited to amplifying extremely weak currents, nor to making the dropping of a pin audible to a public of tens of thousands, but intensifying images invisible to the eye also belongs to the realm of electronics.

## I.16. ELECTRONICS IN ELECTRICAL ENGINEERING: GAS DISCHARGE TUBES

Transmitting tubes for high powers were designed for use in large transmitters.

Such tubes found industrial application too. By means of high-frequency magnetic fields it is possible to heat metal objects very uniformly, a process which is very important in hardening and brazing. It is a special advantage that during the process the metal object can be contained in a glass vessel which may be evacuated or filled with a reducing gas. Shorter waves can be used for the electrostatic heating of insulating materials such as wood and plastic, so that this high-frequency technique has also found a field of application in glueing wood (plywoods), and in the plastics industry. Another noteworthy application is the "magnetron cooker" in which deep-frozen meals are rapidly warmed.

The drawback of the vacuum tube is that the available anode current is at the most equal to the current the cathode can produce. Matters improve when the tube is filled with a gas or metal vapour: the electrons can then "IONIZE" the gas and the resulting electrons and ions can transport a multiple of the cathode current. This was the principle of the gas filled or mercury-vapour rectifiers which were used in X-ray equipment and transmitters, especially for high currents. (For high-frequencies these tubes are useless because the "ions" are too slow).

Under certain conditions the ion current must be initiated by "ignition" of an auxiliary electrode, and so it is possible to make an electronic relay (a weak current on the auxiliary electrode gives rise to a high current towards the anode). If such a tube with ignition electrode is connected to an alternating voltage, the ion current will begin to flow as soon as the auxiliary electrode receives the ignition voltage, to cut out again as soon as the anode voltage passes zero. Fig. 25 shows how the conducted current can be controlled by shifting the point of time in which a short voltage peak is applied to the auxiliary electrode.

The tube by means of which the current can thus be controlled is called a THYRATRON. This tube enables strong alternating currents to be controlled with little energy. Applications are: speed control of electric motors, control of fluorescent lighting, electric furnaces etc.

By means of the same thyratrons it is also possible to accurately adjust the intensity and duration of a current pulse. This is required, for instance, in electric spot welding, where two metal plates are clamped on top of each other between two electrodes. If a short powerful current pulse is sent through these electrodes, the plates will locally be heated to such a high temperature that they weld together without the heat having had the opportunity to spread over the whole surface. This method of SPOT WELDING has found universal use in the motorcar indus-

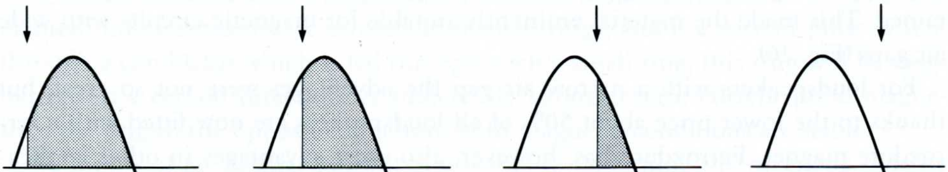


Fig. 25. Current through thyratron, controlled by phase-shifted voltage pulse on the control electrode.

try. Metal tubes with a mercury cathode and carbide igniter, so called IGNITRONS, were developed for very high currents.

All these applications in electrical engineering, although they had been known for quite some time, found a rapidly growing interest in industry during the years 1945-1965. It is not surprizing that manufacturers of electrical engineering equipment, like Siemens Schuckert, and A.E.G. in Germany, and G.E. and Westinghouse in America, who had transferred their high-frequency activities to Telefunken and R.C.A. respectively, were none too pleased to see that "radio" engineering for industrial processing, measuring and "control", penetrated so deeply into their own field that they were compelled to reassume this activity themselves.

Although they are not used in electrical engineering, it is for completeness' sake that we must also mention COLD CATHODE TUBES. In these tubes a high field strength produces a cold discharge which might be observed as a glow discharge. They have the advantage of not needing a heater which consumes power and has a limited life. They are used as electronic relays (TRIGGERS) in switching techniques, and furthermore as voltage stabilizers, counting tubes for α, β, γ and X-rays and so on. Many of these tubes were later on to be succeeded by semiconductors (Chapter I.20). Finally there are the NUMERICAL INDICATOR TUBES in which figures are displayed by a neon discharge.

## I.17. CERAMICS IN ELECTRONICS

Physical research concerning electronics has not been restricted to electron tubes; in other fields too, physical research has contributed to man's mastery over the properties of materials and their application in electronics. A most important aspect in this connection was the research on magnetic ceramics.

Loudspeakers required very powerful permanent MAGNETS and research carried out from 1927-1937 led to an alloy TICONAL, containing Nickel, Aluminium, Cobalt etc., which had a coercive force about 10 times that of the common types of steel at about equal magnetic saturation. However, the ingredients for this material were rather expensive, while moreover, due to the Korean war in 1951, nickel was growing scarce.

Hence it was particularly important that round 1950 when research work was being done on ferrites (see below), the Philips Physical Research Laboratory found a sintered material (FERROXDURE), consisting of mainly cheap iron oxide with which, be it at a lower magnetic saturation, a 5 times better coercitivity was obtained. This made the material eminently suitable for magnetic circuits with wide air gaps (Fig. 26).

For loudspeakers with a narrow air gap the advantages were not so great, but thanks to the lower price about 50% of all loudspeakers are now fitted with a ferroxdure magnet. Ferroxdure has, however, also great advantages in other applications such as: bicycle dynamos, small electric motors, concentration magnets for picture tubes etc. Later on, the quality of the material could be considerably improved by a process of "magnetic orientation" during pressing.

In transformers with iron cores induction gives rise to EDDY CURRENTS in

Fig. 26. A ferroxdure magnet ring is kept floating in the air by the magnetic repellent force of another ferroxdure magnet ring.
(Philips Technical Review).

the iron, which represent considerable losses. Therefore the cores are stacked from thin iron plates, the lamellae. This system is effective for 50 Hz frequency and even for sound frequencies. For higher frequencies as used in multi-channel telephony and T.V. receivers, these lamellae would have to be made extremely thin to keep the losses within reasonable limits. It was, therefore, of prime importance when in about 1943 it was discovered in the Philips Physical Research Laboratory that from a sintered mixture with iron as the main ingredient (FERRITES), crystals could be obtained which at a reasonable magnetic conduction involved but low losses at frequencies up to 500 000 Hz. This material FERROXCUBE was first used in multi-channel telephony, but afterwards also for the magnetic circuits in the deflection system of T.V. receivers (deflection coils and "line output" transformers). Since Ferroxcube was essential for these T.V. applications, large quantities of it were required during the television explosion over the period from 1952-1962. Industry had to make an effort to meet the demand.

As the material could be made suitable for higher frequencies, it was used for intermediate and high-frequency coils, and for built-in aerials. Furthermore such hard wearing material, plus the possibility of filling the narrow air gap with glass, was very suitable for magnetic recording heads in magnetophones (Chapter I.19).

A similar material with a rectangular hysteresis loop introduced an entirely new element into electronics. It is made into small rings, and if a current pulse is sent through a conductor which is led through such a small ring, this ring will be magnetized in a certain direction. A sufficiently strong reverse current pulse magnetizes the ring in the opposite direction; both magnetic conditions are stable.

By means of an auxiliary wire it is possible to determine the magnetic polarity of the ring. By leading a network of wires through a number of these rings it is thus possible to store data (translated into "0" or "1" per ring) and to read it out again with the auxiliary wire. This process develops at a tremendously high speed. And so the MAGNETIC MEMORY came into being. (At the exhibition in

Brussels in 1958 an interesting demonstration of such a memory was given to the school children: when a button marked with a certain date was pushed, the machine immediately gave an account of the events that had taken place in that year).

Ceramic materials of a different nature, based on lead titanates, are PIEZO-ELECTRIC MATERIALS, in which mechanical stress generates an electromotive force or, reversely, which deform under influence of an electric voltage. (This phenomenon had already been used in the quartz crystals for transmitters). The first property can be used in gramophone PICK UPS, microphones and detectors; the second property is turned to use in ULTRA SONIC CLEANING. Here the ultrasonic currents bring the piezoelectric body into mechanical vibration, which causes the surrounding liquid to move.

Another application is that the mechanical vibration of this material, brought about by an alternating voltage, can be transmitted through a piece of glass, for instance. At the other end of the glass body the mechanical vibration is picked up and converted back into electric currents. These currents show a certain time lag with respect to the voltage that was applied at the input: here we have the DELAY LINE. Fig 27.

Finally a plate of such piezoelectric ceramic material has a characteristic mechanical resonance frequency which will amplify an applied alternating voltage of exactly the same frequency. Thus we have a device to replace resonance coils (CERAMIC RESONATORS).

Entirely different ceramic materials are used for the manufacture of non-linear resistors which play such an important part in various kinds of electronic circuits. Silicon carbide (SiC), for instance, forms the basis for VDRs, (Voltage Dependant Resistors). NTCs (resistors with a Negative Temperature Coefficient) are made from iron, cobalt or nickeloxide in which these metal ions are partly replaced by Ti or Li ions. PTCs, (resistors with a Positive Temperature Coefficient) are based on barium or strontiumtitanate in which, again, substitutory ions have been introduced.

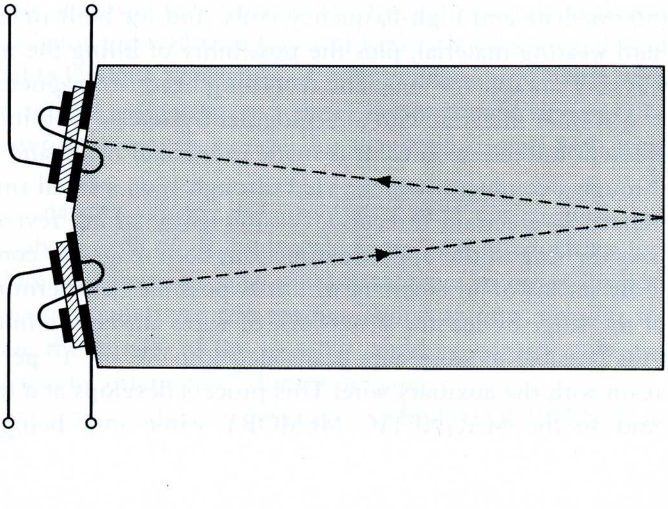


Fig. 27. Delay line.

## I.18. PASSIVE COMPONENTS

Apart from the electron tubes with which the high frequency currents are generated, converted, and amplified, the electronic circuit contains so-called passive components: resistors (R), capacitors (C), and coils (self inductances L), which are decisive for the frequency of these circuits and on which the coupling of the various circuits depends. As already explained in Chapter I.8, the industry began to take an interest in these components round the year 1927. It was then that the original amateurism gradually had to give way to scientific physical and chemical research, manufacturing technologies, and mechanization.

The RESISTORS used in electrical engineering originally consisted of coils of wire with a high specific resistance, wound on an insulating cylinder. Low-power resistors of high value, as needed in radio sets, were made by applying a thin layer of carbon to a porcelain substrate, or by pressing carbon powder mixed with plastic materials (CARBON RESISTORS). The "carbon layer" type won in the end since the required values could be obtained with a higher degree of accuracy, and because the pressed type, owing to insufficient contact between the carbon grains, gave rise to noise interference in the reproduced sound. Also for variable resistors, known as POTENTIOMETERS, a carbon layer was applied to a hard paper base. Since the sensitivity of the ear increases with the logarithm of the power, these potentiometers had to have a logarithmical ratio between angle of displacement and resistance value.

For capacitors used in high-frequency circuits the mineral MICA was eminently suited as an insulating material because of its high dielectric constant and low losses. But at the same time it was rather expensive. Around 1935 it was found possible to multiply the dielectric constant of porcelain many times by adding titanates. From then on the mica capacitors could be replaced by the much cheaper CERAMIC CAPACITORS. Later on the plastics industry brought POLYSTYRENE, which can be applied in much thinner layers.

For low-frequency applications, telephone capacitors for instance, use was made of PAPER CAPACITORS which were already known in wireless telephony. Paper as such is a bad insulator, but by impregnating it with paraffin, the dielectric constant and the loss factor could be improved. For high voltages there remains the drawback that in the pores tiny air bubbles are trapped, in which dark discharges might occur, ultimately leading to complete breakdown of the capacitor. For this application, therefore, impregnation with oil in vacuum was used. After 1950 the plastics industry introduced new materials, such as POLYESTER, which could be made in extremely thin layers. These layers sometimes showed small perforations, but by vaporizing the metal film it was found that in case of a breakdown the metal film burned away, so that the fault was automatically eliminated.

For smoothing the "ripple" in the voltage of tube rectifiers, capacitors capable of standing rather high voltages were needed. Paper capacitors were too expensive and too large for this purpose. For these smoothing capacitors, which need only carry a voltage which is positive with respect to the chassis, a particularly interesting solution was found around 1930 in the form of the ELECTROLYTIC CAPACITOR.

This type of capacitor is based on the fact that aluminium oxide is a very good insulator. By applying a positive voltage to aluminium which is immersed in a certain liquid, a very thin layer of aluminium oxide is produced, so that the aluminium and the liquid, separated by the oxide, form a capacitor. Here too, any breakdown is automatically remedied because a new layer of oxide is formed. Initially these capacitors actually contained a liquid, but in about 1938 the liquid was absorbed in a porous layer of paper, which made it possible to wind electrolytic capacitors like paper capacitors, after which they were fitted in a sealed can.

Fig. 28. Components of a black and white television receiver (left) and of a colour television receiver (right) (from "de Koerier").

In the earliest days a radio set was "tuned" (adjusted to the correct wave length) by means of coils with sliding contacts. Later it was found that better tuning was obtained with variable capacitors consisting of a set of fixed and a set of rotating plates separated by air (ROTATING or VARIABLE CAPACITORS). By driving several variable capacitors with one spindle it proved possible to tune several circuits at a time (see "heterodyne" circuit, Fig. 9f). For that purpose, and to obtain a correct indication of the selected wave length, these capacitors had to be very accurate. After 1960 the Japanese industry succeeded in largely reducing the size of this capacitor by using an intermediate layer of plastic.

The difficulties encountered in coil technology increased with the frequencies used. To keep their capacitance down they had to be wound crosswise (HONEYCOMB). To reduce the eddy current losses the coils had also to be wound of wire consisting of thin strands: STRANDED WIRE. By using cores of a magnetic material, for which purpose Ferroxcube (Chapter I.17) is ideal, the coils could be made smaller so that they involved less losses. By screwing down the Ferroxcube core the coils could be adjusted to the required frequency.

Later on, instead of interconnecting the various components by soldered wires, the connections were photographically produced on synthetic resin bonded paper, to which the different components were automatically soldered (PRINTED WIRING, see Fig. 28).

Perhaps the development of the passive components is less spectacular than that of the electron tubes, ferrites and semiconductors, that will be discussed later. Nevertheless, a tremendous amount of technical scientific work had to be carried out before the passive components with their close tolerances and high degree of reliability could be manufactured at low prices (in terms of dollar cents). Since a colour television set, for instance, is equipped with 1000 passive components (Fig. 28), a breakdown figure of no more than 1/10% per year means one repair per year. This repair figure is absolutely unacceptable as there are many more sources of breakdown or faults.

It is estimated that at the moment about 100 × $10^9$ passive components per year are manufactured in the free world. In view of the high development costs and high degree of mechanization and automation, these components are increasingly manufactured in specialized factories which have their own facilities of manufacturing white and black ceramics.

## I.19. THE MAGNETOPHONE (TAPE RECORDER)

Although not all types of electro or magneto-mechanical equipment will be discussed here, the magnetophone has played such an important part in the development of electronics, that it deserves to be mentioned in this historical survey.

For detection (Chapter I.3), Marconi had already used a so-called "magnetic detector", consisting of a travelling steel wire whose magnetic condition was changed under influence of the received signal. At a later date equipment was made in which such a steel wire was magnetized by microphone currents, so that after-

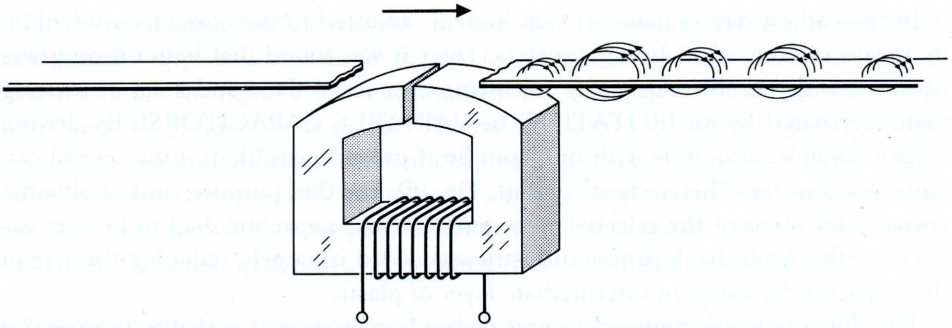


Fig. 29. Principle of the tape recorder.

wards by leading the wire along a coil, these microphone currents could be reproduced in that coil.

During the years 1940-1945 equipment was built in Germany, in which a plastic tape coated with iron powder was used for the same purpose. The result was a much better sound quality. (Fig. 29)

The principle is as follows: the side of the tape coated with a high-coercive magnetic powder is led along a very narrow air gap in an electromagnet through which the microphone currents are fed. As a result each passing iron particle will be magnetized according to the value the microphone current has at the moment that this particular particle passes the air gap. If afterwards the same tape is led along the electromagnet, a current equal to the original microphone current will be generated in the coil. This current can then be converted into sound via a telephone or loudspeaker. If the tape is run along an electromagnet which carries a high frequency alternating current, the iron particles on the tape are "shaken" to such an extent that the magnetism disappears and the tape can be used for a next recording.

Once again an important new element in electronics had been created, that is to say, one to "store" sound, or better still, to store electric oscillations.

As the tape suffers hardly any wear, the recorded sound can be played back many times, but it can also be "erased" again. By using an extremely fine-grained powder and a high tape speed, it was found possible to reproduce rather high notes quite faithfully, and thus a sound quality was obtained that was better than that of gramophone records, while moreover noise produced by the stylus and mechanical damage to the record were out of the question.

At first tape recorders were expensive and used a lot of tape. They were used in broadcasting studios, for film and gramophone recordings, and in places where a "sound archive" was needed.

In 1950, however, industry could already offer amateurs equipment which enabled them to record about two half hours of music on a reel of tape with a diameter of 10 cm. Now amateurs could record music or speech to reproduce it afterwards, or record radio programmes which could be "played back" at any time. (This formed a problem as regards "copy right". Copy rights were included in the price paid for gramophone records, but not so for homerecorded tapes).

In the following years the tape and recording element underwent considerable

Fig. 30. Mini-cassette recorder.

improvement so that the playing time of the tape became much longer (nowadays the market offers HiFi quality at a tape speed of 1 7/8 inch per second, which became possible by making the air gap only 5 microns wide). A way of making more efficient use of the area of the tape is to place the "tracks" close together (in modern equipment there are two tracks on a tape 3.8 mm wide, see Fig. 30).

As it is difficult for the average layman to handle the loose reels, a new method was initiated in 1965: the reels were permanently placed in "CASSETTES" in which they are automatically wound and rewound, the cassettes simply being slid into the apparatus. Thanks to the better sound quality, less danger of mechanical wear and greater ease of handling, these cassettes containing music were more and more supplied instead of gramophone records from 1967 on.

Now the "grain" had become much finer, it was possible to magnetize the tape with currents of higher frequency which determine the video signal (Chapter I.11), so that television images too, could be recorded and stored. After 1965 practically every T.V. studio was equipped with a VIDEO RECORDER, and since 1967 the system has also been used for colour T.V. One might ask whether it would not be simpler to make and store photographic films, but the fact that one need not wait for development is in many cases a great advantage (a scored goal can be shown again almost immediately afterwards).

A tape recorder designed to operate with frequencies from 10 000-100 000 Hz can also be used to "store" information which is expressed in a rapid succession of + and − pulses of which as many as 100 000 per second can be recorded. Since for computers all figures and letter symbols must be translated into this "+ and − language" the magnetic tape represents a rather simple means of storing large quantities of information. Like book printing has already been partly replaced by photography (photocopying and microfilms), the encyclopedia of the future will probably consist of a computer with a magnetic tape memory.

## I.20. THE TRANSISTOR

Not until 40 years after the invention of the crystal detector (Chapter I.3) could physicists explain its operation: they discovered that the crystals of certain semiconductors (germanium and silicon, for instance) contained rather mobile electrons or atoms with one electron short (so-called "holes"), so that a so-called PN layer in these crystals would pass current in one direction only if an alternating voltage was applied between the contact point and the crystal. Industry produced an improved version of these GERMANIUM DIODES sealed into a small glass tube. They were used in television receivers and the first computers.

In 1948 a group of physicists of the Bell Laboratories found on theoretical grounds that in such a diode the current through the two contacts could be controlled by means of a current sent through a third contact. And so an effect corresponding to that in a triode tube was obtained (TRANSISTOR). This invention caused a complete revolution in electronics!

The electron tube had been an obedient servant in electronics. It had performed its functions: rectifying, amplifying, mixing and switching for frequencies up to 50 MHz and powers of no less than 250 kW. It had been generally accepted that such a tube needed a certain amount of energy (for radio tubes about 2 W) to keep the cathode at the required temperature. At the same time the heat produced had to be removed again. The mechanical manufacture of its components as well as the rather complicated processing required expensive machinery. A dozen surviving radio tube industries had such equipment and managed, by manufacturing in large quantities, to sell at a reasonable price. Furthermore it was an established fact that radio tubes had a limited life (evaporation of the cathode) and had to be replaced (according to the guarantee after 1000 hrs, in practice much longer). To make matters complete, the complex system of components mounted in an evacuated glass envelope was still vulnerable as regards breakdown caused by air leaking in, dust particles and mechanical fracture, notwithstanding a rigid process control to reduce such risks to a minimum.

Now it had become possible to achieve a similar effect with such a small crystal: without energy consumption by a heater, with a theoretically eternal life, and without complex mechanical components. Most of the existing radio tube industries began to study this new technique fervently, but also a number of new enterprises were set up which could enter this new field on equal terms. Among the latter, Texas Instruments and Fairchild have developed into very large industries. It is especially this working side by side of so many industries (after all characteristic of free economy) that the transistor made such a brilliant career that, 10 years after its invention, it became a worthy equal of the electron tube in numerous applications.

The transistor of the Bell Laboratories with its two contact points was soon to be replaced by the "alloyed" germanium transistor. The latter was first applied for low frequency amplification in hearing aids (in which the small dimensions and low power consumption were of enormous value), and in portable gramophones. Soon afterwards frequencies up to 1 MHz could be amplified with these transistors, and already in 1953 the first battery operated transistor radio set came on the market (Fig. 31a). No more need for a heater current battery, a much lower voltage for the "anode battery" and far more shockproof: an entirely new concept.

It became a race between the various industries to see who would be the first to manufacture a transistor for shorter waves. To achieve that, the "diffusion" method had first to be introduced. The Philips P(ushed) O(ut) B(ase) transistor was one of

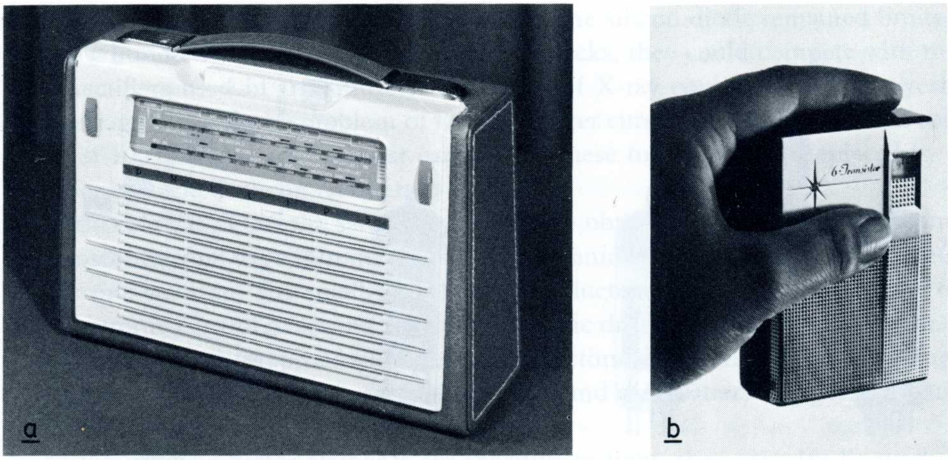


Fig. 31. a. The first Philips transistor radio (1953), b. Transistor pocket radio.

the first of its kind by means of which frequencies up to 10 MHz could be amplified. From then on the handy transportable "transistor radios" which could be played without the electric mains, were sold all over the world. The Japanese industry succeeded in manufacturing transistor radios of small dimensions at equally "small" prices (Fig. 31b).

A great difficulty was maintaining the narrow tolerances in electrical characteristics required for interchangeability. During the initial period 70% reject was no exception, but by exercising an accurate process control system, this reject figure could be brought down.

Also for high power low frequency transistors, which require adequate heat removal, solutions were found and so the "electronic" megaphone, for instance, could be designed.

Germanium, although a suitable material to obtain the required characteristics by adding the proper ingredients, had one drawback: the transistors begin to fail at temperatures higher than 75 °C. With silicon high temperatures formed no problem, but it was not so easily given the required characteristics. The American government realized how important these transistors would be in military projects and space travel, and so millions of dollars were spent on development to be conducted by the industry to advance this new technique. Indeed it did not take science long to master the techniques of purifying, melting, and diffusing this high-melting material.

Finally there came the invention of PLANAR transistors, in which use is made of the circumstance that silicon oxide forms a well-sealing, electrically insulating, layer. This could be used in photographic reproduction methods according to which the various operations, in processing the crystal, became quite reproducible.

Further investigation showed that this crystal could be protected against atmospheric influences so that the originally used glass or metal envelopes were no longer needed. The transistors were pressed into plastic (Fig. 32), and the price could be reduced to $0.10

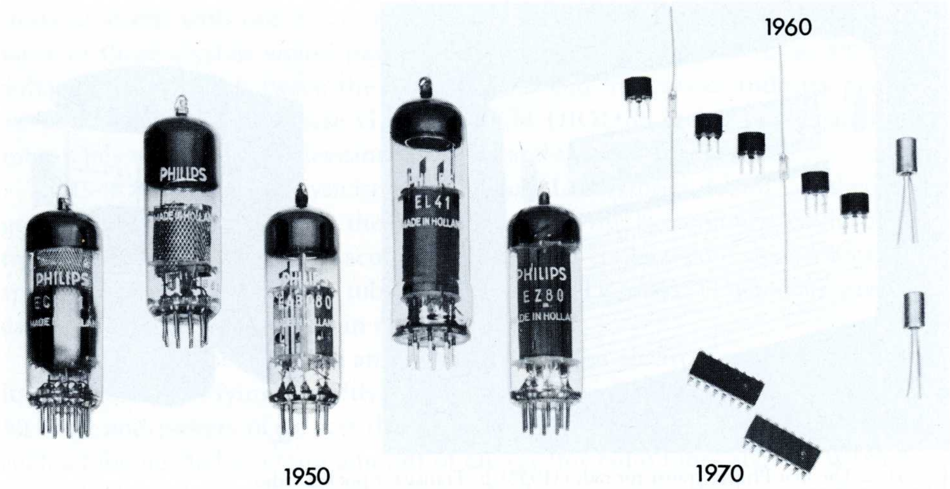


Fig. 32. A set radio tubes for a simple receiver with a sensitivity of 5 μV and an output of 1 W (1950), a set of transistors (1960) and a set of integrated circuits (1970) for a similar receiver.

Gradually it became advantageous to design transistor radios fed from the mains. However, in television there were still requirements that the transistors could not meet (high voltages and powers), and for certain functions electron tubes had still to be used; even more so in colour television. As a consequence of the colour television explosion in 1967/68 the electron tube industry is still in full swing.

With the arrival of the silicon transistor, suitable for high temperatures and more reliable owing to its resistance against atmospheric influences, the transistor became a welcome item in professional applications, and from 1967 on it was no longer possible to sell a piece of electronic measuring equipment, or say a television camera, if not "transistorized".

By means of special configurations the industry succeeded in manufacturing transistors to operate with high powers and high frequencies. Small portable transceivers up to 5 W (WALKIE TALKIES) could now be made, and even larger transmitters in cars (MOBILOPHONES) were transistorized.

Transistors could also be made for still shorter waves, and transistorized beam transmitters for 6000 MHz multi channel telephony are even thought of.

Another attractive aspect of transistors is their eminent suitability for "switching" (the characteristic of the "electronic relay" is not critical), particularly when diffusion techniques made them suitable for high-speed applications. The young computer industry which had to build compact equipment with large numbers of such electronic relays, made ready use of the small transistors which consumed no heater current. Even the cold cathode trigger-tubes (Chapter I.16) could no longer compete. The same happened with electronic organs.

With overwhelming zest and youthful recklesness the transistor industry continued its attack on the positions which the tube industry had won after years of struggle.

The silicon rectifier involved such a low loss that rather great powers could be rectified with a relatively small crystal, soldered onto a block of copper for heat removal (POWER DIODES). The voltage across one silicon diode remained limited to a few hundred volts, but when arranged in stacks, they could compete with the tube rectifiers used in transmitters. Designers of X-ray equipment had for years been struggling with the problem of feeding heater current to rectifier tubes at voltages of 100 000 V or over against earth. Now these tubes could be replaced by a number of stacked silicon rectifiers.

With THYRISTORS the same effect could be obtained as with thyratrons, and so semiconductors were introduced in the electronic control of rectifiers and electronic spot welding. The small, simple semiconductor element opened the way to the electronic control of sewing machines, electric drills and fluorescent lighting.

The sturdy moistureproof silicon semiconductors also found their way to mechanical equipment such as washing machines and cars (battery charges and ignition control).

As the semiconductor elements are sensitive to light, they can also be used to make photo cells. Be it for lower voltages, the ZENER DIODES provided a simpler means to stabilize voltages than stabilizer tubes.

The semiconductor technique opened new possibilities: frequencies could be doubled with a VARACTOR DIODE; these Varactors also provided an electrical means for the remote control of capacitances.

The silicon crystal can be used to make infrared detectors, as well as batteries which are capable of converting solar energy into electric current. These batteries, known as SOLAR CELLS, serve for the current supply in man-made satellites.

There is one field, however, where the vacuum tube is still unrivalled: converting an image into electric currents (camera tube) and, vice versa, converting electric currents into an image (picture tubes). However, screens are being made, which consist of numerous very small planar elements which can take over the function of the grains of photoconductive material in the screen of camera tubes.

In 1967 the Philips catalogue for semiconductor elements already contained as many as 1000 different types.

## I.21. THE INTEGRATED CIRCUIT

Hardly had the electronic industry adapted itself to the enormous change brought about by the arrival of the transistor, when in 1960 a new revolution started which in 1965 began to take effect in the equipment sector. This time it was the position of the "passive elements" industry that was threatened by the fact that a method had been found to equip the silicon crystal with capacitors and resistors.

In electronic circuits, particularly those with a switching function, the capacitors are often loaded with a direct voltage; in this case a semiconductor diode can function as a capacitor too. It is also possible to change the resistance of silicon by "diffusing" it with more or less impurities. In the planar transistor technique the crystals undergo the following treatments: oxidizing the upper layer, applying photo sensitive lacquer, "masking" certain areas of the surface (by means of a photographic negative), exposure, washing off the unexposed lacquer, etching away the protective layer of Si-oxide from these places, diffusing certain materials in the resulting cavities, and finally, vaporizing a metal film for the electric connections. The same process was used to make diodes, transistors, resistors and capacitors on one single piece of crystal. All these components were interconnected by means of vaporized metal strips, thus forming a complete circuit (INTEGRATED CIRCUIT). These techniques owe their quick success to four major factors:

1. The earlier elements, radio tubes and passive components were made considerably larger than strictly required for the powers they handled, the main reason being that it was practically impossible to make them any smaller. (For instance, the high frequency part of a radio set may be of a low power design since the power must be amplified several times anyway to drive the loudspeaker). Now a new manufacturing technique had been found, it became possible, and even necessary, to work with extremely small components which indeed could dissipate little heat, but which on the other hand operated with low powers.
2. Integrated circuits present less tolerance problems than separate components. Usually two resistors, for example, are used to divide a voltage. If one resistor is 20% too high and the other 20% too low, the voltage division will show an error of 40%. If the resistors are manufactured together, and if they are both 50% too high, the voltage division is still correct.
3. Expensive piece operations become cheap when the products can be treated in batches; hence the following procedure is followed for planar transistors and integrated circuits. One mask, magnified many times, is drawn for each operation, the drawing is reduced to the actual size, and a photograph is made of say 1000 of these units on the surface of one slice of silicon cut off a rod (Fig. 33a). The aforementioned operations can be carried out on the whole slice which is afterwards divided into 1000 circuits according to Fig. 33b.

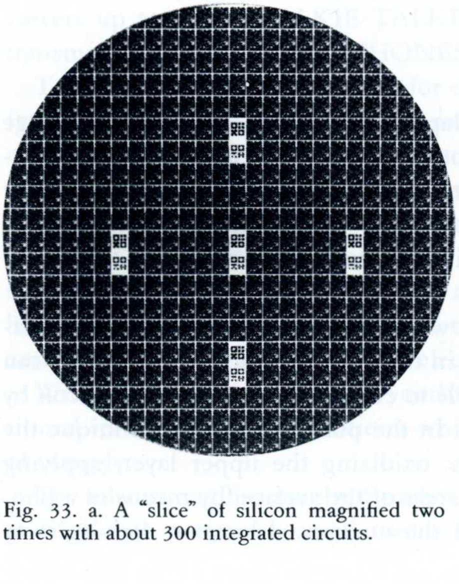


Fig. 33. a. A "slice" of silicon magnified two times with about 300 integrated circuits.

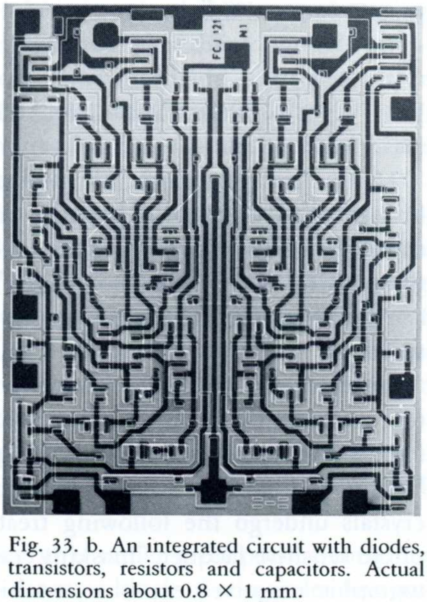


Fig. 33. b. An integrated circuit with diodes, transistors, resistors and capacitors. Actual dimensions about 0.8 × 1 mm.

4. The fault percentage of a circuit built from conventional components on a printed circuit plate is equal to the sum of that of the individual components. The fault percentage of an integrated circuit is practically equal to that of *one* transistor, if it has the same area. The vaporized connecting strips on the crystal are more reliable than soldered ones.

Integrated circuits were easiest to realize for circuit elements applied in the "pulse technique" as required, for instance, for computers, the latter being almost exclusively equipped with diodes, transistors, and capacitors (direct voltage), and because the characteristics need not be critical. Later on, however, the high frequency and low frequency amplifier circuits too, could be made in the form of integrated circuits, but they required the connection of separate self inductances, large capacitors and so on. As a first result a hearing aid was manufactured in 1967, of such small dimensions that it could be worn inside the ear (Fig. 34). For many applications the saving in weight and volume opened new ways.

Integrated circuits have led to circuit designs different from those in which tubes or transistors and discrete components were used. Whereas in the latter circuits a resistor is cheap and a tube or transistor comparatively expensive, the case is quite the opposite with integrated circuits.

Gradually larger and larger circuits are formed on one crystal. This method is known as L(arge) S(cale) I(ntegration), (Fig. 35).

## I.22. THE COMPUTER

Computers can be distinguished into two main types. First, the ANALOGUE computer in which numbers are represented by means of a continuously variable voltage. These voltages can be electronically added, subtracted, multiplied, and di-

Fig. 34. Hearing aid with integrated circuit worn in the ear.

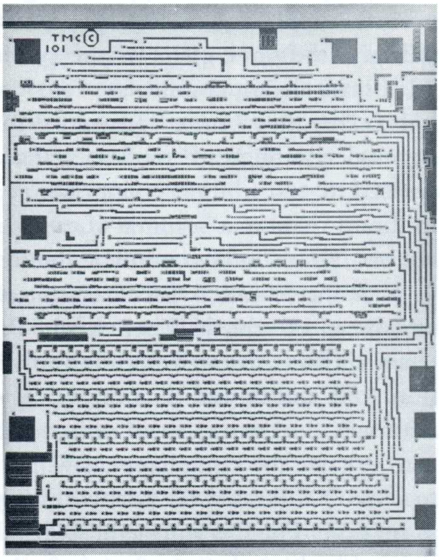


Fig. 35. Example of large-scale integration (4 × 3 mm).

vided. The result of such computations are indicated with electric measuring instruments. Then there are the DIGITAL computers which use discrete numbers for computation with a certain unit as the basis. This system is, of course, to be preferred for administrative purposes (we want every cent to be accounted for).

The mechanical counter is a digital computer in its simplest form: it contains discs which take ten steps per revolution. At the tenth step the unit disc returns to zero, and the second disc counting the tens is shifted 1/10 revolution further, and so on. Endeavours have been made to replace these 10-step discs by electron tubes. As a result gas-filled counting tubes with ten cold cathodes were made, and also vacuum counting tubes in which an electron beam passes through ten positions. When the tenth position is reached, a signal is fed to the next counting tube. These tubes are used to make electronic counters which count the current pulses they receive. In a computer a central "clock pulse" generator produces a series of current pulses which can, for instance, read certain numbers out of a core memory, these numbers to be added in a decade of counting tubes.

For more complex calculations with electronic elements it is far easier to use the binary system instead of the decimal system. In a mechanical counter designed for the binary system the disc would have no more than two positions: "0" and "1", and at the count of "2" the first disc returns to "0" while the second jumps to "1". If, for instance, the number of 91 was to be expressed in the binary system, it would first be divided into 64 + 16 + 8 + 2 + 1, which could be expressed as $1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$. This number is then written as 1011011, and a total of 7 discs would be needed.

An example of an electronic circuit that can be used for binary computation is the FLIP-FLOP or multivibrator circuit. This circuit stabilizes itself in either of the two positions: TR1 conducting, TR2 blocked, and the other way round. A positive pulse to point A will cause the circuit to switch over to the other stable position. Only when TR2 is conducting will Q1 become positive, which means once every two pulses. Connecting point Q1 to point A of a next switching circuit, will give the same effect as the system with binary discs. Thus we have obtained a binary electric counter (see Fig. 36a).

It is true that 7 units are needed for counting to 99 instead of 2 units as in the decimal system. With modern transistors, however, the counting rate can be increased to 100 million counts per second, or it is possible to count to 100 in 1 millionth part of a second, whereas the fastest mechanical disc will always need about 1/10 second to complete the ten revolutions required for counting to 100.

With the so-called GATE circuit of Fig. 36b the series of pulses arriving on A and B can be added. When A and B are at zero voltage, the current from the + terminal flows through the diodes (which are conductive in this particular direction) to A and B, and then the voltage on Q = 0. If A receives a pulse, and B remains 0, the current flows away via the diode connected to B, and Q remains at zero voltage; the same happens when B receives a pulse and A remains connected to zero. Q will receive a positive voltage only when A and B receive a pulse at the same time. For the voltage at Q resulting from the voltages on A and B we can then write: $0 \cdot 0 = 0$; $1 \cdot 0 = 0$; $0 \cdot 1 = 0$; $1 \cdot 1 = 1$. Thus with a positive voltage on B the gate will be "opened" for a pulse on A.

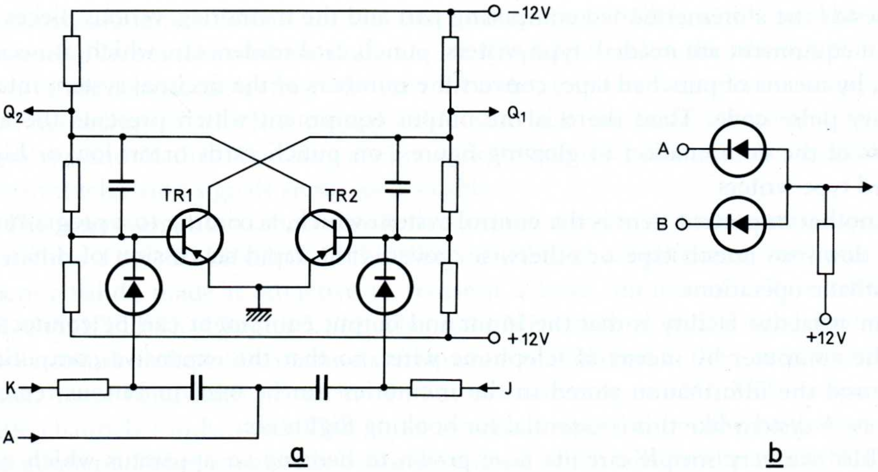


Fig. 36. a. Flip-flop circuit. TR 1, TR 2 = transistors, A = trigger input, Q1, Q2 = pulse outputs; b. "AND" gate. A, B = input, Q = output.

Without going any deeper into these matters in this chapter, it will no doubt be clear that electronics has here created the possibility of adding, subtracting, multiplying and dividing numbers at an unrivalled speed (the numbers expressed in the binary system). The answer obtained can be stored in a core or tape memory for shorter or longer periods, to be read out again for further computation.

Considering that for counting to 100 with a FLIP-FLOP no less than fourteen triodes are needed, it stands to reason that the practical development of computers could not be started until after 1955, when radio tubes could be replaced by the smaller and cheaper high frequency transistors. As Figs. 36a and 36b show, the circuits used contain only diodes, transistors, resistors, and capacitors, so that integrated circuits found a particularly favourable new application, with the result that since 1968 computers are almost exclusively built with integrated circuits.

To mention a few computer applications: the prices of a large number of articles can be recorded on magnetic tape and once the quantities supplied are fed in, the computer calculates the amounts to be paid, taking into account reductions, foreign currency and so on. The stock list can be recorded numerically on tape, mutations can be introduced and the computer will keep the stock lists up to date; the total stock value and the replenishments required can be quickly calculated at any time. The list of workers with their corresponding hourly pay can be recorded on the tape, and by introducing the number of working hours, all weekly wages including allowances and reductions are automatically worked out. The banking deposits of all depositors can be recorded on tape and all mutations can be made, whilst at the same time the interest is calculated. By means of informations stored in the memory it is possible to calculate from the sales figures the "Tax on Value Added".

Furthermore the computer with its many facilities can be used to "try out" the most advantageous traject in a system, to determine the most economical division of card board sheets and so forth. Finally the computer can work out highly complicated mathematical problems.

Beside the aforementioned computing part and the memories, various pieces of input equipment are needed: type writers, punch card readers etc. which, if necessary, by means of punched tape, convert the numbers of the decimal system into a binary pulse code. Then there is the output equipment which presents the answers of the computation in glowing figures, on punch cards or on low or high speed type writers.

Another important item is the control system which, according to a programme laid down on punch tape or otherwise, governs the rapid succession of different automatic operations.

An attractive facility is that the input and output equipment can be connected to the computer by means of telephone wires, so that the expensive computing part and the information stored in the memories can be used in various remote places. A system like this is essential for booking flights etc.

Thus two very simple circuits have grown to become an apparatus which can take over, and sometimes improve on, a part of human memory and calculation.

## I.23. ELECTRONICS IN WIRE TELEPHONY

As already explained in Chapter I.10 the signals in long distance telephony need amplification: at certain distances amplifying stations are built into the lines. For submarine cables these amplifiers will then have to rest on the bottom of the sea. Hence repair and component replacement entails considerable expense since such amplifiers, complete with cable, will have to be hoisted on board. The large telephone companies have, therefore, put quite some effort in developing highly reliable electron tubes with a long working life to be used in these SUBMARINE CABLE AMPLIFIERS. The arrival of the silicon transistor was practically the solution to the whole problem.

The fact that the existing multicore cables between the various exchanges were getting overloaded due to increasing international communication, called for systems in which several calls could be made over one pair of lines only, and so, in 1940, CARRIER TELEPHONY was introduced. In this system several simultaneous calls were modulated onto carrier waves, or just "carriers", which were 4000 Hz apart (twice the maximum frequency of the sound to be transmitted). Initially the maximum carrier frequency was chosen no higher than 60 000 Hz, which was sufficient to conduct twelve simultaneous conversations along one set of wires over a distance of about 50 kilometres between the amplifying stations. By placing the amplifiers closer together (25 km), the maximum frequency used in the Netherlands from 1950 onwards could be increased to 204 000 Hz, which made it possible to use 48 channels simultaneously. Now the transistor technique has made it possible to bury unguarded amplifying stations in the ground, the distances between the amplifiers can be made even shorter. The cable network in the Netherlands is gradually being converted to 120 channels at a maximum frequency of 552 000 Hz, the distances between the amplifiers being 8 km.

If a special so-called COAXIAL CABLE is used, the maximum frequency can be increased to four million Hz (4 MHz), so that 960 channels can be used simultaneously along one cable (later this was increased to 12 MHz with a total of 2700 channels). In countries where there are no telephone cables, the obvious thing to do is to use coaxial cables from the start. In India, for instance, a plan is being realized according to which a coaxial cable will link the large towns of Calcutta, Bombay, and Madras. By using such a wide frequency band it is also possible to transmit television signals along such a cable.

The improvements to shortwave link equipment made in the meantime (particularly the TRAVELLING WAVE TUBES capable of amplifying a very wide frequency band), made it attractive to transmit a large number of calls via BEAM TRANSMITTERS (Chapter I.12.). The 4000 MHz system, widely used for television networks, offers 900 channels (frequency modulation is then used which requires more bandwidth per channel), but by increasing to 6000 MHz as many as 1800 channels can be used. This is the system in operation in the Netherlands at the moment. Tubes are already being developed which will handle 12000 MHz, so that even more channels come available. Thus electronics has enabled "wire telephony" to dispense with the wire (or cable) for trunk calls.

Another field of wire telephony into which electronics is introduced is the TELEPHONE EXCHANGE. In principle a connection is made in the telephone exchange by means of a number of ten-contact "stepping switches" which are driven by a number of pulses produced by the dial of the "caller". The function of a telephone exchange is, however, far more complex: first the caller must be connected to the switch system; for trunk calls the stepping switches are distributed over several exchanges; the "busy line" signal must be given; the call must be registered in time and distance to calculate the cost, and so on. It is fairly obvious that since transistors can play the part of electronic relays in computers, the numerous switching operations in the telephone exchanges too, should be performed by these small, rapid, and cheap electronic relays. Indeed electronic exchanges have been built in different places in the world; Philips Telecommunication Industry built two of them, one in Utrecht and one in Aarhuus (Denmark), which are now in use (Fig. 37).

Fig. 37. Inauguration of the electronic telephone exchange in Aarhuis (dialling by means of push buttons) ("Announcer").

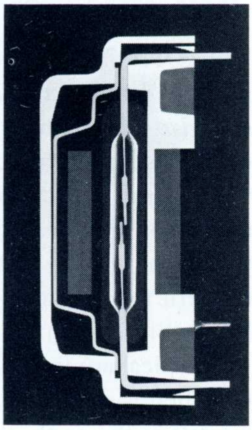


Fig. 38. "Reed" relay.

In principle there are, however, a few difficulties:

1. The currents and voltages needed for the transistors differ from those used in existing systems. Therefore some kind of "translation" equipment must be connected between exchange and subscriber, and also between the electronic exchange and the conventional exchange.
2. An important advantage of the electronic system would be that much time could be saved in dialing as compared with the 1/10 second per pulse, which in the electromechanical exchanges is dictated by the speed of the stepping switches. Special telephones are obtainable from which, by pushing buttons numbered 0 to 9, a corresponding number of high-speed pulses are sent to the telephone exchange (Fig. 37). If such a system is to be connected to a conventional type of exchange, these high-speed pulses must be "stored", afterwards to be released at a speed with which the stepping switches can cope.
3. For a telephone number of, say, six figures, at least six contacts are needed to build up the connection. The minimum resistance of a transistor is still rather high, and several of these resistances in series have too great a "damping" effect on the speech circuit.

Whereas an all-electronic exchange is decidedly advantageous for separate networks with a limited number of subscribers, each with a specially adapted telephone apparatus, these advantages are outweighed by the drawbacks involved in cases where the electronic exchange must link up with conventional exchanges and ditto telephone apparatuses.

According to the latest developments, the electronic system will be used for selecting the connections and performing all subsidiary functions, while the actual connection will be made by means of small relays, consisting of a contact in a glass tube within a coil (REED RELAY), see Fig. 38. The future will show whether or not further development of semiconductors and the necessity of making quicker connections, will in the end lead to all-electronic solutions.

## I.24. COMMUNICATION VIA SATELLITES

RADIO CONTROL of ships and aircraft is a favourite pastime for youngsters interested in electronics. Radio waves can be transmitted to the ship or plane to energize relays, which in turn operate servo-motors which control the steering mechanisms. It is a joy to see the model planes operated by their young owners on the ground. It is surprizing that during the war this method was not used for the attack with pilotless aircraft and the Japanese had to sacrifice their "suicide pilots" to hit the enemy battle ships with their explosive-charged planes. In how far modern missiles are remote controlled or find their targets on their own account is as yet insufficiently known.

A typical application of radio in projectiles is the PROXIMITY FUSE. The head of a shell is equipped with a small transmitter. As soon as this shell approaches an aircraft or the earth, the reflected radio wave causes the shell to explode. A technical masterpiece, all the more so since the complete equipment including the radio tubes must be able to withstand the enormous acceleration which takes place when the shell is fired.

Electronics play a most important part in space travel. The unmanned satellites receive their commands radiographically transmitted from the earth. Manned satellites must remain in continuous radio contact with the earth. Unmanned satellites send information cooncerning weather conditions, cloud formations to the ground stations or direct to aircraft (photographs are often translated into pulse trains which are transmitted in a few seconds) (Fig. 40a). All this is done mainly with V.H.F. (wavelengths of about 1 metre) since these Very High Frequencies are more suitable to penetrate through the higher ionosphere. Of course highly sensitive receivers must be built on the earth because a satellite has but a low transmitting power.

Since 1965 it has been possible to use multiple telephone channels and to transmit television signals respectively, between two remote places on earth via beam links and a satellite in orbit. To achieve this the satellite must first be brought in what is known as a "geostationary orbit", which means that it must describe a circle in the equatorial plane in which, like the earth, it completes a revolution round the axis of the earth in 24 hours.

To the observer on earth the satellite seems to stand still; it is at a distance of about 20 000 miles above the surface of the earth. By means of gas jets which are controlled from the earth, small corrections can be made so that the satellite maintains its correct position with respect to the earth.

Fig. 40. a. Radiographically transmitted photograph of the South pole of the planet Mars, taken and transmitted by the unmanned satellite MARINER 7 (Nasa nr. 69-H-114).

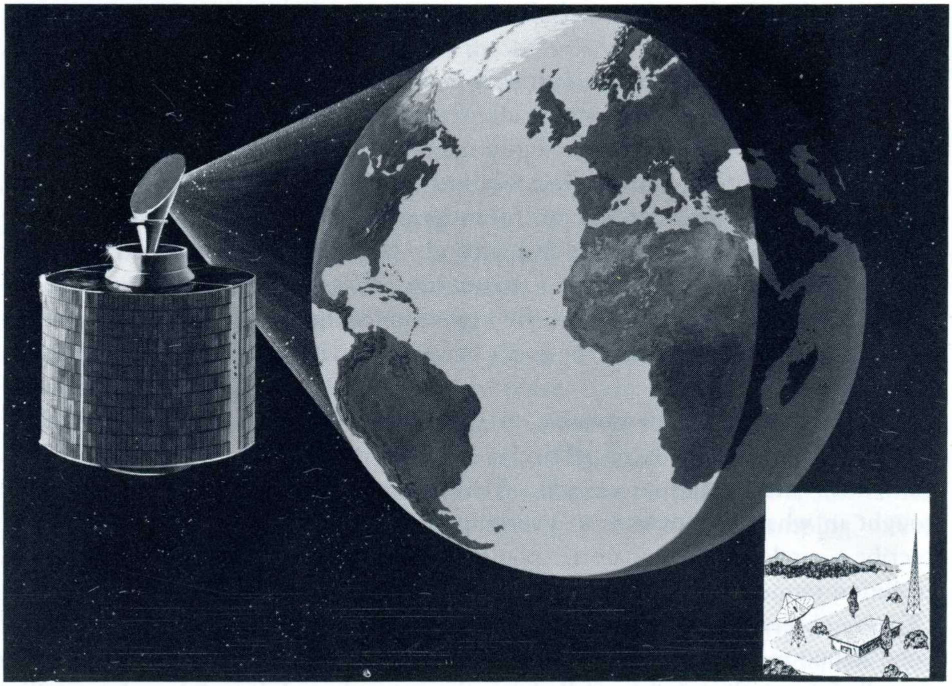


Fig. 39. INTELSAT III and receiving station on earth.

On the 20th of August 1964 the IN(ternational) TEL(ecommunications) SAT(ellite) consortium was founded, of which the U.S.A. owns most of the shares and several other countries what remains.

On the 6th of August 1965 the EARLY BIRD satellite was launched and brought into a geostationary orbit. With it 240 channels could be operated simultaneously between America and Europe, which nearly equals the total capacity of the four submarine cables linking these continents. In 1967 a somewhat larger satellite, the "Intelsat II" was brought into orbit to be followed by a still larger one in 1969: the INTELSAT III (Fig. 39).

The system operates as follows. A ground station with a parabolic aerial with a diameter of about 25 metres, which is aimed at the satellite transmits a signal of a frequency somewhere in the 5920 to 6420 MHz range. With Intelsat III it is possible to use 1200 bi-directional telephone channels or to transmit four television signals or numerous telegraph signals respectively.

The satellite's aerial receives these signals and by means of a wide-band amplifier with a "travelling wave tube" the signal is amplified after the carrier frequency has been changed so that it can be transmitted at a frequency in the 3695 to 4195 MHz range with an equivalent radiated power (EIRP) of 300 W.

The current for this amplifier-transmitter is supplied by a battery of solar cells (Chapter I.20) mounted on the periphery of the satellite; these cells convert the energy received from the sun into electric energy.

Furthermore the satellite is equipped with a reservoir containing a propelling charge which via valves, operated telemetrically from the earth, can be made to escape axially or radially for the purpose of correcting the position of the satellite. The total weight of the satellite is about 125 kg.

The amplified energy is radiated from an aerial as a conical beam of 20°; this cone embraces the surface of the earth. The satellite rotates round an axis (gyroscopic effect) parallel to the axis of the earth. The aerial on the satellite rotates in opposite direction, so that it remains pointing to the earth.

By means of a directed parabolic aerial the ground station (Fig. 39) on the earth receives the signal transmitted by the satellite. Owing to the enormous distance to be covered, and the relatively low transmitting power, the signal is very weak. An amplifier equipped with tubes or transistors can not be used to amplify this weak signal as the characteristic noise of these elements would be stronger than the signal. Therefore "MASERS" or PARAMETRIC AMPLIFIERS are used instead.

With such a system it has been found possible to establish a rather regular telephone communication over great distances, and in October 1968 the whole world was confronted with the fact that it was possible to transmit colour television signals of the Olympic Games in Mexico to Europe and Japan for a full fortnight during eight hours per day.

Television signals have been transmitted over much greater distances during the manned flights round the moon in 1968/9. In December 1968 the "Apollo 9" in

Fig. 40. b. Television picture of the first man on the moon. Photograph taken of the television screen on the earth on 21st July 1969 (Nasa nr. 69-H-136).

orbit round the moon transmitted black-white pictures to the earth. In May 1969 the "Apollo 10" did the same with colour pictures which, at a distance of more than 400 000 km, could be seen by millions of television watchers in their homes, and on the 21st of June, 1969, they witnessed the "moon walk" the astronauts of the "Apollo 11" were taking. Fig. 40b shows a picture received on earth.

# PART II. ELECTRONIC COMPONENTS

## SECTION A. INTRODUCTION

Electronic equipment is built from components, which are either fitted on metal plates (chassis) and connected with wires to form "circuits", or soldered to the plates of insulating material on which the connections have been made in advance ("PRINTED CIRCUITS"). More complex equipment is known as SYSTEMS, which are mostly assemblies of equipment-parts ("BLACK BOXES"), which are interconnected with bundles of flexes or cables.

In the following chapters most of the electronic components will be dealt with.

The requirements for mass production, have of course, led to the need for standardization. Resistors and capacitors, for instance, are made in values following a logarithmic scale; for radio tubes the heater voltages or heater currents, and often the anode voltages, have been standardized; the wavelength of the transmitting stations, allocated according to international agreements, have practically fixed the values of variable capacitors, and so on.

Fig. 41 shows the first Philips catalogue for three types of electron tubes in 1921, and the complete series of DATA HANDBOOKS for 1969, comprizing about 10 000 pages of data on some 100 000 types of standard components all of which are in stock, or can be supplied on short notice, because the manufacturing data have been laid down.

An important problem in this standardization is the question whether the same standardized types for "entertainment" equipment can be used for professional purposes. On the one hand it stands to reason that the necessity for very low cost prices for "entertainment" will probably lead to cheaper materials and production methods, whereas in professional equipment no cost will be spared to achieve the highest possible degree of reliability. On the other hand the following arguments are in favour of using the same components in both fields:

1. "Entertainment" too, must nowadays meet high reliability requirements, since the service costs run rather high and some equipment (colour T.V. sets for instance) contains large numbers of components.
2. Producing large numbers for "entertainment" leads to more experience and better statistical control which, for equal design, yields a greater reliability than would be obtained if smaller numbers were carefully manufactured for professional purposes. This applies particularly where large numbers allow of advanced automation: machines guarantee more accuracy than any manual work.
3. Whereas professional equipment formerly had to meet more stringent demands regarding temperature, humidity and shock resistance, entertainment equipment must nowadays also be made suitable for the tropics, be resistant against high temperatures and shocks (car radios).
4. The differences in cost are becoming so important that they become noticeable for various professional applications.

Although in the years 1950 to 1960, especially in the U.S.A., a great deal of attention was given to improving the reliability of professional components and a complex system of "standards" and "testing methods" was organized, there is a gradual development in which all components must satisfy severe reliability requirements, so that ordinary "entertainment" components can be used, for instance, in measuring instruments, telephony and small transmitters.

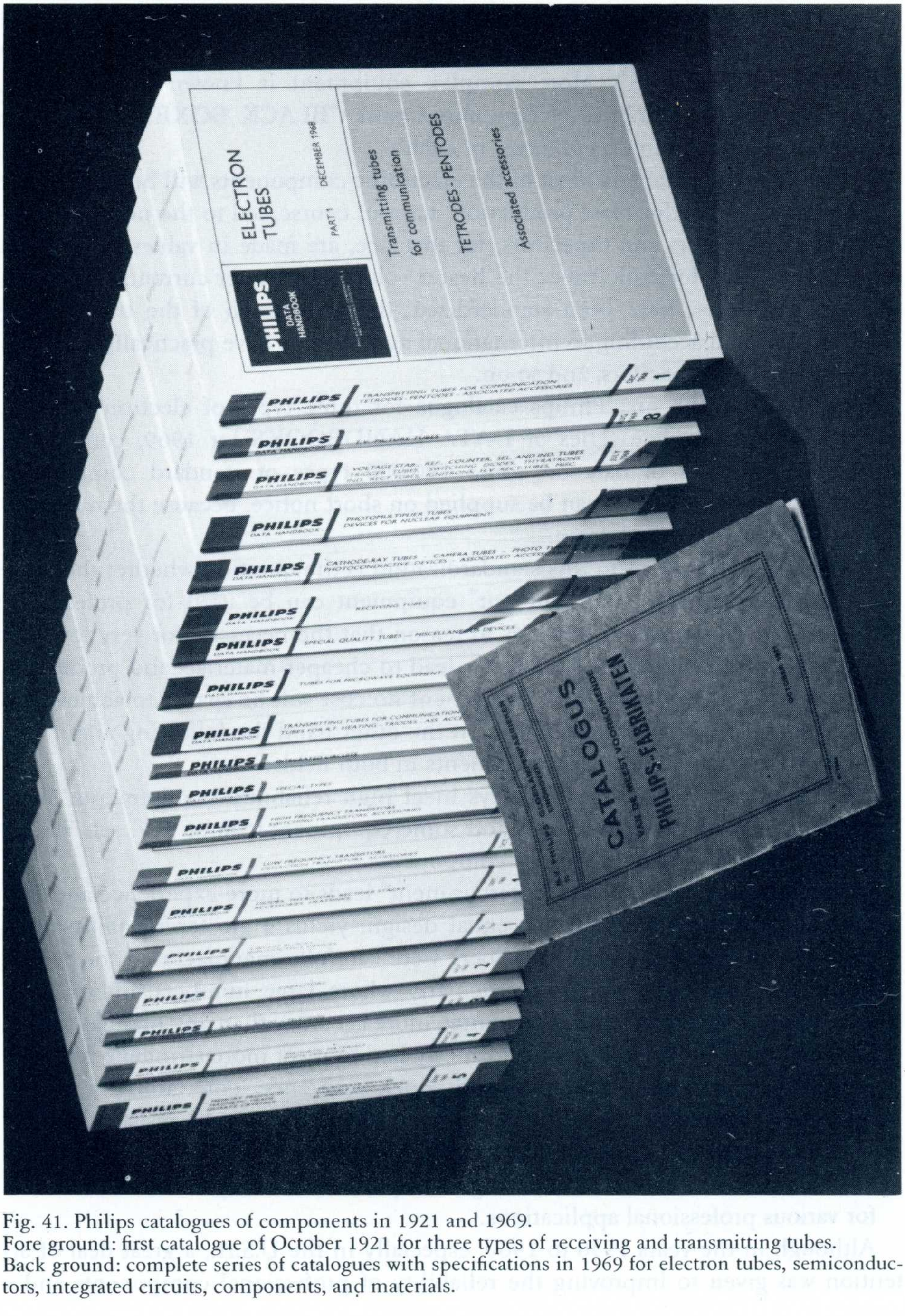


Fig. 41. Philips catalogues of components in 1921 and 1969.
Fore ground: first catalogue of October 1921 for three types of receiving and transmitting tubes.
Back ground: complete series of catalogues with specifications in 1969 for electron tubes, semiconductors, integrated circuits, components, and materials.

If necessary, a certain percentage "underrating" as regards voltage and temperature is used to get an even better reliability.

The components can be divided into three main groups which differ both in production method and application.

1. *The electromechanical components,* which are composed of metal and insulating parts, and which are fitted with coils of insulated wire. Examples are: loudspeakers, high-frequency coils and low-frequency transformers. No special production technologies are involved, while the machines on which these components are manufactured are not too expensive. These components are usually made by the manufacturers of electronic equipment themselves. The magnets and cones for loudspeakers, however, are bought from specializing firms. Further mechanization also leads to the specialized manufacture of electromechanical components: thanks to mass production, the specializing firm can often sell at lower prices.
2. *The so-called standard components,* mainly resistors and capacitors. These are also referred to as PASSIVE COMPONENTS, this in contrast with electron tubes and transistors, which generate and amplify electric currents, and are therefore known as ACTIVE COMPONENTS. Since these "standard" components are defined only by a few parameters, such as resistance or capacitance, maximum voltage or current, their application is rather universal, that is to say, little dependent on the circuit. The consumer can place his order by simply selecting a type with the required parameters and characteristics from the catalogue. On the whole such components are made in large numbers by specialized manufacturers who use specific technologies.
3. A third group of elements: electron tubes and transistors (ACTIVE COMPONENTS) is much stronger related to the circuitry of the equipment. This has been the case for tubes ever since 1930, when different types were developed for different applications, (high-frequency and low-frequency, for instance). The preferred anode voltage and current was quoted by the manufacturer, but often also the design was based on a given circuit, in which the characteristics of such a tube would be done the most justice. Selling these products required a close cooperation in what is known as the "technical-commercial" field between the manufacturer of these elements and the consumers. On the one hand it had to be ensured that the consumer did not use circuitry in which the elements were overloaded, on the other the manufacturer had to inform his customers of the circuits for which these components were intended and with which the characteristics could be turned to full use. Determining the overload too, was of course, far more difficult than for a resistor or capacitor, as more parameters had to be taken into account. To fulfil his purpose, the manufacturer of components needed an "application group" which for a given combination of elements and circuitry sought the best possible solution. On the one hand this called for contact with the component designers who could better judge the technological possibilities, on the other contact was needed with consumers to find out what characteristics they deemed important in their equipment. Large suppliers of elements, such as R.C.A. and Philips have, therefore, a large application and

technical-commercial organization at their disposal to ensure that this essential cooperation actually takes place.

Of course the question may arise as to whether such cooperation might not lead to a certain rigidity in design. For equipment manufacturers are leashed as it were by the component manufacturers, who via this cooperation between circuit development and technological development, have arrived at a certain concept. Should they wish something else, too small series are likely to involve a high cost price. Considering the historical development of radio and television during the period from 1945 to 1968, when the European industry introduced a rather rigorous type-standardization for radio tubes, we may say that this standardization did not impede the rapid progress of the art. On the contrary; it is only fair to say that the whole industry has been more pleased with this standardized type range than with the numerous types developed in America under pressure of keen and widely spread competition. In fact the situation is such that a certain application must be pushed at the right moment. A large components manufacturer in contact with many consumers has the best chance of finding the optimum solution at the right time.

The same holds good for the professional field: at a certain moment there is a need for, say, special shortwave transmitting tubes, certain "wide-band amplifiers" or "low-noise" amplifying tubes. The components industry with its technological experience and its enormous specialized developing power must be expected to provide the proper solution. We have to admit that in certain fields, particularly those of telephony and beam transmitters, some companies have developed their own tubes which they reserve for their private use. In doing so these companies have indeed gained a certain lead. However, as the relevant field expanded considerably, they were often unable to compete with the larger components factories where characteristics and cost price are concerned.

When at the beginning of the semiconductor development the tolerances obtainable were very wide, and it was difficult to produce a transistor with specified characteristics, the choice fell on another system. First transistors were made and then, by measuring, they were selected for the different applications. Understandably this system resulted in a situation where transistor manufacturers could not get rid of large stocks which failed to meet the application requirements. As engineering progressed, ways were found to design transistors which could be given the characteristics specified for a given circuit.

With the introduction of the integrated circuits, a new element came into being which restricted the freedom of action on the part of the equipment designers even more. They were obliged to accept the complete circuit from the components manufacturer, or had to risk the high expenditure and loss of time involved in making their own design. It stands to reason, therefore, that several manufacturers, particularly those making professional equipment, began to develop integrated circuits on their own account. Round 1964 I.B.M. were indeed very successful with a special type of integrated circuit ("thick film").

It should, however, be noted that the technique of crystal circuits is so difficult and changes so quickly that only those, capable of practising this technique on a large scale for a wide range of applications, can expand sufficiently and thus cover their expenses. Even more than in the case of radio tubes and transistors, it is a first requirement here, that there should be excellent cooperation between technological development and application in electronic circuits.

In the above survey, MATERIALS have not been mentioned, particularly the black ceramic materials, as discussed in Chapter I.17. Difficult as they are to manufacture, these materials have been supplied in large quantities by the materials industry to the manufacturers of electromechanical components such as loudspeakers, high-frequency and low-frequency coils, deflection coils, etc. Here too, the properties and characteristics must constantly be tuned between the materials manufacturers and, in this case, the components manufacturers.

Another group of components to be mentioned are the ASSEMBLIES consisting of a number of active and passive components which together have a certain function to fulfil (FUNCTIONAL UNITS). Typical examples are the CIRCUIT BLOCKS developed by Philips, which contain a circuit as shown in Figs 36a and 36b, for instance. They have in fact been the predecessors to the integrated circuits. The function can be further extended into units that can count, convert the binary system into the decimal, memorize, regulate the speed of motors etc.

Since 1921 Philips have been active in both the component and equipment field, but always guided by the thought that components should be developed for the entire market, whilst consumers, both inside and outside their own group, could rely on service being given concerning the relevant application.

We shall now discuss the different electronic components and describe the technology, electrical characteristics and manufacturing method. Radio tubes will receive a great deal of attention although it might be said that they have been outdated for several applications by semiconductor elements. Experience gained in manufacturing methods, circuitry, reliability and life problems in the radio tube industry has, however, had a great influence on the manufacturing techniques of other types of tube, and in many respects even of semiconductor elements.

## SECTION B. RADIO TUBES

### II.B.1 THE ENVELOPE

The envelopes of the first radio tubes were copied from bulbs used for incandescent lamps. The interiors of the tubes were mounted on the "pinch", a glass tube flattened at one end, the round end being sealed to the bulb. The lead-ins, made of copper-clad wire (an iron-alloy core covered with a copper tube) were sealed into the pinch. Since the tube had to be exchangeable in the apparatus, a bakelite base with contact pins was cemented to the bottom of the envelope, the lead-ins being soldered to these contact pins. Thus the base formed a multiple contact plug which was inserted into the "tube socket" (Fig. 42a).

For radio tubes this construction had the following drawbacks:

a. The cement was not resistant against tropical conditions, and the bases of these radio tubes were liable to come off.

b. Due to the large pins and long distance through the glass, the mutual capacitance between the lead-ins was rather great. That is why some types of tube had the grid connection, (which for high frequencies, must have the lowest possible capacitance) fitted at the top of the envelope.

The year 1936 saw the introduction of a new type of tube which had a cylindrical envelope and flat base plate into which the lead-ins were sealed. Thanks to the greater distance between the pins and their shorter length, their mutual capacitance could be considerably reduced. The so-called "exhaust tube" or simply "stem" was also sealed to the tube base.

The Americans (R.C.A.) and the Germans (Telefunken) used this design for a metal tube. A deep-drawn cap and a metal base plate into which the lead-ins were sealed with glass pellets, were welded together. The iron alloys and glass used for the pins (lead-ins), pellets and envelope were selected to have matching coefficients of expansion. The metal envelope provided electrical screening, the construction was mechanically strong (the Americans called it "Ruggedized"). The base could be provided with a stud which fitted accurately into a slot in the tube

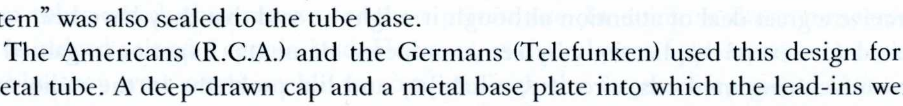


Fig. 42. Various radio tube envelopes: a. pinch tube E-series (Philips), b. metal tube (R.C.A.), c. metal tube (Telefunken), d. Philips B-technique (loctal), e. Philips A-technique (Rimlock), f. miniature (7-pin), g. noval (9-pin), h. subminiature.

socket, to ensure the tube could only be inserted in the correct position and secured there (Fig. 42b, c).

This metal construction was rather expensive and the vacuum tightness of the weld, as well as that of the envelope, was critical. Hence Philips and Sylvania introduced a glass tube of equal design, consisting of a glass envelope, a pressed glass base into which 1.25 mm thick pins were sealed, which at the same time could serve as the contact pins for the tube socket. A metal sleeve round the base of the tube was provided with a stud to ensure correct positioning of the tube (KEY-TUBES, Fig. 42d). It is a noteworthy fact that the receiving tube EF 50, which Philips smuggled to England early in the war in 1940, was an important item in military radar installations, because thanks to this special technique this tube was particularly suitable for short wave reception.

During the war from 1940-1945 the U.S.A. developed a very simple technique for small tubes intended for use in military equipment (MINIATURE TECHNIQUE, Fig. 42f). According to this new technique three-piece electrodes made of copper-clad wire were sealed into the base and welded to one-millimetre thick nickel wires which served as the contact pins. The correct position of the tube with respect to the tube base was found by placing seven pins on a circle equally divided to provide room for eight. Since the tubes were extremely light, no special locking system was needed; the contact pressure of the springs in the tube socket was sufficient to secure the tubes.

After the last war Philips introduced what is known as the "A" technique (Fig. 42e), lightly larger than miniature, with eight one-millimetre thick chromium-

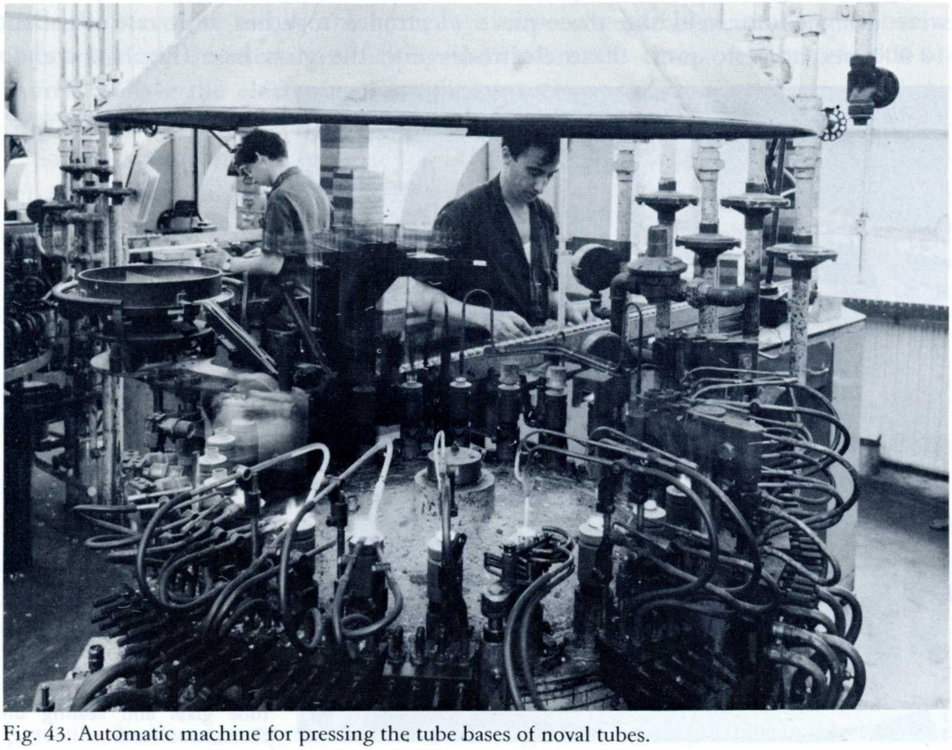


Fig. 43. Automatic machine for pressing the tube bases of noval tubes.

iron pins uniformly distributed along the circumference of a circle. Correct positioning and securing were obtained by means of a "pimple" on the envelope which fitted into a slot in a ring on the tube holder ("RIMLOCK"). The envelope and base were not sealed but "glued" together with a washer of low-melting glass-enamel, so that the shape of the envelope remained unaffected by sealing.

During the Korean war, in 1951, endeavours were made to standardize the radio tube techniques used by the NATO members. The result was the NOVAL technique (nine pins in a circle providing room for ten Fig. 42g). This tube was equipped with three-electrode lead-ins like the "miniature" tubes. Philips, however improved the shape of the envelope by sealing the tube base upwards. The tube was ruggedized (made immune to vibrations) by resting the springy teeth of mica plates against the wall of the tube.

Since 1953 this has developed into the standard technique, and comparison with earlier versions shows the striking simplicity and the resulting low costprice of the present versions. The glass seal between envelope and base as well as envelope and lead-in is reliably air tight. By means of contact springs in the tube socket the tube is adequately secured. Where necessary, screening is provided in the form of a metal screen within the envelope.

It is typical of technical progress in mass production that this technique, owing to its simplicity and logical design, has been universally accepted in spite of advertizing arguments such as "it shows more", "it looks stronger", which could be put forward in favour of other designs.

Once this version had been chosen for nearly all types, the only difference being the length of the envelope, mechanization could be furthered. Machines were designed to weld the three-piece electrodes together at a rate of about 10 000 per hour, to press these electrodes into the glass base (Fig. 43) and to

Fig. 44. Automatic machine for making envelopes of tube glass and sealing on the exhaust stem.

make envelopes of tube glass and to seal on the stems (Fig. 44) at a rate of 2500 per hour. Thus it became possible to turn out a relatively compact unit for a production of about 2 000 000 radio tubes per year. These machines were installed in Philips' and Philips-supported factories in twelve different countries spread over the world.

Thanks to standardization and automation the cost price could be reduced and the limitations of this technique were, therefore, accepted. When in about 1960 R.C.A. introduced a porcelain-metal technique, the so-called NUVISTOR, which undoubtedly had excellent high frequency qualities, it failed to find a wide field of application.

For high powers which could not be dissipated in tubes of NOVAL size, a larger type, also fitted with threepiece electrodes, was brought out in 1962 under the name of "MAGNOVAL".

Still smaller tubes, known as "subminiature tubes" (Fig. 42h) were made for very small electronic equipment, like hearing aids, for instance.

## II.B.2. THE CATHODE

An atom consists of a positively charged core around which a number of electrons travel along paths of different diameters. In metals one or more electrons per atom move about freely between the positive atom rests (FREE ELECTRONS). As the temperature increases the free electrons will travel at higher speed and be inclined to leave the metal. The attraction between negative electrons and positive ions will draw the electrons back again. This means that a certain power is required to have the electrons definitely leave the metal; this power is called the WORK FUNCTION, and depends on the structure of the atom.

So emission from a cathode depends on temperature and a property of the material: the work function. Of course the maximum permissible temperature depends on the melting point and the evaporation rate of the material. Fig. 45 shows the maximum electron current in amperes per square centimetre at various temperatures for some materials that are used for cathodes. The respective temperatures are indicated by dotted lines.

This maximum electron current is known as the "saturation current" which is never reached in a working electron tube since there is always an "electron cloud" (SPACE CHARGE) in the envelope which forces back the emitted electrons.

The power that must be fed to the cathode to keep it at the required temperature depends on the radiation. Fig. 46 shows the maximum emission in amperes, divided by the power supplied for heating, as a function of saturation for some materials. This power represents a loss of electrical energy for the equipment concerned and must be dissipated again. The most favourable condition is obtained when A/W is as high as possible.

Of the materials shown in Fig. 45 tungsten has the most unfavourable ratio between emission current and power supplied. Owing to the high permissible temperature, however, tungsten has the advantage that no impurities inside the tube

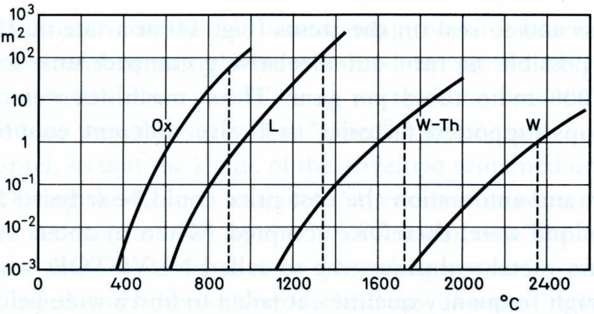


Fig. 45. Saturation emission in A/cm² as a function of temperature in °C for various metals. W = tungsten, W-Th = thoriated tungsten, L = L-cathode, Ox = oxide cathode. The dotted lines give the maximum temperature for a reasonable life. (according to Hermann and Wagener Die Oxycathode).

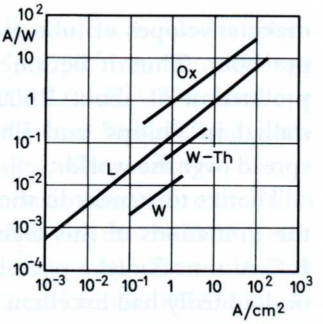


Fig. 46. Theoretical thermal efficiency (in ampere per watt heater current power) as a function of the saturation emission for various kinds of cathode.

will settle on the heater, and therefore tungsten heaters are still widely used in large transmitting tubes.

According to Fig. 45 a lower temperature must be chosen if some thorium is diffused into the tungsten: at this lower temperature, however, emission is even stronger than that of pure tungsten heated to 2350 °C. Thanks to the lower temperature required, less heat will have to be supplied and that is why, as shown in Fig. 46, the A/W ratio is considerably higher. Therefore these tungsten/thorium heaters are widely used in modern transmitting tubes. Like tungsten the material is homogeneous and there is no danger of parts of the heater being torn away by a high anode voltage. This type of heater is, however, more sensitive to impurities and bombardments by gas ions than a pure tungsten heater.

As appears from Figs. 45 and 46, the oxide cathode already has a considerably higher emission at a much lower temperature, and the ratio A/W is almost 100 times as great as that of pure tungsten. This is the kind of cathode used in radio tubes. Oxide cathodes are made of barium-strontium oxide, sometimes with contents of calcium oxide. This material is not suitable to be drawn into wire and is, therefore, applied in a thin layer to a metal basis. Furthermore this emitting layer is not air resistant. Therefore, CARBONATES of these metals are applied to the metal basis, and once the tube has been completely evacuated, the carbonate is decomposed by a process of heating, into oxide and carbonic acid, which is then exhausted. Emission must take place from an extremely thin (and vulnerable) layer of barium at the surface, which is formed after exhausting by having the cathode emit at a high temperature (burning). This oxide cathode is particularly sensitive to contamination inside the tube, and during the first years of radio tube manufacturing so-called "cathode poisoning" was the cause of a high reject percentage in the factory, so high that sometimes production had to be stopped. Difficulties of the same kind were encountered when the tubes were in operation. By a scientific study of the mechanism of emission, a minute process control and by using clean components in a high vacuum, these difficulties were later overcome.

Since the oxide in powder form more or less "clings" to the metal basis, there is the risk that in the presence of high field strengths, particles of the oxide will be torn off the metal. This means that this type of cathode is less suitable for tubes which operate with a high electric field strength, a drawback that can be overcome by using the L-cathode (see Figs. 45 and 46), and which will be discussed in further detail in Chapter II.E.2.

In the original radio tubes the layer of oxide was sprayed direct on the tungsten filament: we then speak of "directly heated" tubes. Nowadays such tubes are only used in battery operated equipment. In order not to drain the battery too soon, the filaments are made of very thin wires; in the latest series of Philips battery tubes the filament has a diameter of no more than about 11 microns and a current consumption of no more than 25 mA at 1.4 V, giving a heater power of 35 mW.

A considerably better adherence between oxide and filament is obtained by using the "cathaphoresis" method, according to which the filament is immersed in a bath containing a suspension of the oxide which is attracted by the filament under influence of a high voltage.

In equipment fed from an alternating current mains, the alternating voltage of the filament would give rise to "hum" (the 50 Hz of the mains would become audible in the sound signal). In those cases INDIRECTLY HEATED CATHODES are used, where the coat of oxide is applied to a small nickel tube containing a heater which is coated with aluminium oxide, so that it is insulated from the cathode proper. This system offers much more freedom in the circuit design since all heaters can be connected in parallel to one transformer winding, and yet each cathode can be given a different voltage. The heater insulation must meet severe requirements if the heaters are series-connected to the mains, which for direct current mains is the only, for alternating current mains the cheaper solution.

Making and maintaining an insulation that can stand an alternating voltage of 220 V at a temperature of 600-800 °C is a difficult task. Here too, cathaphoretic coating of the heater, now with aluminium oxide, brought the solution.

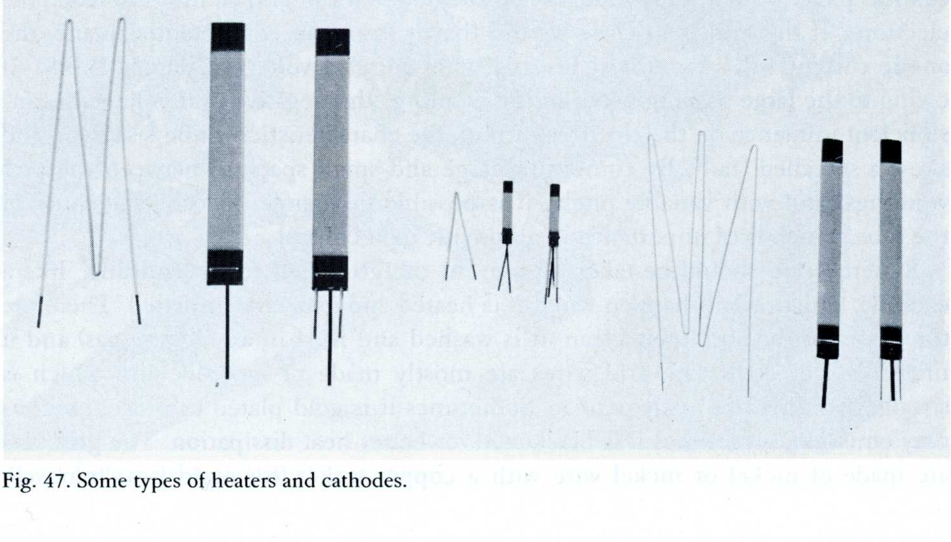


Fig. 47. Some types of heaters and cathodes.

Radio receivers with series-connected heaters required heater wires which consumed no more than 100 mA at a voltage of some tens of volts. The problem of housing such a high resistance in the small space within the cathode was solved by using the double helical filament as known in incandescent lamps.

Let us take a closer look at the construction of the indirectly heated cathode. A tungsten wire, a few tens of microns thick, the resistance of which must be within close tolerances, is wound round a molybdenum core. The resulting helix is coiled to form a second helix after which it is cataphoretically coated with aluminium oxide and fired. Then the molybdenum core is dissolved in acid. A tube is drawn from very pure nickel (sometimes folded from plate) onto which the carbonate layer is sprayed and baked. The heater is inserted into this nickel tube (Fig. 47 shows different versions of this type of cathode).

When this cathode has been mounted in the tube, the carbonate must first be decomposed on the pump, later to be activated on the aging rack. With the cathode surface at a temperature of about 800 °C (constant within close limits), and a voltage of 220 V between the nickel tube and the heater, this critical assembly, must be kept free from the slightest contamination, which might spoil the critical emission process during the expected thousands of hours of operation.

Indeed Fleming would have saved manufacturers a lot of trouble if he had only invented the transistor in 1907, but in the relevant sections we shall see that even the life of a transistor manufacturer is not entirely free from worry.

## II.B.3. THE GRID

The function of the grid is to pass or stop the electrons emitted by the cathode, depending on whether the voltage on the grid is positive or negative with respect to the cathode. The grid consists of a helix (GRID WINDINGS) and two GRID BARS which carry it. These bars, like the electrodes, are supported in two mica plates. The spacing between the windings should not be so small that electrons cannot "pass through", nor should it be so wide that the grid cannot influence the electrons. If the grid is so close wound that it forms an equipotential plane, the anode current will vary almost linearly with the grid voltage (Chapter II.B.8). If owing to the large spacing between the windings the negative grid voltage has insufficient influence on the electron current, the characteristic will be less steep and show a so-called "tail". By combining large and small spacings between the grid windings (grid with variable pitch), it is possible to change the "characteristic" of the tube, which is of importance in automatic gain control.

Extreme care should be taken to prevent the grid itself from "emitting" like a cathode, which might happen when it is heated and gets contaminated. Therefore the grid must be absolutely clean (it is washed and fired in a reducing gas) and it must be kept cool. The grid wires are mostly made of molybdenum which is strong, and does not easily oxidize. Sometimes it is gold plated to prevent secondary emission, sometimes it is blackened for better heat dissipation. The grid bars are made of nickel or nickel wire with a copper jacket (for rapid heat removal).

Frame grids (see below) are made of tungsten wire (which can be drawn extremely thin) with molybdenum bars.

The grid is made by laying windings round a jig which has recesses for the bars. These bars are first notched, then the wire is laid in the notches which are then pressed close again. After 30 years of experimenting it has finally become possible to build a machine which produces complete grids of accurate dimensions (Fig. 48).

The distance between grid and cathode is decisive for the gain of the tube. In order to keep the gain within close limits (which is desirable for interchangeable tubes) it is essential that this distance be accurately maintained. Originally the grid was oval shaped (Fig. 49a), but due to the tension in the wire it was difficult to maintain this form in winding. Each grid was measured for diameter and, if necessary, forced into the required shape. An improved shape is shown in Fig. 49b, where after winding, the wire was stretched under a small press to obtain the required shape. A form still widely used for flat cathodes is shown in Fig. 49c. A much better form was obtained by winding a grid consisting of numerous thin wires over a frame of thick wires (FRAME GRID), Fig. 49d. The width of such type of grid depends on the dimensions of the grid "bars" used. With this method the distance between grid and cathode could be kept in narrow control so that a much smaller distance could be used, which led to higher gain.

The curve and table below give a survey of the results obtained. These results

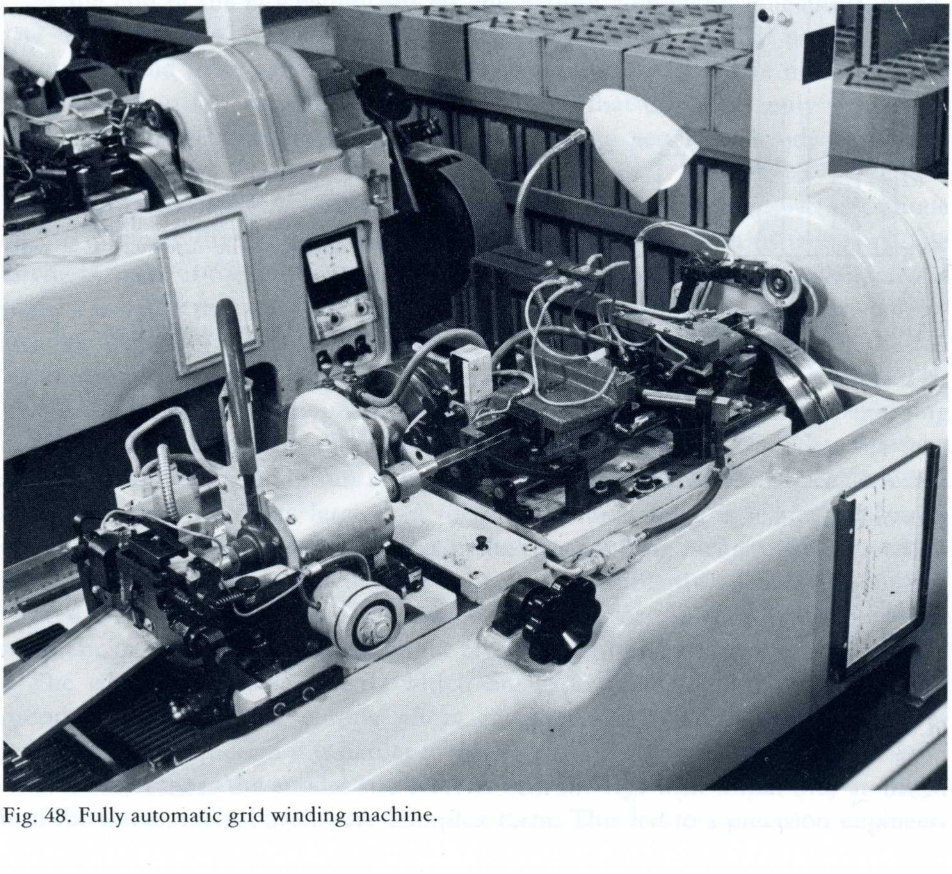


Fig. 48. Fully automatic grid winding machine.

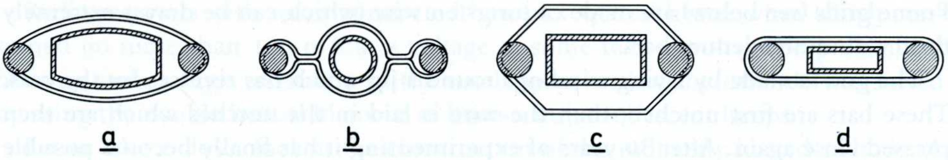


Fig. 49. Different forms of grid: a. elliptical, b. post-formed, c. rectangular cathode, d. frame grid.

bear out the statement in Chapter I.8 that over the years the slope (S) of the electron tubes has improved by a factor of 5.

A phenomenon highly dependent on grid design is MICROPHONY. A change in the distance between grid and cathode, however small, causes the anode current to vary. A tube with a movable grid could then function as a microphone. But that is not the intension: on the contrary, if the oscillations in the loudspeaker react mechanically to the grid, the anode current will vary anew, and hence the current through the loudspeaker, and so on, until the loudspeaker produces a disagreeable roar which rapidly increases in volume. To avoid this, the grid must be rigid (in

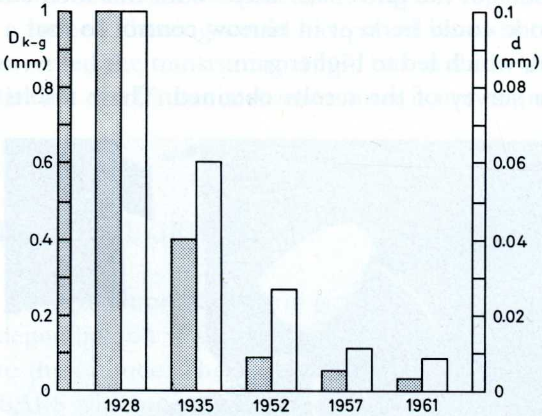


Fig. 50. Historical development of the grid to cathode distance ($D_{k-g}$) and grid wire diameter (d) of radio tubes 1928-1961.

TABLE I. Anode current ($I_a$) and Slope (S) of some Philips tubes constructed in accordance with the data of Fig. 50

| tube type | $I_a$(mA) | S(mA/V) |
|---|---|---|
| with ordinary grid | | |
| A4110 | 6 | 2.4 |
| AC2 | 6 | 2.5 |
| EC92/ECC81 | 10 | 5.6 |
| ECC85 | 10 | 6.2 |
| with frame grid | | |
| PCC88 | 15 | 12.5 |
| PC86 | 12 | 14.0 |
| PC88 | 12.5 | 13.5 |
| PC900 | 11.5 | 14.5 |

this respect the frame grid is quite good), and must be firmly clamped in the micas; the same holds, of course, for the cathode.

The grid, seemingly such a simple little component, is in fact very exacting!

The screen grids 2 and 3, mostly oval, are less critical, and made of molybdenum wire with nickel plated iron bars.

## II.B.4. OTHER PARTS IN RADIO TUBES

In the first place let us consider an important physical-chemical part, the GETTER. It has already been explained that a good vacuum is essential for the reproducible manipulation of the electron beam, and generally for satisfactory operation. Therefore all components are rinsed before mounting and fired at high temperatures in a reducing gas. During evacuation they are heated once again, and the resulting gases are pumped out. And yet, due to heating and the electron bombardment during operation, gases impeding good performance might still be produced inside the tube. This problem is solved by equipping the tube with an element that can absorb a small quantity of gas: the "GETTER". It can be, for instance, a barium mirror deposited on the inner wall of the tube envelope. To that end a "GETTER PELLET" is built in the tube, consisting of some barium aluminate in a small metal cylinder.

At the end of the evacuation process the barium compound is made to evaporate "explosionwise" by means of high-frequency heating, and the barium deposits on the cold tube wall. This getter too, is a critical part in the manufacturing process, and in fact the tube should be so clean that the getter only serves as a safety precaution. For the remainder a part of the getter is already consumed in the aging process.

The function of the ANODE is to collect the electrons emitted by the cathode and to dissipate the energy released in this process in the form of radiation. The anode is made of nickel or nickel-plated iron, but if the power to be dissipated is high, it is better that the material be black. The metal can be blackened by applying fine graphite powder, but in doing so a source of loose particles is created again. A material particularly suitable for this purpose is FERALMA, developed by Philips. It is iron onto which a thin film of aluminium is rolled. At the beginning of the evacuation process the anode is high-frequency heated, so that the iron combines with the aluminium to give a clean, beautifully black surface. Another metal that, thanks to its large area, dissipates much heat, is a kind of "metal déployé" or concertina gauze which is made by stretching nickel plate into which first slots have been punched. This material is eminently suitable for the so-termed "screens" which serve to protect the inside of the tube against external magnetic or electric fields which might give rise to interference.

The edges of all components which are exposed to high voltages must be rounded off to avoid the "corona" effect, and therefore *stainless steel* is to be preferred here, even though it is harder to work.

In modern tubes the anodes and screens with the tags with which they are fitted into the micas, often have a very complex form. This led to a precision engineer-

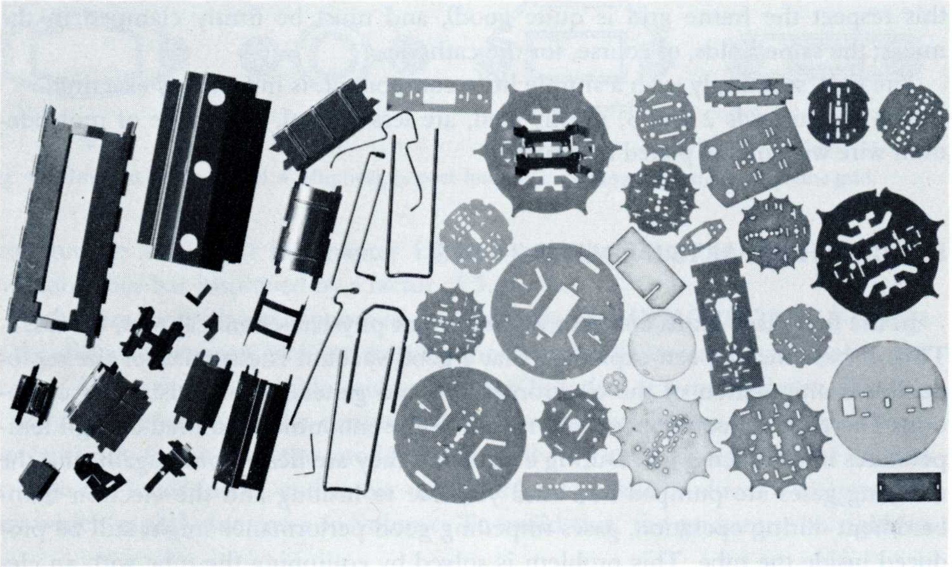


Fig. 51. Some metal components and micas for radio tubes.

ing industry which mass produces these parts with the required precision and at low prices. Two types of automatic machines are needed to do this; the so-called "WRIGHT PRESSES", which punch the components, fold and draw them, all in one operation, and the "FOUR SLIDE" which after punching gives the components a cylindrical shape and makes a folded seam. The output of these machines is greater than that of the machines for the glass components, see Chapter II.B.1, and therefore the production of these components must be centralized in large factories. Fig. 51 shows some metal and mica parts.

After punching, the components must be washed in a grease desolving detergent and then rinsed in boiling water; after that they are fired in a reducing gas.

To mount the components with the required spacing between them, and to support what is known as the "MOUNT", use is made of mica. As yet no suitable synthetic material has been found which can be so accurately punched in a certain shape, is strong and resilient, produces no gas at high temperatures, and is a good electric insulator. To lengthen the "CREEPAGE DISTANCE" across the mica, it is sprayed with an extremely fine magnesium oxide powder.

Natural mica is found in the form of complex crystals; these crystals must be split, and from the resulting thin plate which can have any form, as many as possible mica plates must be stamped. This can only be done manually.

## II.B.5 ASSEMBLY

The highly fragile components described in the preceding chapters must be assembled to form the interior of the radio tube. In this process the cathodes, grid bars, and the tags of the anodes and screens are inserted into the mica, so that a

Fig. 52. Instructing a tube assembly operator by means of a large scale model of a radio tube.

rather rigid "mount" is obtained. The various parts must now be electrically spot welded to the lead-in wires of the tube base, so that the mount is supported on the base, and electrical connections can be made at the same time.

Assembly requires the greatest care: any contamination, from moist hands for instance, may well cause unsatisfactory emission; hence nylon gloves are worn by the workers. The job must, of course, be done in a space absolutely free from dust. Welding takes place by means of a current pulse which is dosaged at the required time by a time switch. Last but not least, the female operators in charge of this work should be carefully trained (Fig. 52), and be devoted to their job.

There are also machines for automatic radio tube assembly. As Fig. 53 shows, such machinery is rather complicated. The lower mica is automatically shifted from one position to another. In each position a part is automatically supplied and inserted in the mica. Finally the upper mica is pushed onto the whole assembly which is then, as the "mount", automatically welded to the tube base. Automatic assembly only pays if one factory can turn out about 3 000 000 tubes of one type per year.

## II.B.6. SEALING, EVACUATING AND MEASURING RADIO TUBES

The tube base complete with mount is now pushed into the envelope in such a

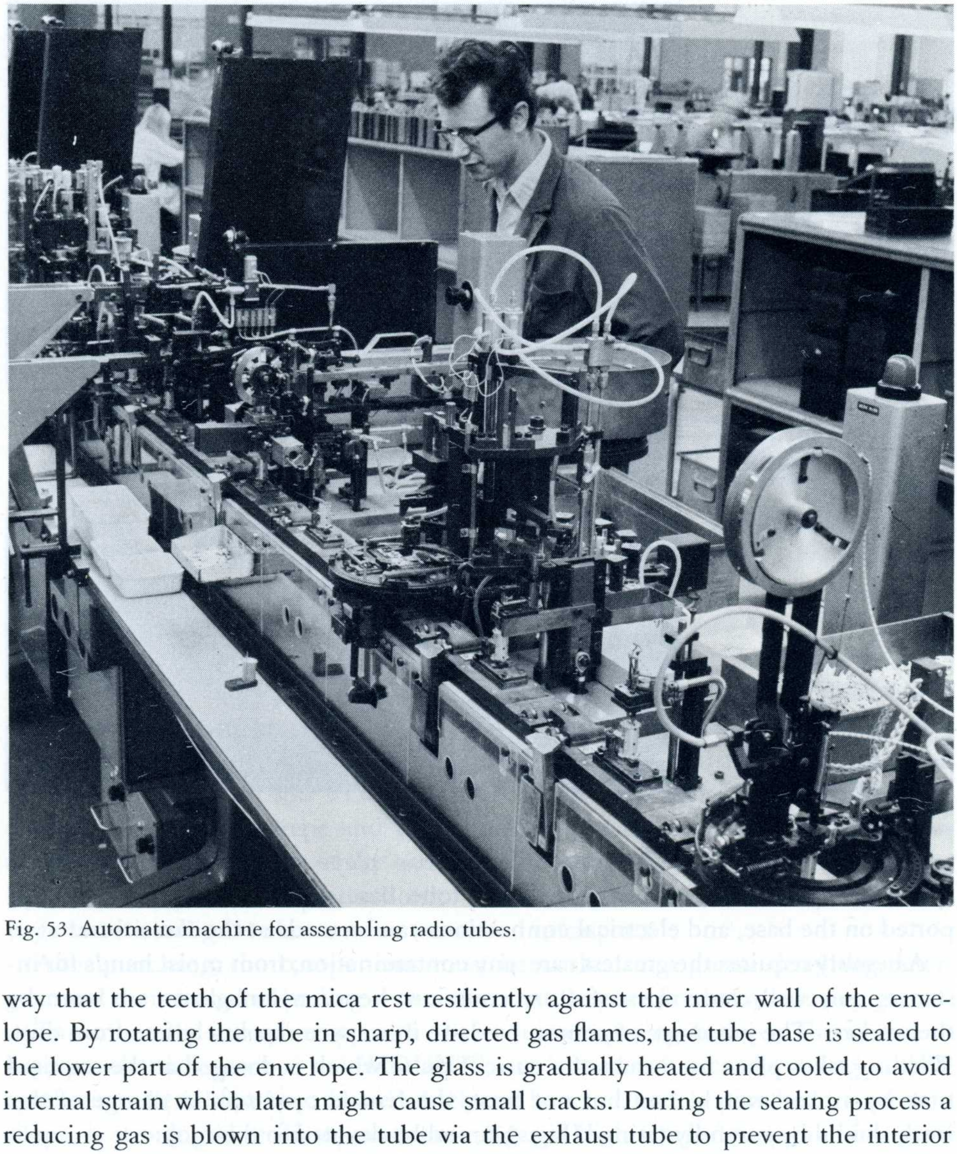


Fig. 53. Automatic machine for assembling radio tubes.

way that the teeth of the mica rest resiliently against the inner wall of the envelope. By rotating the tube in sharp, directed gas flames, the tube base is sealed to the lower part of the envelope. The glass is gradually heated and cooled to avoid internal strain which later might cause small cracks. During the sealing process a reducing gas is blown into the tube via the exhaust tube to prevent the interior from oxidizing at the high sealing temperature.

Now the pumping process can begin; the following sequence of operations must be carried out.

a. The exhaust tube (pump tube) is inserted and pushed firmly into an air tight elastic cylindrical clamp.
b. This clamp is connected to a vacuum pump. Originally Philips used mercury diffusion pumps which can reach a high vacuum against the atmospheric pressure. However, owing to the risk of mercury poisoning, maintenance of these pumps is rather difficult. During the last years oil diffusion pumps have been widely used. They work faster, but must be preceded by a rotating pump (see television tubes).

c. When a certain vacuum has been reached, the envelope is heated in a furnace, and the internal metal parts by the field of high frequency coils so that the gases, which would otherwise be produced in operation, are exhausted.
d. A current is fed through the heater, so that first these parts are outgassed, after which the barium carbonate is converted into barium oxide (Chapter II.B.2).
e. With some tubes a voltage is then also applied to the anodes, so that they are heated and degassed under influence of the electron bombardment.
f. By means of a directed high frequency coil the getter is evaporated, to deposit barium on the wall of the envelope which, where the getter mirror is required, is cooled by a jet of air. Sometimes the getter is evaporated after sealing (g).
g. The exhaust stem is heated in a gas flame, and seals up. Then the tube is removed from the exhaust stem, leaving the sealing pip.

Originally this cycle was run down on a fixed pumping or exhaust unit on which, say, 10 tubes were sealed, all operations having to take place in succession. A much better solution is the rotating pump head (Fig. 54), consisting of a wheel with thirty positions, each containing a connection for the tube and a diffusion pump. This wheel moves stepwise, that is to say, it moves one step forward every three seconds. In each position the tube to be processed finds a furnace, high frequency coils, spring contacts, gas flames, etc. needed for all operations from *a* up to and including *g*. When the wheel has completed one revolution, the tube is re-

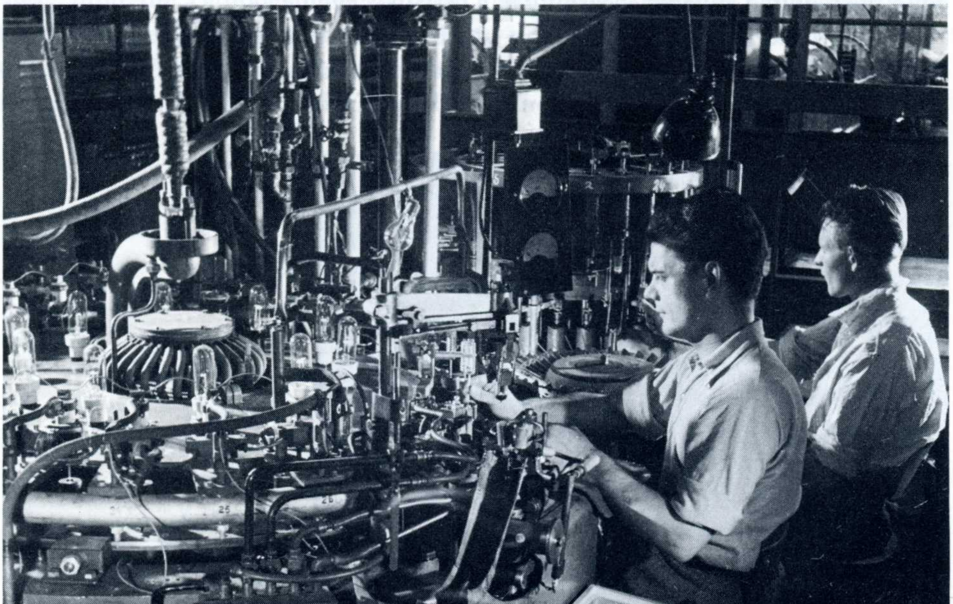


Fig. 54. Rotating pump and sealing machine for radio tubes.

moved and a next sealed tube is inserted. At a total production of 1200 per hour each tube can be subjected to a cycle time of 60 × 30 : 1200 = 1.5 minute (including the indexing time). Furthermore there is the guarantee that all tubes have had exactly the same treatment, and by taking sample tests the setting of the machine for the whole process can be checked.

It should be noted that the manufacture of these types of products involves a major difficulty. During the whole manufacturing process from preparing the materials and components up to and including aging, it is possible to do some mechanical and simple electrical measurements (contacting or insulating), but the result of these numerous operations can in fact only be tested after the tube has been exhausted and aged. Faulty tubes must be thrown away. It is therefore important to ensure a rapid production flow, at least from the mounting and grid-winding stage to final inspection, so that any faults can be corrected before too many faulty tubes are produced.

Whereas in the infancy of radio tube production a reject figure of 50% was by no means exceptional, this figure had already been reduced to 15% in 1945. In modern industry manufacturers are not satisfied if they fail to reach 5% or less. Besides, certain faults are not spotted until after hundreds of hours life testing.

When the sealed tubes leave the pump, they are inserted in tube sockets on the so-called AGING RACKS, where according to a certain cycle, the heater, anode, and screen grid voltages are applied for the correct FORMATION of the cathode. On the ROTATING AGING RACKS (Fig. 55) whole strips of tubes are connected to the required voltages according to a certain schedule.

Finally we come to the TESTING, which ultimately enables the manufacturer to decide whether or not he has turned out a satisfactory product. The measurements cover heater current, emission (SATURATION CURRENT), grid emission

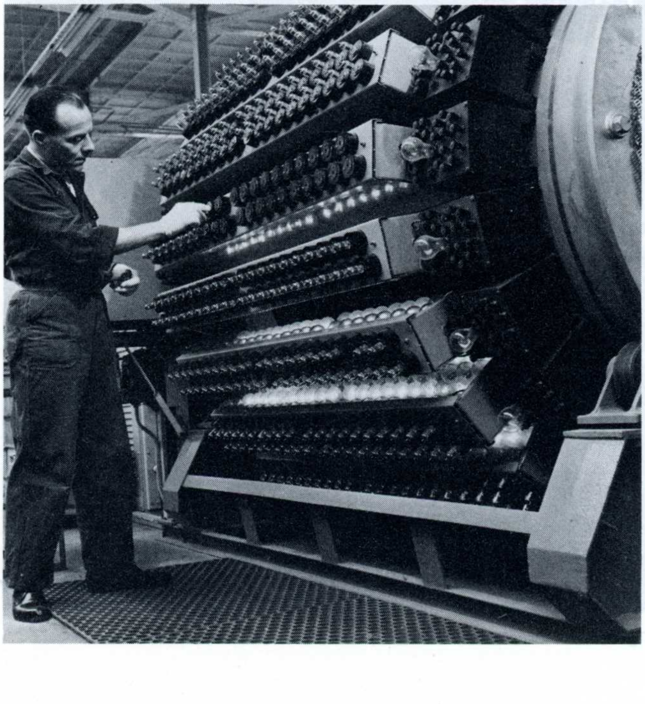


Fig. 55. Rotating aging rack for radio tubes.

(a check on vacuum), insulation, and sometimes capacitances, but also some points of the characteristics (Chapter II.B.8) which indicate the relationship between grid voltage, anode current, anode voltage, etc. In practice it is easiest to carry out these measurements "statically", that is to say, with a direct current. Some of the voltages are then adjusted, and the corresponding current intensities are read off. In certain cases, however, tests have to be carried out under working conditions, viz. high frequencies. It might be stated that on an average a radio tube must satisfy 10 measuring requirements in production testing. By means of sample tests a total of 100 items are checked which might lead to "correction" of the manufacturing process.

## II.B.7. ELECTRICAL CHARACTERISTICS OF THE DIODE RECTIFIER

A rectifier contains a CATHODE which is heated and emits electrons, and an ANODE which must *not* emit electrons, and should therefore, not be heated. When the anode is positive with respect to the cathode, it will attract the electrons emitted by the cathode, and thus there is a flow of current. When the anode is negative with respect to the cathode, the electrons will stay in or near the cathode, and there will be no flow of current.

The graph in Fig. 56a shows the relationship between the anode current ($I_a$) and anode voltage ($V_a$). At a low anode voltage no more than a few electrons are drawn away from the cathode, most of them forming the electron cloud. When the anode voltage increases, there is an almost proportional increase of electron current. For very small variations the small triangle drawn in Fig. 56 can be used to

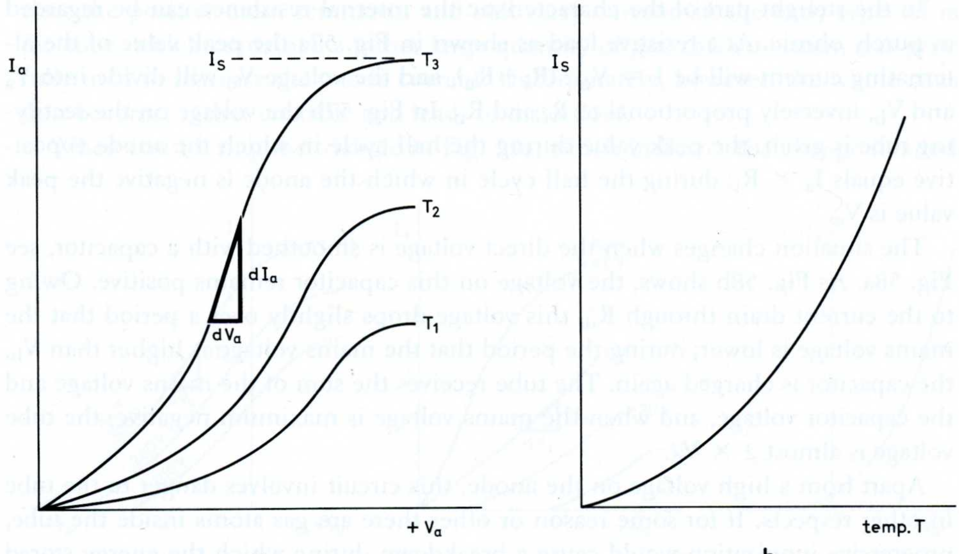


Fig. 56. Rectifier characteristics: a. anode current as a function of the anode voltage at different temperatures. b. saturation current ($I_S$) as a function of the cathode temperature.

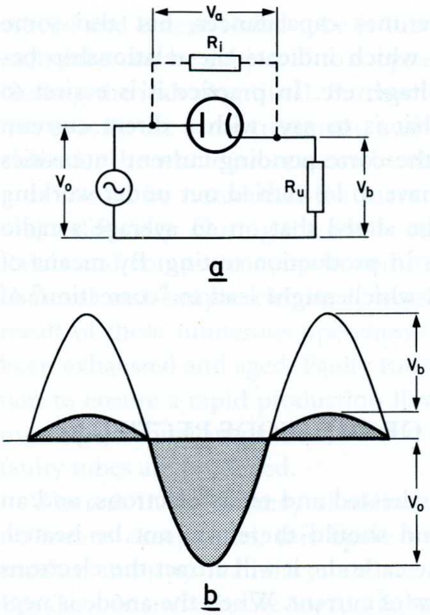


Fig. 57. Rectifier with resistance load:
a. diagram, b. voltage curve.

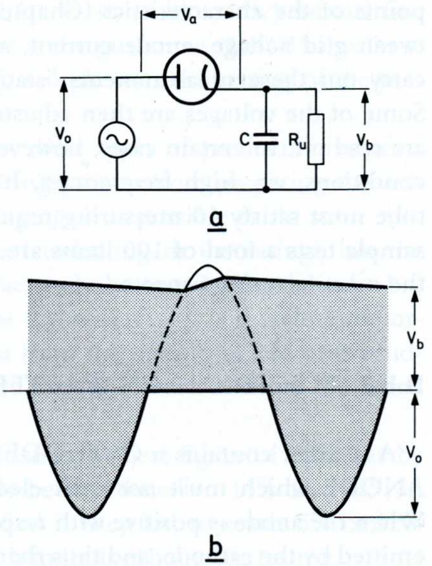


Fig. 58. Rectifier with smoothing capacitor:
a. diagram, b. voltage curve.

define $dV_a : dI_a = R_i$ (INTERNAL RESISTANCE), which in this part of the graph is almost constant. At a still higher anode voltage all the electrons emitted by the cathode are drawn away, and thus the SATURATION CURRENT has been reached. With oxide cathodes this limit cannot be reached; the cathode would be damaged. As already explained in Chapter II.B.2, the saturation current increases with the temperature of the cathode, which in turn depends on the heater current.

In the straight part of the characteristic the internal resistance can be regarded as purely ohmic. At a resistive load as shown in Fig. 57a the peak value of the alternating current will be $I = V_o : (R_i + R_u)$, and the voltage $V_o$ will divide into $V_a$ and $V_b$, inversely proportional to $R_i$ and $R_u$. In Fig. 57b the voltage on the rectifying tube is given: the peak value during the half cycle in which the anode is positive equals $I_a \times R_i$; during the half cycle in which the anode is negative the peak value is $V_o$.

The situation changes when the direct voltage is smoothed with a capacitor, see Fig. 58a. As Fig. 58b shows, the voltage on this capacitor remains positive. Owing to the current drain through $R_u$, this voltage drops slightly over a period that the mains voltage is lower; during the period that the mains voltage is higher than $V_b$, the capacitor is charged again. The tube receives the sum of the mains voltage and the capacitor voltage, and when the mains voltage is maximum negative, the tube voltage is almost $2 \times V_o$.

Apart from a high voltage on the anode, this circuit involves danger to the tube in other respects. If for some reason or other there are gas atoms inside the tube, progressive inonization would cause a breakdown during which the energy stored in the capacitor will be discharged through the tube. The current of positive ions then hitting the cathode may easily damage it (SPUTTERING).

If diodes are used to DETECT (rectify) the signal, it is desirable that the current should be proportional to the voltage to prevent distortion. Therefore, by means of a suitable configuration, the part of the curve starting from $V_a = 0$ must be made as straight as possible.

## II.B.8. TRIODE CHARACTERISTICS

The graphs showing the mutual relationship between anode and grid voltages and currents are known as characteristics.

The most important of these characteristics is shown in Fig. 59a: the anode current as a function of the grid voltage at various (constant) anode voltages. Fig. 59b gives the anode current as a function of the anode voltage at various (constant) grid voltages. At a high positive grid voltage this curve approaches the one of the saturation current (Fig. 56): the grid helps to draw the electrons towards the anode. As the grid voltage approaches zero, a part of the electrons will hover round the cathode as a "cloud", and as the grid becomes more negative, it forces more and more electrons back to the cathode.

Finally Fig. 59c shows the grid current as a function of the grid voltage. As soon as the grid becomes positive, it too, will absorb electrons, and supply current. This implies that energy would be drawn from the circuit which must supply the input signal; a most undesirable situation since this circuit has but little power. So preferably, operation should take place in the area where the grid voltage is negative. Measures should then be taken to keep the grid cool, otherwise it will operate as a cathode. Furthermore it should be ensured that the tube contains no gas, otherwise the electrons accelerated by the anode voltage will collide with the gas atoms, thus releasing more electrons. As a result the negatively charged grid will then attract the positive atom remnants, and still carry current (in the early types of radio tube, which contained a rather large quantity of gas, this ionization was gratefully used to increase the anode current. Unfortunately, however, the whole situation becomes rather unstable and leads to distortion).

When considering the triangle in Fig. 59a, which indicates the change in anode

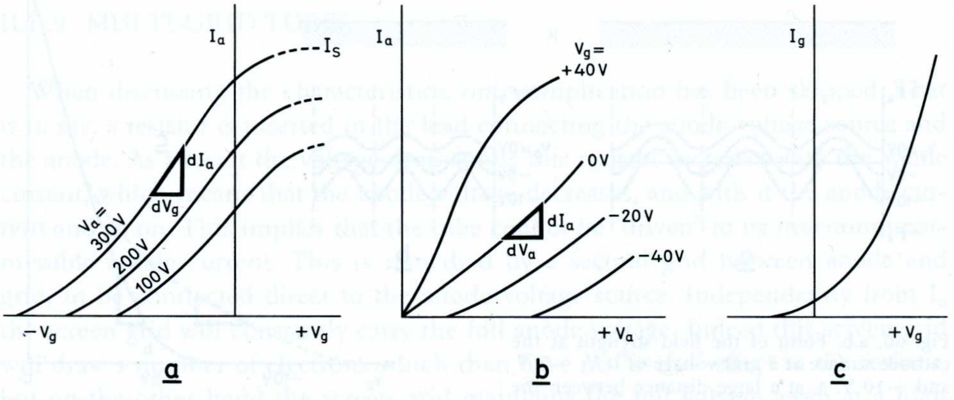


Fig. 59. Triode characteristics. $I_a$ = anode current, $V_a$ = anode voltage, $V_g$ = grid voltage, $I_g$ = grid current.

current, $dI_a$, corresponding to a certain change in grid voltage, $dV_g$, we use the ratio $dI_a/dV_g$ to define S, the SLOPE of the tube (TRANSCONDUCTANCE). Hence S is a measure of amplification related to the anode current.

As appears from Fig. 59b, a certain increase in anode voltage, $dV_a$ causes the anode current to increase by a value $dI_a$. The ratio $dV_a/dI_a$ is defined as an internal resistance $R_i$. The ratio $dV_a/dV_g$ is defined as the (voltage) GAIN:

$$G = dV_a/dV_g = (dI_a/dV_g) \cdot (dV_a/dI_a) = SR_i$$

The slope (S) increases as the distance between grid and cathode becomes less. Moreover, the spacing between the grid wires also affects the slope. If the spacing is wide, the negative voltage between the grid wires will have less influence on the field, so that according to Fig. 60a the field strength on the cathode can become positive even at a grid voltage of $-10$ V. That is why at this high negative grid voltage electrons are still drawn from the cathode (ISLAND EFFECT). The "cut-off" point shifts to the left and $I_a$-$V_g$ characteristic shows a "tail" (Fig. 60c curve a).

To ensure a "distortion free" gain, it is recommendable to operate on the straight part of the characteristic. In case the grid voltage varies but slightly, a small negative grid bias can be applied to shift the working point to the linear part of the characteristic. If the whole "volume" of the tube is used, the curving lower part of the characteristic, the tail, gives rise to distortion, as is still the case in Figs 60c-b. However, the "tail" is quite useful if a tube with a variable slope (S) is required. At a large negative bias the slope is less than at a small one, as shown by the curve a in Fig. 60c. By using grids with variable pitch it is possible to obtain different forms of $I_a$-$V_g$ characteristics. The "tail" of the $I_a$-$V_g$ characteristic is used, for instance, for AUTOMATIC GAIN CONTROL (A.G.C.).

For grid rectification (detection) the bias on the grid will have to be zero. Then the curved part of the $I_a$-$V_g$ characteristic is used, and there will always be distortion. Therefore diode detection is preferred.

The internal resistance $R_i$ depends on the distance between anode and grid. A large voltage gain requires a high value of $R_i$.

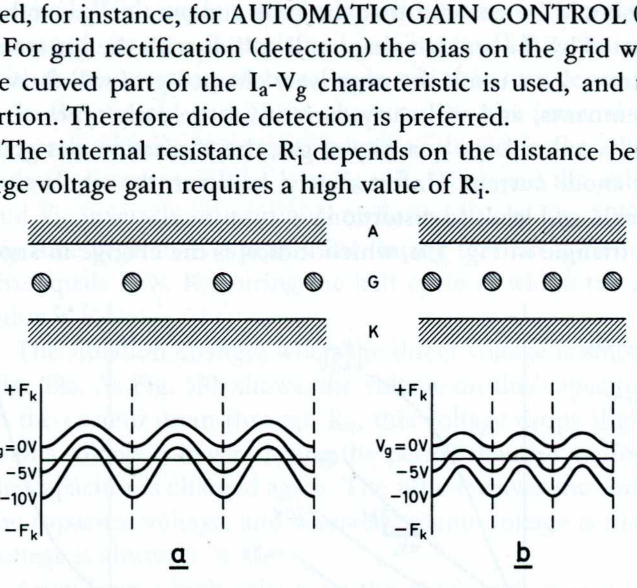


Fig. 60. a.b. Form of the field strength at the cathode surface at a grid voltage of 0 V, $-5$ V, and $-10$ V. a. at a large distance between the grid wires. b. at a small distance between the grid wires.

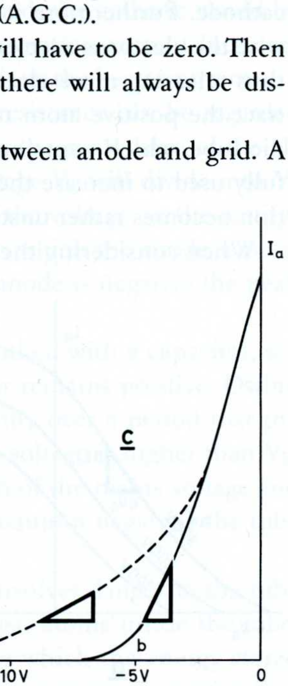


Fig. 60. c. $I_a$-$V_g$ characteristic at a configuration of Fig. 60 a and b.

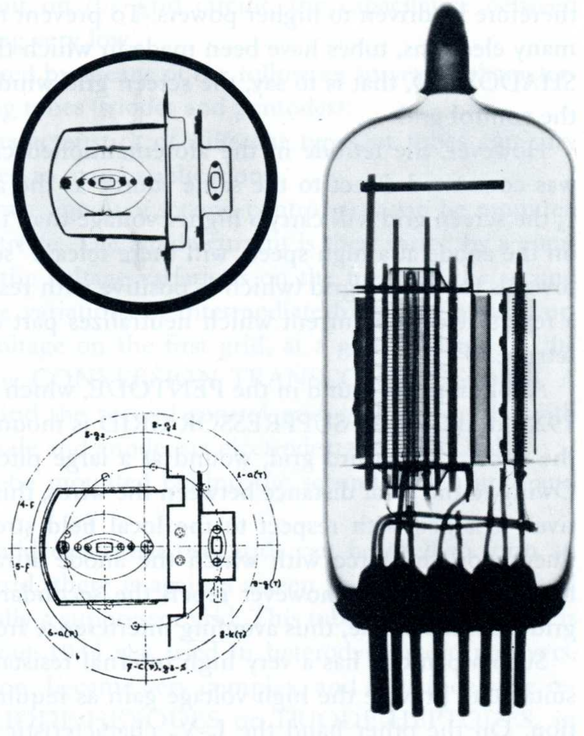


Fig. 61. X-ray photograph and horizontal cross section of the combination tube PCF80.

The power the tube can dissipate is determined by the maximum value of $I_a$, which is strongly related to the emission from the cathode. A high $I_a$ requires more heater current.

Finally there is another important tube quantity which is not directly related to the $I_a$-$V_g$ characteristic, namely the capacitance between anode and grid: $C_{ag}$. In many instances the anode and the grid will form part of oscillating circuitry whose output signals must be amplified without distortion. Owing to the capacitance $C_{ag}$, these oscillating circuits may influence each other, thus giving rise to distortion. Furthermore this coupling might cause undesirable oscillations.

## II.B.9. MULTI-GRID TUBES

When discussing the characteristics, one complication has been skipped. That is to say, a resistor is inserted in the lead connecting the anode voltage source and the anode. As a result the voltage drop across this resistor increases with the anode current, which means that the anode voltage decreases, and with it the anode current and so on. This implies that the tube cannot be "driven" to its maximum permissible anode current. This is remedied by a second grid between anode and grid, to be connected direct to the anode voltage source. Independently from $I_a$ the screen grid will constantly carry the full anode voltage. Indeed this screen grid will draw a number of electrons which then have no "useful" effect in the resistor, but on the other hand the screen grid maintains the full voltage, even at a high anode current, and accordingly draws electrons. This so-called TETRODE can therefore be driven to higher powers. To prevent the screen grid from drawing too many electrons, tubes have been made in which the control and screen grids were SHADOWED, that is to say, the screen grid winding was placed precisely behind the control grid.

However, the tetrode in the aforementioned circuit, in which the screen grid was connected direct to the same source as the anode, has a drawback. At high $I_a$ the screen grid will carry a higher voltage than the anode. The electrons landing on the anode at a high speed, will there release "secondary" electrons which move towards the screen grid (which is positive with respect to the anode). The result is a reverse electron current which neutralizes part of the advantages offered by the tetrode.

A solution was found in the PENTODE, which was invented by TELLEGEN in 1925. In this tube a SUPPRESSOR GRID is mounted between the screen grid and the anode. This third grid, wound at a large pitch, is connected to the cathode. Owing to the great distance between the wires, this suppressor grid, which is negatively charged with respect to the local field strength, will not have a great influence on the force with which the anode attracts the electrons coming from the cathode. It will however repell the secondary electrons reaching the screen grid from the anode, thus avoiding interference from these secondary electrons.

Such a pentode has a very high internal resistance and is, therefore, eminently suitable to provide the high voltage gain as required in high frequency amplification. On the other hand the $I_a$-$V_g$ characteristic is rather flat, which makes the tube also suitable for amplifying high powers.

A secondary advantage (highly important for several applications) is that the suppressor grid, connected to the cathode, provides an excellent screening against

TABLE II. Main data on some typical radio tubes

| tube type | | ECC85 | EF184 | EF183 | | EL84 | EL503 | unit |
|---|---|---|---|---|---|---|---|---|
| function | | r.f. ampl. | i.f. ampl.<br>for T.V.<br>S constant | i.f. ampl.<br>for T.V.<br>S variable | | output<br>tube | output<br>tube | |
| anode voltage | $V_a$ | 230 | 230 | 230 | | 250 | 265 | V |
| screen grid voltage | $V_{g2}$ | - | 230 | 230 | | 250 | 265 | V |
| control grid voltage | $V_{g1}$ | − 2.2 | − 2.5 | − 2.1 | − 12 | − 7.3 | − 13.2 | V |
| anode current | $I_a$ | 10.8 | 10 | 10.5 | 2.4 | 50 | 118 | mA |
| max. output | $W_o$ | - | - | - | - | 6.0 | 20 | W |
| slope | S | 6.8 | 15.6 | 10.6 | 0.5 | 11.3 | 23 | mA/V |
| internal resistance | $R_i$ | 8.3 | 380 | 500 | | 38 | 7.3 | kΩ |
| capacitance | $C_{ag1}$ | 1.5 | 0.005 | 0.005 | | 0.5 | 2.2 | pF |
| heater voltage | $V_f$ | 6.3 | 6.3 | 6.3 | | 6.3 | 6.3 | V |
| heater current | $I_f$ | 435 | 300 | 300 | | 760 | 1050 | mA |

the reaction of the anode circuit on the grid circuit; the capacitance between anode and control grid, $C_{ag}$, being very low.

The foregoing may be illustrated by means of the following survey of characteristics of some modern amplifying tubes (triodes and pentodes):

According to table II the characteristics of different types of tubes can vary widely so that they can be adapted to many applications.

For the purpose of "mixing" two signals, a second control grid can be mounted behind the screen grid of the tetrode. The anode current is then varied by a combination (usual the product) of the voltage variations on the first and the second control grid. The quotient of the variations in intermediate frequency anode current, and the high frequency voltage on the first grid, at a given voltage on the second control grid, is called the CONVERSION TRANSCONDUCTANCE. A second screen grid is fitted behind the second control grid, yielding a tube with four grids or, including the cathode and anode, six electrodes: the HEXODE.

An extra suppressor grid can be mounted behind the second screen grid, and then we have the HEPTODE.

Finally another grid can be added, so that two grids can be used to form an oscillator. Behind this second grid, there is again a screen grid, then the control grid, again a screen grid and finally a suppressor grid. This tube with its six grids is the OCTODE which round about 1934 was used in heterodyne radio receivers. These tubes, also their application, became very complex, and later they were replaced for this purpose by TRIODE-HEXODES or TRIODE-HEPTODES, in which a separate oscillator triode and a mixer hexode are contained in one envelope.

## II.B.10. COMBINATION TUBES

As the reject figure in manufacture could be brought down, it became possible to combine several systems into one envelope. This is, of course, a simple matter if the cathodes of both systems receive the same voltage, for then the two systems can be built side by side or on top of each other facing the common cathode. The simplest method was to mount one or two diode plates over the triode or pentode system (duodiode-triode etc.).

With miniature tubes with seven connecting pins it was already possible to build two triodes with a common cathode in one envelope. The A-technique, with eight pins, even offered the possibility of building two triodes with separate cathodes in one envelope. With the Noval technique, employing nine pins, it became feasible to build a triode-pentode with separate cathodes in one envelope. Fig. 61 gives an X-ray picture and horizontal cross section of such a triode-pentode PCF80, clearly showing that quite a number of components must be housed in the envelope with its diameter of 22 mm.

In about 1964 the Americans introduced a range of tubes with twelve pins, which allowed much more complex combinations to be made (COMPACTRON). The advantages obtained from the fact that for several systems no more than *one* envelope and *one* evacuation process were needed, were outweighed, however, by

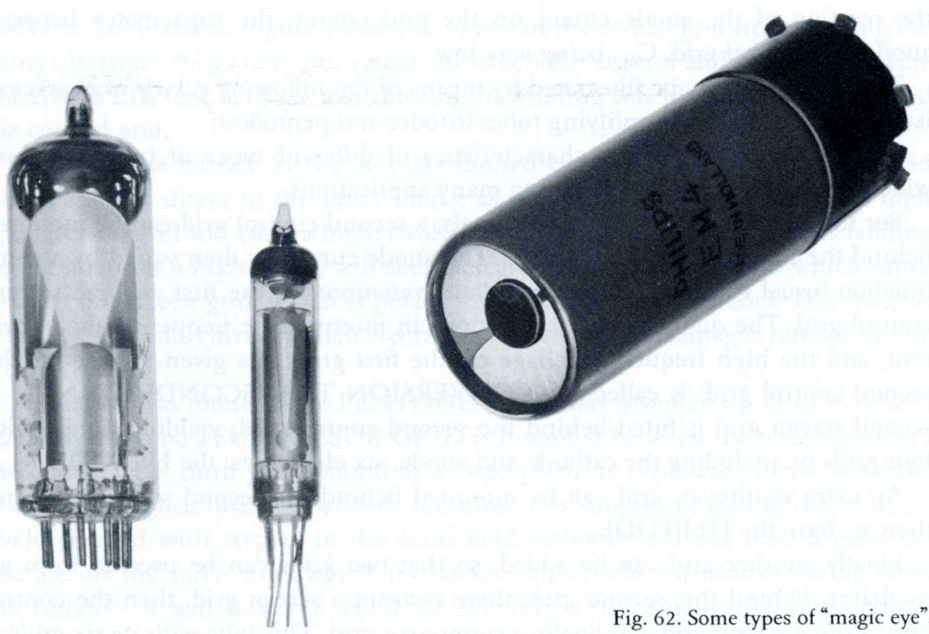


Fig. 62. Some types of "magic eye".

the drawbacks resulting from the complexity of these tubes and the fact that they deviated from the standard Noval envelope.

## II.B.11. THE MAGIC EYE

We have yet to discuss one type of tube which is manufactured in a radio tube envelope, but in principle is a cathode ray tube.

This tube is used to indicate whether the receiver is correctly tuned, and shows a light spot whose dimensions change as the low frequency signal is stronger or weaker (at maximum signal the set is correctly tuned). In practice the A(utomatic) G(ain) C(ontrol) signal supplies the suitable control voltage.

The tube consists of a radio tube cathode and a metal anode which is coated with a fluorescent material. Under influence of the electron current landing on the anode, this screen will light up. An electrode of a special shape is mounted between the cathode and anode, and will focus the electron beam depending on the voltage applied to it; the area of the light spot increases or decreases accordingly.

The A.G.C. voltage can be applied direct to the focussing electrode or it can be fed to the grid of a built-in amplifying triode.

Advantages of this tuning tube are the inertia-free operation and the attractive pattern (say a flower pattern) that can be displayed.

## II.B.12. QUALITY AND RELIABILITY.

It did take some time before designers of measuring instruments used in power current systems, and of telephone and telecommunication equipment, were prepared to accept the radio tube in installations where its failure might cause interruption of work or even endanger human life. The highly complex mechanical, physical-chemical nature of the radio tube initially restricted the normal life to 1000 hours at the most, whilst sometimes failures might occur much earlier.

The failure risk and the life time to be expected could hardly be determined on the basis of data supplied by consumers here and there. Industry had to prepare itself for carrying out life and shock tests under accurately controlled and recorded conditions. Fig. 63 shows a photograph of a life test installation, in which the heater current, anode voltage and grid voltage can be accurately adjusted and kept constant. Fig. 64 shows equipment in which the radio tubes are subjected to vibration tests at frequencies of 0-5000 Hz with adjustable amplitude.

When considering the reject occurring in these life tests, they can be divided into three groups:

a. During the first hours (EARLY FAILURES) the reject figure is much higher

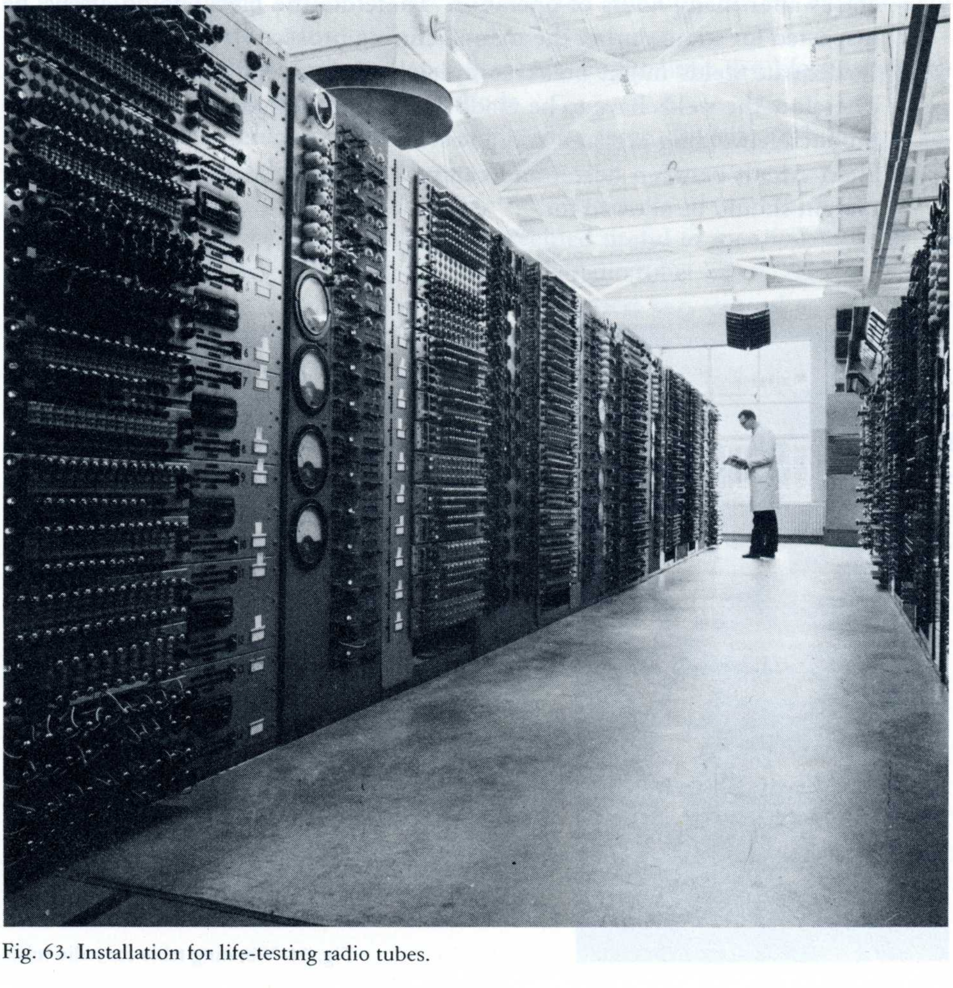


Fig. 63. Installation for life-testing radio tubes.

than in the following period of hundreds of hours. These initial failures are due to production faults which did not manifest themselves in the tests carried out during the production process. To find such faults it might be desirable to age all tubes, say, 48 hours. The faults occurring during this aging test supply rather quick information concerning production faults, and if the reject figure is too high, the whole batch must be destroyed.

b. INCIDENTAL FAILURES occur quite regularly during the first hundreds of hours, particularly under conditions of vibration, impact and high temperature. They can have the following causes:

- Dust particles in the tube or particles released from one of the components might lodge somewhere in the small insulation distances thus giving rise to noise. The remedy: work under clean conditions and wash properly.
- Development of gas by pollution or otherwise might lead to breakdown before this gas is absorbed by the getter. The remedy: fire the components before assembly and degas them on the rotating pump.
- Strain in the glass round the sealings of the lead-ins might cause small cracks after many hours of operation. Therefore the glass should always be inspected for strain during the manufacturing process.
- Inadequate welds might break completely after several hours of operation. Therefore the welds have to be made by trained operators working with a reproducible welding pressure and power.
- Short circuit between heater and cathode. In this respect a sufficient safety margin should be allowed for.

Fig. 64. Vibrating table for radio tubes.

c. GRADUAL FAILURES do not bear an incidental, but a principle character and restrict the life of any tube.

- End of life cathode. As explained earlier, the surface of the cathode is coated with a thin layer of pure barium, which must be maintained. It is found that certain (reducing) pollutions in the nickel of the cathode contribute to the formation of this barium coat, whereas other pollutions cause a kind of "poisoning". Hence the cathode material must be of a very accurate chemical composition. Certain additions to the nickel produce electrolysis in the emitting layer, so that an insulating layer is formed therein (INTERFACE). This process intensifies at higher cathode temperatures, but at lower temperatures the cathode is more sensitive to poisoning.
- Evaporation. The cathode as well as other heated components produce metal vapour which might settle on other components, thus causing secondary emission or short circuit. The design must be such that this phenomenon is avoided.
- Life time of the heater. By means of an appropriate thermal contact between cathode and heater the temperature can be chosen so low that evaporation of the tungsten, which is decisive for the working life of incandescent lamps, plays no part.

  Certain electrolysis phenomena, however, might limit the life time of the insulation between heater and cathode.
- Getter. The getter must have a sufficiently large capacity to bind any released gases for thousands of hours.

Although this survey is not complete, it seemed useful to mention some of the phenomena which are decisive for the life of the radio tube.

Competition from transistors have made consumers very critical as regards life time, and even GRADUAL FAILURES after 1000 hours are not tolerated.

A thorough study into these phenomena, and an extensive process control, manufacturing discipline, and testing system, has led to a range of SPECIAL QUALITY TUBES with a life time of 10 000 hours or over. The study required to arrive at these "special" tubes also led to improvements in reliability and life time of mass produced tubes for radio and television.

# SECTION C. CATHODE RAY TUBES

## II.C.1. GENERAL DESCRIPTION OF TELEVISION PICTURE TUBES

As already explained in Chapter I.11, the picture in the television tube is produced by having an electron beam, whose intensity varies according to "light" and "dark", describe a line pattern across a screen coated with a powder that lights up under the bombardment of electrons.

Therefore the cathode must be able to emit an electron beam of sufficient strength, this strength to be controlled with a grid in accordance with the video signal. Furthermore the beam must be "focussed" in such a way that over the whole area of the screen a small spot is produced. The area of this spot is a measure of picture sharpness: it should not be larger than the picture height divided by the number of lines (625 in the C.C.I.R. system). The electron beam is attracted by a cylindrical anode which carries such a high potential with respect to the cathode, that the beam shoots through the anode and towards the screen. The voltage controls the speed of the electrons, the current their number, and the product of the two sets the luminous intensity, which also depends on the ability of the powder on the screen to convert electrical energy into light.

Behind the anode the electron beam passes between two pairs of perpendicular coils which carry "sawtooth" currents that deflect the electron beam in such a way that it describes the full sequence of 625 lines from left to right, and from top to bottom (Fig. 16a). The standardized ratio between picture width and picture

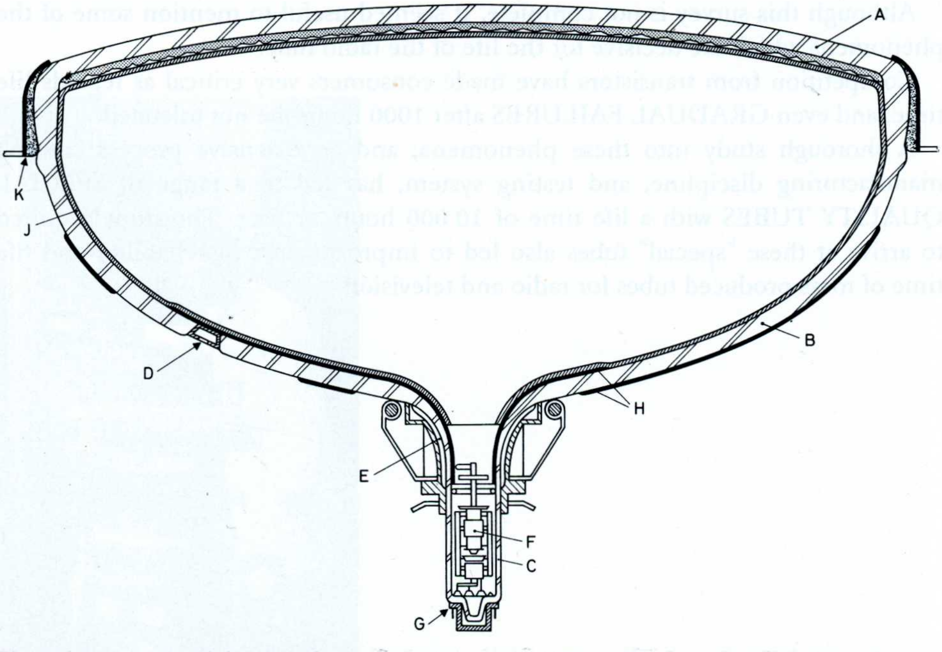


Fig. 65. Cross sectional drawing of a modern picture tube (A59-11 W). A = glass front (screen) coated with a fluorescent powder and a layer of aluminium, B = cone, C = neck, D = high-voltage terminal, E = deflection coil, F = electron gun, G = socket with connection pins, H = aquadag layer, J = aluminium layer, K = rim profile with plastic filling for compressive force (P-tube).

height is 4 : 3, but to facilitate matters for the tube, the picture width was slightly reduced so that a ratio of 5 : 4 was used until 1968. In view of the difficulties involved in making a glass envelope with a perfectly flat rectangular screen and square corners, the screen is made slightly convex, the corners are slightly rounded, and the sides are slightly barrelled.

To offer several television watchers a comfortable view of the picture, it should have rather large dimensions (at the moment a 59 cm diagonal is normal). In order to keep the set as shallow as possible, the gun (combination of cathode, grids, and anode) and the deflection coil must be kept as short as possible, whilst the "deflection angle" must be as wide as possible. Starting from 70° it was increased to 90°, and ultimately to 110° (measured diagonally), now a tube with a diagonal of about 60 cm has a maximum depth of 38 cm. It should be borne in mind that the electron beam must be "focussed" (maintain the same small diameter) across the entire screen and that the electron beam must be deflected horizontally and vertically along straight lines and proportionally to time. Finally the tube must be housed in a decorative cabinet in the living room, and be safeguarded against high voltages, x-rays and explosions. Fig. 65 gives a cross sectional view of a modern picture tube.

Let us now investigate how the required results can be obtained.

## II.C.2. THE ENVELOPE

In the early years of television glass envelopes with a diameter of 9 inches were blown. Experiments have also been made with round tubes with a metal cone and a flat glass screen, but for the rectangular format pressed glass was the only solution. It should be noted that the rather flat screen must be able to withstand a vacuum on the inside, and atmospheric pressure on the outside, which, at a screen area of 2000 cm$^2$, amounts to a pressure of 2000 kg. This means that a rather thick glass is needed (the envelope of a 23 inch tube weighs about 12 kg).

The screens are pressed between an upper and lower die (Fig. 20) which are machined of a special steel and ground into the accurate shape. From a large furnace the molten glass flows through a FEEDER to the die which periodically opens to receive a portion of molten glass in the lower part. To prevent the glass from solidifying it must flow continuously through the feeder. Therefore such a glass factory must work non-stop (four shift system) and a few portions must be fed per minute, so that the minimum production may be about 1 000 000 pressings per year. Measures must be taken to keep the glass free from bubbles and stones, which would spoil the picture. In the press the molten glass is pressed into the required shape between the upper and lower die, after which it is transferred to a tunnel oven in which it is cooled slowly to prevent residual strain in the finished product.

Initially the cone was made according to a "centrifuging" process, in which molten glass is dropped into a rotating lower die so that it crawls up against the wall of this die. Then the upper die moves down to give the cone the correct shape. By increasing the deflection angle the cones became shallower, and so it

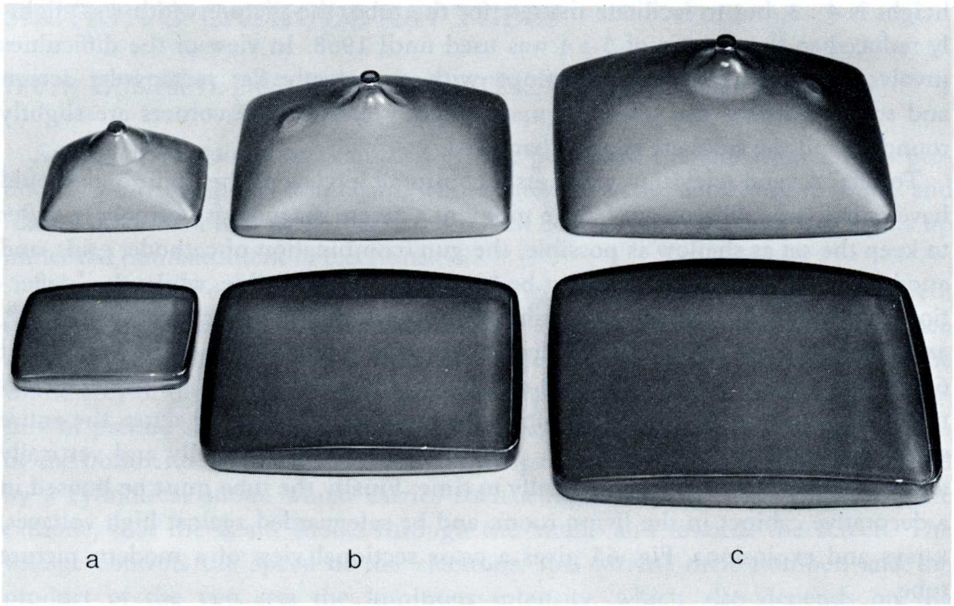


Fig. 66. Some examples of glass pressings for picture tube envelopes; glass front screens at bottom, cones on top: a. 11″ 90°, b. 17″ 110°, c. 23″ 110°.

was easier to make cones and screens on the same press. Fig. 66 shows a few samples of glass pressings used for picture tubes.

The "neck" of tube glass, into which the "gun" is to be fitted, is sealed to the cone. The funnel shaped transition between cone and neck requires particular care. On the inside there should be sufficient room for the electron beam to deflect; on the outside the magnetic circuit of the deflection coil must be mounted as close as possible to this beam to limit the power required for deflection to a minimum. All this implies that the funnel must be pressed with the highest possible degree of accuracy, and in some cases a grinding operation is required to obtain the correct form.

Reliable sealing of screen and cone whilst maintaining the correct shape presented quite a problem, for it is neither possible to apply heat nor to give mechanical support from the inside. The best solution was found by CORNING in having the seam rotate between directed gas flames to which a high-frequency voltage is applied. The gas flames heat the seam between cone and screen so that it becomes conductive for the high-frequency current which is supplied via the gas flames to heat the glass uniformly at the place where the seal must be made. Then the screen and cone are pressed together, and rollers ensure that the correct external shape is obtained. Then again into the "cooling" oven, in which the glass is relieved from all internal strain. Philips have succeeded in building a rotating sealing machine which operates fully automatic.

As already explained in Chapter I.13, such a factory in which the glass parts for the envelopes of picture tubes are pressed and assembled, requires an investment of about 10 million dollars, and can only yield profits by turning out very large quantities. In the free world there are at the most 20 of such factories where these envelopes are made.

## II.C.3. THE REINFORCED TUBE

As already said before, the total pressure on the screen of a picture tube amounts to 2000 kg, which implies that every television set contained some kind of "bomb". One crack in the screen would cause it to be pushed in at a tremendous force while at the same time pieces of glass would fly in all directions. Even more dangerous, however, is the fact that in case of a mechanical fault at the cone end, the gun would be pushed forward with an enormous force to be "shot" through the screen into the room. As a protective measure the sets of before 1969 were equipped with a glass plate in front of the screen to serve as a shield if IMPLOSION of the tube should occur. In America a pressed protective glass plate was later glued to the screen (TWIN PANEL). In 1960 Philips found a way of making the tube "implosion-proof", and the protective glass was no longer needed: this was the so-called "P-TUBE". The screen of this tube was provided with a radial compressive strain directed inwards (tensile strain leads to cracking), while at the same time the sealing ridge (where there is the gravest danger of cracking) was held together by a layer of plastic. This was done by surrounding the seam by a frame of profiled sheet steel, and pouring a thick layer of polyester resin in the groove formed by this frame and the envelope (Fig. 65). Fig. 67 illustrates how this operation was initially done. When hardening, the layer expands, and the compressive stress is produced whilst the resin holds the glass together in case the rim

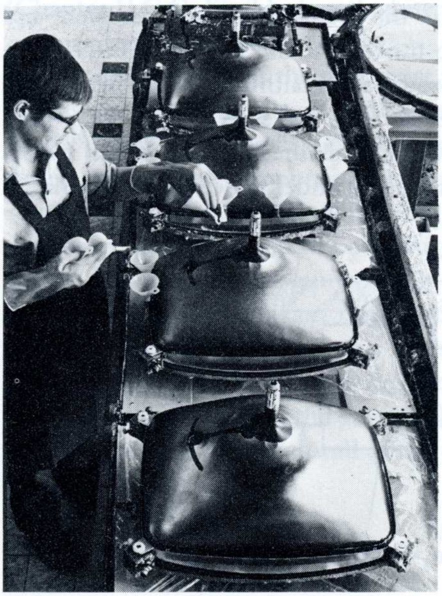


Fig. 67. Manufacture of the reinforced tube. Polyester resin is poured between the metal rim and the envelope.

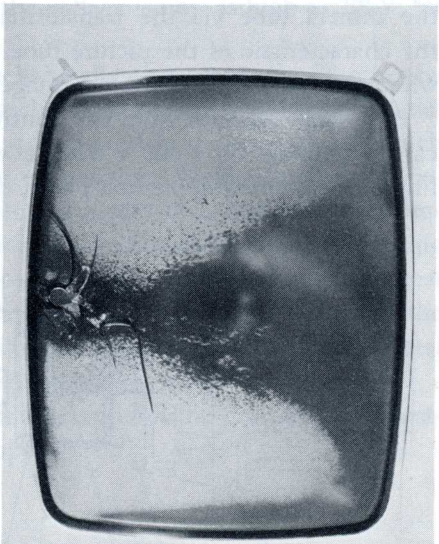


Fig. 68. Reinforced tube with a hole hammered in near the edge.

should crack. The effect is shown in Fig. 68: with a firm hammer blow, given in the critical spot, a hole was struck in the envelope quite near the ridge: the cracks have not extended and no pieces of glass flew around. The outer air entered through the hole at such an enormous speed that all fluorescent powder was blown away.

Later on other methods were used to obtain the same effect, namely by using a metal strap that is drawn round the glass with a certain stress.

## II.C.4. THE GUN

The assembly of components responsible for emission, acceleration, and focussing is known as the gun, a modern version of which is shown in Fig. 69.

In the first place it contains a cathode, consisting of a small nickel tube, with inside it the heater. At one end this nickel tube carries a cap on which the emitter is fitted. The specific emission must be very high (say 1.5 A/cm$^2$), since the electron beam must be very narrow. To keep the heater power low (2 W), the cathode is supported by thin strips of metal of low heat conductivity.

The part of the control grid is played by the WEHNELT cylinder, G1, which contains a very small and accurately shaped hole (about 0.6 mm diameter). At zero grid voltage with respect to the cathode, the anode will draw the maximum number of electrons (about 1 mA) from the cathode (resulting in "white" on the screen). As the grid becomes more negative the current decreases, showing "grey" on the screen, to become "black" at a high negative voltage of the grid. The voltage variations on the grid are known as the VIDEO signal. Due care must be taken that the addition of all transformations from the current/light curve of the camera tube via the transformation in the transmitter and receiver, and the characteristic of the picture tube, are so tuned that the GRADATION curve white-grey-black of the original image is reproduced as faithfully as can be.

At a small distance from the control grid we have the screen grid, G2, which carries a voltage of + 400 V and draws electrons from the cathode. When travelling through an equipotential plane, electron beams undergo deflection, similar to

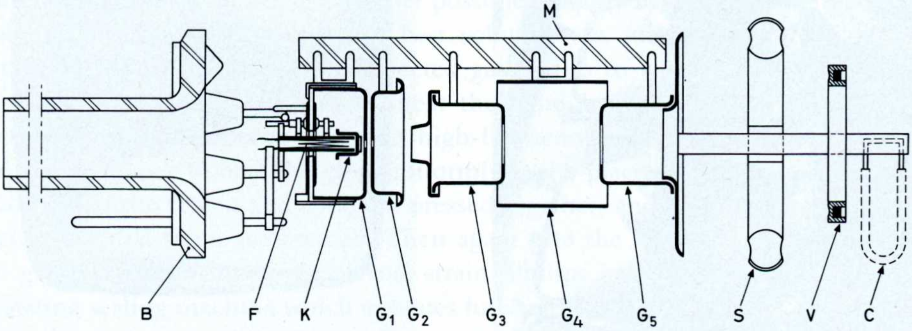


Fig. 69. Gun of a television picture tube. B = tube base, F = heater, K = cathode, $G_1$ = control grid, $G_2$ = screen grid, $G_3$ $G_4$ $G_5$ = anode plus electron lens, M = multiform connecting piece, S = supporting spring, V = getter, C = contact spring against the envelope wall.

"REFRACTION" of light-rays. As we know, a bundle of light can be "focussed" by a lens. The same happens to a bundle of electrons which are led through rotation symmetrical equipotential planes of definite form. The bundle is already focussed in the field between G1 and G2; then the electrons are accelerated by anode G3 which is connected to the high tension supply. To "focus" the electrons to a beam which produces a small light spot on the screen, this aceleration anode is again designed to form an electrostatic lens, G3-G4-G5, of which G3 and G5 are connected to a high voltage, whereas G4 is given a lower voltage with respect to the cathode. Small focus adjustments can be made by varying the voltage on G4. Outside the tube there is, furthermore, a ring with a small magnet. By turning the ring it is possible to compensate slight asymmetries and to aim the electron beam, exactly at the centre of the screen if there is no deflection.

Immediately behind the gun, outside the tube, there is the deflection coil which must cause the electron beam to swerve across the screen at deflection angles of 110°. Measures must be taken to prevent the electron beam from spreading during deflection (DEFLECTION-DEFOCUSSING). It is therefore necessary that the electron beam is accurately round, as thin as possible, and runs exactly through the centre of the deflection coil. This, and the correct design of the deflection coil (see Chapter II.C.5) must ensure that the deflection in one direction, during deflection in the other, remains constant (there should be no PINCUSHION or BARREL DISTORTION).

By means of a few contact springs, the cylinder G5 is electrically connected with the aquadag layer on the inner wall of the envelope, and via the aquadag layer also to the high tension contact (Fig. 65).

The wish for higher light intensities at a small illuminated area (SPOT) leads to higher anode voltages, which were already needed because of the metal backing (Chapter II.C.6). For black and white tubes the voltage is now as high as 20kV; for colour picture tubes even 25kV.

The various parts of the gun must be mechanically connected to each other by means of connecting pieces that can stand such a high voltage. A suitable material is MULTIFORM (powder glass). The gun is mounted in a precision jig which keeps the various components in the required position. Welded to the components are small pins into which the connecting pieces of a powder glass compound are pushed. Then the powder glass is melted, after which the mounting jig can be removed and the gun is firmly fitted together. The gun, in which such great potential differences occur, can function satisfactorily only if the surface of the metal components is absolutely smooth (sharp points cause breakdown), metal and glass surfaces are perfectly free from dust, and no gas is developed.

The gun is fitted with the getter (V) which at the end of the process is evaporated to deposit on the inner wall of the envelope.

## II.C.5. THE DEFLECTION COIL

Deflection must take place in the conical part of the envelope, as otherwise the electron beam, at a large deflection angle, would touch the glass. On the other

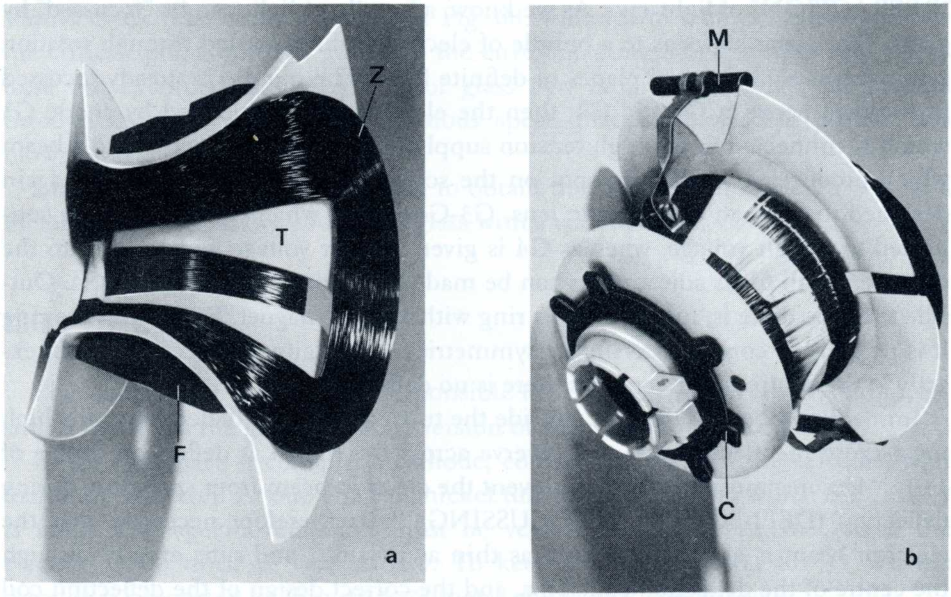


Fig. 70. Deflection coil for 110°: a. one half seen from the centre of the tube. b. complete coil seen from the gun. Z = saddle coil for horizontal deflection, T = toroid coil for vertical deflection, F = ferroxcube ring, M = correction magnet for pin-cushion distortion, C = adjustable ring with magnets for centring correction.

hand the deflection coil must be placed as close as possible to the electron beam, that is where the tube diameter is smallest. Thus we arrive at the form shown in Fig. 70, where the coil is slid over the transition between neck and cone. The magnetic circuit consists of ferroxcube (Chapter I.17) which has a good permeability and involves hardly any losses at the high frequencies used.

The neck is fitted with two so-called "saddle" coils (Fig. 70a), for the large and also rapid horizontal deflection. The coils for vertical deflection are wound on the core (TOROID WINDING). They can be wound only if the ferroxcube rings are split.

As it would be difficult to have the half rings match accurately, a neat solution was found in making the ferroxcube ring in one piece which is afterwards cracked in halves by means of a pilot flame. These two halves cannot but match perfectly. Fig. 70b shows the complete deflection coil; small linearity corrections can be made by shifting the magnet rods M. At the end of the gun we see a rotatable ring by means of which the centring can be corrected (see above).

It has been considered to mount the deflection coil inside the tube: the power supply could then be less. However, as yet no suitable solution has been found for wire insulation, nor a magnetic material that will not produce gas inside the tube.

## II.C.6. THE FLUORESCENT SCREEN

The phenomenon that certain materials exposed to cathode rays will radiate light is known as LUMINESCENCE. It can be explained as caused by electrons in the atom being transferred to another orbit by electrons which collide against the atom at a high velocity. This means that energy from the electron beam is transferred to the atom. Afterwards the atom will reassume its original condition while releasing energy in the form of light. Some materials reassume their original state very quickly, and then we speak of FLUORESCENCE. With other materials it takes longer, and then we have PHOSPHORESCENCE; the powders producing light are called PHOSPHORS.

Different materials produce light of a different wavelength (colour), and also the light efficiency (quantity of light per unit power received from the electrons) is different. A widely used material is zinc sulphide, which emits green light with a short afterglow and high efficiency (up to 100 lm/Watt); "silicates" too, are suitable. Small additions (ACTIVATORS) to the crystal lattice can have an important effect on luminescence. Extensive research has been carried out to find phosphors suitable for the screens of picture tubes; their highest light efficiency must coincide with the electron velocity as occuring in these tubes, and they must react very quickly. Usually green and yellow phosphors are mixed to obtain a colour nearest to white.

The phosphors must, in the form of a fine powder, be spread over the glass screen with the highest possible degree of uniformity. To avoid that light, radiated on the inside from the grains first hit by electrons, is intercepted by the grains behind, the phosphor layer must be extremely thin. Generally this layer is applied by using a method known as SETTLING, according to which a suspension of the powder mixed with potassium silicate is brought into a liquid to be allowed to settle gradually in the envelope with the screen pointing down. Then the liquid is carefully poured out, and owing to the silica-gel formed, the layer of powder clings to the screen where it is dried. In this process any "lumping" of the powder particles can be avoided by previously COATING them with a thin layer of silicate.

A grain struck by electrons will emit light in all directions (Fig. 71), wheras only the light directed outwards is required. Therefore a thin layer of aluminium (METAL BACKING) is applied across the inside of the powder layer. This aluminium works as a reflector and returns all light beams directed inwards. Moreover, there is the advantage that the powder layer, which is not a good conductor, will not be irregularly charged: the whole layer is now kept at anode potential.

Applying the aluminium layer direct to the grains would yield unsatisfactory results, for the aluminium would then form a mountainous landscape, and reflect light in all directions. Hence the "mountainous landscape" presented by the grains must first be smoothed, and that is done by spraying on a coat of nitrocellulose, which is then coated with aluminium by evaporisation. Since the nitrocellulose would produce gas in the vacuum, it is made to evaporate once the aluminium layer has been applied, and so the aluminium layer rests on the tops of the multitude of grain (Fig. 71).

Indeed the aluminium layer too absorbs a part of the electron current, but this loss is more than outweighed by the advantage of light reflection. This absorption is reduced by increasing the voltage (Fig. 72); by aluminizing the light intensity per Watt can be doubled, to reach a value of 50 lm/W in modern tubes.

The aluminium layer is deposited on an automatic machine on which the en-

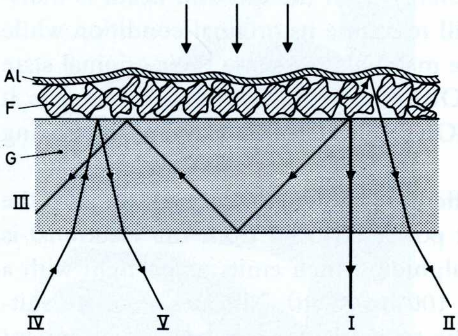


Fig. 71. Reflection and diffraction of light beams in the screen of a picture tube. G = glass, F = fluorescent powder grains, Al = aluminium layer, I = useful direct light beam, II = light beam reflected by Al, III = light beam diffracted and reflected by the glass surface, IV = light beam entering from the room, V = reflected light beam.

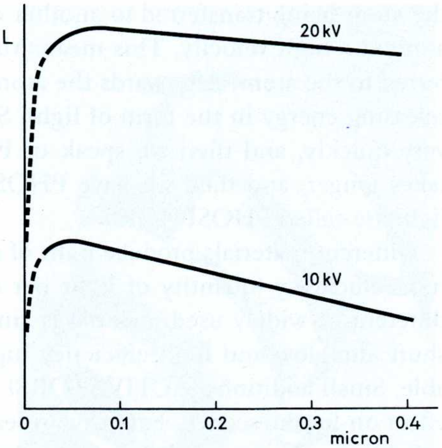


Fig. 72. Luminous intensity (L) at the front of the screen as a function of the thickness of the aluminium layer at anode voltages of 10 and 20 kV, respectively. (Philips Technical Review).

velope is evacuated and a piece of aluminium mounted on a tungsten helix is evaporated; the thickness of the deposited aluminium layer is measured with a high-frequency coil. The current through the heater is switched off the moment the maximum effect is obtained (Fig. 72). After aluminizing the inside of the envelope is given a coat of graphite mixture (AQUADAG) to ensure that the voltage is uniformly distributed over the entire envelope. Then the envelope is heated in a tunnel oven so that less gas needs to be removed by pumping.

The process described in this chapter is a very critical one. It is a mechanical requirement that the layer of powder should adhere firmly to the glass; chemically the fluorescent powder is extremely sensitive to contamination, which may result in discoloration. A minor discoloration can cause a noticeable picture fault which can only be discovered after evacuation. The neck must the be cut off, the screen washed off, and the whole procedure must be repeated from the beginning. A sharp inspection of the basic materials and a rigorous process control have made it possible to increase the yield from 50% in 1950 to 95% to day.

A final complication is the colour of the glass. A picture tube in a lighted room is in principle an illogical system. The light in the room is incident on the white powder of the screen and is reflected (Fig. 71,IV,V). The watcher has to observe the black and white image produced by the fluorescent powder superimposed on the reflected ambient light. The luminescent grains emit light partially in outward direction, a portion of which light is reflected inside at the outer surface of the glass screen (Fig. 71,III). Therefore pigments are added to the glass which colour the screen grey, so that no more than 43% of the light can pass through. Indeed about 50% of the "effective" light of the picture is lost, but reflected daylight and diffracted light must pass at least twice through the glass, and are thus further attenuated.

## II.C.7. SEALING AND EVACUATION

Once the envelope has been fired and the screen has been roughly inspected with ultraviolet light, the gun assembly (Fig. 69) mounted on the base with the exhaust tube must be sealed in the neck. This is done the same way as for radio tubes, the only difference being that centring is now required. At a deflection zero, the cathode ray (SPOT) must arrive exactly at the centre of the screen. That is why for the mounting and alignment of the gun very accurate "jigs" must be used.

This done, the tube must be evacuated, a procedure which in principle involves the same operations as needed for the radio tube (Chapter II.B.6): connecting the exhaust tube, pumping, heating the envelope in an oven, degassing metal components by high-frequency heating, formation of the cathode, evaporating the getter etc. Of all these, degassing the envelope is the most important operation, for the large area coated with phosphors and aquadag will, unless properly degassed, form a source of gas when the tube is put into operation. Here the rotating pump has been replaced by a "STRAIGHT LINE", (an obsolete name since the original straight line has assumed the form of an oval) formed by 95 carriages slowly travelling through an oven. Each carriage, to which one envelope is connected, is equipped with a diffusion pump, a rotating evacuation pump, a circulating pump for the cooling water needed for the base of the tube, and transformers plus controls to supply the heater, the getter, etc. Furthermore each carriage is fitted with an electronic manometer to measure the vacuum. In about one hour the tube has completed a whole cycle, the vacuum is measured, and if that is correct, the tube is "sealed off" by means of a small electric oven.

After that the tube is "aged" on an aging rack similar to the one used for radio tubes (but here in the form of a conveyor, suspended from the ceiling), afterwards to be measured on a large measuring table. Finally the outer aquadag layer is applied and the P-ring fitted.

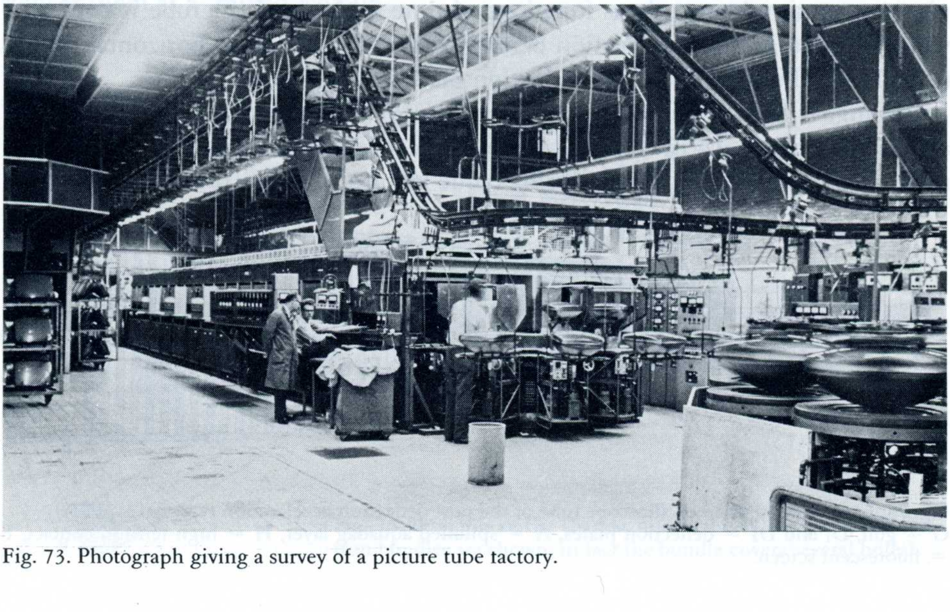


Fig. 73. Photograph giving a survey of a picture tube factory.

## II.C.8. OSCILLOSCOPE TUBES

In deviation from the historical development we have dealt with the television picture tube before the oscilloscope tube because the former takes up the most important position on the market.

Originally the oscilloscope tube was considerably less complex than the television picture tube. The consumer was content with a round tube of small diameter (7, 10 and 13 cm), the envelope being blown complete with screen. The green colour of the zinc sulfide, which produces so much light, was generally accepted (for photographic purposes the blue zinc silicate was better). For rapidly developing aperiodic phenomena which required a delayed display for observation, a second layer of fluorescent powder was used whose light intensity decreased but slowly (this is an even more stringent requirement in Radar tubes, Chapter I.12, where for P.P.I. an image must still be visible when the aerial has completed a revolution).

The "spot" must meet severe demands. The deflection system is electrostatic; hence the input signal can be weaker, and it is possible to reproduce high frequencies without distortion. Since the tube was used as a measuring instrument it was necessary that there should be proportionality between the deflection of the electron beam and the voltage on the deflection plate. This can be achieved only at a small deflection angle (20°) so that the tube automatically became long.

The "sweep" of the electron beam on the screen can be roughly expressed as:

$$Y = (L - l)V_y / 2dV_a$$

where L = distance from deflection point to screen, l = length of the deflection plate, d = distance between the deflection plates, $V_y$ = the voltage on these plates, and $V_a$ = the anode voltage.

To observe an electrical phenomenon as a function of time, it is customary to apply a "saw tooth" voltage to the plates for horizontal deflection, and the voltage under test to those for vertical deflection. A very useful method of watching two phenomena simultaneously as functions of time is obtained if a tube is equipped with two guns, the time deflection being provided by one set of horizontal deflection plates, whereas the vertical deflection plates are separated.

If a higher light intensity of the image is required, increasing the anode voltage entails a drawback in that the sensitivity decreased according to the above formula. A better method is to limit the anode voltage to about 1000 V, but to con-

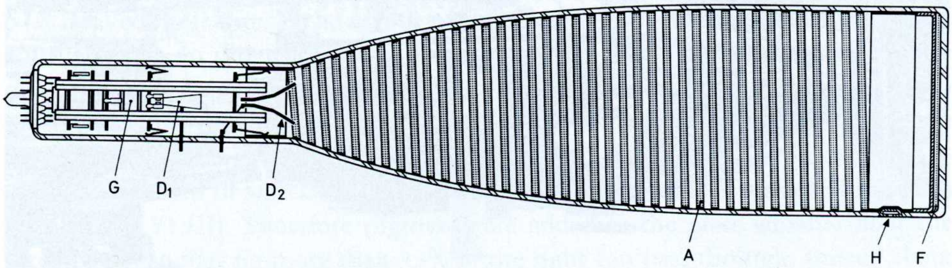


Fig. 74. Cross section of an oscilloscope tube of the post deflection acceleration type (D13-23GH). G = gun, $D_1$ and $D_2$ = deflection plates, A = spiralled aquadag layer, H = high-tension contact, F = fluorescent screen.

nect the (aluminized) screen to a voltage of about 5000 V, and to cover the inside of the envelope with a spiral of aquadag which ensures a gradual transition from 1000 to 5000 V. Another method is placing a high voltage metal mesh between the deflection plates and the screen. Thus the electrons are accelerated behind the anode of the gun and after deflection (POST DEFLECTION ACCELERATION). Fig. 74 gives a cross sectional view of such a tube with aquadag spiral. This tube is already suitable for frequencies of 200 MHz; to minimize the capacitance the connections of the deflection plates are led direct through the glass.

At higher frequencies the "transit" time of the electrons begins to play a part, since the voltage on the deflection electrodes changes in the period that the electron passes by. The deflection electrodes are then made in the form of a flat helix, which provides a corresponding delay. In this way it has become possible to reproduce frequencies up to 900 MHz with the required accuracy.

Furthermore it should be noted that the screens of modern oscilloscopes are pressed, so that they are flat, and often rectangular, and that a raster is applied between the fluorescent screen and the glass to get a scale division.

## II.C.9. THE PRINCIPLE OF COLOUR TELEVISION TUBES

In the television camera the colour image is split up by means of filters into the red, the green, and the blue component. According to the "simultaneous" system as used to day, the video signals of these three components are transmitted simultaneously. The receiver is able to separate these three signals and could lead them to three tubes with red, green, and blue fluorescent phosphors respectively, via which the image could be mixed again by optical means (projection, for instance). The feature of the colour reproduction tube is, however, that the images are mixed in only one tube. This is done by reproducing the images in the three basic colours in dots or lines which are positioned so close together that the watcher experiences them as a "mixed" colour.

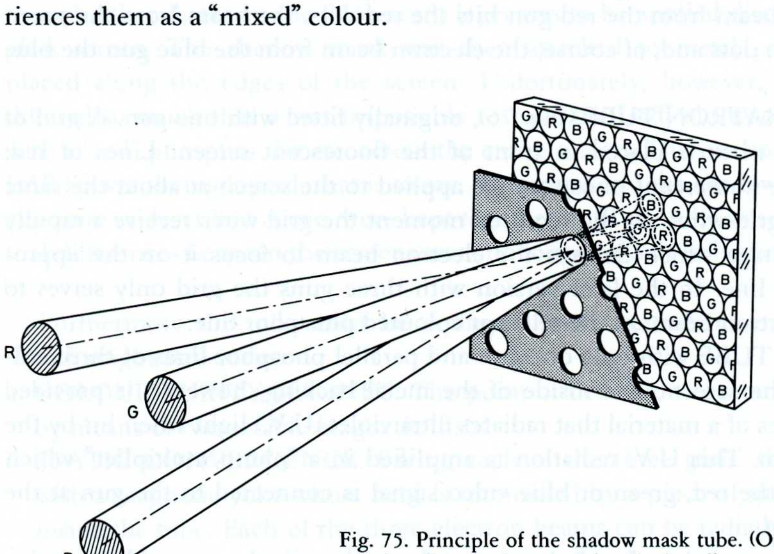


Fig. 75. Principle of the shadow mask tube. (Only partly the electron bundles are shown; in fact the bundle covers several holes).

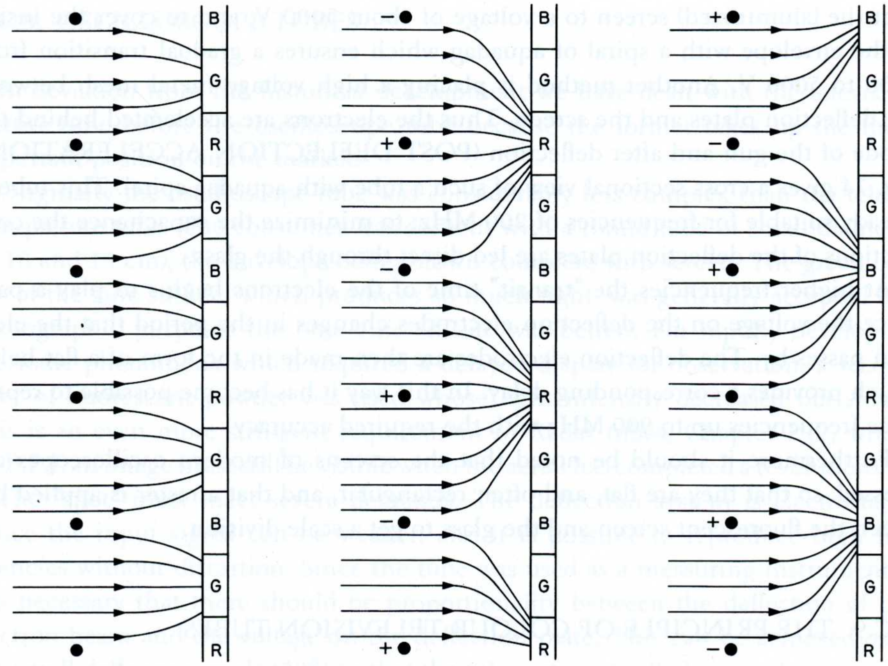


Fig. 76. Principle of the "Chromatron" tube with one cathode. When the voltage on the grid wires are equal, the electron beams are focussed on green, when it is + − the beams are focussed on red, and at − + on blue.

Based on this principle there are three systems:

a. The SHADOW MASK TUBE (Fig. 75), which was brought out by R.C.A. In this tube three guns, for the red, the green, and the blue signal respectively, are mounted side by side to form a triangle. A metal plate with small holes is mounted a few millimetres in front of the fluorescent screen. On the screen small dots of red, green, and blue fluorescent phosphor are so positioned that the electron beam from the red gun hits the red dots, the beam from the green gun the green dots and, of course, the electron beam from the blue gun the blue dots.

b. The CHROMATRON TUBE (Fig. 76), originally fitted with one gun. A grid of thin parallel wires is placed in front of the fluorescent screen. Lines of red, green or blue fluorescent phosphor are applied to the screen at about the same pitch as the grid wires. At the required moment the grid wires receive a rapidly changing voltage which deflects the electron beam to focus it on the appropriate colour line. In the later version with three guns the grid only serves to focus the electron beam on the relevant coloured phosphor line.

c. The INDEX TUBE with only one gun and parallel phosphor lines of three colours across the screen. The inside of the metal backing, however, is provided with thin lines of a material that radiates ultra violet (U.V.) light when hit by the electron beam. This U.V. radiation is amplified in a "photo multiplier" which ensures that the red, green or blue video signal is connected to the gun at the right moment.

It might be noted that the chromatron tube with a single gun and the index tube are both based on the "colour sequential" system (three colour signals in sequential order).

A drawback of the shadow mask tube is that the mask intercepts about 83% of all electrons. In the chromatron tube the grid area is much smaller. In the index tube no electrons are intercepted, but a certain distance between the coloured phosphor lines must be maintained to allow for sufficient switching-over time.

The line pattern of the tubes mentioned under b. and c. is however, less agreeable to the eye than the dots in the shadow mask tube described under a. For various reasons it is simpler to realize the shadow mask tube, so that at present it is the only colour picture tube that is manufactured on an industrial scale. We shall, therefore, restrict ourselves to this type.

## II.C.10. REALIZATION OF THE SHADOW MASK TUBE.

Realization of the requirement: "red" electron beam on the red dot etc. became possible thanks to the photographic method. The phosphor is mixed with a lacquer soluble in water; this lacquer hardens only when exposed to light. First the green phosphor mixed with lacquer is sprayed over the entire screen, then the shadow mask is placed in front of this screen and a light source is positioned in the place where the electron beam from the "green" gun passes through the deflection plane of the deflection coil. Then the screen is rinsed, and the green phosphor remains on the screen only in places where the light beam (and thus the "green" electron beam) has been incident. Then the procedure is repeated for the "blue" and the "red" electron beams with their corresponding phosphor colours.

Since the electron beams do not follow a perfectly straight path, the exposure system described here calls for a correction lens.

Originally the fluorescent layer was applied to a flat glass plate within the tube, and this plate and the shadow mask could be kept accurately centered with respect to each other. Later on the fluorescent layer was to be applied direct to the curved glass screen. The shadow mask was then attached to metal pins which were placed along the edges of the screen. Unfortunately, however, various screens differed so much that it was impossible to have the "red" electron beam coincide with a "red" phosphor dot etc. across the entire screen. The method of MARRIED PARTS was then adopted, that is to say, the screen and the shadow mask, used for exposure as described above, were kept together until the tube was assembled. Mutual differences between the screens were no longer of any influence.

Furthermore the tube must be equipped with controls to correct deviations which might still occur.

a. COLOUR PURITY MAGNET. The point of deflection can be laterally shifted by means of two rotable magnetic discs.

b. STATIC CONVERGENCE. Facing each of the three guns, the tube is fitted with a magnetic yoke whose field is intensified by a piece of magnetic material inside the tube. Each of the three electron beams can be radially shifted by displacing a permanent magnet in this yoke. A fourth magnet serves to shift the blue beam tangentially. Thus the three electron beams can be adjusted to form the corners of a certain equilateral triangle.

c. DYNAMIC CONVERGENCE. Since the mask is almost flat, the distance from mask to the deflection point increases as the electron beam is deflected. Because the three beams must intersect on the mask, the coils round the above-mentioned three magnets in the yoke are fed with currents that vary synchronously with the deflection to obtain the necessary correction.

The effect of the three settings must be read from the indirect effects displayed in the colour picture, and since the different settings affect each other, adjusting a colour television tube is no simple matter. It may even take an experienced person over an hour to obtain satisfactory results. Although practice has shown that perfect pictures can indeed be obtained, it should be borne in mind that the service mechanic as well as the final inspector in the factory must be well trained and have special equipment at his disposal.

Here too, a way has been found to shorten the tūbe by using a wider deflection angle. From 70° deflection in the first round tube a change over took place to the 90° angle in the rectangular tube, and on 2nd May 1969 Mr. Philips could announce at the stockholders meeting the arrival of the 110° colour picture tube.

## II.C.11. MECHANICAL PROBLEMS OF THE SHADOW MASK TUBE

The shadow mask is a metal plate with about 500 000 small holes which are so positioned that in spite of the wide deflection angle all colour dots on the screen are equidistant. These holes are etched after a photographic process; the negative used for that purpose is the result of complex calculation. The shadow mask must be given about the same curvature as the screen. Afterwards it is welded to a metal frame and is then sufficiently rigid.

Metal pins are sealed in the glass screen to which the frame is resiliently mounted without any play. The springs used for mounting are equipped with a bimetal which compensates for the heat expansion of the mask. During the process of applying the three colours, the mask must be placed in front of the screen and removed for rinsing three times in succession. Finally the screen is aluminized and fired, after which the mask is definitely mounted.

Now screen plus mask must be sealed to the cone. A glass-to-glass seal is out of the question, for the fluorescent layer cannot stand the high melting temperature used. Therefore the edges of cone and screen are ground flat and coated with a layer of low-melting enamel. Precision ground cams keep the cone and screen in their exact positions, and in an oven the enamel is made to melt to produce a vacuum tight seal. This done the gun is sealed accurately into the neck of the cone.

Evacuation takes place on the "straight line" as already described with the black-white tubes. Fig. 77 gives an impression of the complete tube.

The shadow mask colour tube is a typical example of a cooperation between different branches of science and engineering, of cooperation between theory and practical experience, which makes it possible to mass produce a highly complex product.

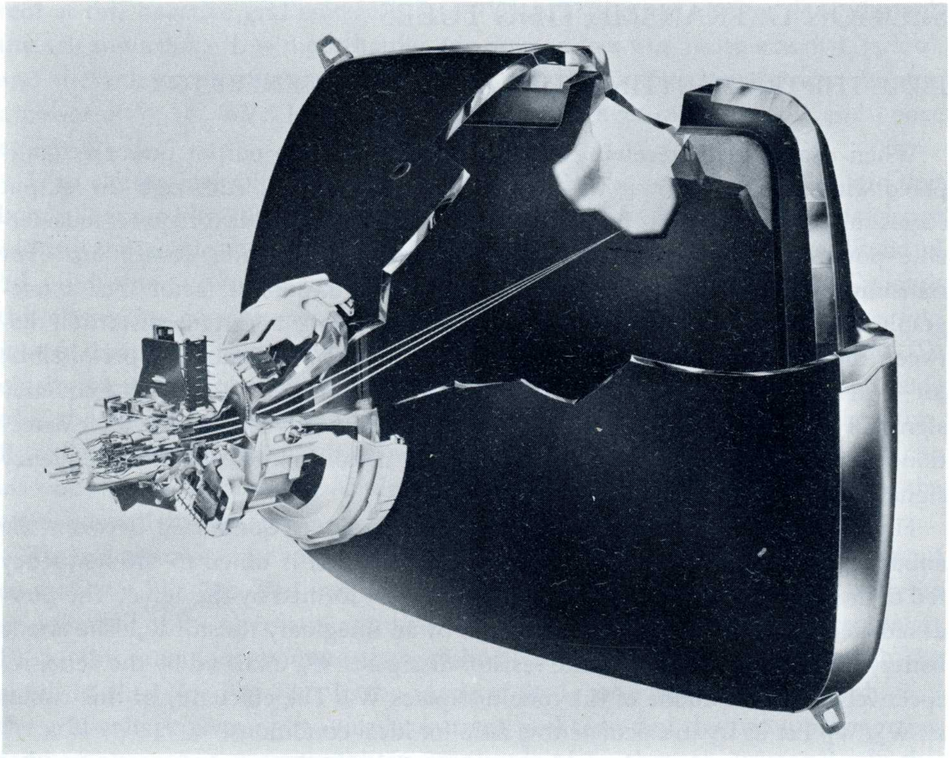


Fig. 77. Shadow mask tube type A56-120X.

## SECTION D. TRANSMITTING TUBES

### II.D.1. THE TRANSMITTING TUBE AS A POWER GENERATOR

When dealing with receiving tubes, calculation of the output power as compared with the input power was left out of consideration. Although the output stages in receivers and low frequency amplifiers sometimes also require considerable power, the requirement of distortion-free amplification comes first. For transmitting tubes the output power is the most important factor. For a telegraph transmitter (or high-frequency heating system), a certain distortion between the alternating voltage on the grid and that on the anode is permissible; for a telephone transmitter the "modulation" added to the high-frequency signal in the transmitter must be reproduced as faithfully as possible after detection in the receiver, but within these limits distortion of the high-frequency signal is acceptable.

Fig. 78a shows the circuit diagram of a transmitter. Connected between the anode and anode battery there is an LC circuit which is tuned to the frequency fed to the grid. Capacitor C is partly or completely formed by the aerial. The power radiated from the aerial can be replaced by an imaginary resistor $R_u$. The anode battery supplies the power $W_i$, the resistor dissipates $W_o$ (radiated by the aerial respectively), and the anode of the tube dissipates $W_a$. The efficiency of this circuit is: $W_o/W_i$. Let us try to calculate this ratio for ideal conditions.

Our starting point is a tube with a horizontal $I_a$-$V_a$ characteristic, meaning that the anode current will remain constant even if the voltage changes while other parameters do not vary. Furthermore the $I_a$-$V_g$ characteristic (Fig. 78c) is assumed to be linear. At a grid voltage $-V_{go}$ the anode current is 0, at a grid voltage 0 the anode current is $I_{ao}$. To make full use of the "grid space", the negative grid voltage is adjusted to $\frac{1}{2}V_{go}$. With no signal fed to the grid, the anode cur-

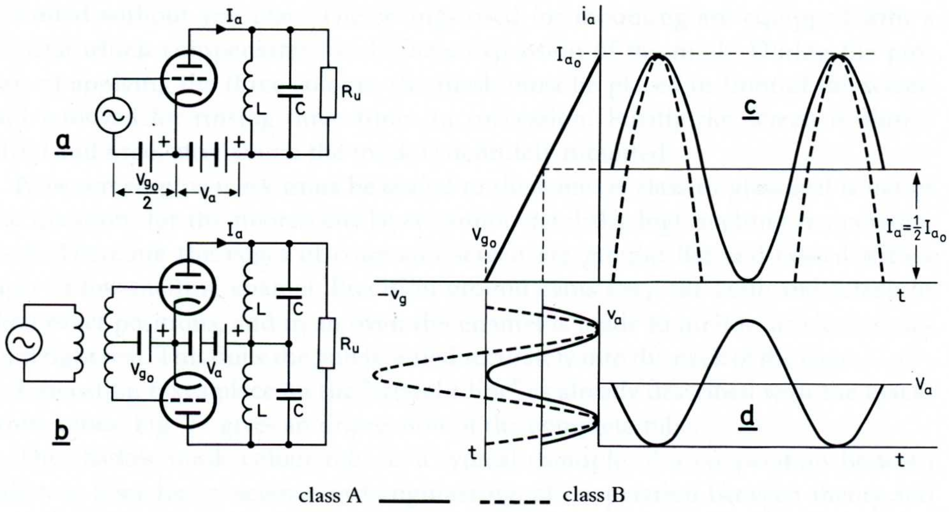


Fig. 78. a. Circuit diagram of a class A transmitter, b. Circuit diagram of a class B, C transmitter, c. Grid voltage and anode current, d. Anode voltage (class A).

rent at this negative grid voltage is $I_a = \frac{1}{2}I_{ao}$. This direct current flows through the self inductance L without hardly any voltage loss (the resistance of L is low), and the voltage on the anode is then $V_a$ (Fig. 78d). The battery now supplies a power of $W_i = \frac{1}{2}V_a . I_{ao}$ which is dissipated as heat in the anode, and hence equals $W_a$.

If an alternating voltage with an amplitude of $\frac{1}{2}V_{go}$ is applied to the grid (left bottom corner Fig. 78c) an alternating current will flow through the anode circuit causing the anode current to vary between 0 and $I_{ao}$, or with an amplitude of $\frac{1}{2}I_{ao}$ (full sine line to the right in Fig. 78c). In the appropriately tuned LC circuit this alternating current encounters a very high impedance and will, therefore, find its way through $R_u$ which will then dissipate a power of $\frac{1}{2}I_a{}^2R_u = 1/8\ I_{ao}{}^2R_u$. To obtain the maximum output, $R_u$ will be so selected that at maximum amplitude of $I_a$ the anode voltage is exactly zero (full sine line in Fig. 78d); to achieve this, $R_u \times \frac{1}{2}I_{ao}$ must equal $V_a$. The power produced in the resistor, $1/8I_{ao}{}^2R_u$, can also be written as $1/4V_aI_{ao}$. Under optimum conditions as described above, the efficiency will then be 50%.

The efficiency can be improved by connecting two tubes in "push pull" circuit as shown in Fig. 78b, and making the negative grid voltage so high that only the positive half of the alternating grid voltage causes a change in anode current (class B). In this situation no more than a short current pulse will flow through the anode (dashed line Fig. 78c). In the anode of the second tube the negative phase of the grid voltage gives rise to similar current peaks, so that the two tubes together send an alternating current with an amplitude of $I_{ao}$ through resistor $R_u$. In this way it is possible to obtain an efficiency of 75%. In class C adjustment the negative bias and the alternating voltage on the grid are increased still further, so that the total grid voltage can become positive. A higher efficiency is obtained (theoretically up to 85%), but then the grid begins to draw current. Of course these considerations will work out only if the cathode is indeed capable of emitting a peak current $I_{ao}$.

## II.D.2. TRIODES, TETRODES, PENTODES, DOUBLE-TETRODES

Fig. 79a gives the $I_a$-$V_a$ characteristic for various grid voltages of a triode whose anode can dissipate about 300 W. Here it is seen that the requirement "$I_a$ independent of $V_a$" is not at all complied with, and also that at a grid voltage of 0 V the anode current is but weak. Hence this type of tube is used only in a push-pull circuit "B" or "C". If this tube is used as a class B amplifier, and the adjusted grid voltage is −135 V, an alternating grid voltage peak of about 240 V is permissible so that an output of 750 W per tube at an efficiency of 72% and a distortion of less than 2,5% is obtained. In the class C configuration a grid bias of −350 V and a grid peak voltage of 535 V yield 1200 W per tube at an efficiency of 79%.

Tetrodes are not only used as receiving tubes, but also as transmitting tubes; a screen grid then compensates the drop in anode current brought about by the voltage drop across the load resistor. To avoid excessive heating of the screen grid, it receives no more than about 1/5 of the anode voltage.

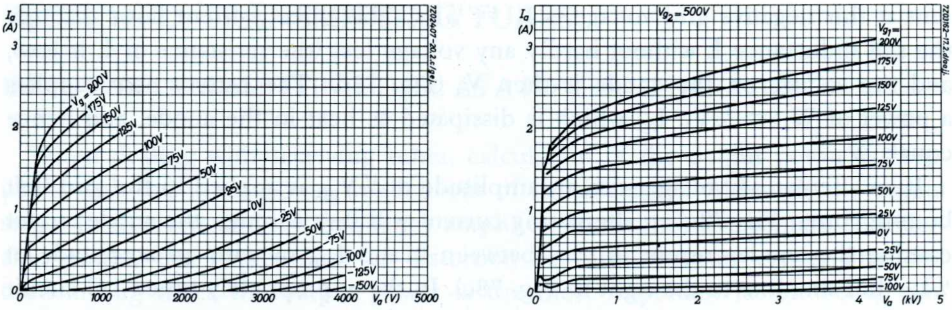


Fig. 79. Left Ia-Va characteristics of the transmitting TRIODE type TB3/750, right Ia-Va characteristic of the transmitting TETRODE type QB4/1100.

Fig. 79b shows the $I_a$-$V_a$ characteristic of a TETRODE of about the same power as the triode of Fig. 79a. Here the anode current is far less dependent on the anode voltage, and even at zero grid voltage there is a reasonable anode current.

As a class B amplifier it is even possible without a positive voltage on the control grid, that is with $I_{g1} = 0$, to reach an efficiency of 68% at 2.5% distortion. When using the tube in class C, an output power of 1000 W and an efficiency of 80% are obtainable. However, as the graph in Fig. 79b shows, here too there is a limit; when the anode voltage drops below the screen grid voltage, the anode current falls steeply. This is due to the fact that the secondary electrons emitted by the anode are attracted by the screen grid which then carries a higher positive voltage than the anode. Consequently the tetrode can only be operated in the region where the anode voltage is higher than the screen grid voltage. With a triode the positive voltage peaks on the control grid can be used for "full drive" of the tube when the control grid current, and thus distortion, plays no part. This results in a higher output. That is why in high-frequency heating systems triodes are preferred.

To meet the drawbacks of the tetrode, transmitting tubes have also been designed as PENTODES. For great powers, however, it is difficult to support the three grids properly, to insulate them from each other, and to provide adequate heat removal.

Improving the tetrode gave better results. In the first place by "shadowing" the first and second grid (as already discussed with radio tubes), owing to which the electrons shooting through the control grid are prevented from landing on the screen grid. As a result the screen grid current is reduced, and efficiency improves. The disadvantage of secondary electrons from the anode landing on the screen grid can be remedied by making the distance between anode and G2 large. A certain space charge will then be formed between anode and screen grid, which stops these secondary electrons. Another method is also fitting "beam plates" at cathode potential beside the anode.

The widely spread application of transmitting tubes in push-pull circuit has often led to combining two low-power triodes or tetrodes in one envelope, the DOUBLE-TETRODE (Chapter II.D.4).

## II.D.3. TRANSMITTING TUBES FOR SHORT WAVE

As already explained in the historical introduction, wave lengths of more than 100 metres were originally used for broadcast and telecommunication. Round the year 1925 the use of wave lengths for A.M. transmitters was expanded to 10 metres (30 MHz). With the arrival of F.M. transmitters the frequencies became as high as 105 MHz and the initially applied T.V. frequencies in the channels I - III even reached 215 MHz. At such high frequencies the self inductance of the tube begins to play a part. Of the total self inductance and capacitance of the oscillating circuit (frequency $= 1 : 2\pi\sqrt{LC}$), the greater part should preferably be outside the tube where it is easier to "manipulate". Individual differences in internal tube capacitances would otherwise call for readjustments to be made each time a tube is replaced.

This is particularly important for F.M. and T.V. signals for which a larger band width is required. For F.M the requirement is a bandwidth of 200 kHz at a carrier frequency of 100 MHz (here the bandwidth determines the modulation depth). With T.V. signals, where the "video" signal is "amplitude modulated", the frequency of this signal is very high (Chapter I.11) and the requirement is a band width of 5 MHz. It can be shown that for the circuit of Fig. 78a the bandwidth is inversely proportional to $CR_u$. Thus to generate sufficient power in a low resistance, the current must be high. Furthermore the capacitance should also be low. A high current intensity requires a cathode with large area and considerable specific emission. Therefore oxide cathodes are used for small powers, whereas tungsten-thorium heaters are used for great powers (Chapter II.B.2 Figs. 45 and 46). If the capacitance of the anode is to be kept low, it should be of small dimensions; during heat dissipation it will then assume a high temperature.

The self inductance outside the tube for these short wave lengths is formed by a U-shaped conductor; the self inductance of the connection between the lead-in and the active electrode inside this tube must be small. Whereas this so-called "Lecher" system is an excellent solution for double tubes, "coaxial" circuits are used for some tubes. In this case the connections of the tube must be in the form of coaxial rings.

Since the high frequency currents are concentrated mainly at the surface of the metal, it is recommendable that the connecting rings be silver plated and make contact with the internal circuit over a large area.

Finally it is impossible to use insulating materials to support the electrodes; under influence of the high field strength and high temperature these materials would entail high dielectric losses. Hence the electrodes are only supported by the wall of the envelope where temperature and field strength are lowest.

In the following a few typical examples of modern transmitting tubes will be discussed in which the aforementioned considerations are put into practice.

## II.D.4. TRANSMITTING TUBES WITH OXIDE CATHODES

The low power consumption of the oxide cathode is, of course, of particular interest for portable equipment. Care should be taken, however, that no contamination to which the cathode is so sensitive takes place due to the high temperatures occurring in the tube.

As explained earlier, efficiency improves when two transmitting tubes are used in a push-pull circuit. The cathodes and the screen grids of the two tubes are then interconnected. A very neat solution is combining two transmitting tubes in one envelope with a common cathode and a common screen grid. This design ensures that the self inductance of the connections between the elements of the two systems is reduced to a minimum. Fig. 80a shows a photograph, and Fig. 80b a cross section of such a tube (Philips type QQE06/40), which has an output of 100 W at 200 MHz, a heater power consumption of no more than 10 W, and an input of 1 W. Hence this type of tube is particularly suitable for mobile transmitters operating over rather great distances. The plan shows the combined cathode, two separated control grids, a common screen grid and two anodes, and furthermore

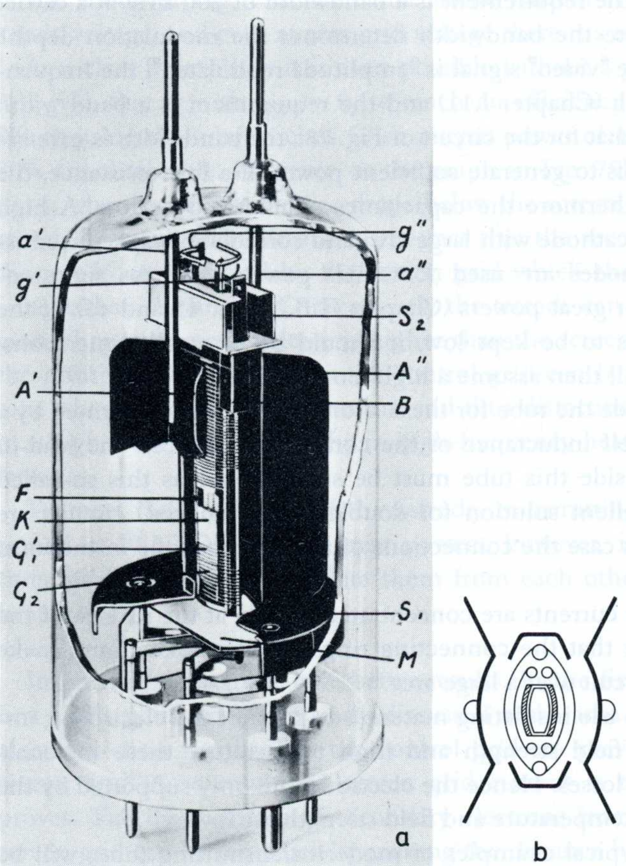


Fig. 80. a. Cut-away photograph of a double tetrode QQE06/40. K = cathode, $G'_1$ = control grid, $G_2$ = common screen grid, A', A'' = anodes, B = beam plates, M = mica plates, $S_1$, $S_2$ = screen plates, g', g'' = brackets forming a neutralizing capacitance between the control grids and the anodes of the other system. b. Cross section of the QQE06/40.

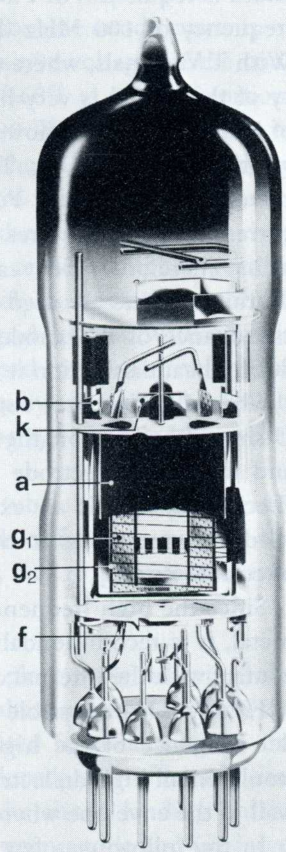


Fig. 81. Cut-away photograph of a double tetrode QQE02/5 in radio tube technique, a = zircon-coated anode, b = beam plates, $g_1$ = frame grid, $g_2$ = screen grid, k = cathode, f = heater.

the so-called "beam" plates B connected to the cathode and serving to intensify the space charge between screen grid and anode. The two control grids are fitted with small brackets which form a capacitance with the anodes of the other system. This simple form of "NEUTRALIZATION" prevents oscillation caused by the parasitic capacitance between anode and grid of any system.

The cathode and the grids are kept in position by two micas and are supported by the connecting pins which are sealed in the powder glass bases. The anodes are only supported by the lead-ins at the top of the envelope. The glass used for the envelope and the powder glass are of a high melting quality whose coefficient of expansion matches that of the molybdenum lead-ins. The latter, after being sealed in, are silver plated outside the glass.

The anodes are made of molybdenum coated with zirconium, which improves radiation and secondary emission, the zirconium functioning as the getter.

Fig. 81 shows a smaller version of such a double tetrode, in which the control grid is a frame grid, in a conventional radio tube envelope. The output of this tube is 7.2 W at 500 MHz at a heater power consumption of only 2 W and an anode voltage of 250 V. It is used, for instance, in transmitters mounted on motor cycles, but later on transistors were developed for such applications.

## II.D.5. COTTON-REEL TUBES

This range of transmitting tubes up to 2 kW owe their name to the reel-shaped graphite anode which has a greater heat radiation than molybdenum, and is connected to the envelope by means of a cylinder made of a material of low heat conductance. Here, the other electrodes are supported by the hard glass envelope by means of molybdenum pins and powder glass seals. To meet the requirements for U.S.W. the dimensions must be small, with the result that the anode assumes a

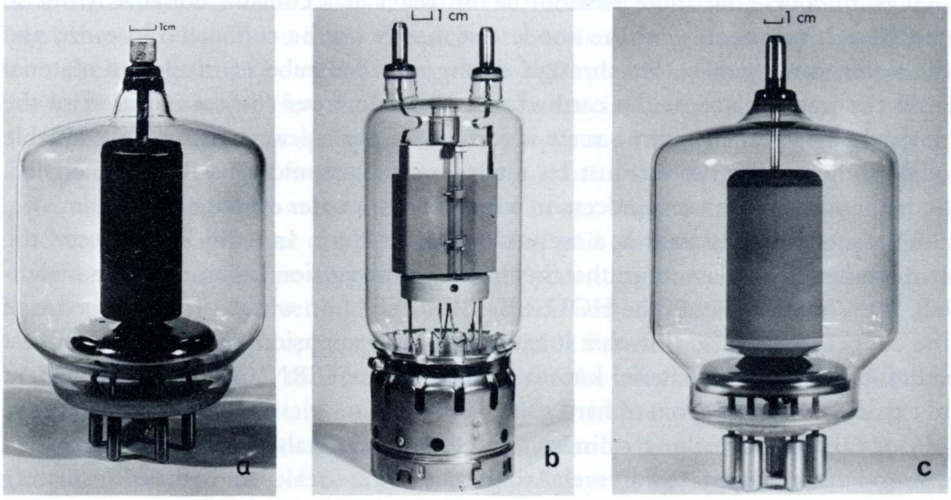


Fig. 82. Modern (a) and older (b) type of transmitting tube for about equal powers, and a modern transmitting tube (c) for twice that power: a. cotton-reel tube tetrode QB3.5/750 for a max output of 1000 W, b. older type of transmitting tube pentode PB3/800 for a max output of 1000 W, c. cotton-reel tube tetrode QB5/1750 for a max output of 1750 W.

rather high temperature. Consequently the oxide cathode would not be able to stand the radiated heat, and therefore its heater was made of THORIATED TUNGSTEN (Chapter II.B.2). Due to the high operating temperatures it is no longer possible to use mica supports for the grids and cathode; the grids are now supported only by the lead-ins through the base.

Fig. 82a shows a photograph of the Philips tetrode QB3.5/750 which can put out 1000 W at 75 MHz.

Fig. 82b shows a pentode for 1200 W at 10 MHz, constructed according to conventional methods, and Fig. 82c a cotton-reel tube for an output of maximum 1750 W at 60 MHz. From the pictures it can be seen that careful design and the proper use of suitable materials have led to a considerable reduction in size.

The cotton-reel tubes have contributed substantially to the development of economic local T.V. and F.M. transmitters as well as simple equipment for dielectric heating.

## II.D.6. TRANSMITTING TUBES WITH AN EXTERNAL ANODE

If the output power is to exceed 2000 W, it is not quite possible to build tubes in which the heat produced by the anode is removed by radiation in the vacuum. For larger powers a part of the envelope is of metal and forms the anode which can be cooled with different means outside the vacuum. It has thus become possible to cool so abundantly that even the heat radiated from the cathode and grids can be removed via the anode. For smaller powers a radiator with cooling ribs can be soldered on to the anode; a flow of air then provides sufficient cooling. For larger powers a honey comb shaped cooling body is soldered on to the radiator; cooling is then provided by compressed air (MULTIFIN). This method has been used to build tubes up to 100 kW (TBL12/100).

It is, of course, far more efficient to use water as a coolant, but that is not so simple as it may seem. For the anode can mostly not be connected to earth, and the water must then be led through a long spiralled tube of insulating material from the water mains or the earthed reservoir. Moreover the composition of the water should be such that no scale is formed. If the tap water is not of a suitable quality, the cooling water must be made to circulate and must itself be cooled. Some manufacturers use a process in which cooling water evaporates to steam.

The anode-to-glass seal is a technological problem. Initially Philips used the chrome-iron-lead-glass joint (the coefficients of expansion being carefully matched). The Americans used the HOUSEKEEPER SEAL in which the sealing edge of the copper anode is so thin that it can follow the expansion of the glass. The best solution is using the material known as KOVAR or FERNICO, whose coefficient of expansion matches that of hard glass.

Once the art of making cylindrical metal-to-glass seals had been mastered, all electrodes were supported on metal rings which were sealed in between insulating glass rings. The tube was thus suitable to be assembled according to a so-called COAXIAL system in which the self inductances and capacitances were formed by concentrical cylinders.

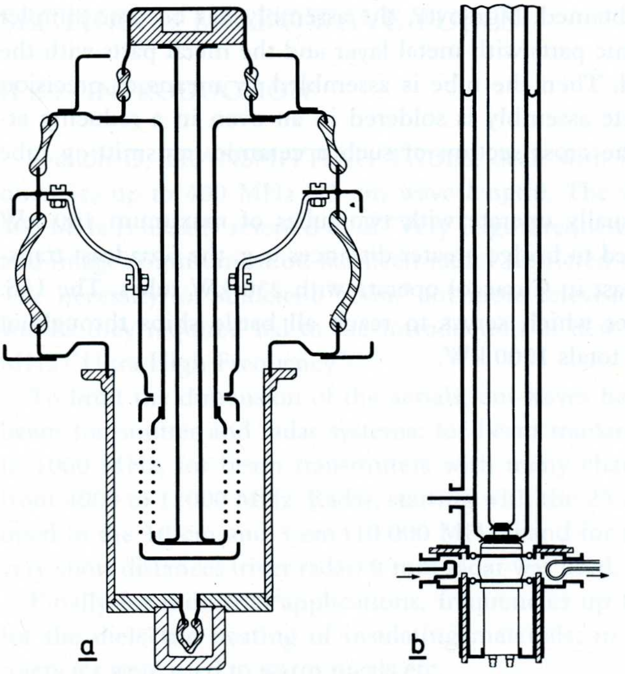


Fig. 83.a. Cross section of a disc seal transmitting tube with external cooling TBW6/20, b. Fitting the type TBW6/20 tube in a coaxial wave guide system.

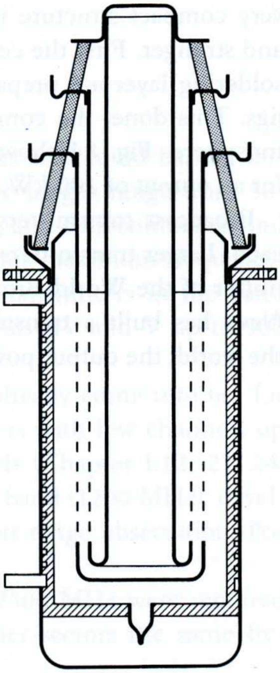


Fig. 84. Ceramic transmitting tube YD1212 for 240 W output.

Fig. 83a shows a cross section of a triode TBW 6/20 which can put out 15 kW at 110 MHz and is suited for either water or air cooling. One recognizes the different rings supporting the electrodes and separated by glass rings; this principle is even used for the two heater connections. To withstand the high voltage, the glass rings must have a certain length. Beside the "active" part of the tube, located inside the anode, a second, passive, part is then formed, consisting of the metal and glass rings. To be able to make the successive seals without damaging the adjacent one, these rings must be rather long, which in turn calls for long supporting tubes for grids and cathode. Here the cathode is of a special design, the so-called MESH CATHODE, a cylindrical plane formed by a wound mesh of thoriated tungsten. So the cathode surface forms a cylinder at a short distance from the grid, which gives a high gain, while at the same time the heater being concentrated on this cylindrical plane saves a lot of heater power. The grid consists of a helix of platinum coated molybdenum wire supported by a molybdenum tube on a metal disc which forms the electrical screening between cathode and anode. In view of heat conduction the anode is a copper cylinder to which a Fernico seal is soldered.

Fig. 83b shows how this tube is inserted in the coaxial system. The different rings and partitions contain holes through which the cooling air can be blown.

In later designs of transmitting tubes the insulation between the rings that carry the electrodes is made of ceramic material. A layer of molybdenum-manganese oxide is applied to the ceramic which can then be soldered to the metal rings. The advantage is that no length needed for sealing is lost and that the height of the ceramic rings is determined only by the required flash-over distance; thus a very compact structure is obtained. Moreover, the assembly has become simpler and stronger. First the ceramic parts with metal layer and the metal parts with the soldering layer are prepared. Then the tube is assembled by means of precision jigs. This done, the complete assembly is soldered in an oven in a reducing atmosphere. Fig. 84 shows the cross section of such a ceramic transmitting tube for an output of 100 kW.

Broadcast transmitters usually operate with two tubes of maximum 100 kW each. Larger transmitters used to bridge greater distances (e.g. the broadcast transmitter of the World Broadcast in Curacao) operate with 250 kW tubes. The U.S. Navy has built a transmitter which serves to reach all battle ships throughout the world; the output power totals 1000 kW.

# SECTION E. MICROWAVE TUBES

## II.E.1. INTRODUCTION

Section D, TRANSMITTING TUBES, dealt with tubes that could handle frequencies up to 400 MHz (75 cm wave length). The wave length range from 30-400 MHz is usually referred to as "Very High Frequency", and conventional sound and image communication has been rather restricted to this maximum frequency. The necessity for sufficient "room" for more television transmitters in the band widths they required led to the introduction of new bands: IV and V up to 900 MHz ("Ultra High Frequency").

To limit the dimension of the aerials, cm-waves had already come into use for beam transmitter and radar systems; for beam transmitters with few channels up to 1000 MHz, for beam transmitters with many channels (Chapter I.12, 23, 24) from 4000 to 12000 MHz. Radar, starting with the 25 cm band (1200 MHz), developed to the 10 cm and 3 cm (10 000 MHz) band for short range observation. For very short distances (river radar) 8 mm radar was used.

Finally for industrial applications, frequencies up to 2500 MHz were required for the dielectric heating of insulating materials; in other sectors the same frequencies were used to warm meals etc.

Furthermore, physicists wished to experiment with still higher frequencies to explore the field between radio waves and infrared radiation of millions of MHz.

Of course it was first tried to use the well known triode to obtain shorter waves, and the next chapter will deal with a triode which could reach 4000 MHz or ten times the abovementioned maximum frequency of the V.H.F. range. This triode has been of excellent use in beam transmitters used for television broadcast. Yet the very rapid electrons need a certain time to cross the short distance between cathode and grid and because of this, there are interference phenomena. Furthermore the ever present capacitance and self inductance pose limits to the application of "density" modulation, and so we come to new types of tube which are based on the principle of "velocity" modulation. Over the years 1940 to 1960 a rather random development took place of electron tubes based on different principles and of different types, so that even for insiders it was difficult to find their way between all these "what-have-you"-trons. Various types of tube were proposed for a certain application, and entirely different versions were introduced under one name. With the danger of new developments changing the picture again, it might be stated that for the most common applications the following tubes are used:

— beam transmitters to 4000 MHz : TRIODES
— high power T.V. transmitters : KLYSTRONS
— beam transmitters over 4000 MHz : TRAVELLING WAVE TUBES
— radar : PULSE MAGNETRONS
— "microwave cooking" : CONTINUOUS WAVE MAGNETRONS

Of each of these applications and tube types we shall discuss one in the following chapters.

## II.E.2. THE MICROWAVE TRIODE EC157/158

Fig. 85a shows a photograph of this tube, partly cut away. The principles for obtaining short waves, discussed in chapter II.D.3, are turned to full use here.

For the greater band width it is necessary to have a high current at a relatively low anode voltage (200 V). To keep the capacitance down, the cathode area must be small: a large emission per unit area is required. A solution was found in the form of the L-cathode, already discussed in Chapter II.B.2; the design principle is given in Fig. 85b. The heater filament is enclosed in a molybdenum cap which is connected to the base of the tube by means of a tantalum tube of low heat conductance. In the top of this cap there is a cavity which contains a pellet of barium strontium aluminate. This cavity in turn is sealed off a porous disc of sintered tungsten. When heated by the filament, the Ba-Sr oxide will evaporate, its vapour penetrating through the sintered tungsten, and after the "activating" process, as discussed with the radio tubes, an emissive layer of barium strontium oxide will have formed on the surface of the tungsten pellet. The advantages over the conventional oxide cathode is that a large store of Ba-Sr remains behind in the tungsten pellet, and will constantly replenish the emissive layer as soon as, for whatever reason, a shortage of Ba-Sr should occur. It is therefore possible to use a much higher specific emission, and as much as 1.5 A/cm$^2$ is permitted, with which lifes of 10 000 hours have been reached.

With the EC157 type the grid-cathode distance is 30 microns, and to avoid the island effect the grid is wound of wire with a diameter of 7.5 and a pitch of 48 microns.

In the construction ("disc seal"), all is done to keep the internal self inductance and capacitance low. The circuit most suitable for that purpose is the one with

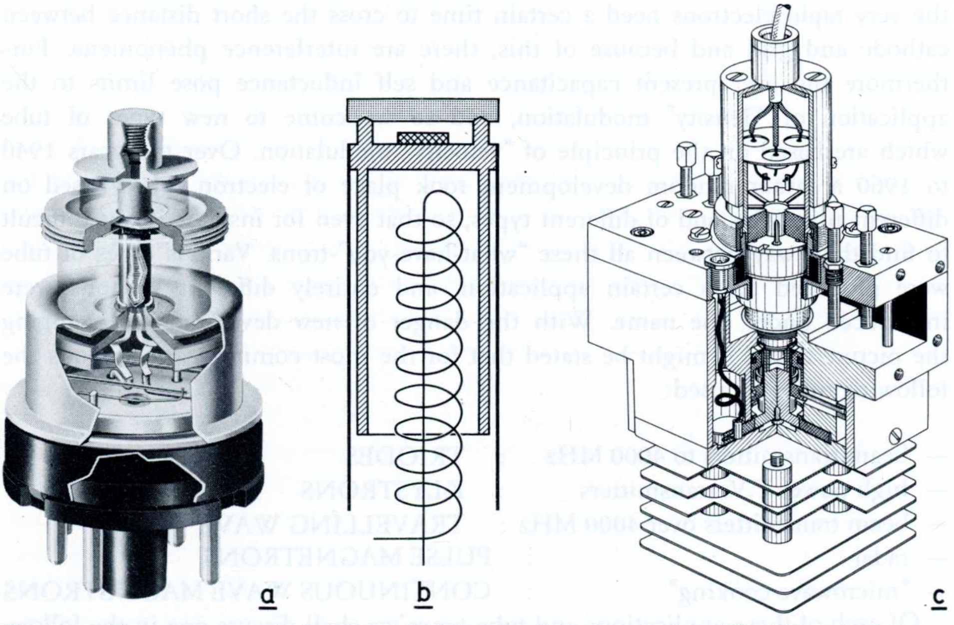


Fig. 85. a. The microwave triode EC158, b. Principle of the L-cathode, c. The EC158 built into a resonator cavity, which serves as an amplifier.

earthed cathode. The result for the EC158 type is an output of 5 W at a band width of 50 MHz and a "power gain" of 6 dB at a heater power consumption of about 5 W.

Its simplicity, easy replacement, low anode voltage, low heater current, and its long life make this tube particularly suitable for unguarded amplifier stations used for beam transmitters. Fig. 85c shows this built into a resonance cavity system which serves as an amplifier circuit.

When considering the grid dimensions, grid to cathode distance and cathode load, it becomes clear that this tube design was not suitable for a further increase of frequency and power.

## II.E.3. KLYSTRONS

There are several basic differences between the triodes and multigrid tubes dealt with so far, and the microwave tubes discussed in the following chapters.

In the first place it is now no longer the amount of electrons, travelling from the cathode to the anode, that is varied, but the velocity of these electrons. All electrons leaving the cathode land on the anode, so that difficulties entailed by the "electron cloud" belong to the past. The electrons are subjected to velocity variations which are removed after amplification, a process taking place independently of cathode and anode. The latter two form no part of the resonance circuit so that the characteristic capacitance and self inductance of these parts and their connections have no effect on the resonance circuit. The resonance circuits are formed by "cavity resonators" in which the electrons move and which have a "resonance frequency" determined by their dimension and form, in analogy with the self inductance and capacitance of the "oscillating circuit", referred to earlier. The main difference being that the resonance frequency of the cavities is much higher.

Fig. 85c already shows a cavity resonator in which a triode EC158 is fitted; in some types of tube the cavities are built into the tube.

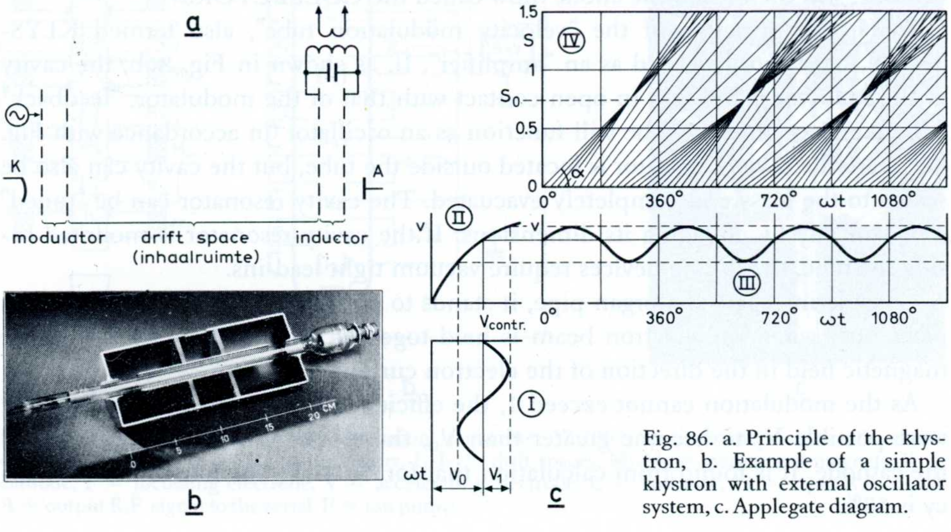


Fig. 86. a. Principle of the klystron, b. Example of a simple klystron with external oscillator system, c. Applegate diagram.

The high frequency power can be fed to, or obtained from, these resonance circuits by means of a lead-in loop.

Finally it is a requirement for velocity modulation that the electron current be bundled, whereas the electrons tend to repel each other and drift apart. Therefore the tube is provided with a constant internal magnetic field, which bundles the electrons closely together to form a narrow beam (during the discussion of the magnetron it will be explained that the effect of the magnetic field is slightly more complex here). Another method is using an electrostatic focussing system.

Let us first give a very simple explanation of the KLYSTRON by means of Fig. 86. Fig. 86a shows the operating principle. On their way from cathode to anode the electrons first pass a cavity resonator, the MODULATOR, which receives the incoming signal of a frequency to which it is tuned. Apart from the fieldstrength $V_0$, the electron beam, according to Fig. 86c, is subjected to a sinusoidally changing fieldstrength of amplitude $V_1$ and, according to graph 86c II, the velocity of the electrons is thus "modulated" in accordance with Fig. 86c III. After that the electrons travel through the field-free "driftspace" (Fig. 86a). The APPLEGATE diagram of Fig. 86c IV gives the path of the electrons as a function of time (tan $\alpha$ is velocity). We now see that the accelerated electrons overtake the non-accelerated electrons with the result that at a distance $S_0$ from the modulator there are subsequent moments of increased or decreased density of electron concentration. This phenomenon can be compared with the periodic compression of air in an organ pipe. The alternating electric field in the modulator has given rise to velocity variations which in turn, owing to the "overtaking effect", result in current variations (more or less electrons flow through per unit time), which are maximum at a distance $S_0$. The energy of these current variations can be a multiple of the energy fed to the modulator because energy is supplied via the anode. If a cavity resonator, tuned to the frequency of the signal which is fed to the modulator, is placed at a distance $S_0$, an intensive electron movement will be produced in this cavity (INDUCTOR) by inductance, and the resulting high frequency energy can be fed to the aerial. Only the direct current component, that is the current produced by the cathode, will then reach the anode (now called the COLLECTOR).

So far the principle of the "velocity modulation tube", also termed KLYSTRON, has been discussed as an "amplifier". If, as shown in Fig. 86b, the cavity of the inductor is brought in open contact with that of the modulator, "feedback" will take place, and the tube will function as an oscillator (in accordance with Fig. 9c). Here the cavity resonator is located outside the tube, but the cavity can also be sealed to the glass and completely evacuated. The cavity resonator can be "tuned" mechanically by changing its dimensions. If the cavity resonator is mounted inside the tube, the tuning devices require vacuum tight lead-ins.

Considering again the organ pipe, it stands to reason that the above effect takes place only when the electron beam is held together. This is done by a constant magnetic field in the direction of the electron current.

As the modulation cannot exceed 1, the efficiency is limited. If in Fig. 86c I it were possible $V_1$ to become greater than $V_0$, the electrons would be returned to the cathode. It is found from calculation that for $V_1 = V_0$ the maximum efficiency is 58%.

Klystrons (particularly the reflection klystrons discussed in Chapter II.E.4) are widely used as local-oscillators for beam transmitters and radar; efficiency and gain are not very high. When used as a high-power output tube the gain can be improved by having the electron beam pass through several cavity resonators, each functioning as an inductor for the preceding part, and as a modulator for the following part. The different cavities can, to a certain extent, be detuned with respect to each other so that a greater band width is obtained.

An interesting example of such a "MULTI CAVITY KLYSTRON" has been made by Valvo for the T.V. transmitters operating in channels IV and V in Germany and other countries. It was a requirement that they could be tuned to frequencies from 400-960 MHz and have a bandwidth of 6 MHz. Initially 10 kW was sufficient for the T.V. network in Germany (the range was already limited by the aerial height and the curvature of the earth). Later the power of a few transmitters was increased to 25 kW. For an output of 10 kW the anode voltage is about 18 kV, and the anode current 1.85 A (efficiency about 33%). By using several amplifying stages, the gain could be made substantially higher than that of a simple klystron; the power gain is about 40 dB.

Fig. 87a gives a diagrammatical survey of the structure of this tube. It contains four RESONATORS (R) flanked on each side by five "drift spaces", in which the electrons accelerated by the resonators overtake each other. The magnetic focussing field is provided here by cylindrical Ferroxdure permanent magnets positioned round the drift spaces. The magnetic field can be controlled by means of magnetic shunts. As can be seen from Fig. 87a, the magnetic fields in the resona-

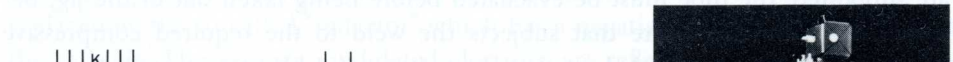


Fig. 87. Multi-cavity klystron for a television transmitter YL1002: a. schematic structure, b. circuit diagram, c. front view. $R_1$-$R_4$ = resonators, $I_1$-$I_5$ = drift spaces, $M_1$-$M_5$ = permanent magnets, K = cathode, F = focussing electrode, V = accelerating electrode, C = collector, H = input R.F. signal, A = output R.F. signal to the aerial, P = ion pump.

tors have a polarity opposite to that in the drift spaces, but that does not form an objection: as long as the magnetic field is directed axially the electron current will be focussed.

Fig. 87b shows how the various parts are connected. The walls of the drift spaces and the magnets are connected to earth, so that the latter can be adjusted without personal risk. The resonators are earthed via resistors; one can freely adjust their R.F. voltages within certain limits. Although at first also the anode (called COLLECTOR here) was connected to earth, it was later given a potential of —5 kV ("depressed collector"). Therefore, after having transferred their high frequency energy in the last resonator, the electrons are slowed down a little before they land on the anode, a procedure which reduces the anode loss, thus improving efficiency. The input signal (H) is fed to the cavity resonator nearest the cathode, and the output signal (A) is picked up from the cavity nearest the collector. The cathode carries a voltage of —18 kV, the electron current is focussed by the focussing electrode F, which is negatively charged, and accelerated by acceleration anode V whose voltage can be controlled during operation. Fig. 85c gives a photograph of the complete tube.

As regards technology the following may be noted: At the locations of the cavity resonators the wall of the tube consists of a special ceramic material (aluminium oxide), which is resistant against powerful high frequency fields. Outside the tube the cavity resonators are fitted in two parts. Rings are soldered to these ceramic cylinders, to be welded together thus forming the field-free drift spaces. An interesting particular is that after assembly, during which the different components are supported, the tube must be evacuated before being taken out of the jig, because it is the air pressure that subjects the weld to the required compressive force. The anode is air cooled, so that the voltage of 5 kV against earth forms no problem.

The cathode is dish-shaped so that a certain focussing is initiated, and is made of tungsten saturated with barium strontium aluminate. In contrast with the L-cathode, where a "pellet" was placed behind the porous tungsten plate (Fig. 85b) "replenishment" now takes place from the material stored in the porous pellet ("impregnated cathode").

Behind the cathode there is, furthermore, an ion pump in which any gas ions are led along a surface of sputtered titanium which functions as a getter. The ion pump can also provide a means of measuring the vacuum.

This tube has been described in slightly more detail as it really concerns a product which requires "all registers of electronics to be opened". It is typical of the sometimes unexpectedly rapid developments needed in electronics that in a few years time about 40 transmitters using this tube had to be put into operation. There was a demand for a second (in some countries a third) programme, the bands I - III had no sufficient room left, so band IV - V transmitters had to be built. Indeed, some initial difficulties had to be overcome, but the second and third programmes were started in time, and quality and range were so good that, given a free choice of two transmitters for one programme, preference went to the band IV - V transmitters.

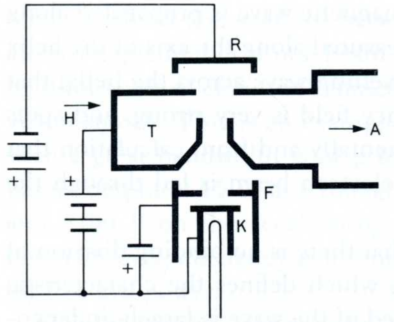


Fig. 88. Diagram of a reflex klystron. K = cathode, F = focussing electrode, T = resonance cavity or resonator, R = reflector, H = input R.F. signal, A = output R.F. signal to the aerial.

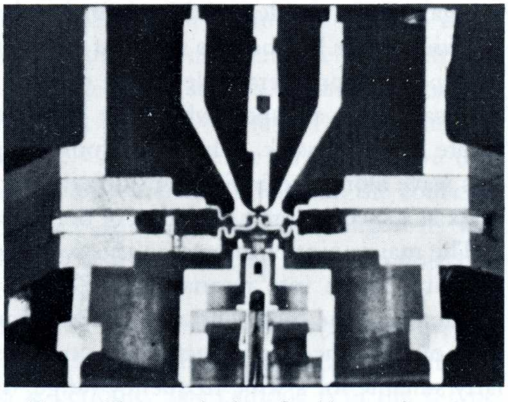


Fig. 89. Photograph of a reflex klystron for a 4 mm wavelength and 200 mW.

## II.E.4. REFLEX KLYSTRONS

Instead of amplifying the signal in several successive cavity resonators the idea was conceived to have the electron current pass several times through only one cavity.

Fig. 88 shows the diagram of this principle. The cathode sends an electron current to the cavity resonator which carries a positive potential with respect to the cathode. The electrons propelled through the cavity resonator are velocity modulated. The positively charged anode or collector behind the cavity resonator is now replaced by the so-called "reflector" which has a negative potential with respect to the cathode. The velocity modulated electrons are reflected by this reflector (REFLEX KLYSTRON), and arrive once more in the cavity resonator. If the dimensions are correctly chosen (the distance between reflector and cavity resonator can be adjusted outside the tube), the reflected electrons will be able to intensify the original alternating field in the cavity, thus leading to a kind of "boosting" as a result of which considerably amplified power can be drawn from the cavity resonator.

Such reflex klystrons are made for extremely high frequencies. Fig. 89 shows a cross section of a reflex klystron for a wave length of 4 mm. The dimensions then become very small indeed: the central apertures, in the cavity resonators have a diameter of 0.25 mm, and once again the L-cathode comes in as a handy means to obtain sufficient power (200 mW) at such a small beam diameter.

## II.E.5. TRAVELLING WAVE TUBES

During the inventivity explosion which after the second world war was aimed at microwave tubes, another type of tube came into existence: the "travelling wave tube". The idea was conceived to replace the sequence of "resonators" and "drift spaces" in the "multicavity klystron" by a metal helix mounted between cathode and collector, the high frequency signal being fed to the cathode end to be picked up again at the anode end (Fig. 90a). An electromagnetic wave is propagated along the wire of this helix at the speed of light, but measured along the axis of the helix this speed is considerably less. The result is a travelling wave across the helix, that is to say, there are spots where the high frequency field is very strong, and spots where it is very weak. It has been found experimentally and from calculation that this wave along the helix is amplified when an electron beam is led through the helix at a speed greater than that of the wave.

The main point of interest in this solution is that there is no binding division of the tube length into resonators and drift spaces which defines the characteristic frequency of the tube. The axial propagation speed of the wave is largely independent of frequency, an aspect which makes such a travelling wave tube so particularly suitable for a wide frequency range, (that is to say for a very wide frequency "band" if the carrier wave lies in the middle of this frequency range). Whereas the klystron discussed in Chapter II.E.3 could be adjusted to a basic frequency between 400 and 960 MHz, the band width this tube can amplify is "only" 6 MHz, and that figure can be reached only by detuning the different resonators with respect to each other. The travelling wave tube illustrated in Fig. 90, (YH1170),

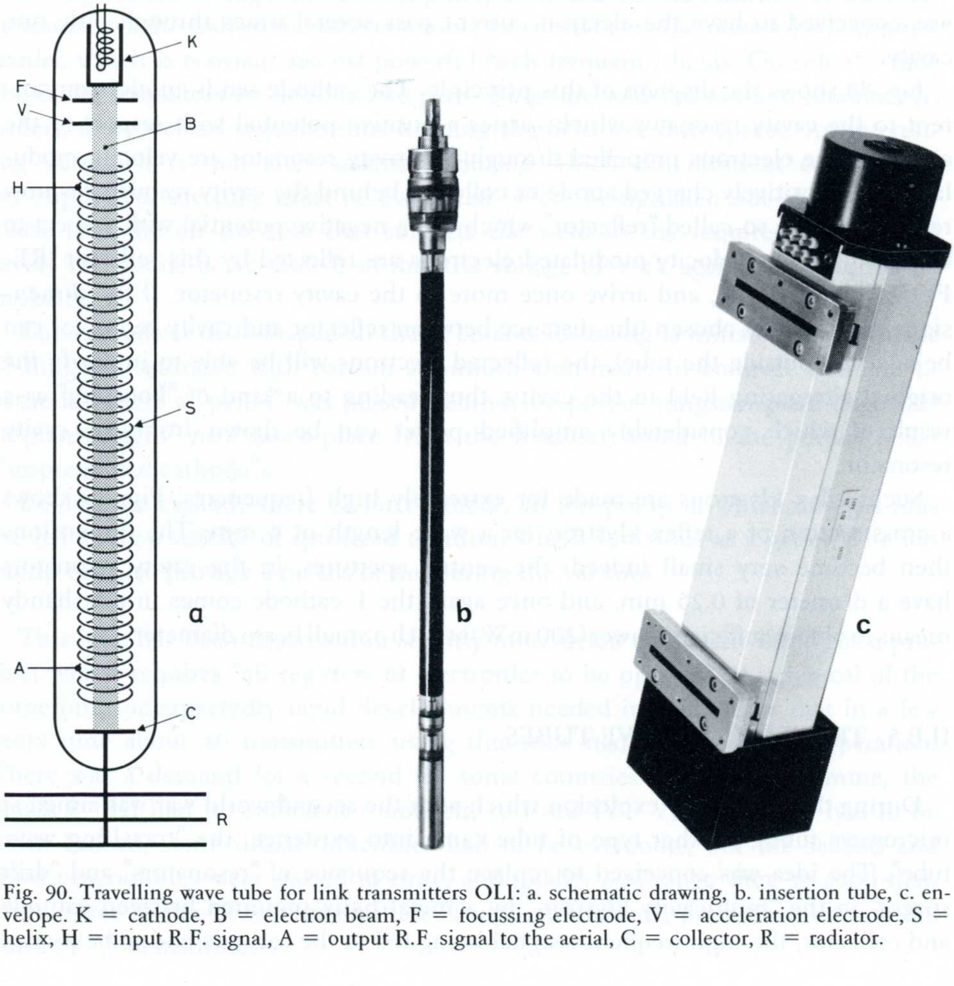


Fig. 90. Travelling wave tube for link transmitters OLI: a. schematic drawing, b. insertion tube, c. envelope. K = cathode, B = electron beam, F = focussing electrode, V = acceleration electrode, S = helix, H = input R.F. signal, A = output R.F. signal to the aerial, C = collector, R = radiator.

offers a gain of about 40 dB between 5800 and 8500 MHz. In Chapter I.23 it was already explained how one beam link can be used for multichannel F.M. telephony (thousands of channels), sometimes with the simultaneous transmission of television signals. This is an example where the capability of the YH1170 to amplify such a great bandwidth is of priceless value. This particular travelling wave tube has an output of 15 W at a voltage of 2800 V on the helix, 1300 V on the collector, and 2100 V on the accelerating electrode. Furthermore the noise figure is very low. However, efficiency is no more than 20%, which in itself forms no objection for this low power.

Figs. 90(b) and (c) show a photograph of the insert tube and the magnet system. As described with the klystron in Chapter II.E.3, this system is an assembly of permanent magnets; the magnetic field is periodic. The cathode is of the L-type. Mechanical centring requires due care, for the electron beam should nowhere come into contact with the helix. The heat generated in the helix must be dissipated by radiation; the collector dissipates its heat via a radiator. The tube insert is placed in the envelope consisting of the magnet system, waveguide connection, and electrical connections. Thanks to narrow mechanical and electrical tolerances it has been made possible to replace the tube without the need for subsequent tuning. Now this problem has been solved, the travelling wave tube will replace the triodes for 4000 MHz as one single tube offers a higher gain.

## II.E.6. THE MAGNETRON.

The magnetron is the one longest known among the microwave tubes, but it is difficult to understand and rather difficult to operate. Particularly in pulse operation this type of tube combines high output powers and high efficiency.

Basically the magnetron consists of a copper cylinder which functions as the anode, and a cathode situated concentrically in this copper cylinder (Fig. 91a). Without influence from the magnetic field the electrons moving from the cathode to the anode will describe a straight line. (1 in Fig. 91a). However, a magnetic field is applied in the direction of the centre line of the cathode, so that the electron beam is subjected to a force perpendicular to the velocity; it will then follow the course indicated as 2. If the magnetic field is further intensified, the electron beam is deflected to such an extent that it fails to reach the anode (3). In this situation

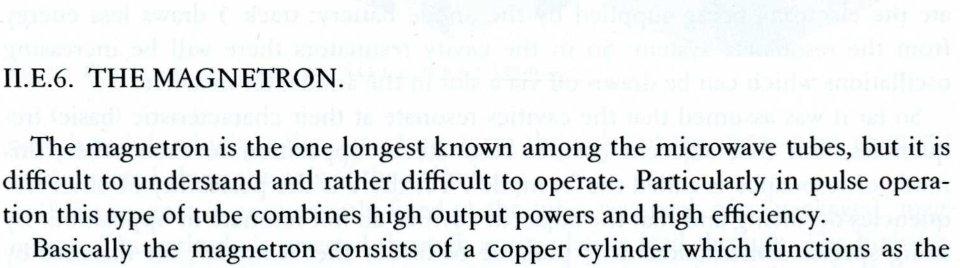


Fig. 91. Electron paths in a magnetron: a. without H.F., b. with H.F.

the magnetron is cut off. The magnetron is driven in the condition where it is just not cut off.

The anode cylinder is provided with axial slots (Fig. 91b), which form cavity resonators in which the oscillations of their resonance frequency can occur. The cavity resonators can be compared with a self inductance, and the slot with a capacitor with an alternating voltage at its terminals. If an electron with a track 4 passes the slot at a moment when the field between the teeth opposes the speed of the electron, it will be slowed down and transfer energy to the resonance circuit in which the oscillation is amplified. Under influence of the magnetic field the electron is deflected to the right to be attracted by the anode again. It will reach the next slot just when the field there opposes the velocity of the electron again, and once more the electron will transfer energy to the resonance circuit. When the electron passes the teeth while the field has the same direction as the electron speed, energy will be drawn from the resonance circuit, and the electron is accelerated, deflected to the right under influence of the magnetic field to land on the cathode at an enormous speed (5 in Fig. 91b), with the result that secondary electrons are emitted (it is indeed possible to drive an operating magnetron with little or no heater power). It is typical that the electrons following path 4 transfer more energy to the high frequency field than those of track 5 draw from it. In the case of track 4 energy transfer takes place several times, the power needed to accelerate the electrons being supplied by the anode battery; track 5 draws less energy from the resonance system. So in the cavity resonators there will be increasing oscillations which can be drawn off via a slot in the anode, for instance.

So far it was assumed that the cavities resonate at their characteristic (basic) frequencies, and that adjacent cavities resonate in opposition to each other; this mode of resonance is called the "π mode". Yet there is the possibility of other frequencies occurring and that the adjacent cavities do not resonate in opposition. By using special constructions it is possible to favour the "π mode", for instance by using connecting STRAPS or using alternate larger and smaller cavity resonators (RISING SUN).

Since the frequency is determined by the dimensions of the cavities, these dimensions must be extremely small for high frequencies. It is then no longer possible to make the cavities by conventional mechanical means (drilling, milling), but instead a precision ground die is pushed through a copper block ("hobbing").

One of the reasons why the magnetron with its relatively small volume can put out so much power is that the central cylindrical cathode can have such a large area (mostly oxide cathodes are used, sometimes the L-cathode or the Wo-Th cathodes). Furthermore the copper cylindrical anode can easily dissipate heat. Usually a radiator is used for cooling; for continuous power of 5 kW or more water cooling is required.

In radar the magnetron is operated for short periods ("pulses") by switching on the anode voltage. This method calls for rapid-control high voltage rectifiers, usually HYDROGEN-THYRATRONS, described in chapter II.H.2.

Fig. 92 shows the design of a magnetron. The copper cylinder with the cavities is equipped with cooling ribs. At the top and the bottom this cylinder is closed

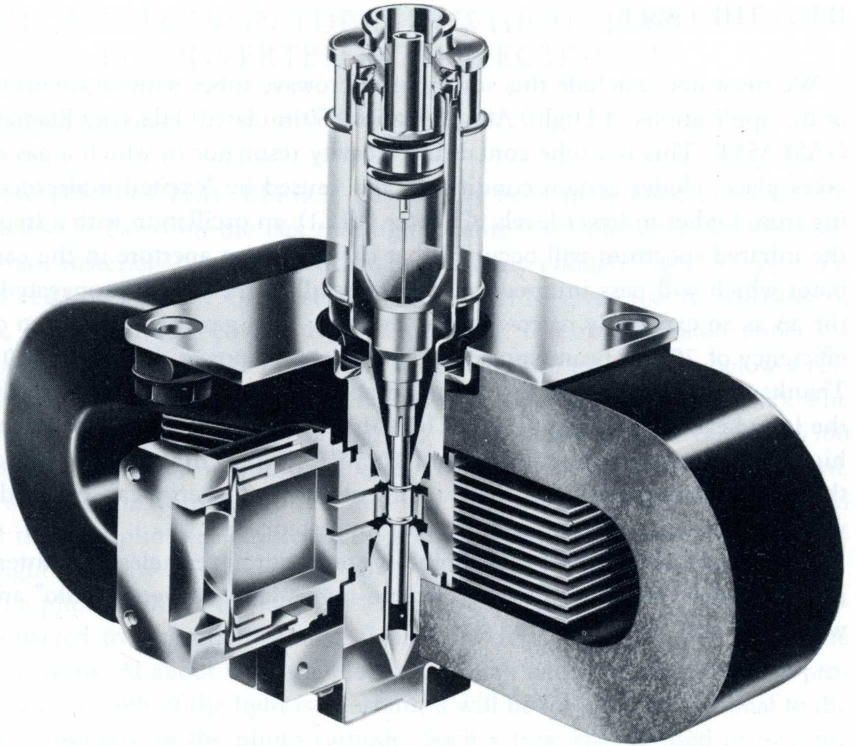


Fig. 92. Photograph of a magnetron 4J50, partly cut open.

with iron lids which at the same time form the poles of the permanent magnetic field directed along the axis of the cylinder.

If the magnet is permanently fixed to the tube, we speak of a "packaged" magnetron. The cathode is inserted through an insulating lead-through. The high-frequency energy is led away via a wave guide connected to a slot in the cylinder.

Magnetrons are used for wavelengths ranging from 30 cm to 6 mm, continuous powers up to 5 kW, and pulsating powers up to 5 MW.

Finally, an interesting development that took place in the Swedish Philips laboratories is the SPIN-TUNED-MAGNETRON. Here, a cylinder with equally-spaced holes (the spinner), driven as in the Rotalix X-ray tube (Chapter II.G.3), rotates in the cavity so that a sinusoidal frequency variation of about 450 MHz with respect to the basic frequency of 9000 MHz takes place 1000 times per second. The receiver must be tuned to the same frequency variation.

Several advantages are thus obtained:

a. Secrecy has improved (an incorrectly variably tuned receiver cannot receive the radar signal).
b. Interfering with the signal (JAMMING) has become much more difficult.
c. Cross talk from other transmitters, as well as atmospheric interference will have less effect on the returning signal.
d. It can be shown that the "signal-to-noise" ratio, particularly for moving objects, is almost doubled.

## II.E.7. THE LASER

We must not conclude this survey of microwave tubes without mentioning one of the applications of L(ight) A(mplification) S(timulated) E(lectric) R(adiation), the GASLASER. This is a tube containing a cavity resonator in which a gas discharge takes place. Under certain conditions, and caused by "excited molecules" returning from higher to lower levels (Chapter II.H.1), an oscillation with a frequency in the infrared spectrum will occur in that cavity. Via an aperture in the cavity resonator which will pass infrared rays, these oscillations will be propagated through the air as an extremely narrow beam. By using $CO_2$ gas it is possible to obtain an efficiency of 20%, a beam spread of 2 mrad, and a power density of 600 $W/cm^2$. Thanks to these values we have here an application for telecommunication along the laser beam. With a germanium lens the beam can be concentrated to an even higher degree, and a power density of $10^6$ $W/cm^2$ (ten times that of the surface of the sun) is obtainable. This heavy concentration of energy can be used to burn very small holes in quartz and tungsten, for instance.

This technique is still in the laboratory stage, but nevertheless it is interesting to note that electronics has now crossed no-man's land between "radio" and "light" waves ($10^5$ and $10^8$ MHz).

# SECTION F. ELECTRON TUBES IN WHICH LIGHT IS CONVERTED INTO ELECTRICITY

## II.F.1. PHOTOCELLS

The PHOTO-ELECTRIC EFFECT has already been mentioned in Chapter I. 15. This effect is based on the fact that light beams (electromagnetic oscillations) cause certain materials to emit electrons. Surprisingly enough, light energy is radiated in certain "quanta". Of these light quanta 1/10 gives rise to emission of electrons, the remainder being converted into heat. Electron emission under influence of light will take place only if the wave length of this light is below a certain "red level". as required for the work function (see also Chapter II.B.2). The work function of caesium is low (at 1.9 V), so the infrared level is high (6500 Angstrom); in this case a photo emission of 50 μA per lumen is obtained, whereas with S20-S25 photo emitters 200 and 400 μA/lm are reached. It should be noted that the photo-electric current is able to follow the very rapid fluctuations in light intensity (in the order of $10^{-12}$s).

If such a photo cathode is placed in a vacuum envelope together with an electrode connected to a positive voltage, this electrode (the anode), will attract the emitted electrons. At about 50 V the electron current becomes saturated, and provided the wave length of the light is constant, it will be directly proportional to the light flux in lumens on the photo cathode. Such a tube can be used to measure light intensities by measuring electric currents; this tube is known as the PHOTOCELL.

In the earliest photocells the photo emissive layer was applied to a metal plate. Later, to obtain the shortest possible distance between photo cathode and light source, the photo emissive layer was deposited on the glass wall. For that purpose the glass wall was first given a silver coat; this coat is oxidized in oxygen by means of a glow discharge afterwards to be exposed to the effect from caesium-antimony vapour. The result is a layer of caesium-antimony oxide and the free silver restored in the process provides a conductive layer from which the emitted electrons can be replenished. The available photo current (50 μA/lm) is, of course, very weak. It can be intensified by giving the tube a gas filling, so that on their way to the anode the electrons emitted from the photo cathode are accelerated and collide against gas molecules from which several electrons are then released (IMPACT IONIZATION); thus it is possible to obtain 200 μA/lm at voltages up to 200 V. However, the photo current is then no longer proportional to the light flux, and as a result of impact ionization there is a certain inertia. An important application of gas filled photocells is the automatic control of flames in oil combustion installations used in central heating systems.

## II.F.2. PHOTO MULTIPLIERS

A modern application of photocells is the measuring of radio active radiation by SCINTILLATION (certain materials generate a short weak light pulse per quan-

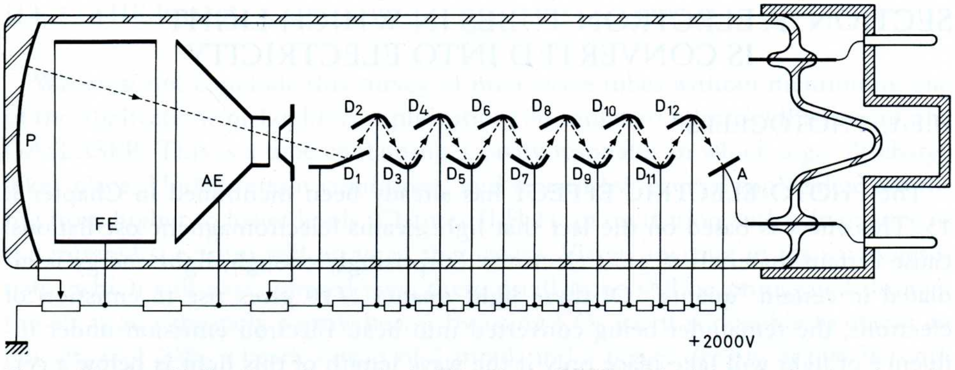


Fig. 93. Photo multiplier. P = photo cathode, FE = focussing electrode, AE = accelerating electrode, $D_1$ = first dynode, $D_2$-$D_{12}$ = amplifying dynodes, A = anode.

tum of incident radiation). The photocell, however, is not sensitive enough to measure these weak light pulses. It has been found possible to amplify the photo electric current by means of secondary emission in a cascade arrangement of suitable electrodes (PHOTO MULTIPLIER).

The principle of such a tube is shown in Fig. 93. Photo cathode P is situated on the face of the tube. By means of an electrostatic lens, formed by the concave surface of the photo cathode, the focussing electrode and the accelerating electrode, the electrons are concentrated on the first dynode. The latter is coated with a material that under influence of the electron bombardment emits a multiple of secondary electrons which are attracted by the second dynode (with a positive voltage with respect to the first), where again a multiple of secondary electrons is released to be attracted by the third dynode. Thus each dynode stage multiplies the number of electrons until finally a signal current can be picked up from the anode millions of times as strong as the original photoelectric current. Most sensitive photo multipliers have a sensitivity of 1000 A/lm, with the possibility of reproducing light fluctuations of no more than a few nano seconds.

## II.F.3. IMAGE TRANSFORMERS AND IMAGE INTENSIFIERS

In the photo multipliers discussed in the foregoing, the average light intensity on the photo cathode was determined by the extent to which electrons, emitted over the whole area, were electrostatically focussed on the first dynode. However, the electrostatic lens can be so designed that the electrons emitted from a particular spot of the photo cathode are directed to a corresponding spot of a screen on the other side of the tube (Fig. 94). If a black-white image is projected on the flat photo cathode P, the second screen, F, will display the same image in which the light intensity has been translated into electron current. By simply applying a fluorescent layer with aluminium coat (as in a cathode ray tube) to the second screen, and observing the screen from the right, a faithful reproduction of the image on the photo cathode can be seen; the image of the photo cathode has been "converted" to the fluorescent screen.

Fig. 95 gives one of the earliest types of IMAGE CONVERTER. Here the elec-

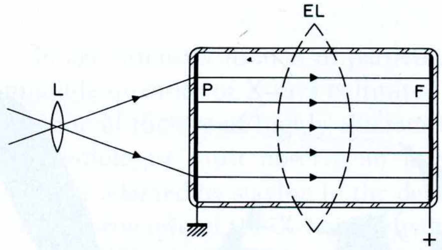


Fig. 94. Principle of the image transformer. P = photo cathode, F = fluorescent screen, L = lens, EL = electrostatic lens.

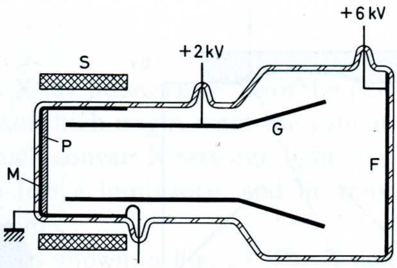


Fig. 95. Construction of the image converter (ME1201). M = metal layer, G = grid electrode, S = focussing coil (other letters as in Fig. 94).

tron-optical lens is formed by coil S, grid electrode G, and the fluorescent screen F. At the proper configuration, and the correct current and voltage settings, fluorescent screen F will display an accurate reproduction of the image on photo cathode P. This tube is used for high-speed photography. With electronic equipment very short voltage pulses can be applied to the grid and anode of the tube, at the same time igniting a flash light. The voltage pulses used for this purpose can be initiated by the object to be photographed (e.g. a bullet hitting an armour plate), and so the (after glowing) screen can be used to make instantaneous photographs of a rapidly developing phenomenon.

Even more important is the application in which invisible light, to which the photo cathode is sensitive, is converted into visible light. This system is used, for instance, in INFRA RED TELESCOPES as used by the armed forces. Unknowingly the enemy is caught in an infra red spot light. The infra red image (to which the photo cathode is sensitive) is converted by the image converter into a visible image on the fluorescent screen.

To obtain more light amplification, several image converters can be connected in series, an objective lens being used to project the image on the fluorescent screen of the first tube on the photo cathode of the second tube, and so on. Such an objective can have a maximum angular aperture of f : 1, at which it will use no more than 60% of the light radiated from the fluorescent screen. An interesting optical system can be made with FIBRE OPTICS, a disc consisting of glass rods with a diameter of 5 microns, each consisting of concentrical layers of glass with different refractive indices, surrounded by a layer of black glass to prevent diffraction. Each light ray entering through the end plane of such a fibre is reflected by the boundary layers between two different kinds of glass to emerge at the other end almost perpendicular to the plane of the disc. Here we have an optical system with an aperture angle of 180°; the glass fibres do, however, involve some light loss. Fig. 96 gives the principle of such a cascade tube consisting of a number of image converters the partition between fluorescent screen and photo cathode of the next converter consisting of fibre optics. A further advantage of fibre optics is that the form can be made to match the concave form of the screens, which is favourable for the electron optics.

So far we have paid no attention to the light "amplification" in such image converters. In Chapter II.F.2 we saw that only one in ten light quanta releases elec-

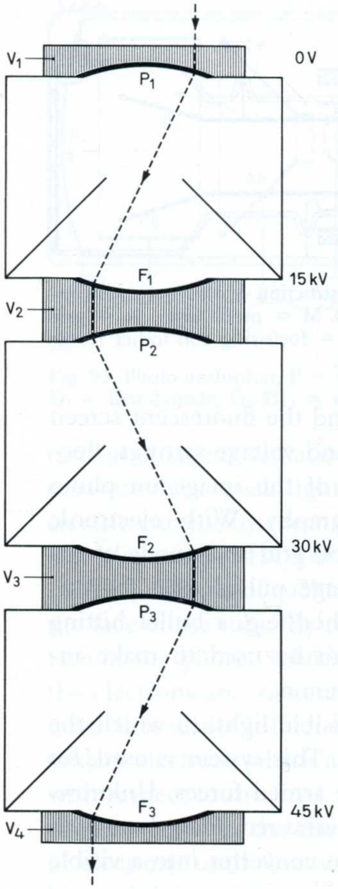


Fig. 96. Cascade circuit of three image intensifiers by means of fibre optics. P1, P2, P3 = photo cathodes, F1, F2, F3 = fluorescent screens, V1-V4 = fibre optics.

Fig. 97. Telescope with image intensifier (10 000), with which the scenery can be observed in the light from the stars.

trons from the photo cathode. Under influence of the anode voltage on the photo electrode, the electrons are however accelerated, and when discussing picture tubes we already saw that the light production of the fluorescent screen increases considerably if the voltage is raised to about 25 000 V. Under these conditions one electron can give rise to 1000 light quanta so that the total light amplification is 100. Light amplification can also be obtained since the same rule holds for both electron optics and light optics that the light intensity is proportional to the reduction squared. If the fluorescent screen is made 9 times as small as the photo cathode, light amplification will be 81 times and combined with the earlier mentioned effect of electron acceleration the total light amplification is in theory 100 × 81 = 8100 times (IMAGE INTENSIFIER).

With the above means and a large optical mirror system the Mullard Research Laboratory have succeeded in building a telescope (Fig. 97), which gives a light amplification of 10 000 times, and through which the scenery can be observed by the light of the stars.

## II.F.4. X-RAY IMAGE INTENSIFIERS

Image intensification is of particular value in X-ray fluoroscopy. Here the permissible quantity of X-rays is limited by the dose which might harm the patient. In spite of the use of highly efficient screens which convert X-rays into light rays, the radiologist must observe an image of too low a luminance, and he must become adapted by staying in the dark for some time.

The principle of the X-ray intensifier was already shown in Fig. 23. The X-rays pass the object under investigation (O) and the glass wall of the tube to land on a screen that consists of a thin aluminium carrier D, a layer R that converts X-rays into light rays, and a photo cathode K. (Unfortunately no photo cathode has been found yet which can directly convert X-rays into electrons with a sufficient output). The pierced anode A forms part of the electron lens, and the fluorescent screen F receives an "electron pattern" which is a precise replica of the X-ray image reduced by a factor of about 9. This electron pattern in turn is converted by the fluorescent screen F into a light image which can be observed through a microscope M. This method yields a light amplification of 1000 times with respect to the image on the fluoroscopy screen, which is partly sacrificed in order to protect the patient against to intensive radiation. A fortunate circumstance is that the fluorescent screen can be given a greater definition than the screen that converts X-rays into light, so the image does not lose too much of its sharpness in the reduction process.

Fig. 98 shows a cross section of a modern version of a 9″ X-ray image intensifier. The three parts: glass screen with photo cathode bearer, cylindrical envelope with electron optical system, and fluorescent screen are manufactured and inspected individually. They are provided with metal rings that are welded air tight together according to the "argon-arc" process. A special feature of this image intensifier is its variable magnification, the electronic ZOOM LENS. Magnification is varied by controlling the voltage on a pre-anode, enabling the radiologist to have a part of the image occupy the whole fluorescent screen so that he can observe relatively enlarged images.

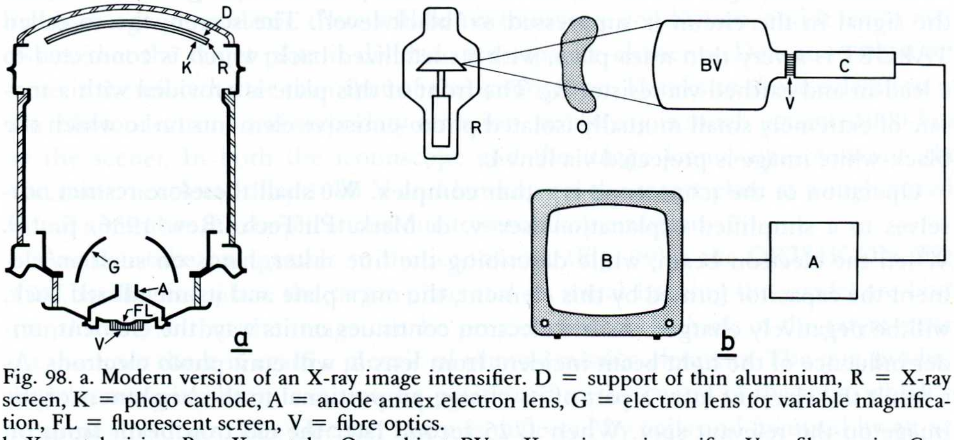


Fig. 98. a. Modern version of an X-ray image intensifier. D = support of thin aluminium, R = X-ray screen, K = photo cathode, A = anode annex electron lens, G = electron lens for variable magnification, FL = fluorescent screen, V = fibre optics.
b. X-ray television. R = X-ray tube, O = object, BV = X-ray image intensifier, V = fibre optics, C = camera tube, A = amplifier, B = picture tube.

Thanks to modern developments of television cameras this image can be transferred to a television picture tube (Fig. 98b). The optical device ideal for the purpose, because of its wide aperture angle and the little space it takes up, is the fibre optic. The fluorescent screen of the X-ray image intensifier is applied to a fibre optics plate. Similarly the photo conductive layer of the camera tube (vidicon or plumbicon) is applied to a fibre optics plate. The signal from the camera tube can be electrically amplified and reproduced on the picture tube at any required luminance level and contrast. This is no less than a revolution in radiology; without any danger from radiation the unadapted physician can now observe the X-ray image anywhere and in broad day light.

## II.F.5. CAMERA TUBES WITH A PHOTO CATHODE

In Chapter I.11 (Fig. 16b) it was already explained that in the earliest camera tube, the ICONOSCOPE, an electron beam "describes a zig-zag across a screen (on which the image to be transferred is projected) which is coated with a material that emits more or less current under influence of accordingly more or less incident light". The screen might be regarded as consisting of a photo cathode as used in the abovementioned image intensifiers. Matters are not that simple; for electrons are picked up only when the electron beam passes over a certain spot on the screen. During the time that the electron beam is completing its zig-zag pattern, this particular spot on the screen "does not know what to do with the photo-electrons". Hence an accumulating element must be added which accumulates the electrons released from the photo cathode by the light image, until the electron beam picks them up.

A constructional diagram of the iconoscope is shown in Fig. 99. From gun K a very thin electron beam emerges which is deflected by deflection coils D and describes complete zig-zag patterns across the screen at a rate of 50 times per second. The "scribing" speed of the electron beam is, however, so adjusted that the speed from left to right is low (this is the movement in which the image is formed); the movement from right to left ("fly back") is very rapid, and during this movement the signal in the circuit is suppressed to "black level". The screen, the so-called TARGET is a very thin mica plate, with a metallized back, which is connected to a lead-in and earthed via resistor $R_S$. The front of this plate is provided with a mosaic of extremely small mutually isolated photo-emissive elements on to which the black-white image is projected via lens L.

Operation of the iconoscope is rather complex. We shall therefore restrict ourselves to a simplified explanation (see v. d. Mark, Ph.Techn.Rev. 1936, p. 18). When the electron beam, while describing the line raster, lands on such an element the capacitor formed by this element, the mica plate and its metallized back, will be negatively charged. As the electron continues on its way, the element, under influence of the light beam incident from lens L, will emit photo electrons. As a result the element loses a part of its charge proportional to the brightness of the image on the relevant spot. When 1/25 second later the electron beam lands on the same element, the original charge will be restored, and from the voltage source

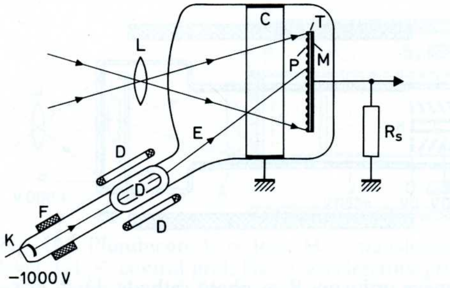


Fig. 99. Iconoscope. L = lens projecting the scene on the mosaic P of photo-emitting elements, T = target plate, M = metal layer earthed via signal resistor $R_s$, C = collector, E = electron scanning beam, D = deflection coil, F = focussing coil, K = cathode.

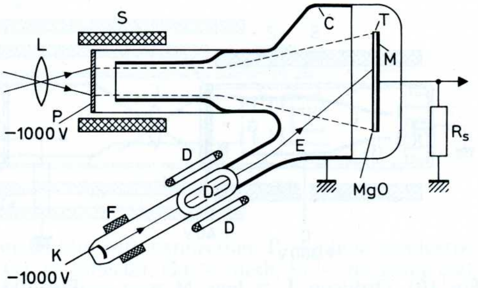


Fig. 100. Image iconoscope. P = photo cathode, S = coil of the magnetic electron optics (all other letters as in Fig. 99).

(and so through resistor $R_S$) there flows a current pulse proportional to the brightness of the image on the relevant spot. Across $R_S$ there will then arise a corresponding signal voltage which is fed to the modulator of the transmitter.

The iconoscope has formed the first solution in the realization of "electronic" television owing to the fact that photo emission is such a high-speed phenomenon. Sensitivity is, however, limited so that a great deal of light is required at the scene.

In the so-called IMAGE ICONOSCOPE (Fig. 100) the iconoscope is combined with the previously described image intensifier. The black-white image is projected on photo cathode P which emits an electron "pattern" in accordance with this image. By means of the electron optical system, formed by magnet coil S and the suitably formed collector C, the electron pattern is projected on target plate T. The latter is a mica plate covered with magnesium oxide which has a great secondary emission, so that each particle emits secondary electrons proportional to the brightness of the image at that particular spot. When passing the relevant particle, the electron beam describing the line raster has to replenish the charge lost during secondary emission. When this happens, a signal current occurs (as in the iconoscope) which is proportional to the light intensity at this point. Since magnesium oxide is a good insulator, there will be no transverse conduction, and it is not necessary that the target plate is divided into separate elements. Definition obtained here is therefore better than that of the iconoscope. Thanks to the acceleration of the photo electrons and secondary emission, sensitivity is much greater (1000 Lux at the scene). In both the iconoscope and the image iconoscope, however, the electrons, accelerated by 1000 V and landing on the target plate at a high speed, entail a number of complications due to secondary emission.

Another tube comparable to the iconoscope (Fig. 99) is the ORTHICON (Fig. 101). In the first place the target plate and its metal backing (tin oxide) are both translucent, so that the image can be projected from the back of the target plate. At the front there is again a mosaic of photo-emissive elements. The gun producing the scanning beam can now be positioned in the centre of the tube, which is of great advantage. The electron beam is deflected by coils and focussed by an accelerating grid $G_2$ and a collector C which at its end is provided with a mesh G to

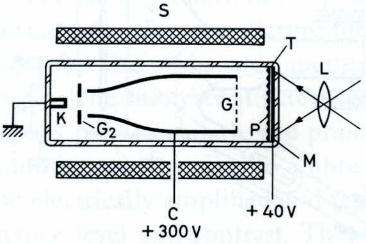


Fig. 101. Orthicon. L = lens, M = translucent metal film, T = target plate, P = mosaic of photo-emitting elements, C = collector, G = mesh, K = cathode, S = deflecting and concentrating coil.

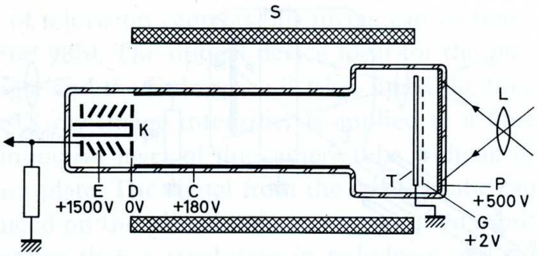


Fig. 102. Image orthicon. P = photo cathode, D = dynode, E = electron multiplier (other letters as in Fig. 101).

which a voltage of 300 V with respect to the cathode is applied. This high voltage is required to obtain a narrow beam. The target plate, however, carries a voltage of no more than 40 V; the electron beam lands on the target plate at a lower speed, and although charge lost due to photo emission is restored, there is much less secondary emission.

During the second world war the R.C.A. developed a camera tube which for military purposes had to operate at low light intensities, the so-called IMAGE ORTHICON (Fig. 102). It would lead too far to describe this complex tube in detail. In brief it functions as follows: the photo cathode P is fitted to the inside glass wall. Through a very fine mesh the electron pattern, focussed by a coil, is drawn to the target plate which consists of a glass membrane. As in the orthicon the gun in the centre of the tube sends the electron beam, which describes the line raster, to this target plate. Only the electrons required to compensate the photo current arrive on the target plate. The other electrons return to the cathode and land on a disc round it, the dynode, which radiates secondary emission to be amplified in a "multiplier". The result is a camera tube with good definition and low inertia, and highly light sensitive (100 Lux at the scene). This tube has, therefore, become the standard camera tube for black-white television, but has, owing to its complexity in adjustment and use, been somewhat difficult to users.

## II.F.6. CAMERA TUBES IN WHICH PHOTO CONDUCTIVITY IS USED

The demand for simpler and cheaper tubes not having to satisfy such stringent sensitivity requirements (CLOSED CIRCUIT applications) led to development of the VIDICON. As regards mechanical design it resembles the orthicon (Fig. 101), but now the target plate covered with photo emissive elements was replaced by a thin film of material (antimony sulphide) which becomes conductive under influence from light. As more or less light is incident on a spot of this photo conductive film, the electrons leaving the cathode will, to a greater or lesser extent, pass on to the tin oxide layer behind the photo conductive film. This photo conductivity has a certain inertia owing to which it takes the electrons, transferred to a cer-

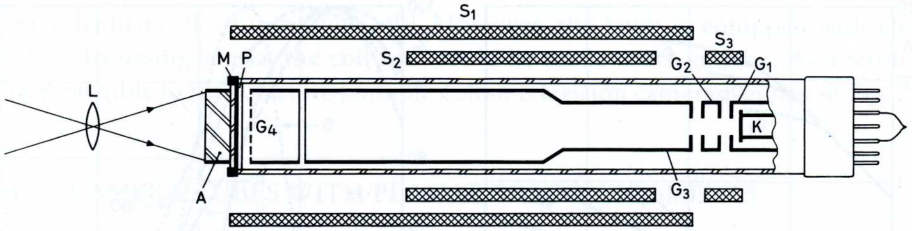


Fig. 103. Plumbicon. L = lens, M = translucent metal film with connection, P = photo conductive layer, G1 = control grid, G2 = accelerating grid, G3 = collector, G4 = mesh, S1 = focussing coil, S2 = deflecting coil, S3 = correction coil, A = anti diffraction lens.

tain spot on the target plate by the moving electron beam, some time to flow away. This phenomenon gives the "accumulating effect" as already mentioned in Chapter II.F.5.

With this tube the useful signal is determined by the difference in layer resistance at full light and no light; this resistance variation is relatively small, or in other words: the "dark" current is rather high as compared with the "light" current.

Another drawback of this type of tube is that we are not dealing with a normal ohmic resistance, but a rather complex phenomenon in a semiconductor, which gives the resistance variation a certain inertia with respect to a rapid light variation. Suppose the reactive speed of the human eye is 60 milliseconds, then the percentage original "light" minus "dark" current, remaining 60 milliseconds after a sudden transition from light to dark, is a measure of this inertia. In a vidicon this inertia percentage is 25% when operating with the normal signal current. This implies that rapidly moving white objects are trailed by a white line, and therefore this type of tube is unsuitable for television.

Matters are considerably improved by using a similar tube in which the photo conductive layer is made of a modified lead oxide: the PLUMBICON (Fig. 103). Here the residual signal current after 60 milliseconds is 4%, so that the tube is very suitable for television. The lead oxide forms an excellent diode so that the dark current is extremely low. Due to the crystalline form consisting of rods perpendicular to the image plane, the resistance of this layer when exposed to light is low at a greater thickness than that in the vidicon. The extremely small diameter of the rods (0.1 $\mu$) ensures such a good definition that even at an image diameter of no more than 20 mm 600 lines can be sharply displayed. In the orthicon definition is limited by the mesh, and screen diameters of at least 40 mm are required. This means that at equal definition the plumbicon is thinner and also shorter, so that the optical system is cheaper. Furthermore it is known in camera engineering that the "depth of focus" of the optical system at a given light intensity increases as the image plane becomes smaller.

Fig. 104 gives the signal current as a function of the luminous flux plotted logarithmically for several types of tube. For various reasons the image orthicon must be adjusted in the knee of this curve; the plumbicon can be used in the straight part of the curve. The signal current in the image orthicon is much larger, but so

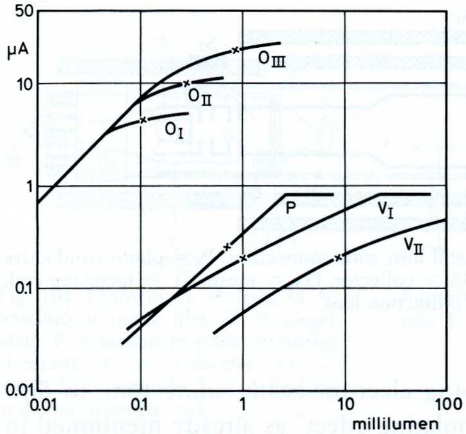


Fig. 104. Signal-light characteristics for various camera tubes, $O_{I}$, $O_{II}$, $O_{III}$ different types of image orthicon. P = plumbicon, $V_{I}$ = vidicon, $V_{II}$ = vidicon at a lower signal current and hence greater velocity (Philips Techn. Review).

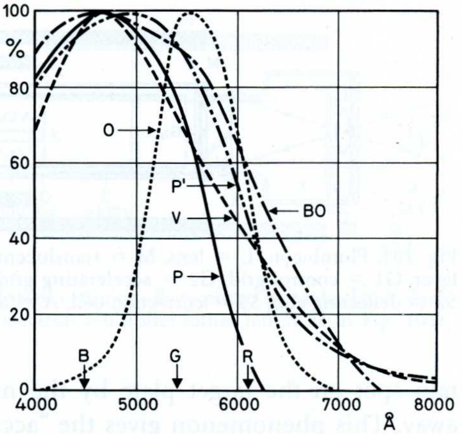


Fig. 105. Colour rendition of different types of camera tubes. BO = image orthicon, V = vidicon $Sb_2S_3$, P = plumbicon, P′ = red-sensitive plumbicon, O = sensitivity of the human eye (Philips Technical Review).

is noise, consequently the plumbicon is quite suitable to operate with a lower signal current (300 lux at the scene).

With the plumbicon the light level required at the working point is higher by a factor of 5 than that of the image orthicon, slightly lower than for the vidicon, but many times lower when the vidicon is to be adjusted to low inertia.

Fig. 105 illustrates the colour reproduction of the different tubes as compared with the sensitivity of the eye. It is seen that the plumbicon comes nearest this sensitivity. Later it was found possible to construct a plumbicon (P′) with a PbO-PbS layer particularly sensitive to red, which is even better in this respect. Fig. 105 also shows the wavelengths of the three colours of which the image is composed: B(lue), G(reen) and R(ed), and it is found that for the reproduction of "red" the red-sensitive plumbicon is a great improvement. This is also of special importance since the efficiency of the red phosphor of the reproduction tube is lower than that of the green and blue ones, so that the red colour reproduction is decisive for the light intensity throughout the circuit.

A major advantage of the plumbicon in colour television is, however, the linear relationship between signal strength and light intensity as illustrated by the graph in Fig. 104. To ensure faithful colour reproduction the gain in the three-colour channel throughout the circuit from camera tube to reproduction tube must be so adjusted that the proper colour mixture (particularly "white") is obtained. This adjustment having been made at one particular light intensity must also hold for the other light intensities, which requires a straight signal current versus light characteristic plotted along a logarithmic scale at an angle of 45°. The plumbicon, combined with the compact colour splitting prism, has therefore contributed substantially to a successful realization of colour television.

As an additional proof of the particularly good definition of the lead oxide layer it might be noted that it has been found possible to build a plumbicon with sufficient definition at 625 lines with a screen diameter of 16 mm, and even with a screen diameter of 10 mm (Fig. 24b). Moreover, the latter is equipped with electrostatic focussing so that the coil unit could be made much smaller. As a result it is now possible to manufacture portable colour television cameras, shown in Fig. 22.

## II.F.7. CAMERA TUBES WITH PLANAR SILICON ELEMENTS

Here too, a new development is presented in which tubes are going to be replaced by semiconductors.

As will be explained in Chapter II.J.3 a silicon semiconductor diode can be made whose leakage current varies linearly with the incident luminous flux. The planar technique offers the possibility of applying a raster of numerous small diodes of that type on a silicon crystal. Such a crystal can then be used as the target plate in a camera tube which mechanically resembles a plumbicon. Such type of tube would in the first place be used in applications where definition need not meet severe requirements; e.g. television via telephone networks (PICTURE PHONE), but there is also the possibility that it may be used to control larger numbers of lines. For the time being there is the difficulty of making a target plate without one single faulty diode.

## SECTION G. X-RAY TUBES

Although X-ray equipment, usually operated from a 50 Hz alternating current source, does not come under "electronics", the X-ray tube itself is certainly an "electron tube". The X-ray tube is older than the radio tube (its invention by RÖNTGEN dates from 1895), but its development into an industrial and practical product took place within the 50 years chosen as the title of this book. During that period there has undoubtedly been technological interaction between X-ray tubes and (the other) electron tubes. It is for that reason that X-ray tubes too, will be (very briefly) discussed in the following chapters.

### II.G.1. GENERATION, CHARACTER AND APPLICATION OF X-RAYS

If electrons travelling at a high speed collide against a heavy metal, electromagnetic oscillations are radiated at a wavelength below 1 Angstrom ($10^{-8}$ cm), that is to say at a frequency many times that of light. In principle the mechanism of this phenomenon is the same as that of fluorescence discussed in Chapter II.C.6. In the atoms which are hit by high-speed electrons, electrons are driven into another orbit whilst radiating the aforementioned oscillations. In general the frequencies of these oscillations form a continuous spectrum, the maximum frequency of which is about proportional to the voltage that accelerates the electrons. The intensity of the X-rays is proportional to the square of this voltage and to the atomic number of the material of the radiation source.

It is characteristic of X-RAYS that they penetrate through all material. In doing so, a large part is lost due to absorption which increases with the thickness, the atomic number of the material, and with the wavelength. If X-rays are led through an object consisting of different materials or one material with different thicknesses, a silhouette will become visible behind such object. This silhouette will show the more contrast as the difference in absorption between the various parts is greater. It can be seen from Fig. 106 that a difference in thickness, d′, for iron gives a considerable difference in contrast at 100 kV, whereas for aluminium the difference in contrast is better at 50 kV. To display iron parts in a piece of aluminium, the greatest difference in contrast is obtained at 50 kV. By selecting the proper voltage it is possible to find the maximum contrast, but it should be borne in mind that at lower voltages the intensity per mA electron current decreases rapidly. In practice the voltages used for the human body vary from 50 kV (extremities, chest) to 150 kV (stomach, and thick bone); for material inspection these voltages range from 30 kV (metals in plastic) to 300 kV (iron castings).

X-rays can be made visible since they can cause a certain fluorescent screen to light up or blacken a photographic plate. This offers the possibility of observing the shadow image or making a photograph of it.

As X-rays show neither refraction nor reflection so that they cannot be controlled by "optics", the only way of obtaining a shadow image is by means of what is known as "central projection" of the smallest possible area that radiates X-rays

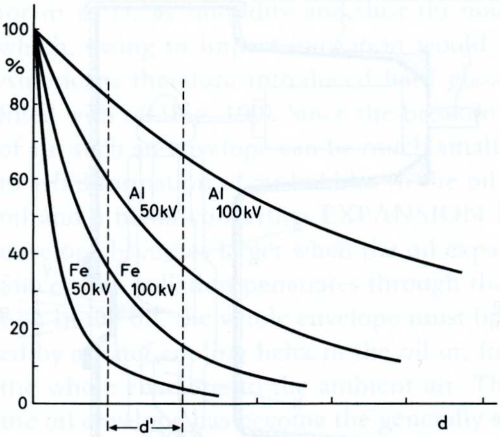


Fig. 106. Percentage of X-rays passed by iron (Fe) and aluminium (Al) during radiation generated at 100 kV and 50 kV. d = thickness of the object.

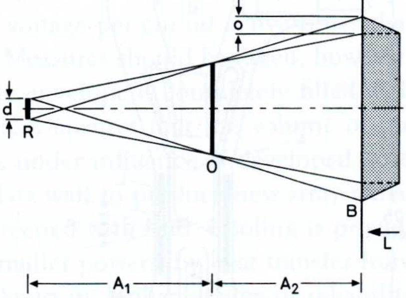


Fig. 107. Formation of the X-ray shadow image. R = X-ray source with diameter d, O = object, $A_1$ and $A_2$ = distances between tube — object, and screen — object respectively, B = image, L = luminous intensity. The unsharpness $o = dA_2/A_1$

(Fig. 107). According to this figure the shadow image will, due to the finite dimension of the X-ray source, show a black-white transition at its periphery, known as "unsharpness". To keep this unsharpness at a minimum, the area of the X-ray source should be as small as possible, the distance between tube and object as large as possible, and the distance between object and viewing screen again as small as possible. There is, however, no free choice of these respective distances.

Finally it should be borne in mind that the photographing of moving objects requires short exposure times, and thus a high intensity which in turn requires a large-area radiation source (Chapter II.G.3).

## II.G.2. THE CONSTRUCTION OF X-RAY TUBES, HIGH VOLTAGE AND RADIATION PROTECTION, HEAT REMOVAL

From what has been discussed so far it will be clear that an X-ray tube consists of a cathode with heater, which emits electrons, and an anode at a high positive voltage by which the electrons emitted from the cathode are attracted. The collision of these electrons against the anode gives rise to X-rays. Apart from that, the greater part of the energy in the anode is dissipated as heat.

Originally the cathode and the anode were sealed in a glass tube which was operated in air. Those who can imagine insulators as used for high voltage cables will be astounded at the idea that a tube about 50 cm long must be able to withstand a voltage of about 80 kV. It is not a vain instruction that was given with this tubes supplied by Philips in 1920: "Before the tube is allowed to pass current, it should be carefully dusted and dried". Furthermore the presence of these high voltages held a great personal danger for both patient and physician. It was not until after the first radiologists had fallen victim that it was found that the stray X-

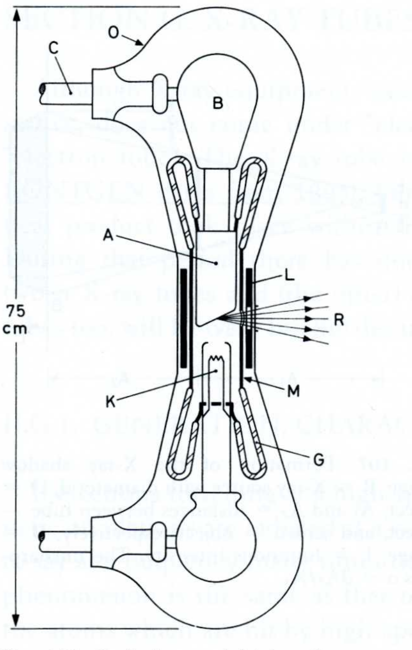


Fig. 108. Radiation and high voltage protected tube of the METALIX type. K = cathode, L = lead jacket, G = glass sleeves, O = earthed envelope, B = metal sphere accumulating and radiating heat, C = high voltage cable in earthed sheath leading to the high voltage source.

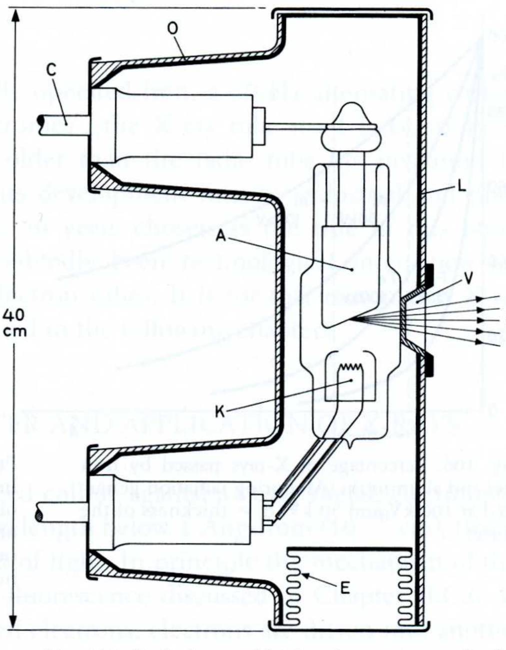


Fig. 109. Radiation and high voltage protected tube with oil insulation. E = metal expansion piece, V = plastic window preventing X-rays from going through a layer of oil (other letters as in Fig. 108).

rays radiated from the tube in all directions could also cause "X-ray burns" which involved peril of life. To turn the X-ray tube into an instrument suitable for medical diagnostics as well as therapy, three requirements had to be met: high voltage security, protection against high voltage, and protection against radiation.

A particularly neat solution to this problem was found in 1925, by BOUWERS: the METALIX TUBE (Fig. 108). The centre of the tube, which is connected to earth, consists of a chrome-iron cylinder with a glass window through which the rays can escape to the outside. The cylinder is surrounded by a lead jacket which stops undesirable radiation. The anode and cathode, with a symmetrical voltage to earth, are sealed to glass sleeves which form a long "creepage path" for the high voltage. The lead-ins of anode and cathode are also surrounded by metal caps which are fitted with inward insulators through which, via rubber insulated cables, the high positive and negative voltages from the X-ray generator are supplied. The result is a fully earthed contact-safe ENVELOPE round the X-ray tube. In this system the heat dissipated in the anode was removed by radiation and conduction of a sphere which was mounted on the anode lead-in. For short exposure times this sphere forms a heat reservoir. For greater powers a fan was used to blow air through the anode. Now the X-ray tube had become a medical instrument that presented no personal danger from high voltage contact and radiation for physician and patient.

A drawback of this ingenious construction was that air is an unreliable conductor in so far as humidity and dust do not get a chance to introduce discharges which, owing to impact ionization would lead to breakdown. In about 1940 the Americans therefore introduced hard glass tubes that are placed in an envelope filled with oil (Fig. 109). Since the breakdown voltage per cm oil is five times that of air, such an envelope can be much smaller. Measures should be taken, however, to avoid formation of air bubbles in the oil: the envelope is completely filled with oil and a metal concertina EXPANSION PIECE ensures that the volume of the envelope becomes larger when the oil expands under influence of developed heat. Since stray radiation penetrates through the glass wall to produce new stray radiation in the oil, the whole envelope must be screened with lead. Cooling is provided by a water cooling helix in the oil or, for smaller powers, by heat transfer from the whole envelope to the ambient air. Thanks to its higher degree of reliability the oil envelope has become the generally applied version.

For smaller equipment the tube together with the transformer could be mounted in one oil container, and so we have an X-RAY UNIT without cables.

## II.G.3. THE FOCUS

As already shown in Fig. 107, the X-ray source, the FOCUS, must be as small as possible to avoid unsharpness round the edges. Therefore the electron beam must be "focussed" on a small area of the anode. This is done by placing the heater in a cavity in the cathode which together with the anode forms the "electron optical system" that focusses the electrons on a small area of the anode. The heater should not be positioned too deep in its cavity to prevent this cavity from slightly functioning as a grid that would cause the anode current to decrease too much at a low voltage.

So the electron beam lands on the anode on the smallest possible FOCUS,

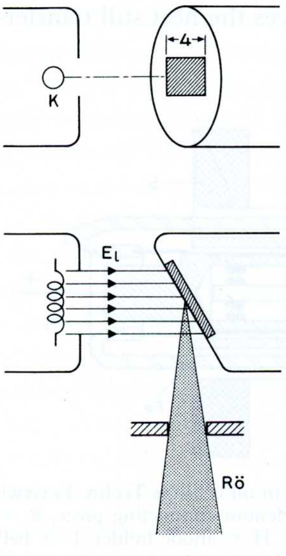


Fig. 110. Principle of the line focus.

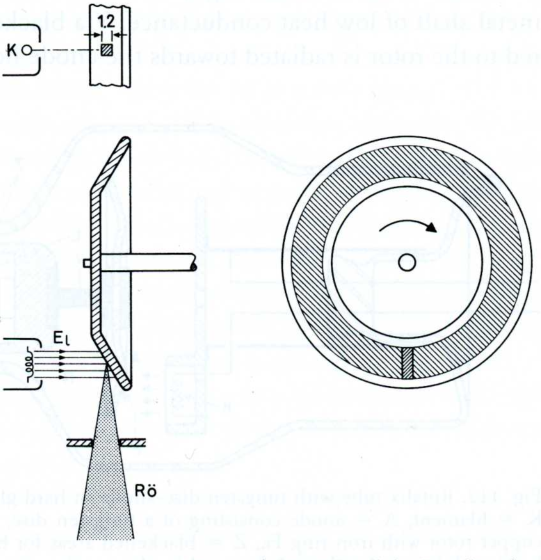


Fig. 111. Principle of the Rotalix tube.

where a small part of the energy is converted into X-rays, the greater part being dissipated as heat. The material most suitable for this "focus" is tungsten; it has a high atomic number (high X-ray efficiency) and a high melting point. Heat conduction of tungsten is, however, low. To compensate for that, a rather thin tungsten pellet is cast in a copper anode. Heat removal can be further improved by casting the copper round the tungsten pellet in a special way so as to have it solidify as one crystal. Furthermore the area on which the electrons land can be made larger than the area that is decisive for the unsharpness of the X-ray image by having the anode plane form an angle of 80° with the centre line of the electron beam, and by using only a bundle of the X-rays perpendicular to this centre line (Fig. 110, LINE FOCUS). By using this method it is possible to photograph a stomach (e.g. at 100 kV, 300 mA, 0.1 sec) by means of a line focus of 12 × 4 mm which for the X-ray beam means an area of 4.0 × 4.0 mm. The unsharpness of the image is, however, still considerable.

A most interesting solution was found in the form of the ROTALIX TUBE. Here the anode is a disc rotating at a high speed. According to Fig. 111 the focus obtained to form the X-ray image is of the same dimensions as the line focus, but as regards cooling, the area of the focus is equal to the hatched ring. The anode is mounted on one shaft together with an assembly of copper and iron which functions as the ROTOR of a synchronous motor; the STATOR outside the tube drives the rotor and the ROTATING ANODE. It has been a particularly difficult task to find bearings which would function well in this high voltage tube (in which any contamination in the vacuum leads to breakdown). The ultimate solution was using ball bearings, "lubricated" with very fine powdered lead or silver.

Fig. 112 shows a cross section of a Rotalix tube as generally used nowadays. The tube is made of hard glass in an oil envelope. The anode is a somewhat conically shaped tungsten disc that transfers its heat by radiation to the glass wall of the tube, and from there to the oil. To ensure adequate heat removal, the temperature of the disc must be rather high, so the disc is connected to the rotor by means of a metal shaft of low heat conductance. Via blackened surfaces the heat still transferred to the rotor is radiated towards the anode holder.

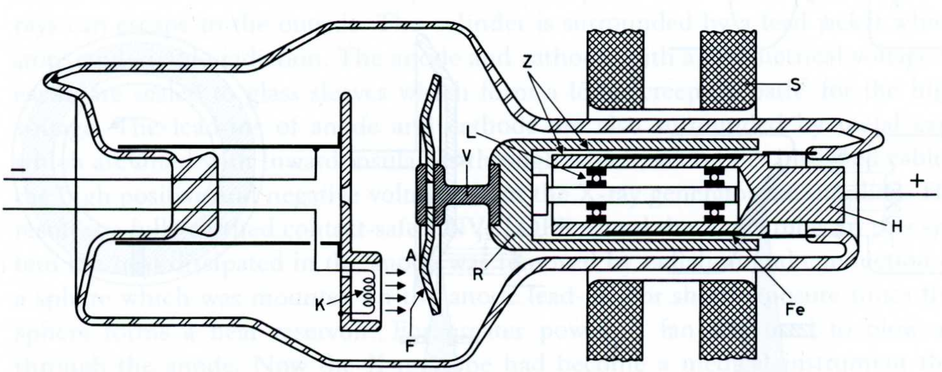


Fig. 112. Rotalix tube with tungsten disc-anode in hard glass envelope in oil (Philips Techn. Review). K = filament, A = anode consisting of a tungsten disc, V = molybdenum connecting piece, R = copper rotor with iron ring Fe, Z = blackened areas for heat radiation, H = anode holder, L = ball bearings lubricated with lead, S = coil for driving the rotor.

Such a Rotalix tube is suitable to photograph the stomach with a focus of 1.2 mm, and these photographs are much sharper than those made with the line focus.

The realization of a tube for such high voltages and currents, in which the anode completes 3000 revolutions per minute in a vacuum, the electrons being focussed on an area of 1.2 mm, can be regarded as another prime accomplishment in electronics. Few people who witness how an "X-ray" is made in a hospital will realize how much had to be done to arrive at these results.

This must be the end of our small excursion through X-ray tube engineering, and we shall have to do without description of particularly interesting tube designs which were developed for surface radiation, depth radiation, and material testing.

## II.G.4. X-RAY VALVES

X-ray tubes are usually driven with a high direct voltage, so that the alternating voltage of a high voltage transformer must be rectified. The tube rectifiers used for that purpose are referred to as VALVES. For the abovementioned tubes the maximum reverse voltage will be 150 kV. They are preferably placed in an oil container together with the transformer. Here too, a hard glass envelope is used. To minimize the heater current consumption, a tungsten-thorium heater is used. The anode is made of molybdenum plate. The high voltage requires polished electrode surfaces, a very high vacuum, and absolute absence of any contamination inside the tube so as to avoid breakdowns which might cause damage to the tube.

Obtaining the heater voltage for the cathode which carries a high voltage with respect to earth, is a difficult and expensive matter. In modern equipment, therefore, these valves have been replaced by series connected silicon rectifiers.

# SECTION H. GAS DISCHARGE TUBES

## II.H.1. CHARACTERISTICS OF THE GAS DISCHARGE

In vacuum electron tubes the electrons, influenced by the electric field, are free to find their own way. If gas molecules are let into the vacuum, the moving electrons will collide against them. The so much lighter electron will hardly move the gas molecule, but bounce back instead. Apart from the field strength, the speed of the electron depends on the length of the "free path": i.e. the distance the electron can cover before it collides against a next gas molecule. The longer the free path (the lower the gas pressure), the more speed the electron can develop before it hits another gas molecule. At a certain velocity (EXCITATION VOLTAGE) the electron will be able to force electrons rotating in a certain orbit round the atom nucleus in the gas atom, to another orbit. After a short time (about $10^{-8}$ sec) the displaced electron drops back to its original orbit, and in doing so produces energy in the form of light radiation (see also Chapter II.C.6). This phenomenon is observed as a GLOW DISCHARGE. At a still greater velocity (IONIZATION VOLTAGE) the electron is able to remove electrons from the gas atom which is then divided into an electron and a positively charged ION. These electrons will be attracted by the anode, so that the anode current increases; the positively charged ions are attracted by the cathode. If there is an "electron cloud" near the cathode (as is the case with a heated cathode), the ions combine with these electrons to form atoms without any charge. If the cathode does not emit surplus electrons, the ions will collide against the cathode at a high velocity, releasing electrons by impact and heat development (as with the "mercury pool" cathode).

Fig. 113 shows the voltage between two flat electrodes in a rarified gas as a function of the current. The two electrodes are supposed to be connected to the + and − pole of a direct current source via a variable resistor which limits the currents and is gradually decreased.

When the breakdown voltage $V_d$, which depends on the ionization voltage of the gas filling, is exceeded, a dark discharge occurs at a low current (AB in Fig. 113) to change into a glow discharge above $10^{-5}$A (DE); many gas atoms are "excited" and some are ionized. If the external resistance is reduced further so that the current in the tube can increase sufficiently, an ARC DISCHARGE sets in at point F. There are then so many free electrons and ions, that the latter can release ever more electrons at the cathode by development of heat. The result is an explosion-like multiplication of the number of electrons and ions already present, and according to GH in Fig. 113 we then have a high current (ARC) at a very low voltage across the electrodes (ARC VOLTAGE).

If this phenomenon takes place in a rectifier, the positive and negative phase of the alternating voltage being alternatingly applied to the anode with respect to the cathode, we notice the following differences with the vacuum rectifier:

a. the cycle shown in Fig. 113 is completed 50 times per second. Current will flow only when the anode voltage rises above the ignition voltage (breakdown voltage) to drop again to the much lower arc voltage as soon as the arc is formed.

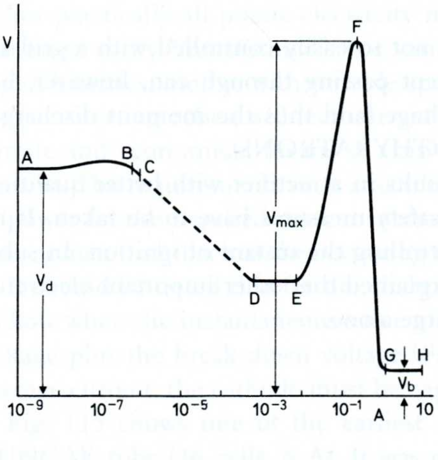


Fig. 113. Current-voltage characteristic of a gas discharge between parallel straight plates. $V_d$ = breakdown voltage, $V_{max}$ = voltage at which the glow discharge changes into an arc, $V_b$ = arc voltage, AB = dark discharge, DE = glow discharge, GH = arc discharge.

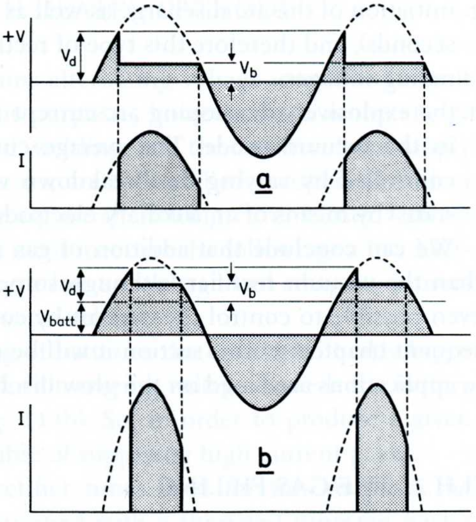


Fig. 114. Current as a function of time with a half wave rectifier a. resistive load, b. battery charge (during battery charge the maximum current intensity at equal average current values is considerably higher).

At an anode voltage zero the arc extinguishes. The current and voltage will vary according to Fig. 114a.

b. in contrast with the vacuum rectifier, where a rather high voltage is required to obtain sufficient current (to overcome the space discharge or electron cloud), the voltage between anode and cathode, after ignition, is equal to the arc voltage, that is to say, very low (about 10 V). In other words: during conduction the internal resistance $R_i$ is very low, and the efficiency $R_u/(R_i + R_u)$ is much higher than that of a vacuum rectifier.

c. since the electrons (slowed down by collision with the gas atoms) land on the anode at a much lower velocity, the anode is heated to a lesser extent than in a vacuum rectifier for equal output power.

d. in the rectifying phase the arc discharge can be stimulated by a dense electron emission from the cathode (cathode heated with built-in heater or by electron bombardment). In the phase at which the anode is negative there must be no arc discharge. An arc discharge in this particular phase is the dangerous backfire, a short circuit, which may well damage the installation and the tube. Hence measures are taken to suppress the glow discharge and electron emission from the anode by cooling the whole area of the anode and making it of a metal that does not readily emit electrons.

e. the characteristic of the tube is highly dependent on the gas pressure. It should be ensured that the quantity of gas stored in the tube remains constant and that neither tube wall nor electrodes produce gas.

f. the cathode, and particularly an oxide cathode, must be resistent against ion bombardments. The oxide cathodes of these tubes are provided with numerous cavities into which the ions can indeed enter, but without velocity.

g. initiation of the arc discharge as well as its termination takes time (tens of micro seconds), and therefore this type of rectifier cannot handle high-frequency alternating voltages.

h. the explosively developing arc current is not so easily controlled with a grid as in the vacuum triode. The average current passing through can, however, be controlled by varying the breakdown voltage (and thus the moment discharge starts) by means of an auxiliary electrode (THYRATRONS).

We can conclude that addition of gas results in a rectifier with better qualities than the vacuum rectifier, although some safety measures have to be taken. It is even possible to control the current by controlling the instant of ignition. In subsequent chapters of this section it will be explained that other important electronic applications are based on the glow discharge alone.

## II.H.2. THE GAS FILLING

The gas filling should preferably be a rare gas which does not easily react with the glass wall or electrodes of the tube. ARGON has the lowest ionization voltage (15.7 V). NEON has a higher ionization voltage, but at 16.55 V a metastable condition is reached in which atoms are excited so that their electrons are made to follow another orbit. When dropping back to their original orbit, they can ionize an argon atom. So a mixture of neon with 0.1% argon produces an arc at a lower voltage (PENNING EFFECT). In high-voltage rectifiers the expensive XENON, with an ionizing voltage of 12.1 V, is sometimes used. This gas has large atoms which land on the cathode at a lower velocity, thus causing less damage there. Tubes containing xenon can be operated at temperatures ranging from −75 to +90 °C.

A drawback of the gas filling is that during operation some gas is always absorbed by the glass envelope of the tube and by its electrodes, particularly by metal layers which settle on the glass wall. As a result the characteristics of the tube change. Like in vacuum tubes, getters are used to absorb foreign gases that are produced during operation.

Thyratrons for high frequency radar use hydrogen, which ionizes more quickly. However, it soon disappears, so the tube must be equipped with a REPLENISHER, a material that produces hydrogen when the gas pressure drops below a certain value.

A simple “replenisher” is obtained by inserting a few drops of mercury into the tube which is then filled with mercury vapour. At 45 °C saturated mercury vapour has a pressure of about 0.1 mm mercury. The tube is often filled with a rare gas to introduce the discharge; as the temperature increases, the mercury vapour takes over. It is desirable to collect the mercury in a suitable place (usually below the cathode); the temperature in that spot is decisive for the pressure. If other components of the tube cool off, the vapour pressure is locally reduced. A certain amount of mercury then condenses to flow back to the lowest position. Mercury vapour rectifiers must always be operated in the same position (vertically) and they can work only at an ambient temperature ranging from 25-75 °C.

## II.H.3. GAS FILLED RECTIFIERS WITH HEATED CATHODE

For practically all public electricity mains alternating voltage is used because it can be so easily "transformed". For various purposes, however, such as the charging of batteries, motors with variable speed, welding, and the anode voltage supply for transmitting tubes, a direct current is required. Gas filled rectifiers provide a simple and economical means of converting an alternating current into a direct current. The requirements such a rectifier must meet are the following: a low heater power for the required current intensity, a low arc voltage and adequate protection against flash-back at the occurring reverse voltage.

It should be noted that for the charging of batteries the current will only begin to flow when the instantaneous value of the alternating voltage exceeds the battery voltage plus the break-down voltage (Fig. 114b). So, in order to produce a given average current, the cathode must be capable of supplying high current peaks.

Fig. 115 shows one of the earliest rectifier tubes used for this purpose: the TUNGAR tube (36 cells, 6 A). It was equipped with a thoriated tungsten heater and an argon filling of 10-15 cm pressure, so that evaporation of the heater was reduced to a minimum.

Fig. 116 shows the classical Philips type 367 (50 years in use now) for the same current intensity (12 cells, 6 A). It contains a double helix oxide cathode, to reduce the heater current consumption, a rare gas filling, and graphite anodes. Here the two anodes are, however, housed in one envelope, so that one tube can be used to rectify two phases. Surprisingly enough no arc is formed between the anodes which carry a voltage with respect to each other that is twice the reverse voltage

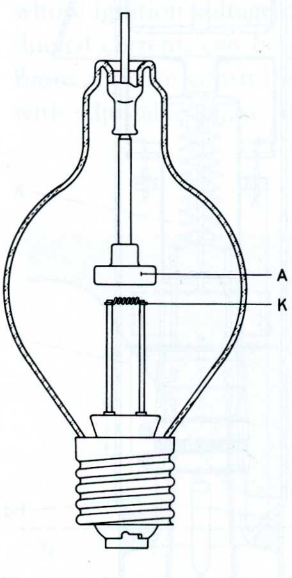


Fig. 115. Tungar rectifying tube with argon filling, 100 V, 6 A. A = graphite anode, K = thoriated tungsten heater (v.d. Horst Gas Discharge Tubes, Philips Technical Library, 1964).

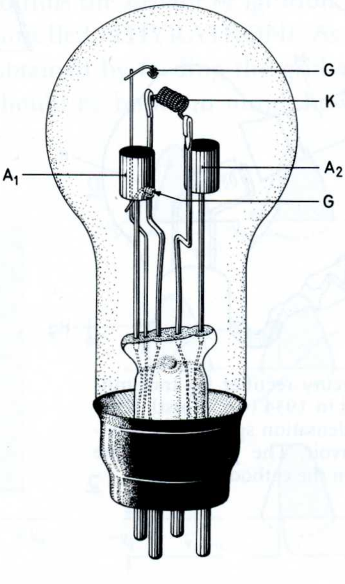


Fig. 116. Full wave rectifying tube. Philips type 367 (designed in 1922) for 60 V, 6 A (v.d. Horst). K = oxyde-cathode-heater, A1, A2 = graphite cathode, G = getter.

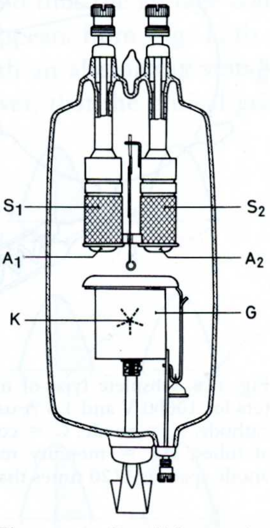


Fig. 117. Rectifying tube Philips type 1859 for 110 V, 50 A, mercury-argon filling (v.d. Horst). K = cathode, G = anode screen annex ignition electrode, A1, A2 = anodes, S1, S2 = gauze screens to prevent anode emission.

between anode and cathode. The explanation is that the anode with the negative voltage is screened by an ion cloud.

Fig. 117 shows a Philips type 1859 for 50 A and 115 V (a.c.), in which more rigorous measures had to be taken to prevent flash-back; a screen at cathode potential is mounted between the anodes. The anodes are surrounded by cylinders of gauze, which prevent a glow discharge there, but allow radiated heat to pass through. The anode leads are screened by glass. The cathode is surrounded by a cylinder to which a positive voltage is applied, and which serves as the ignition electrode. So the functions of electron emission and ion formation take place inside this hood, and the particles leaving the oxide cathode also remain inside it.

The arc is formed between this hood and the face of the anodes. The gas filling is a rare gas plus mercury vapour.

For rectifying tubes for a higher voltage (220 V) the anodes are mounted in separate side arms of the envelope so that the length of the glow discharge — and thus its voltage — is increased.

For still higher voltages the gas pressure must be chosen much lower to increase the back-fire voltage (glow discharge voltage). In the vicinity of the heater the pressure must, however, be high to prevent evaporation. Fig. 118 shows an inter-

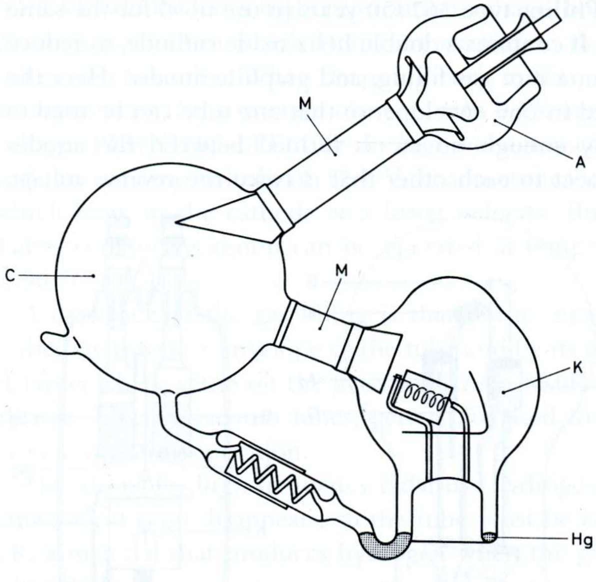


Fig. 118. Obsolete type of mercury rectifier for transmitters for 10000 V and 1.5 A used in 1934 (v.d. Horst). K = cathode, A = anode, C = condensation space, M = metal tubes, Hg = mercury reservoir. The pressure in the anode space is 1/20 times that in the cathode space.

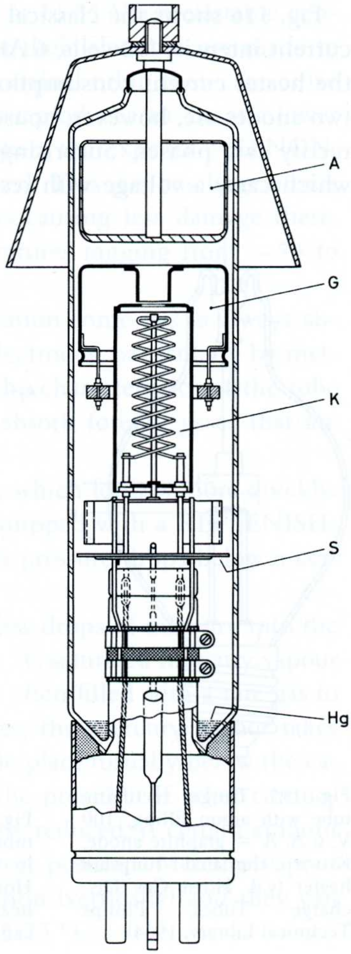


Fig. 119. Controlled high voltage rectifier. Philips type DCG 12/30, 27 kV, 2.5 A (v.d. Horst). K = cathode, G = grid, A = anode, S = heat screen to keep mercury reservoir (Hg) cool. The discharge takes place along the short discharge path G-A where no flash-back arc can develop.

esting solution for a 10 000 V, 1.5 A rectifier which is now obsolete (MULDER 1934). Here the pressure in the hot cathode space (K) is high. In the cooler condensation space (C) the mercury condenses to flow back to the cathode space. Notwithstanding the high temperature of the anode, the pressure in the tube to the anode (C-A) is also low, so that the breakdown voltage of the gas discharge is high there.

The modern rectifier DCG 12/30 for 2.5 A at 27 kV (Fig. 119) is greatly in contrast with the design shown in Fig. 118, particularly since the distance between anode and grid G, which has almost cathode potential, is kept very small. Thanks to this short distance and the long free path of the electrons at the applied low pressure, there is no glow discharge. When the anode is positive, the multitude of electrons emitted by the heater can easily reach the cathode. The condensed mercury gathers at the bottom of the tube in a space which is screened from the cathode by a screen S, and so remains cool (15° C).

## II.H.4. CONTROLLED RECTIFIERS (THYRATRONS)

In a gas-filled rectifier a grid (forming a hood or cylinder round the cathode) with a negative voltage with respect to the cathode will increase the ignition voltage. Ignition will then occur only in the hatched area of the graph shown in Fig. 120. For a sinusoidally alternating voltage on the anode this graph can be translated into the curve $V_{g\,crit}$ in Fig. 121a. As the grid voltage becomes more negative, the time during which the current flows becomes shorter. Here we have a rectifier whose ignition voltage, and thus the instant of ignition, and thus the average conducted current, can be controlled (THYRATRON). As appears from Fig. 121b a more accurate control is obtained by feeding the grid with an alternating voltage with adjustable phase. It should be borne in mind, however, that the critical grid

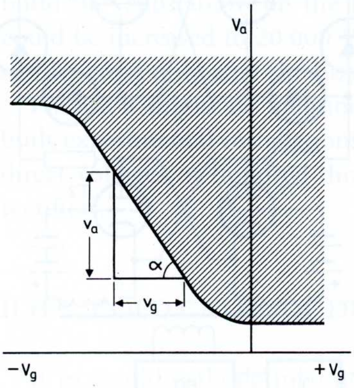


Fig. 120. The ignition voltage of a thyratron is influenced by the grid voltage; $\tan\alpha = V_a/V_g$ can be defined as the slope. Ignition takes place in the hatched area.

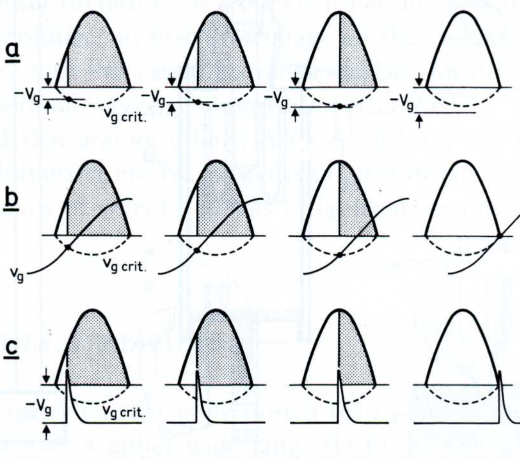


Fig. 121. Various methods of adjusting the ignition moment of a thyratron: a. with an adjustable negative direct grid voltage. b. with an alternating voltage of adjustable phase. c. with a positive voltage pulse, whose instant of occurrence can be varied, on a constant negative grid voltage.

voltage varies with temperature, gas pressure etc. Better reproducibility of control is therefore obtained by using a very steep positive voltage peak (the instant of occurrence being varied) on a negative direct voltage, (Fig. 121c). With modern electronic means, diodes and transistors, such a voltage peak can easily be produced.

Finally there is the possibility of using a second grid to fulfil several functions. Since in many applications the thyratrons have been replaced by variable silicon diodes (thyristors) we shall refrain from giving a description of all existing types.

## II.H.5. IGNITRONS

In Chapter II.H.1 it was explained that ions landing on the cathode can cause electron emission due to heat development. For the heater cathode it involves a danger that the hottest point emits electrons and attracts ions so that the temperature rises even more. This phenomenon need not be feared when the cathode is formed by a mercury pool. If mercury evaporates from the hottest spot in this pool, the "liquid cathode" immediately replenishes the missing quantity, which means that the pool serves as cathode as well as "replenisher" for mercury vapour. The mercury pressure depends on the temperature in the tube, and if the tube wall is cooled, the mercury will condense there to flow back to its source of origin; the pool.

Ignition at a particular spot on the mercury pool is, however, hardly reproducible, and therefore an IGNITOR is required to introduce a reproducible ignition at a certain point. This can be done with an auxiliary electrode which is coated with

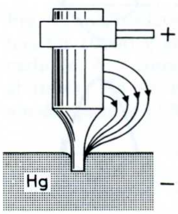


Fig. 122. Principle of ignitron ignition.

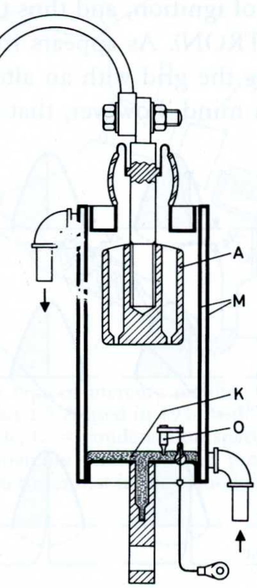


Fig. 123. Cross section of an ignition used for welding (v.d. Horst). A = anode, M = double cylinder wall with cooling water, K = mercury pool cathode.

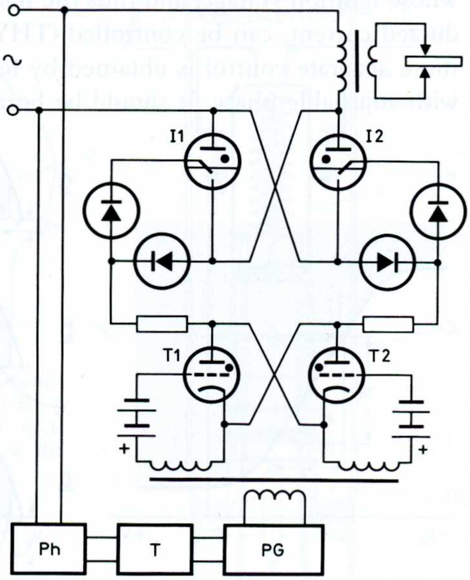


Fig. 124. Circuit for controlling an ignitron spot welding machine (full wave, connected direct to the mains). $I_1$, $I_2$ = ignitrons, $T_1$, $T_2$ = thyratrons, PG = pulse generator, T = time switch, Ph = phase control.

an insulating layer (glass). When a high voltage is applied to this electrode a glow discharge will occur at the interface between glass and mercury; such a tube is known as a SENDYTRON. Even better ignition is obtained by using an electrode of semiconducting material of a special shape (Fig. 122) so that an extremely high field strength is produced at the surface of the mercury. This tube is called IGNITRON.

Fig. 123 shows a cross section of an ignitron used for spot welding. The tube is made of sheet steel and has a double jacket used for water cooling. The graphite anode is fitted to the top by means of a glass lead-through. The anode dissipates its heat by radiation towards the tube wall, and will therefore always have a higher temperature than this wall. As a result the mercury will not condense on the anode, but on the tube wall to flow back to the pool at the bottom where the igniter is fitted.

Fig. 124 shows the circuit diagram for a full-wave spot welding machine with two ignitrons operated direct from the mains. The ignition current is supplied by a pulse generator whose output is amplified with two thyratrons. Apart from controlling the number of ignition pulses, it is also possible to control their phase (current control as with thyratrons). Such a simple installation is capable of generating powers up to 2000 kVA for spot welding. For this particular application the ignitron is more economical than the silicon controlled rectifier because the peak value of the current can be very high (10 000 A).

An important application in which ignitrons are used as rectifiers is electrical traction. Here direct current motors are used to facilitate control. Alternating current is then converted into direct current in substations. In the twenties, these substations for tramways were fitted out with glass mercury rectifiers. For the electrification of railways large metal rectifier vessels with vacuum pump were used. However, the motor did not permit more than 1500 V voltage for the overhead contact wire, so that a considerable wire diameter and many substations were required. By using the so much lighter thyratrons as rectifiers it became possible to build the "substation" on the locomotive, so that the voltage on the contact wire could be increased to 20 000 V (a.c.). In this case electronics was late, for this possibility arose only after the European railways had almost been electrified.

Finally it might be mentioned that among others, A.S.E.A., B.B.C., and G.E. built experimental installations to transfer electric power over great distances with direct voltages of hundred thousands of volts by means of ignitrons or other gas rectifiers.

## II.H.6. COLD CATHODE DIODES: STABILIZERS

If in a cold cathode tube no greater current is permitted than point E in Fig. 113, we see from this graph that over a rather wide range (D-E) the voltage remains constant when the current is varied. If the tube is used in a circuit as shown in Fig. 125, the voltage across the load resistor $R_o$ ($V_b$) will remain constant if its value is varied within certain limits, because the load resistor and the glow discharge together will draw so much current that the mains voltage minus the volt-

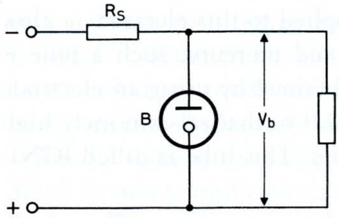


Fig. 125. Circuit for connecting a stabilizing tube.

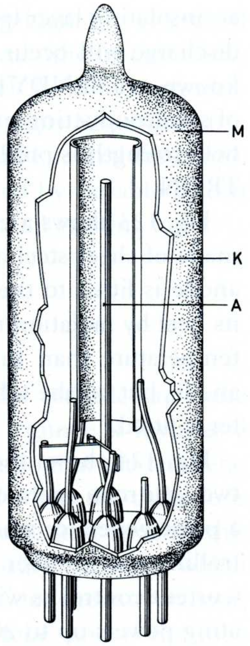


Fig. 126. Reference tube 85A2 with cut-open envelope (v.d. Horst). A = anode, K = cathode, M = layer of molybdenum sputtered on the envelope wall.

age loss across $R_S$ is equal to the glow discharge voltage. At the proper choice of mains voltage and $R_S$ the voltage across $R_0$ will remain constant when the current through $R_0$ varies between 0 and 50 mA. The tube functions as a voltage STABILIZER. Furthermore the voltage across the glow-discharge tube can be used as a VOLTAGE REFERENCE.

Fig. 126 shows the construction of such a tube consisting of a cylindrical cathode with a rod anode mounted inside. It is important that the ignition voltage be slightly higher than the glow voltage, and therefore the tube is filled with a mixture of helium or neon and argon (Chapter II.B.2). During the pumping process the molybdenum cathode is partly evaporated towards the glass envelope. The result is a very pure cathode surface, whilst the molybdenum layer on the glass wall will bind any released gases. Thus it has been achieved that the voltage of this tube (about 100 V) does not drift more than 0.1 V in 1000 hours. By using a larger cathode area, stabilizer tubes can be made for up to 200 mA.

## II.H.7. THE PENNING MANOMETER

Of the many other versions of gas discharge tube be it sufficient to mention the RADIATION COUNTERS by means of which $\alpha$-, $\beta$-, and $\gamma$-rays can be measured, and furthermore the SURGE ARRESTERS.

However, a subject requiring more attention is the PENNING MANOMETER, which is used to measure gas pressures from $10^{-5}$ to $10^{-9}$ mm mercury in combination with a simple micro-ampere meter. Hence this tube is of prime importance in the manufacture of electron tubes. This tube is brought in connection with the space in which the gas pressure is to be measured. The latest version consists of two cathodes on either side of a cylindrical anode, in the center of which a magnetic field is generated. Influenced by this magnetic field, the electrons emitted by the cathode follow a helical path towards and through the anode. Owing to their helical movement there is a fair chance that, even at a very low gas pressure, the electrons will collide against sufficient gas molecules which are then ionized, giving rise to a discharge current. It is an interesting feature that throughout the range from $10^{-9}$ to $10^{-5}$ mm mercury pressure this current varies linearly with the gas pressure, so that this pressure can be measured with an ammeter indicating 2 μA.

It should be noted that the ions are trapped by the cathode, so that the manometer tube improves the vacuum. This principle has been used to manufacture a tube by means of which a very high vacuum can be obtained: the so-called ION PUMP.

## II.H.8. RELAY (TRIGGER) TUBES

Like thyratrons, glow discharge diodes can be equipped with an extra electrode to introduce the ignition. The anode voltage is so chosen that no ignition takes place; then a voltage applied to the auxiliary (TRIGGER) electrode starts the discharge. Although the current of the glow discharge (40-100 mA) is much weaker than that of a thyratron, there is still amplification, for the current on the trigger electrode is but of the order of 50 μA. This is a true example of a relay-tube; the device is switched on with the trigger voltage, and switched off by the cutting out of the anode voltage.

Of course, when operating on an alternating voltage, the usual measures will have to be taken to prevent back-fire.

The ignition time is of the order of 50 μs (some radio active material is added to accelerate the ignition); the time needed to extinguish the arc is about 200 μs.

Trigger tubes have played an important part in the first years of electronic counting and computing. Nowadays their functions have been completely taken over by transistors. The only advantage of trigger tubes is that the glow light "shows" whether the relevant circuit is switched on or off.

## II.H.9. INDICATOR TUBES

Although their functions are quite simple, tubes in which figures, characters etc. are displayed by the neon discharge ought to be mentioned here. According to Fig. 127, the figures 0-9, for instance, are cut out of metal and connected to 10 lead-ins. The figures are placed in a row behind a gauze anode. If one of the lead-ins receives a negative voltage with respect to this anode, the corresponding figure will light up as a result of the glow discharge. This type of tube presents an inexpensive means of immediate indication of calculated results without mechanical moving parts, and at low power consumption.

Fig. 128. 14 Systems of numerical tubes contained in one common envelope.

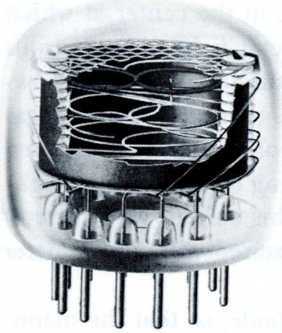


Fig. 127. Cut away view of a numerical tube.

Fig. 128 shows fourteen of such systems sealed together into one envelope (PANDICON).

These indicator tubes are the last gas discharge tubes still used in "switching".

# SECTION J. SEMICONDUCTORS

## II.J.1. PHYSICAL PRINCIPLES

The atom structure of the most used semiconductor materials GERMANIUM and SILICON is such that the outer orbit contains four (VALENCY) electrons. These electrons will be inclined to leave the atom and to combine with other nearby atoms. In the crystal lattice of these materials the valency atoms will link with nearby atoms as shown in Fig. 129a (for the sake of simplicity drawn in a flat plane). In contrast with metals like copper and iron, which contain large numbers of free electrons ensuring electric conduction, a pure (INTRINSIC) semiconductor crystal of low temperature contains no free electrons.

As the temperature rises, the movement within the atom is accelerated so that a few electrons (FREE ELECTRONS) can escape, which leave HOLES in the atom. If an electric field (F in Fig. 129a) is applied, a "hole" will be formed in atom A to be filled again by an electron b from atom B which is directed there under influence of the field. The hole left open in B is filled again by an electron c from atom C, and so on. Here we see that the electrons move in the direction of the field, whereas the holes (apparently) move in the opposite direction. At room temperature the number of "free" electrons is still very small (for germanium one per $10^9$, and for silicon one per $10^{12}$ of the total amount of atoms). The number of free electrons increases with the temperature, which means that the resistance decreases. At temperatures of 100 °C for Ge, and 200 °C for Si, the functioning of the diode or transistor is noticeably affected by the electrons splitting from the atoms (LEAK CURRENT; at a higher voltage: breakdown).

There are also materials like phosphor, arsenic or antimony which have five electrons in their outer orbit. If a small quantity of antimony is added to the pure germanium, the antimony atoms will try to fit in the crystal lattice as shown in Fig. 129a, but they have one "free" electron too many. Whereas the nucleus with four electrons must remain stationary in the crystal lattice, the free electron can move about. The mixture with a surplus of free electrons is called the N(egative) germanium, the materials supplying the free electrons are referred to as DONORS.

Other metals, such as borium, aluminium, indium or gallium, have atoms with three electrons in the outer orbit. If these atoms settle in the crystal lattice, there is a shortage of electrons: "holes" are left open. In passing through, free electrons will fill these holes, and the result is a "hole current" as shown in Fig. 129a. This material has a shortage of electrons and is known as P(ositive) germanium. The impurities supplying holes are called ACCEPTORS, the general term for both types of impurities being "DOPE".

Such "doped" material will be a better electric conductor than the intrinsic material. As explained in the foregoing, intrinsic germanium has but one free electron per $10^9$ atoms: addition of one millionth part impurity to the germanium is sufficient to multiply the number of free electrons or holes by a factor of 1000. In silicon, which has one free electron per $10^{12}$ atoms, the number of free electrons is multiplied by 1 000 000. Other impurities, already present in the intrinsic mate-

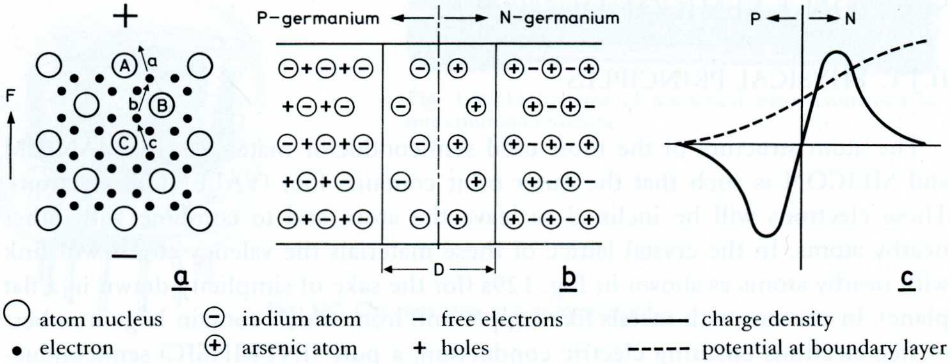


Fig. 129. Principle of the diode: a. crystal structure, b. P-N junction, D = depletion layer, c. functions.

rial, are likely to have an uncontrollable and irreproducible effect. Hence the starting point should be an extremely pure intrinsic material.

The "doped" material is conductive in all directions. The electrons and holes will follow and oppose the field respectively, and change direction with the polarity of the field. Here, however, the resistance increases at high temperatures.

If a P and N germanium are sealed together (Fig. 129b), a peculiar phenomenon occurs at the boundary layer (JUNCTION). Under normal conditions the free electrons and holes would, by "diffusion", be distributed uniformly throughout the crystal. However, on the left side of the P-N juction the crystal absorbs so many electrons from the N material that no holes are left; on the right side of the junction no electrons are left (the area, D in Fig. 129b, is called the DEPLETION LAYER). A surplus of electrons of the acceptor crystal (four instead of three) causes a negative charge left of the junction; a shortage of electrons (four instead of five) causes a positive charge right of the junction (Fig. 129c). These charges will repel the electrons and holes respectively, and prevent further diffusion of electrons from right to left, and holes from left to right. The result is a potential difference (Fig. 129c) of the order of 0.2 V for Ge, and 0.5 V for Si (THRESHOLD VOLTAGE). If a positive voltage with respect to P is applied to N, this potential difference is increased, and there is no flow of current. If a negative voltage is applied to N, the potential difference is neutralized and, provided the voltage exceeds 0.2 or 0.5 V, the electrons and holes can once again move to the left and right respectively. At a higher voltage the transfer of electrons and holes is intensified. Thus we come to the conclusion that such a P-N junction functions like a rectifier, whose characteristics will be described in further detail in Chapter II.J.2.

It stands to reason that the scientists in the Bell Laboratory, once they could explain the mechanism of the crystal diode in 1948, did not want to make the same mistake as Fleming in 1904. (He knew the electron tube as a rectifier, but had to leave the honour of having invented the amplifier to de Forest). They started in investigating how the flow of electrons and holes could be controlled with a third electrode consuming low energy. And so the TRANSISTOR was invented.

As will be further explained below, it is essential for the functioning of a transistor that the dope concentrations in the P and N layers are different.

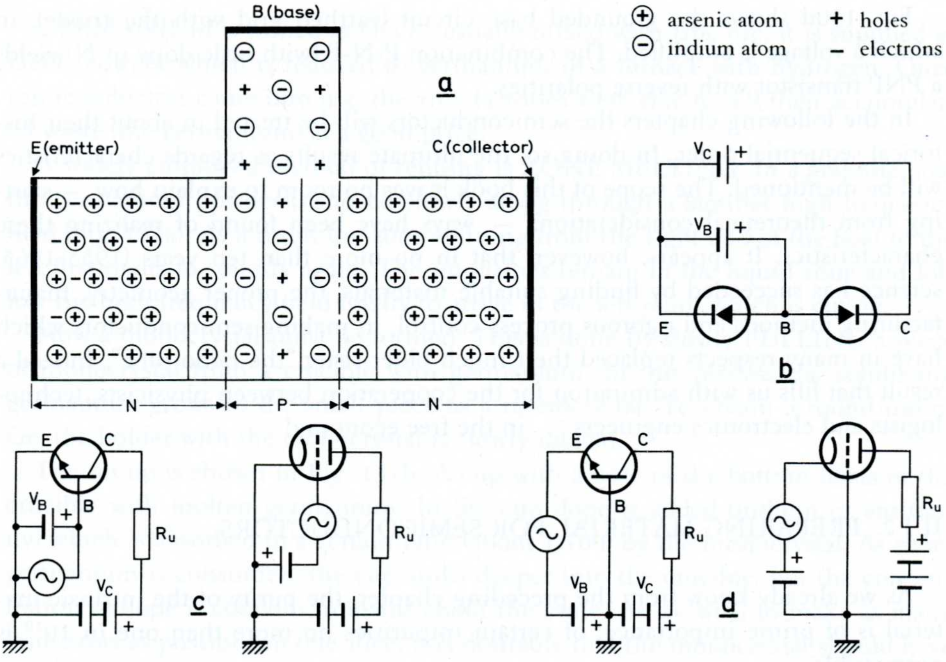


Fig. 130. Principle of the transistor: a. NPN transistor, b. equivalent circuit, c. grounded emitter or grounded cathode circuit, respectively, d. grounded base or grounded grid circuit, respectively.

Fig. 130a shows the structure of an N-P-N transistor. It consists of an N-layer to which the EMITTER (E) is connected, a P-layer to which the BASE (B) is connected, and a second N-layer to which the COLLECTOR (C) is connected. In terms of circuitry we now have two diodes in opposition with each other (Fig. 130b). In the neutral condition there is a shortage of electrons left of junction EB, whereas on its right there is a shortage of free holes. If a positive voltage with respect to E is applied to B, the potential difference (Fig. 129c) will be overcome, and the diode E-B becomes conductive in that free electrons flow towards the base. This current increases with the voltage $V_B$. However, the P-material has intentially been given far less dope than the N-material, with the result that the electrons fail to find enough holes to be neutralized in the base, and will of necessity accumulate there. If a positive voltage with respect to E is applied to C, the diode B-C will be blocked (Fig. 130b) (as long as the diode E-B is non-conductive) and there will be no flow of current. If E-B is conducting, and provided the voltage $V_C$ is high enough, the electrons accumulated in B will be attracted by this external voltage. So a small part of the electrons supplied by E will flow to B, and a large part will flow to C, but the number of electrons supplied by E will depend on the voltage between B and C. Hence there is "amplification": the current towards the collector is many times that towards the base, but depends on the voltage or current through the base respectively.

The circuit discussed here is called "grounded emitter", to be compared with a triode with earthed cathode (Fig. 130c). By means of load resistor $R_u$ the variations in $I_C$ brought about by variations in $I_B$ can be converted into voltage variations across $R_u$.

Fig. 130d shows the grounded base circuit (earthed grid with the triode), in which a voltage is amplified. The combination P-N-P (with little dope in N) yields a PNP transistor with reverse polarities.

In the following chapters the semiconductors will be treated in about their historical sequential order. In doing so, the ultimate results as regards characteristics will be mentioned. The scope of this book leaves no room to explain how — starting from theoretical considerations — ways have been found of realizing these characteristics. It appears, however, that in no more than ten years (1955-1965) science has succeeded by finding suitable materials, the proper geometry, manufacturing methods and rigorous process control, in making semiconductors which have in many respects replaced their much older sister "the radio tube". Indeed a result that fills us with admiration for the cooperation between physicists, technologists and electronics engineers . . . in the free economy!

## II.J.2. PREPARING MATERIAL FOR SEMICONDUCTORS

As we already know from the preceding chapter, the purity of the intrinsic material is of prime importance; of certain impurities no more than one in $10^{10}$ is permissible.

Furthermore the dope must be dosaged with extreme care, since a difference of a few millionth parts may already have considerable effect on the characteristic.

The discussion in Chapter II.J.1 was based on a completely regular crystal lattice. Irregularities (DISLOCATIONS) in this crystal lattice might prevent the transistor from functioning.

Although some chemical factories specialize in manufacturing this pure material, large semiconductor factories have started their own production, because the quality of the material has such a great influence on the characteristics of the final product, and even on the possibility of making it.

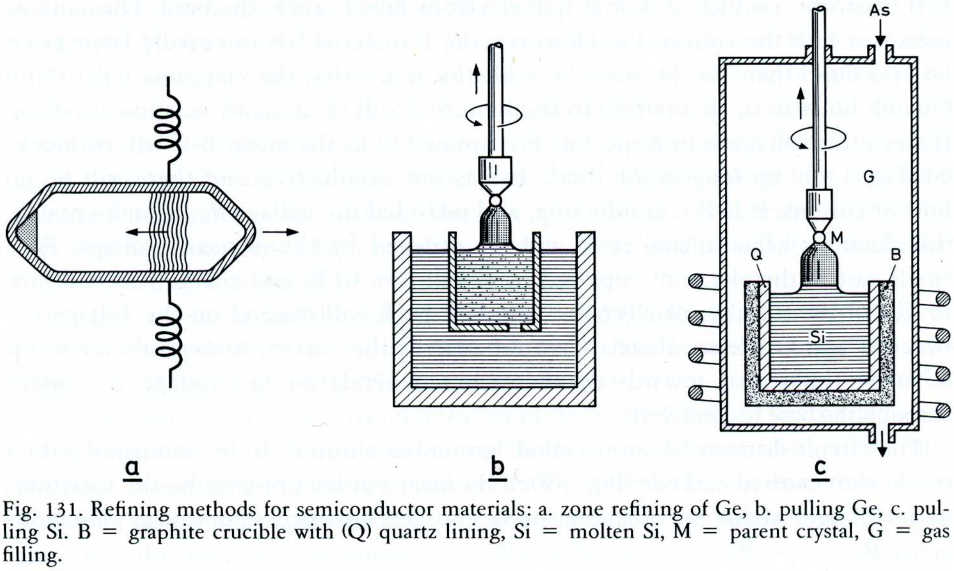


Fig. 131. Refining methods for semiconductor materials: a. zone refining of Ge, b. pulling Ge, c. pulling Si. B = graphite crucible with (Q) quartz lining, Si = molten Si, M = parent crystal, G = gas filling.

GERMANIUM is mined as $GeO_2$, usually mixed with zinc ore, it is supplied as $GeO_2$ powder which is reduced to germanium in a furnace with hydrogen. Once semiconductors came into use, the zinc factories were able to sell their accumulated waste (the germanium) at a good price.

A widely employed method of refining is ZONE MELTING. In a graphite boat the germanium powder is moved slowly along through a laminar high-frequency field (Fig. 131a). As a result the zone running from the right end of the boat to the left first melts to solidify again. The impurities remain in the liquid zone and follow the high frequency field finally to arrive in the left point which is cut off.

Now a monocrystal must be formed. This is done by slowly PULLING a piece of monocrystal from a crucible with germanium. In this process the solidifying germanium grows to the small piece as a monocrystal. To obtain a round rod of Ge, the holder with the monocrystal is slowly rotated.

The set up is shown in Fig. 131b. A cup with a hole in the bottom floats in the crucible with molten germanium. In the cup dope is added (indium or antimony) which is absorbed in a certain ratio (about 1:100) by the monocrystal. As more germanium is consumed, the cup sinks deeper into the crucible, but the concentration of dope it contains, remains about the same. As we wish to make as many transistors as possible on one slice, it is desirable that the monocrystal should have the largest possible diameter (25-50 mm). Fig. 133 shows the growth of the rod of germanium.

SILICON is supplied in rods of rather pure polycrystalline material. For high-ohmic material the FLOATING ZONE process, as shown in Fig. 131a, is sometimes used, but then vertically, Fig. 132; for low-ohmic material the pulling process shown in Fig. 131c is preferred. Here the dope in the form of a melting ma-

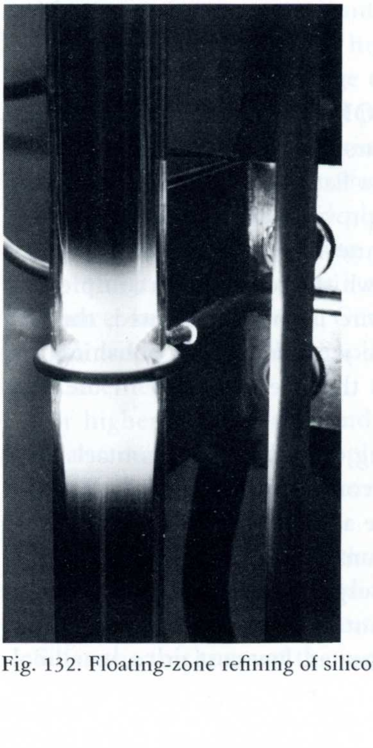


Fig. 132. Floating-zone refining of silicon.

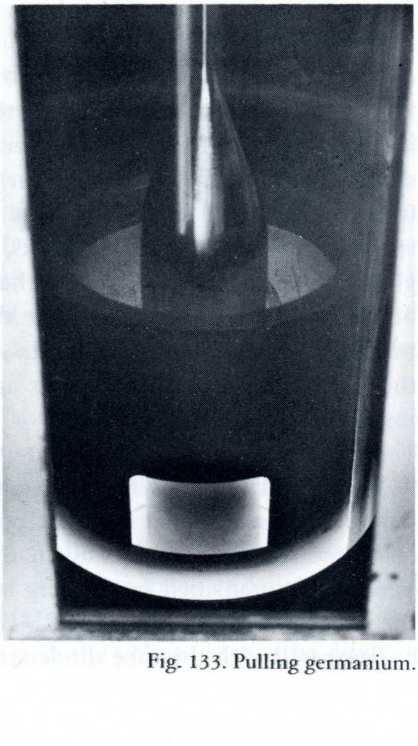


Fig. 133. Pulling germanium.

terial (aluminium, borium) can be added to the liquid silicon or, alternatively, it can be added from the ambient space in the form of vapour (arsenic, phosphor).

With a diamond circular saw the rod of monocrystal material, obtained in the aforementioned processes, is cut into SLICES about 0.3 mm thick, which are then carefully polished and finally etched and rinsed to remove all mechanical contamination. The correct composition can be checked by resistance measuring; by inspecting the etched surface at obliquely incident light, one gains an impression of the crystal structure. Here too, it applies that "the proof of the pudding is in the eating", and only after the transistors made of the material have been measured, will it be possible to say something definite as to the quality of the material made.

It should be noted that with a large number of silicon transistors the original crystal material does not participate at all in the transistor function which, instead, takes place in an EPITAXIAL layer grown on the crystal. The crystal must have a certain thickness to enable processing in manufacture (0.2 - 0.4 mm), and if a material of high resistance is required to obtain specific characteristics, this may lead to an impermissible increase in collector resistance. In that case the transistor is made of a low-resistance material to which the collector is connected at a later stage, and the crystal is coated with a layer of high-resistance material. This can be done by heating a slice to 1200 °C in an atmosphere of hydrogen mixed with silicon-tetrachloride and the required dope. The vapour then generates monocrystal silicon of the required composition (usually of high resistance) in a layer a few microns thick. Since the characteristics of the transistor are now no longer determined by the parent crystal but by this expitaxial layer, this layer must meet severe requirements as regards purity and crystal structure.

## II.J.3. SMALL DIODES

In principle the earliest form, the POINT CONTACT DIODE, was almost identical to that of the crystal detectors of the years round 1910. The structure is shown in Fig. 134a. A pellet of germanium with a flat polished plane is soldered to one of the connecting wires, the other wire is provided with a sharply pointed springy tungsten end (WHISKER), and both connecting wires are sealed into a glass tube in such a manner that the point of the whisker exerts a certain pressure on the crystal. It is essential that the tungsten wire is sharply pointed, the germanium is a pure monocrystal, the crystal surface is cleaned after polishing and then protected by a thin lacquer against vapours that are produced in the sealing process.

It is quite difficult to investigate the functioning of such a point contact diode (it was still in the experimental stage of the semiconductor technique!). We may confidently assume that from the point of the wire acceptors diffuse into the crystal to produce a P-N junction in the N-germanium. The formation of this P-N junction is intensified by sending a few current surges through the diode during the manufacturing process (FORMING). The advantages of this germanium diode as compared with the tube diode are evident: no heater, low cost price, small vol-

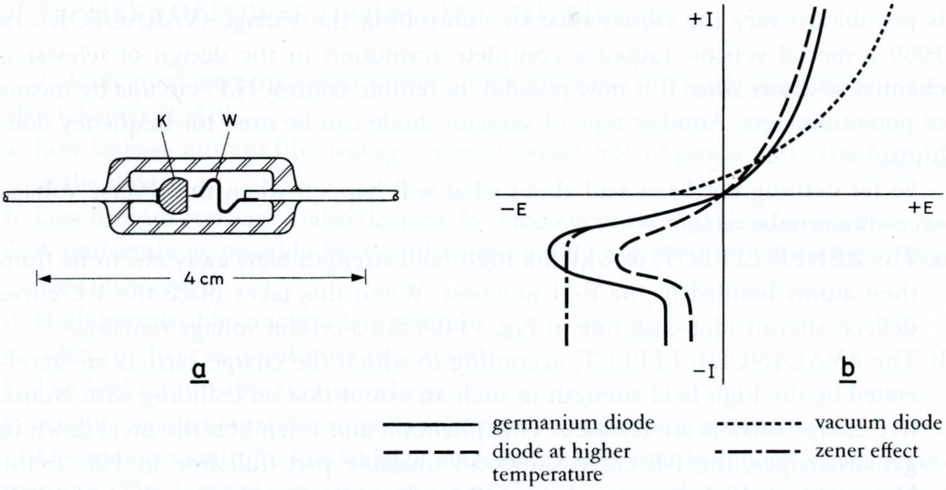


Fig. 134. Point-contact diode: a. cross section K = germanium crystal, W = whisker, b. characteristics.

ume; the low capacitance caused by the small active area is an additional advantage.

Fig. 134b gives the current versus voltage characteristic of such a diode as compared with a tube diode. At a positive voltage there is a gradually increasing current conductance, the forward resistance is lower than that of the tube diode, with the result that the detection efficiency, for instance, particularly at a low load impedance, is better. Therefore the point contact diode is eminently suitable for detection. In contrast with the tube diode, however, a negative voltage still produces a certain leakage current which increases with temperature. This implies that the ambient temperature sets a limit to the voltage used.

Detection and switching at higher frequencies were impeded by another unwelcome effect. When the voltage changes from the forward to the reverse direction, the charge carriers at the P-N junction require some time to disperse, and so the diode remains conductive during the initial period of the negative phase (HOLE STORAGE EFFECT). This effect can be reduced by certain additions to the germanium. As a result of gold plating the wire tip, a quantity of gold diffuses into the crystal during the forming process (GOLD BONDED DIODES), and so the area of the junction becomes larger, the forward current increases, and the forward resistance becomes lower. Germanium point-contact diodes have found, and maintained, a wide field of application, particularly where detection and switching are concerned.

For higher temperatures and improved reliability silicon point contact diodes were also made. They were replaced however, by WHISKERLESS DIODES made according to the diffusion process which will be dealt with later. Here the contact wire is soldered to the crystal.

Some diodes have noteworthy properties which allow of special applications.

Particularly in germanium diodes free electrons are produced under influence of incident light. This implies that PHOTO DIODES can be used as luxmeters.

Since the thickness of the junction changes with the voltage across the diode, it is possible to vary the capacitance by controlling the voltage (VARACTOR). In 1969 a special version caused a complete revolution in the design of television channel selectors since it is now possible to remote control H.F. circuits by means of potentiometers. Another type of varactor diode can be used for frequency doubling.

So far nothing has been said about what will happen when the reverse voltage exceeds a certain value:

a. The ZENER EFFECT, in which a high field strength tears away electrons from their atoms bonded in the P-N junction. When this takes place the I-E curve deflects sharply (dot-dash line in Fig. 134b), but a certain voltage remains.
b. The AVALANCHE EFFECT, according to which the charge carriers are accelerated by the high field strength to such an extent that on colliding with atoms, free charge carriers are released. This phenomenon resembles the breakdown in gas discharges; the I-E curve shows an unstable part (full line in Fig. 134b). Maintaining a high reverse voltage will lead to a short-circuit in which the P-N junction is destroyed. Provided the voltage peak is short, the P-N junction will be able to recover.

A higher concentration of the dope yields a thinner junction; the fieldstrength in this junction will then be high. The Zener effect then occurs at a lower voltage than the avalanche effect, and can be controlled. This is the principle of the ZENER DIODES which function as voltage stabilizers or provide voltage references.

Without going into detail we should also mention the TUNNEL diodes and GUNN effect diodes which are used for extremely high frequencies (up to 5000 MHz).

Fig. 135 gives a survey of a factory for small diodes. It appears that, thanks to the multilateral field of application, the manufacture of these relative simple semiconductor elements, sometimes costing no more than a few cents each, has developed into a large industry.

Fig. 135. A factory of small diodes.

## II.J.4. SILICON POWER DIODES AND THYRISTORS

A diode used to rectify great powers must meet the following requirements (see also Chapter II.H.1)

a. Low leakage current (the leakage current causes loss of power and extra heating of the diode).
b. Low forward resistance (also decisive for efficiency and heat development).
c. A uniformly as possible heat, and consequently temperature, distribution over the entire crystal.
d. High permissible junction temperature.
e. Adequate heat removal from the crystal.
f. The highest possible reverse voltage.
g. Security against breakdown (avalanche effect) at excessive voltages.

The requirement of a low leakage current is better met by silicon than by germanium. The forward resistance of germanium is lower than that of silicon, but can be made very low for silicon too, by using the proper doping. The maximum permissible temperature for silicon is, however, so much higher that higher current intensities can be used. This is the reason why silicon, once it could be made in the right way, is used almost exlusively for power rectifiers.

Requirement c, that of a uniform heat distribution, is easier satisfied by adding dope according to the diffusion method than by using the alloying method as discussed in Chapter II.J.5. With the diffusion method the dope gradually and constantly diffuses from the surrounding gas into the surface of the monocrystal, whereas in the alloying method incidental circumstances can often give rise to irregularities. So the type almost exclusively used for POWER DIODES is the diffused silicon diode.

Fig. 136a shows the distribution of the charge carriers. During the diffusion process the borium atoms gradually penetrate into the N-crystal, so their concentration is densest at the surface to reduce rather quickly deeper in the crystal until past point J the negative charge carriers are in the majority. If a strong current is

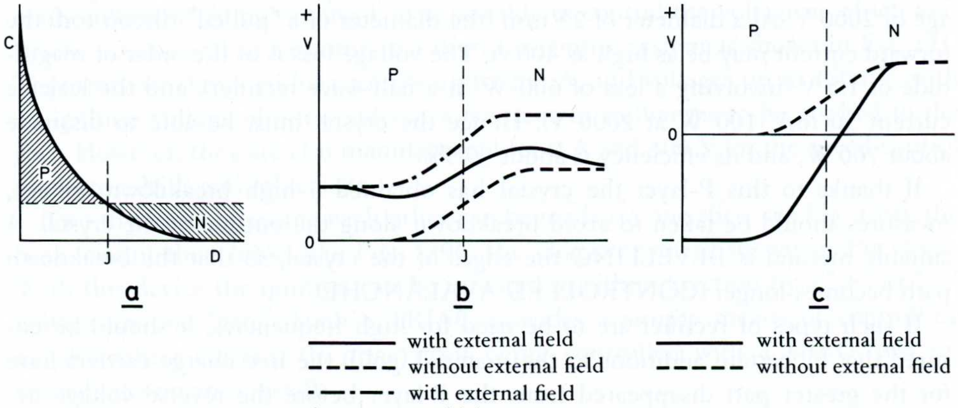


Fig. 136. Two-layer diode at a high forward current and a high reverse voltage: a. concentration of charge carriers at diffusion J = junction, b. potential distribution during the forward phase, c. potential distribution during the reverse phase.

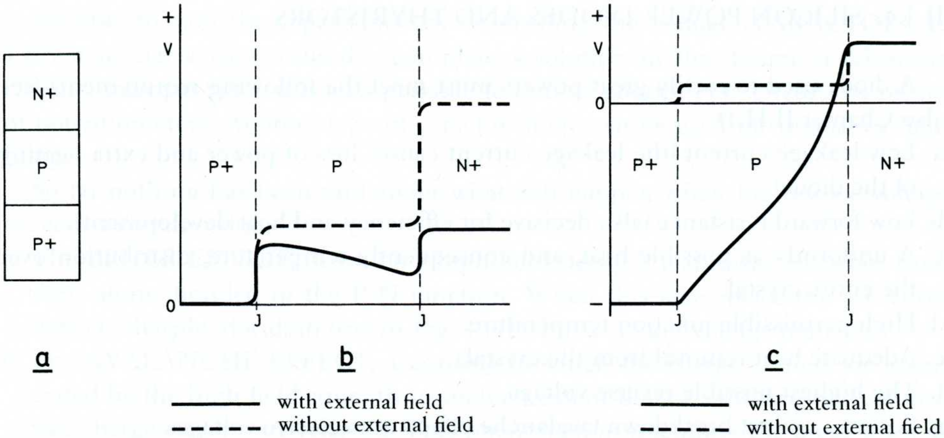


Fig. 137. Three-layer diode: a. principle, b. potential distribution during the forward phase, c. potential distribution during the reverse phase.

made to flow in the forward direction, the current of positive charge carriers, emerging from the surface in large numbers, will lower the threshold voltage in the junction (Fig. 136). In the reverse direction the external voltage will increase the threshold voltage (Fig. 136c) until the avalanche effect occurs. To obtain a low forward resistance a lot of dope must be added to the material, thus increasing the risk of breakdown in the reverse direction.

The permissible reverse voltage can be increased by applying a slightly doped P-layer between the P- and N-layer (indicated as P+ and N+) as shown in Fig. 137a. During the forward half cycle the voltage applied will cause the holes and electrons to travel from the left and right respectively towards the P-zone, thus ensuring good conductivity there (Fig. 137b). During the reverse half cycle the holes will flow to the left, and the electrons to the right, thus causing a high resistance in the P-layer, which contains but few charge carriers (Fig. 137c). As a result the reverse voltage can become rather high before avalanche sets in. It has been found possible to make such TRIPLE DIFFUSED diodes capable of withstanding a voltage of 2000 V. At a diameter of 25 mm (the diameter of a "pulled" silicon rod) the forward current may be as high as 400 A. The voltage loss is of the order of magnitude of 1.5 V (involving a loss of 600 W in a half-wave rectifier), and the leakage current 50 mA (100 W at 2000 V). Hence the crystal must be able to dissipate about 700 W, and its efficiency is about 99.9%.

If thanks to this P-layer the crystal has obtained a high breakdown voltage, measures should be taken to avoid breakdown along the outside of the crystal. A suitable method is BEVELLING the edges of the crystal, so that the breakdown path becomes longer (CONTROLLED AVALANCHE).

If such types of rectifier are to be used for high frequencies, it should be ensured that by certain additions to the material (gold) the free charge carriers have for the greater part disappeared from the P-layer before the reverse voltage occurs (FAST RECOVERY). If the resistance in the forward direction need not meet such severe requirements, the P-layer can be doped to an even lesser extent, and reverse voltages of some thousands of volts are obtained. By stacking such

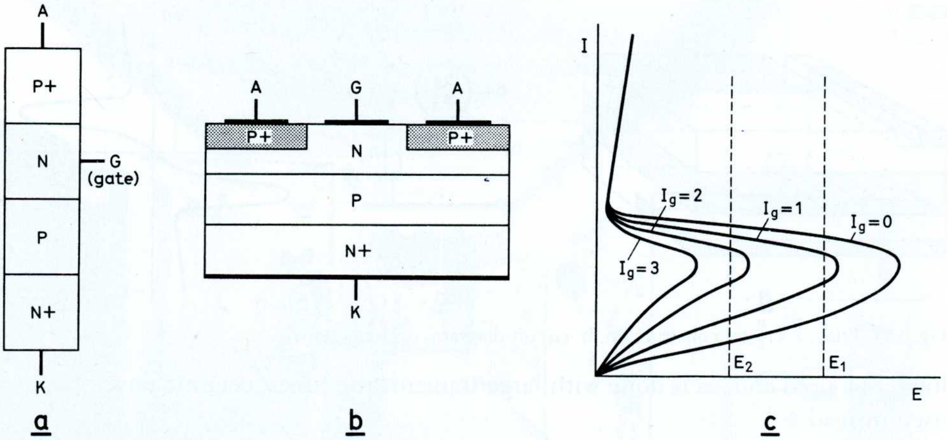


Fig. 138. Thyristor: a. principle, b. cross section, c. characteristic.

diodes it is possible to arrive at very high voltages. Rectifiers up to 30 kV are made for television sets, and up to 150 kV for X-ray equipment.

If the semiconductor technique is to be used for designing a substitute for the thyratron, it is particularly the avalanche effect that must be turned to use. As already explained before, this effect can be compared with an arc discharge, and this discharge is characteristic of the thyratron function. An auxiliary voltage carrying a small power serves to ignite the arc, which extinguishes only when the voltage applied passes zero. This phenomenon is reproduced by a THYRISTOR as shown in Fig. 138a. It can be regarded as consisting of two transistors PNP and NPN, connected partly in parallel and partly in series with a terminal G(ate) in one of the central layers. In the reverse half cycle (when P receives a negative voltage), this combination shows a current versus voltage curve as drawn in Fig. 138c. We recognize the characteristic of the arc discharge: if a certain current value is exceeded the resistance drops, and the current can assume a very high value. The voltage at which this instability occurs can be controlled by varying the "gate" current $I_g$. At $I_g = 1$, a voltage E1 is needed to cause breakdown; at $I_g = 2$ a voltage E2 is already sufficient. So by varying $I_g$ it is possible to control the voltage at which avalanche occurs, and so we come to a thyristor control system as shown in Fig. 121. Thyristors are designed for currents up to 200 A, and voltages up to 1600 V, with an ignition voltage of a few volts at a few tens of milliamps to be applied to the gate. However, they are also manufactured for 2 A and 400 V for the speed control of electric drills and mixers etc.

Two thyristors opposing each other can be made on one slice, see Fig. 139a; the basic circuit then looks like Fig. 139b, the characteristic is shown in Fig. 139c. With this device the ignition can be initiated in either direction by a "+" or "—" voltage on one "gate". Such a TRIAC provides a simple means of controlling an alternating current, the triac itself being controlled with a DIAC (two inverse diodes on one crystal).

For power diodes and thyristors it is essential that there be adequate heat removal, at the same time allowing the highest possible crystal temperature (190 °C). At this temperature, also in view of mechanical stress, glass insulation can no

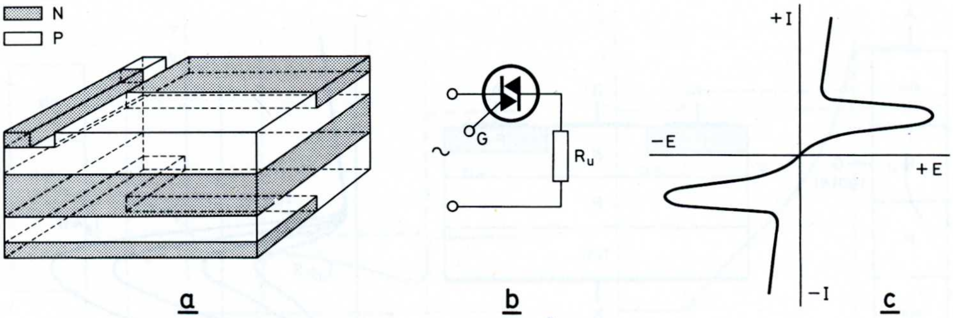


Fig. 139. Triac: a. crystal construction, b. circuit diagram, c. characteristic.

longer be used and, as is done with large transmitting tubes, ceramic envelopes are used instead.

Fig. 140a shows the envelope of the largest diode (BYX33), Fig. 140b shows the largest thyristor (BTX38). The thermal resistance of BXY33 is 0.2 °C/W, which means that as much as 700 W can be removed.

For smaller powers the diodes and thyristors are pressed in plastic, the crystal being soldered to a plate which serves at the same time as the heatsink. Fig. 140c gives a cross section of the BT100.

The thyristors and triacs have opened new ways towards all kinds of power control. High power electric motors, furnaces etc. can now be controlled without loss of power, and with a minimum of control energy. The smaller and cheaper ver-

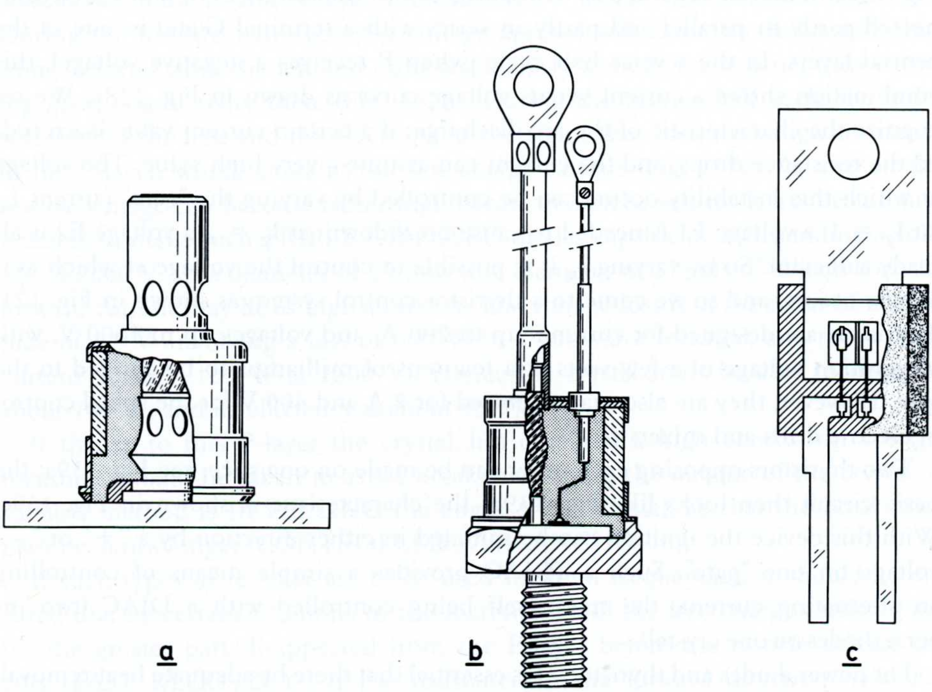
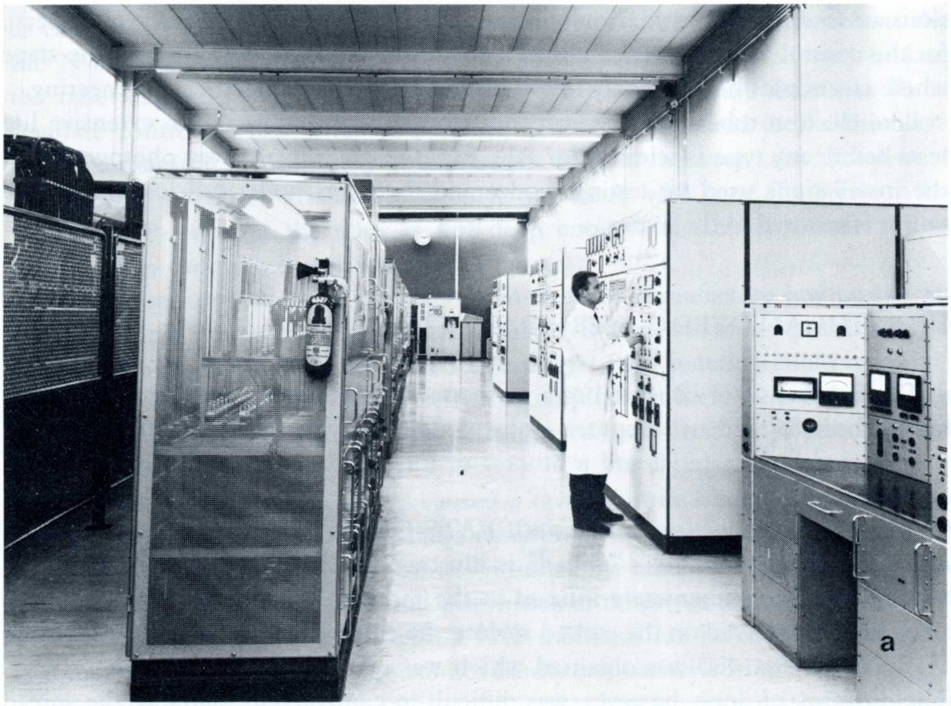


Fig. 140. Envelopes for power diodes and thyristors: a. power diode BYX33 (1600 V, 400 A), b. thyristor BTX38 (800 V, 70 A), c. thyristor BT100 in plastic (500 V, 2 A).



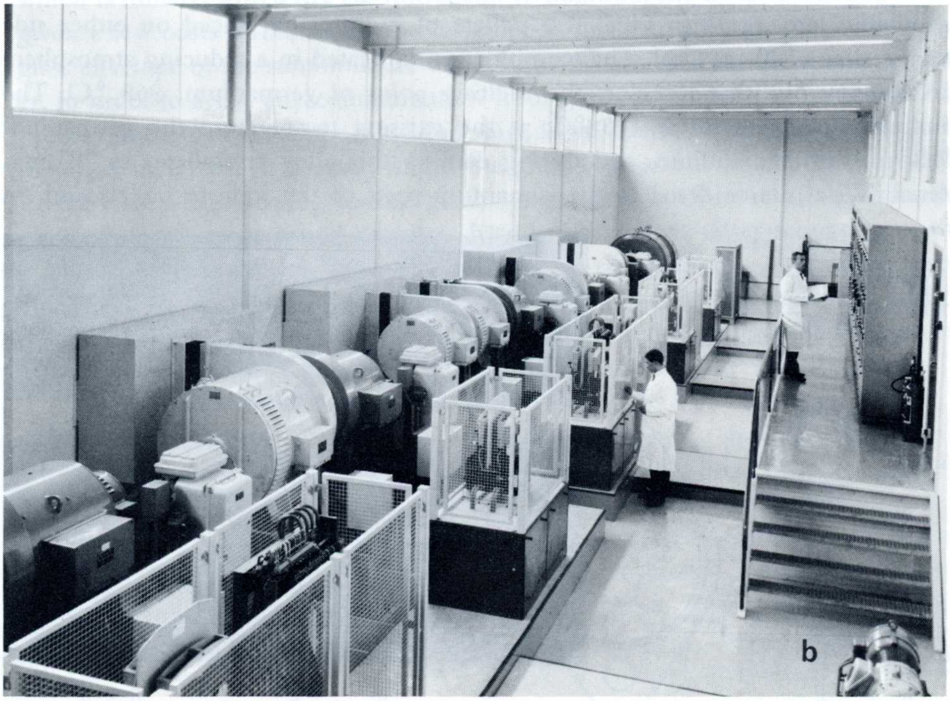


Fig. 141. Life testing installation for power diodes and thyristors: a. switching device and life testing racks, b. machines for the current supply.

sions are finding application in domestic appliances, in motor cars (screen wipers), for the control of fluorescent lighting and so on. We have now reached the stage where electronics has definitely found its way into general electrical engineering.

Like electron tubes, power semiconductors must be subjected to extensive life tests before any type is released for sale. Fig. 141 gives two survey photographs of the installations used for testing diodes and thyristors under full load whilst the power is restored to the mains.

## II.J.5. THE ALLOYED (LAYER) TRANSISTOR

The first transistor consisted of a germanium crystal (base) with two point contacts. These formed two opposed point contact diodes, see Fig. 130b, and by a proper choice of material and a process of forming, it proved possible to obtain some amplification.

Large scale production started with RATE GROWN transistors. According to this system germanium was "pulled" as illustrated in Fig. 131b, donor and acceptor material being alternately infused in the molten germanium. The P-N structures can be observed on the etched surface. By cutting the pulled rod through the P and N parts, a disc was obtained which was cut into rods forming transistors. The dosaging of dope, however, was difficult to control, and therefore the spread in characteristics considerable.

Better controllable electrical characteristics were obtained with the ALLOY method. Here, see Fig. 142a, two pellets of indium are placed on either side of the slice of N-material. This combination is heated in a reducing atmosphere to 500-650 °C (that is below the melting point of germanium: 968 °C). The indium already melts at 155 °C; at the existing temperature the germanium "alloys" with the indium, and during cooling this alloy crystallizes as "P" material. Wires are soldered to the remaining parts of the indium pellets and on

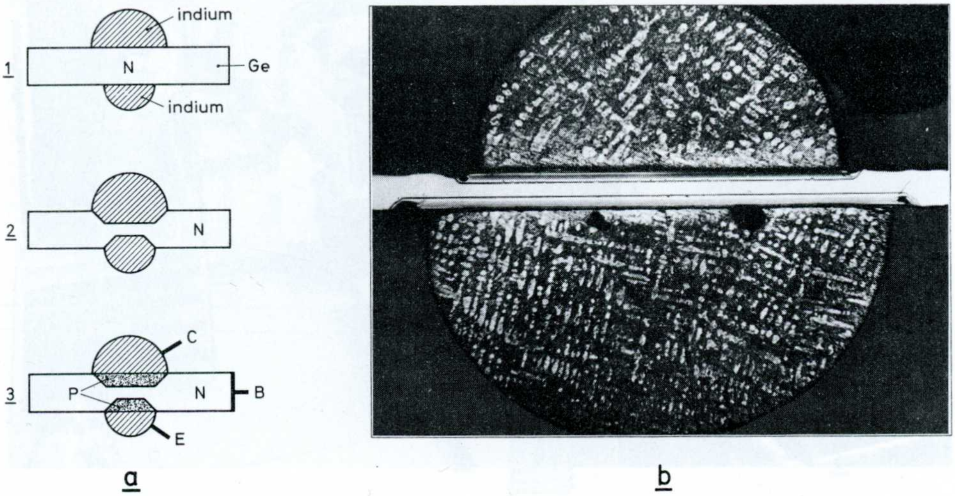


Fig. 142. Alloy-diffused transistor: a. the three stages of the alloy-diffusion process, b. cross section of a beautiful alloy-diffusion obtained with the U.L.T. method.

to the crystal, and thus a PNP transistor is obtained. The difficulty now is that the N-material with its few donors has a high specific resistance, and to keep the base resistance low, this layer must be extremely thin (20 microns). By an accurate control of time and temperature it must be ensured that a thin layer of N-germanium of uniform thickness is formed between the P-layers. If the alloying effect penetrates too deeply, the result will be a short circuit between the P-layers; if the alloy is not deep enough, the characteristic will be unfavourable owing to excessive base resistance.

Industry soon succeeded in using this method to manufacture low power amplifiers. By using a better alloying technique (U-L-T) it was found possible to obtain a uniform junction (Fig. 142b). As a result the PNP junction could be given a larger area so that greater powers could be amplified which were fed to a large loudspeaker. The characteristic could be improved considerably by adding gallium to the indium.

After low power transistors had caused a revolution in hearing aids, it became possible to build battery operated radio receivers which used a voltage of 6 to 12 V (without heater) and conquered the world under the name of TRANSISTOR RADIOS. By starting with a P-material into which antimony is diffused we obtain an NPN transistor which, together with a PNP transistor can be used to build a very simple "push-pull circuit" that consumes hardly any battery current.

The outer surface of the PN junction of germanium is very sensitive to contamination, and therefore the crystal must be encapsulated in an airtight envelope which is inflated with a dry gas (nitrogen). In addition, the transistor is sometimes given a few coats of laquer, or the envelope is filled up with grease. Initially the glass envelope of the subminiature tube (Fig. 42h) was used for the transistor. Later, in order to avoid any contamination from gases developed during the sealing process, the SOT1 envelope was used (Fig. 143a). Here the envelope is made of metal. A metal rim is pressed round the glass base and is known as the COMPRESSION SEAL (owing to the greater expansion coefficient of the iron the rim will shrink tight round the glass base when cooled). After mounting of the crystal the envelope is pressed round this iron rim with the addition of a soft metal as the bonding medium. Better results are obtained, however, with a MATCHED SEAL lead-in (metal and glass having the same coefficient of ex-

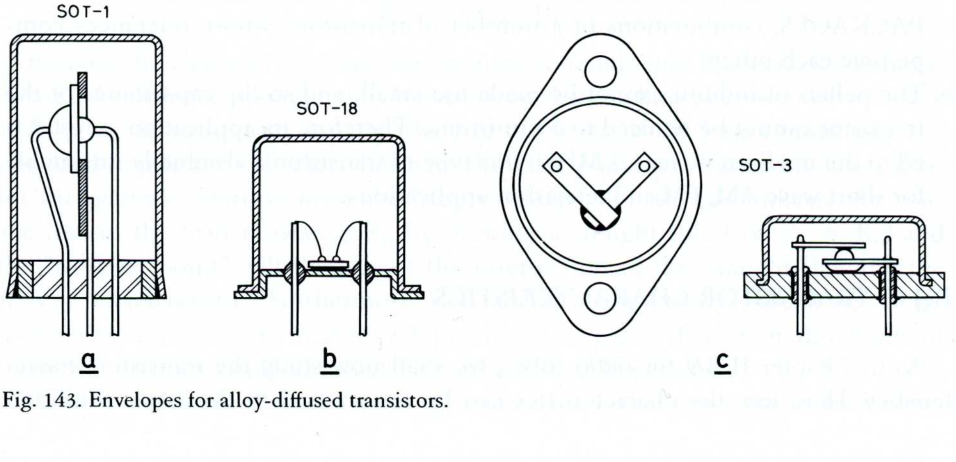


Fig. 143. Envelopes for alloy-diffused transistors.

pansion) which, as in transmitting tubes, is obtained from a combination of FERNICO and matched glass. The base consists of a fernico plate with glass-insulated lead-ins; by means of a short current pulse the rim of the envelope is welded to the edge of the metal base plate (Fig. 143b). The fernico can be tinned or gold plated to ensure a clean surface and better current conduction. In Fig. 143a the base on a metal plate is soldered to one of the lead-ins; the thin wires sealed to the indium contacts are welded to the other two lead-ins. Heat removal is, of course, very low here. In this respect it is preferable to solder the collector to the metal base of the SOT18 envelope. For greater powers, as handled by output transistors, the can should be given a special shape to ensure satisfactory heat removal. Fig. 143c shows the principle of what is known as the SOT3 envelope. Here, after alloying, the collector pellet is cut flat and soldered to a solid copper base which can transfer the heat to the chassis (if required, with an insulating thin layer of mica between) or to a radiator. Thus a transitional heat resistance of 2 °C/W can be obtained between the junction and the heatsink, so that at a temperature difference of 50 °C 25 W can be removed. The metal can is pressed tight to the base (the copper is loaded beyond the yield point) or welded (copper base with iron rim).

Alloy transistors are still used in the low frequency part of battery radio sets (see the type AC 127 and AD 149 in Table III). At the low voltage available here, they have indeed even better characteristics than the silicon transistor which will be discussed later. However, these germanium alloy transistors also entail a number of drawbacks:

a. The PNP junction is very sensitive to contamination, so that notwithstanding the precautions mentioned above, reliability can be endangered.

b. The maximum permissible temperature is rather low (75 °C), which sets a limit for power and ambient temperature.

c. The permissible voltage is relatively low (maximum 50 V), so that equipment fed direct from the mains must be fitted out with a supply transformer.

d. The process, in which success depends to a certain extent on the dispersion of the molten indium, causes a rather considerable spread in characteristic which is worstened by temperature differences. By artificial means in the circuit, such as negative feedback, pre-set potentiometers and temperature dependent resistors, this difficulty can be more or less dealt with. Furthermore the market offers PACKAGES, combinations of a number of transistors, whose tolerances compensate each other.

e. The pellets of indium cannot be made too small, and so the capacitance of the transistor cannot be reduced to a minimum. Therefore its application is restricted to the medium wave (1.5 MHz); this type of transistor is absolutely unsuitable for short wave AM, FM and television applications.

## II.J.6. TRANSISTOR CHARACTERISTICS

As in Chapter II.B.9 for radio tubes, we shall now study the transistor characteristics. Here too, the characteristics can be determined in the static condition

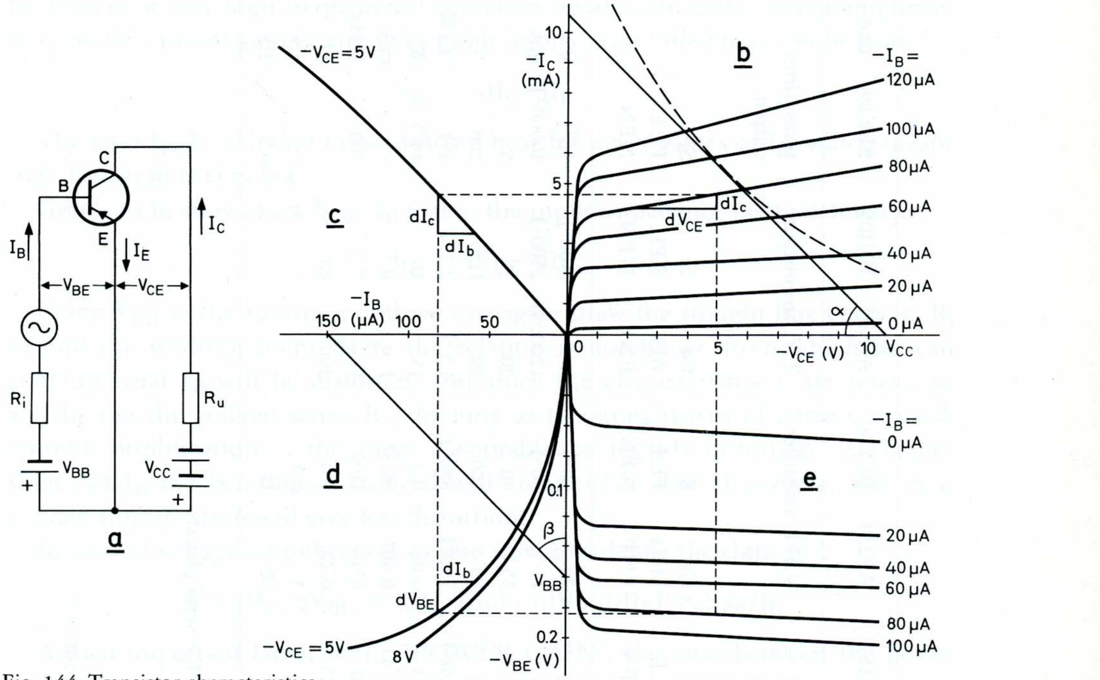


Fig. 144. Transistor characteristics.

(with a direct current). This is done with the circuit shown in Fig. 144a in which a certain voltage (or current) is varied, two others are kept constant, whilst the variation of a fourth voltage (or current) is measured. There is, however, always the danger of the transistor soon being overloaded when the breakdown voltage or the maximum permissible temperature is exceeded. With modern measuring equipment the characteristics can be measured "dynamically" (with alternating voltages), which has the advantage that, owing to the short duration of the measuring process, the transistor cannot be damaged. The complete set of characteristics can be displayed on the oscilloscope.

Fig. 144b shows $I_C$ as a function of $V_{CE}$ at different constant values of $I_B$. The output resistance $R_o$ (called the "internal resistance" with radio tubes) can be read from the triangle drawn in the figure:

$$R_o = dV_{CE}/dI_C$$

Because the characteristics fan out, we may conclude that $R_o$ decreases as $I_B$ increases. As with the pentode, the characteristic shows a knee; below a certain value of $V_{CE}$, $I_C$ decreases very rapidly (the free electrons in the base cannot take the potential step to the collector), but here the knee point is considerably lower than for the pentode. Starting from the battery voltage $V_{CC}$ we can indicate the voltage loss in the load resistance $R_u$ by drawing a straight line ($\cot \alpha' = R_u$) and the "working point" will then lie at the intersection of this straight line and the $I_C$-$V_{CE}$ characteristic. Furthermore $V_{CC}$ and $R_u$ respectively should be so selected that no excessive heat development takes place. This heat development limits the product $V_{CE}I_C$ as indicated by the dashed hyperbola.

Curve 144c shows the ratio $I_C$ to $I_B$, a most important transistor characteristic,

TABLE III. Main data on some typical transistors

| transistor type | | AC127 | BC109 | AD149 | BD124 | AF121 | BF167 | AF239 | BFY90 | AU104 | BU105 | BLY89 | unit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| function | | a.f. preamplifier | | a.f. output | | i.f. amplifier (F.M.-T.V.) | | V.H.F. amplifier | | T.V. line output | | transmitter output | |
| material | | Ge | Si | Ge | Si | Ge | Si | Ge | Si | Ge | Si | Si | |
| technique[1] | | A | PL.E. | A | PL.E. | P.O.B. | PL.E. | Mesa | PL.E. | AD | S.PL. | PL.E. | |
| polarity | | NPN | NPN | PNP | NPN | PNP | NPN | PNP | NPN | PNP | NPN | NPN | |
| max. coll. voltage | $V_{CBO}$ | 32 | 30 | 50 | 70 | 25 | 40 | 20 | 30[2] | 185 | 1500[2] | 36[2] | V |
| max. coll. current | $I_C$ | 500 | 200[2] | 3500 | 4000[2] | 15 | 25 | 15 | 50[2] | 12000[2] | 2500[2] | 10000[2] | mA |
| current gain | $h_{FE}$ | 100 | 400 | 50 | 50 | 75 | 60 | 35 | 50 | 20 | 3 | 5 | |
| max. junction temp. | $T_j$ | 100 | 175 | 100 | 175 | 90 | 175 | 90 | 200 | 90 | 115 | 200 | °C |
| thermal resistance | $R_{th}$ | 370[4] | 500[4] | 2[6] | 7.5[6] | 450[4] | 1000[4] | 750[4] | 880[4] | 1.5[6] | 2.5[4] | 4[6] | °C/W |
| power dissipation | $P_{tot}$ | 0.34 | 0.3 | 32.5 | 15 | 0.14 | 0.13 | 0.06 | 0.2 | 15 | 10 | 44[3] | W |
| coll. leak current | $I_{CBO}$ | 0.01 | $\ll$[7] | 3 | 0.001 | 0.001 | $\ll$[7] | 0.001 | $\ll$[7] | 1 | 0.1 | - | mA |
| trans. frequency[5] | $f_t$ | 2.5 | 300 | 0.5 | 120 | 270 | 350 | 650 | 1100 | 15 | 7.5 | 700 | MHz |
| reactive cap. | $C_{re}$ | 70 | 2.5 | 220 | 55 | 0.45 | 0.25 | 0.23 | 0.6 | 300 | 65 | - | pF |
| noise | F | 4 | 1.8 | - | - | 4.5 | 3 | 5 | 5.5 | - | - | - | dB |
| power gain | $G_p$ | - | - | - | - | 19 | 26 | 12 | 11 | - | - | 6.2 | dB |

[1] A = alloy; AD = alloy diffused; PL.E. = planar epitaxial; P.O.B. = pushed out base; S.PL. = semi planar

[2] peak value

[3] V.H.F. output power is 25 W at $V_{CE} = 13.5$ V and 175 MHz

[4] from junction to ambient

[5] frequency at which $h_{FE}$ drops to the value at low frequency divided by the square root of 2

[6] from junction to heatsink

[7] current is too low to be measured

for (except at very high frequencies) operation usually concerns current amplification. So the current gain ($\alpha'$ in older publications, now called $h_{FE}$) can be read:

$$h_{FE} = dI_C/dI_B$$

The ratio $I_B/I_C$ is found to be almost linear for lower values of $I_B$, which means that distortion is very low.

Graph 144d shows how $V_{BE}/I_B$ varies, the input impedance being defined as:

$$h_i = dV_{BE}/dI_B$$

Using $V_{BB}$ as the starting point we can again draw the straight line cot $\beta = R_i$ to find the working point. Here the relation is not linear, from which we can conclude that $I_B$ will be distorted, and since the characteristics c are linear, so will $I_C$ and the voltage across $R_u$. As long as the straight part of curve c is used, current amplification is the more favourable as regards distortion. At higher values of $I_B$ the bending of curve d will compensate that of curve c, and then voltage amplification will give less distortion.

In view of voltage amplification we can now also define the slope S.

$$S = dI_C/dV_{BE} = (dI_C/dI_B) \;/\; (dV_{BE}/dI_B) = h_{FE}/h_i$$

A final important factor is the "POWER GAIN", the ratio between the power supplied by the alternating current source and the output power $I_C^2R_u$.

For high frequencies it is essential to keep the input capacitance small, and for this application it is recommendable that the area of the junction be made as small as possible, and its thickness as great as conditions will allow.

Without going into details it will be clear that the characteristics of the transistor are dependent on the choice of N-material, the concentration of the alloyed acceptor, the area and thickness of the junction. The most effective use for a certain application is obtained by choosing the correct circuit design and a proper "matching" of the input and output circuits. We shall confine ourselves to referring to results given in Table III, in which two types of alloy transistor occur.

Type AC127 is a preamplifier with high current gain (up to 2.5 MHz).

Type AD149 is an output amplifier with a high maximum collector current and, thanks to the SOT3 envelope, a maximum permissible dissipation of 32.5 W. This requires a junction of a rather large area. The resulting leakage current of 3 mA at room temperature has no effect on the load applied. The large internal capacitance and the low cut-off frequency do not impede satisfactory functioning at low frequencies.

## II.J.7. GERMANIUM TRANSISTORS FOR HIGHER FREQUENCIES

Although the manufacturing process of alloy transistors has been improved in many respects, the alloying process remains an art difficult to master. It will be clear that if the dope diffuses gradually into the surface of a crystal from the surrounding gas, a better dosaging and, consequently, a layer of more uniform thickness is obtained, which is of particular importance for the base-collector junction.

Furthermore the active part of the transistor area can be made smaller by employing the "masking" method. It should be noted that diffusion gives no clear-cut junction between a constant concentration of N-material and P-material, but a gradually decreasing N-concentration. The result is that a field, the so-called "DRIFT FIELD" is incorporated in the base (see also Fig. 136). This drift field accelerates the flow of charge carriers towards the collector. Thanks to their higher speed higher frequencies can be used. Owing to the thicker junction, the breakdown voltages increase and the capacitance decreases.

An early type of DIFFUSED BASE or ALLOY DIFFUSED transistor (originally known as DRIFT transistor) is shown in Fig. 145a. Diffusion of gaseous arsenic causes an N-layer to be formed in the P-crystal. The bottom of this crystal is ground down, so that the collector can be connected to the original P-material. A piece of aluminium is placed on the diffused N-layer, and when heated it "alloys" with this N-layer, thus providing the P-N junction, the emitter being connected to the aluminium. Now the N-surface is provided with a gold film to which the base can be connected.

A simpler type to manufacture is the P(ushed) O(ut) B(ase) transistor as introduced by Philips in about 1959. In accordance with Fig. 145b two pellets of lead, mixed with the donor antimony, are placed on a crystal of P-germanium. These pellets have a diameter of 150 microns and are positioned with a distance of 50 microns between them. Such pellets are easily made by leading a jet of molten lead through a capillary into a liquid. A small quantity of aluminium is applied to the pellet on the right. It is typical of the process that diffusion and alloying take place in one operation at a temperature of about 780 °C. First the antimony diffuses into the germanium to form a N-layer about three microns thick. After cooling, some aluminium from the pellet on the right will be absorbed in the crystal lattice, thus forming the P-N junction whilst the pellet serves for the emitter contact. The left pellet seals to the N-germanium to form the base. Finally the N-layer diffused round the pellets is sprayed with lacquer and the N-germanium outside the junction is etched away.

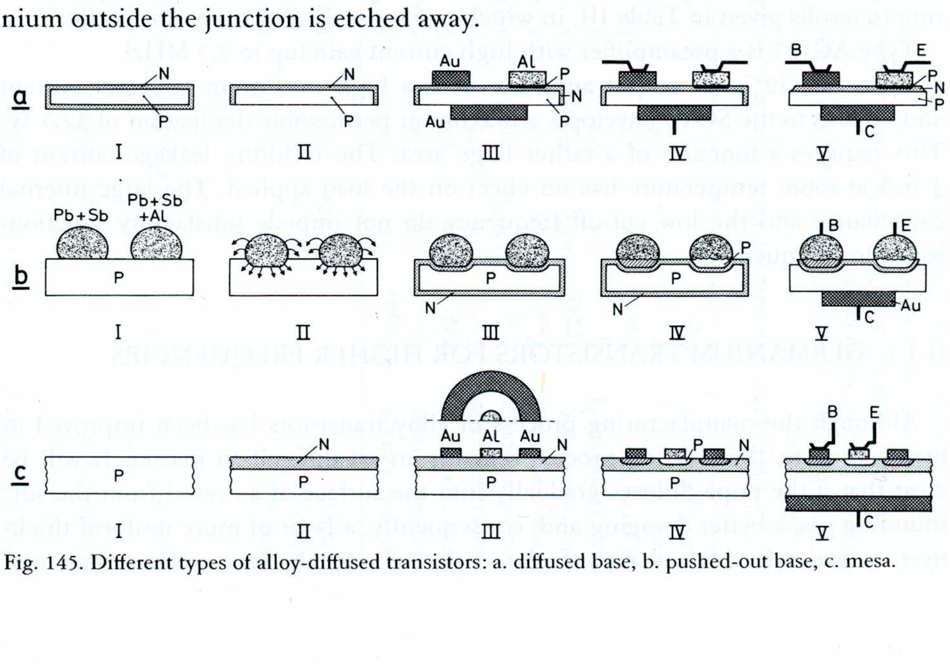


Fig. 145. Different types of alloy-diffused transistors: a. diffused base, b. pushed-out base, c. mesa.

Soldering the terminal wires to pellets of 150 microns is, of course, no simple matter, and the alloy process seems difficult to control. Yet this method has been found suitable for the economical manufacture of high-frequency transistors. The P.O.B. transistor has opened the way towards transistorization in the field of shortwaves, F.M., and even television. Table III (Chapter II.J.6) contains the type AF 121 with a transition frequency of 270 MHz and a power gain of 19. Later on, however, the Philips type for U.H.F. television was given the MESA form.

The manufacturing stages of the so-called MESA transistor are shown in Fig. 145c (the name MESA refers to the table model). In principle the structure is equal to that of the drift transistor (Fig. 145a), the only difference being that now several transistors are formed on one slice of P-material with a diameter ranging from 25 to 50 mm.

First an N-layer is diffused into the entire surface. Then the gold and aluminium plaques are vapour deposited through apertures in a mask.

On the one hand the P-material to be alloyed later can be properly dosaged so that a thin emitter can be obtained. Furthermore the area can be accurately determined by means of the mask, and finally the form of emitter and base can be chosen to yield favourable characteristics (the annular base configuration of Fig. 145c III offers a low base resistance). When stage IV is completed, the separate transistors are cut out and further finished.

In Table III, under AF239, we find a MESA transistor with a transition frequency of 650 MHz and a reactive capacitance of 0.23 pF, which has served excellently as an amplifier in U.H.F. television, producing far less noise than tubes. And yet, in this field, more than in the low frequency field, the germanium transistors have been replaced by silicon transistors. To begin with, silicon allows higher temperatures and is less sensitive to humidity and vapours. Furthermore the photographic process to be described in the next chapter, enables us to obtain even smaller and more accurate junction areas.

## II.J.8. PLANAR SILICON TRANSISTORS

As already explained in the foregoing, the advantages of silicon over germanium are: a higher temperature resistance, lower leakage current and higher breakdown voltage. There is, however, one more property of silicon that is of prime importance in transistor engineering, being that SILICON OXIDE (found in nature as "quartz") forms a hard layer impervious to humidity. By means of a simple oxidation process the silicon crystal can be coated with such layer which serves to protect the crystal (and particularly the highly sensitive junction) against moisture and contamination. The silicon oxide can easily be removed locally since it is soluble in hydrofluoric acid in which the pure silicon will not dissolve.

Another material playing an important part in the manufacture of transistors is "PHOTORESIST", a film of lacquer that polymerizes when exposed to ultraviolet light. When the silicon diode, coated with a photoresist, has been exposed through a photographic mask, the parts actually exposed will harden: they remain behind on the silicon as a protective layer whilst the unexposed parts can be

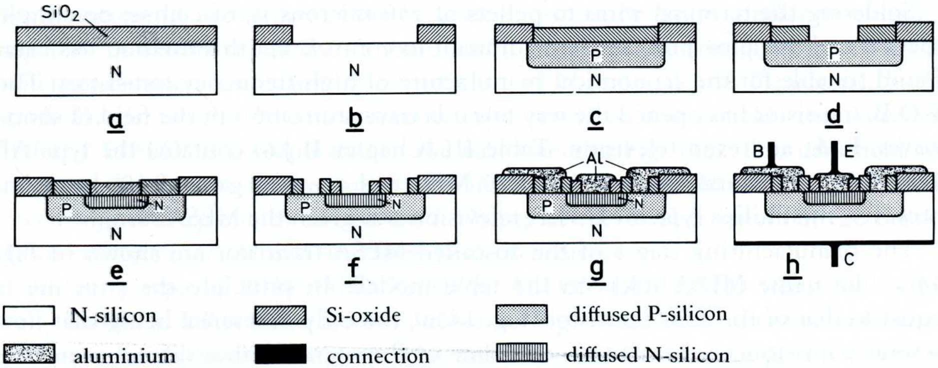


Fig. 146. Sequence of operations in making a planar transistor.

rinsed off. In a subsequent etching process the silicon oxide is then etched away in places where the photo mask was black. It is important that silicon oxide and photoresist have no grain and that films used nowadays are extremely fine grained. Thus it is possible to reproduce very small planes and small figures as holes in the silicon oxide, so that far more accurate images are obtained than would ever be possible by using the mechanical masks as for germanium mesa transistors.

It stands to reason that the manufacturer of photo cameras, FAIRCHILD, was the first to conceive the idea of employing this PLANAR technique. By means of Fig. 145 we shall now explain the manufacture of planar transistors (the name "planar" originates from the fact that all components are located in one plane).

The basis is the silicon N monocrystal. By heating to 1000 °C in steam, a layer of $SiO_2$ about 1 micron thick is applied to this crystal (a). This layer in turn is given a coat of photoresist lacquer which is then exposed to U.V. light through a photographic plate on which the areas to be etched are black. The exposed parts of the lacquer harden, and the unexposed parts can be rinsed off. This done, the silicon oxide of the unexposed areas (b) is etched away with fluoric acid. Then the crystal is placed in a furnace to be diffused with boron at about 1200 °C. Thus P-silicon is produced in the opening behind the window in the $SiO_2$ (c). It is important that during this diffusion the boron penetrates slightly under the edge of the $SiO_2$. The PN-junction, across which there will occur a considerable voltage gradient, and on which the transistor characteristic depends, is then protected by the $SiO_2$-layer against humidity and contamination. After diffusion the P-layer is again coated with $SiO_2$ in the same furnace to provide protection against direct contamination. Then another coat of photoresist is applied and exposed through a mask containing a smaller black area. After rinsing off the unexposed photoresist (d), the $SiO_2$ is etched away again. The crystal is now heated in a furnace in a phosphor atmosphere. As a result a layer of N-silicon is produced behind this window, and thus the NPN structure is obtained which will form the ultimate transistor. In terms of order of magnitude the original P-layer will be four microns thick, and if the N-layer becomes two microns, the remaining P-layer (the base) will be two microns. Thanks to the slowly developing diffusion process the penetration depth and degree of doping can be accurately controlled. According to the latest techniques used, the width of the various areas can be three microns.

The N diffusion too, is immediately sealed off with $SiO_2$ (e), and now a third photoresist treatment is to follow, as a result of which the windows for the terminal contacts are etched away (Fig. 146f). In the openings thus made in the $SiO_2$, aluminium is deposited for the emitter and base connections. The aluminium film can be made to continue over the $SiO_2$-layer to provide sufficient space to fasten the terminal wire. The bottom of the crystal is given a metal coating which forms the connection for the collector terminal (h). It goes without saying that handling planes of a few microns requires an atmosphere absolutely free from dust. Furthermore the location of the successive masks must be accurate to a high degree.

As we already know from Chapter II.J.2, the starting point is often a crystal already provided with an epitaxial layer.

The manufacturing method described here, with its many successive operations, is of course expensive. These operations are, however, carried out on a SLICE of "pulled" silicon with a diameter of 25-50 mm. If one transistor geometry can be placed on a crystal of 1 × 1 mm, one such slice yields 500-2000 transistors.

Photographic reproduction must meet very stringent requirements. First drawings, magnified 100 times, are made of the areas to be etched away. Of these drawings, side by side and on top of each other, multiple small-scale photographs are then made on a fine-grained film which has the dimensions of the slice (Fig. 33a). Markings both on the film and the crystal ensure accurate positioning, so that the successive negatives coincide as required.

The round 1000 transistors as shown in Fig. 145g can easily be subjected to a preliminary measurement once the bottoms of the crystals have been provided with a layer of metal. An ingenious machine "feels" all transistor elements in turn, measures the most important characteristic values, and marks the faulty elements.

In a following operation a grid pattern is scribed between the transistor elements; now the "slice" can be broken up into 1000 separate transistor crystals.

The next operation is soldering the crystal (the collector) to a metal strip, and then the terminal wires must be welded to the aluminium planes. This operation is called BONDING. As yet no automatic method has been found. The bonder must place the point of contact under the cross lines of a microscope and in this spot a very thin wire is pressed against the contact plane where it is heated and welded. A new method used to ensure satisfactory welds is ultra sound vibration.

The silicon crystals are fitted in the SOT18 envelope (Fig. 143b), and for large powers in the SOT3 envelope (Fig. 143c). Since the Si junction is hardly sensitive to humidity, and the fact that it can be coated with $SiO_2$ has made it also possible to press these transistors into plastic, with the combined result that the size as well as the price of the transistor could be reduced. This had called for an extensive search for a kind of plastic that is elastic, insulates, can easily be pressed, and is moisture-proof. A change-over takes place from the originally used epoxy resin to silicons with their greater temperature resistance. The collector of the transistor crystal is then soldered to one terminal, the emitter and base being connected to two other terminals by means of the thin "bonded" wires. These three terminals must be held in position in a jig until they are

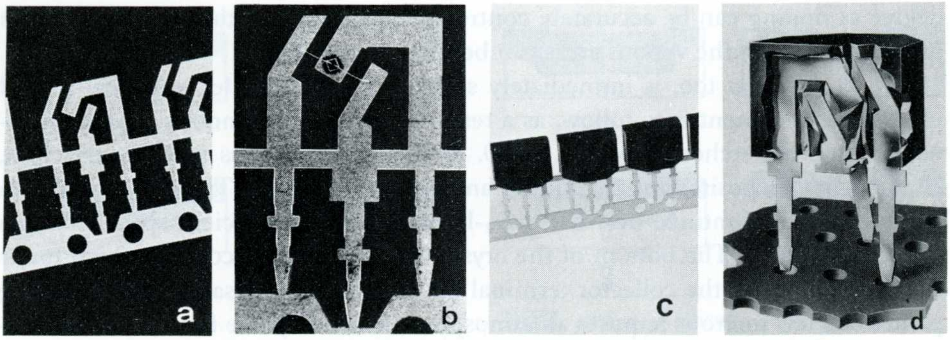


Fig. 147. Lock-Fit transistor (pressed into plastic): a. comb, b. comb with crystal and connecting wires, c. pressed, d. connecting pins bent.

fixed in the plastic during pressing.

A neat solution was found in the form of the Philips LOCK-FIT transistor (Fig. 147). Here the starting point is a comb of gold plated metal with adjacent groups of three terminal strips (Fig. 147a), which are easily clamped and fed through automatic machines. The crystal is soldered on one of the three strips and in a following operation the terminal wires for the emitter and base are bonded to the crystal and to the other two strips (b). It has been found possible to make four bonds at a time (MULTIWEDGE BONDING). Afterwards the terminal strips are placed under the plastic press and the transistors are pressed into plastic (c). Then the pieces linking the combs are cut away and the terminal strips are bent to form a triangle (d). Owing to their special form, these terminal strips are "self-locking" when inserted in the printed wiring board. Since the price of such transistors, when made in large numbers, comes down to $ 0.10, and one minute mounting time costs $ 0.05, the saving on mounting costs (by automatic mounting) is quite considerable as compared with the price of the transistor.

Table III (Chapter II.J.6) gives six important types of planar transistor; Fig. 148 shows configuration photographs of four types:

- the BC109; an L.F. amplifier with a very high current gain and very low noise figure, excellent for a magnetophone amplifier (the output power of the magnetophone is very low);
- the BD124; an output amplifier with characteristics not much better than those of the comparable germanium type, only the collector voltage is higher. There are already output amplifiers which can be connected direct to 110 V mains via a rectifier diode; as yet this has not been possible for 220 V mains;
- the BF167; an I.F. television amplifier of very low capacitance and high power gain. The low capacitance has been obtained by making a diode in parallel with the transistor on the same crystal as shown in Fig. 148a;
- the BFY90; a very high frequency amplifier with a low noise figure serving well in television aerial amplifiers (photograph Fig. 148b). Thanks to the zigzag pattern of the emitter and a large collector area it is possible to reach a high emitter current;
- the BU105; the high permissible collector voltage and heavy current makes this type suitable to be used as the line output amplifier in television, meaning

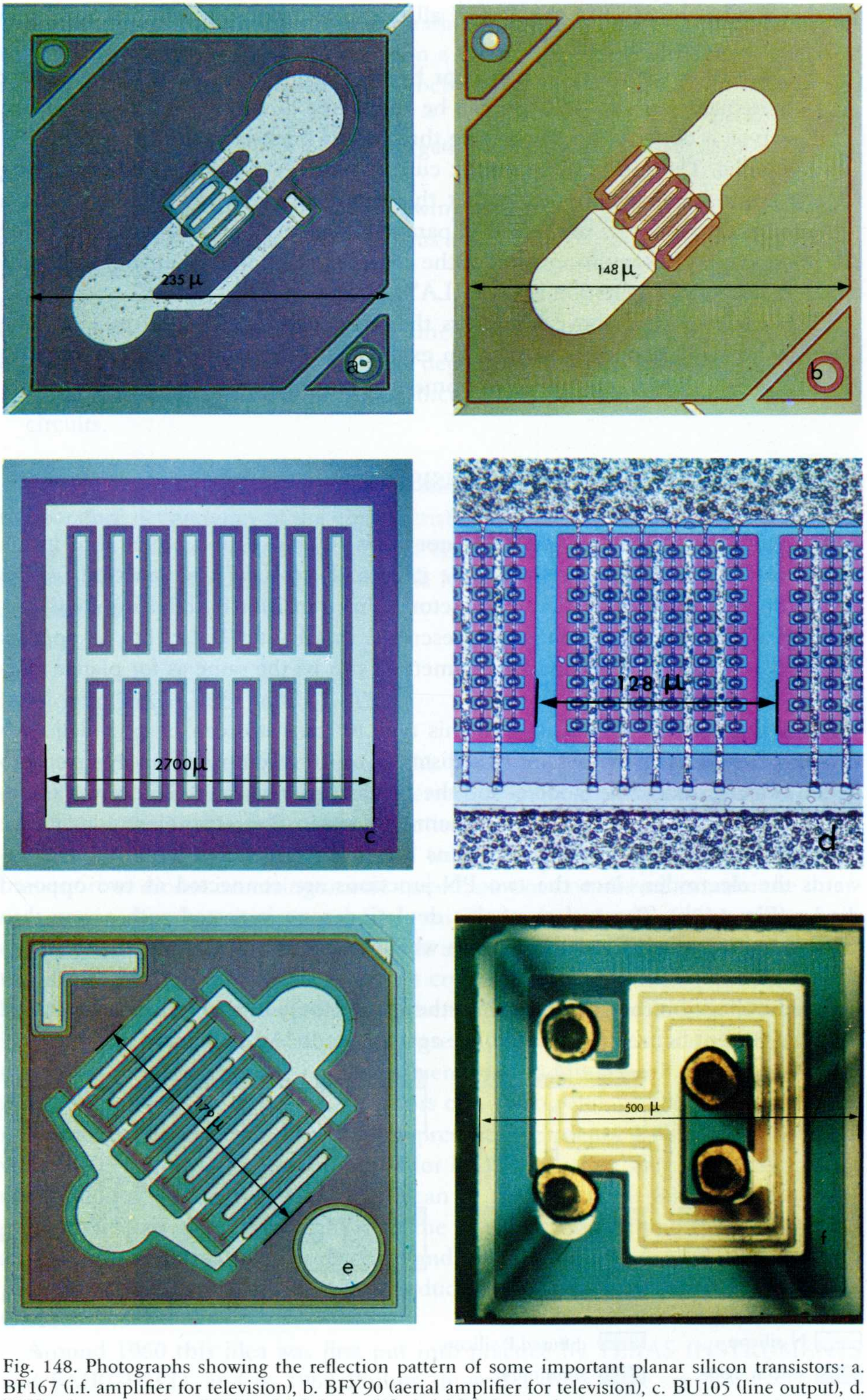


Fig. 148. Photographs showing the reflection pattern of some important planar silicon transistors: a. BF167 (i.f. amplifier for television), b. BFY90 (aerial amplifier for television), c. BU105 (line output), d. BLY89 (transmitter), e. BFW10 (M.O.S. amplifier), f. M.O.S. tetrode.

that the last obstacle in the way of all-transistor television sets has now been removed (photograph Fig. 148c);

- the BLY89; a transmitting transistor by means of which a 25 W transmitter for a frequency up to 175 MHz can be connected direct to a (12 V) car battery. This type is obtained by connecting three of the crystals as shown in Fig. 148d in parallel. The high collector peak current of 10 A is obtained by employing the emitter-grid technique, so that this transistor is in fact built up from a number of transistors connected in parallel by means of aluminium strips. The photograph gives an impression of the complex geometry. The American term for a similar configuration is OVERLAY.

From what was said above it appears that ways have been found to "drill" the electrons in semiconductors to such an extent that semiconductor elements perform their functions in electronic equipment as required.

## II.J.9. M.O.S. (FIELD EFFECT) TRANSISTORS

The patent on the field effect transistor dates from as early as 1928, but it was the silicon planar technique that made it possible to turn it to practical use by means of a M(etal) O(xide) S(emiconductor). This transistor functions quite differently than the three-layer transistor described in Chapter II.J.1, but as appears from Figs. 149 a-g, the manufacturing method can be the same as for planar transistors.

According to Fig. 149g and 149h this type of transistor can be provided with two P-diffused zones (with a mutual distance of 10 microns) in an N-crystal to which two electrodes: the S(ource) and the D(rain) are connected. D receives a negative voltage with respect to S. The channel between these two electrodes in the N-zone (DEPLETION ZONE) contains free electrons which cannot travel towards the electrodes since the two PN-junctions are connected as two opposed diodes (Fig. 149h). The surface of this depletion zone is coated with a very thin layer (0.1 micron) of silicon oxide onto which the control electrode, the G(ate), is deposited.

A low voltage on the gate influences the conductivity of the depletion zone and thus the current between S and D; once again there is "Amplification".

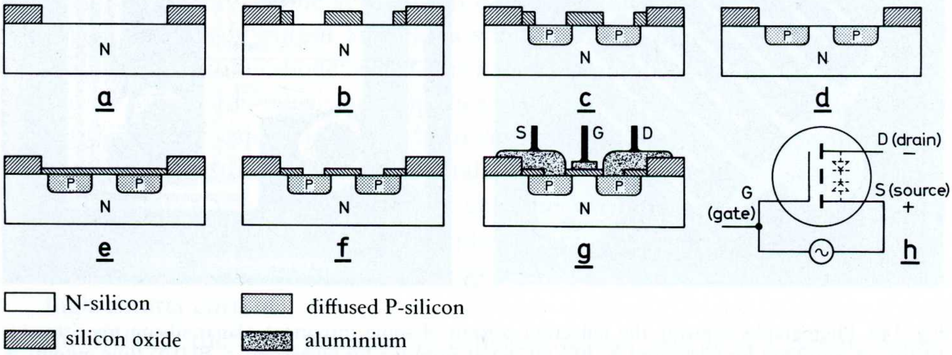


Fig. 149. Sequence of operations for making MOS-transistors: h. circuit diagram.

Without going any deeper into the rather complex phenomenon occurring in this transistor, we would like to mention a few noteworthy characteristic values of the BFW10 type: at a voltage of 15 V between D and S the cut-off voltage ($V_{GS}$ at $I_D = 0$) is 5 V; at $V_{GS} = 0$, $I_D = 12$ mA, and the slope between these points varies from 1 to 5 mA/V. The gate current is practically zero (capacitive current only).

This type of transistor offers the following advantages:

a. the gate carries hardly any current, thus a very small control power is needed;
b. the input impedance is very high;
c. the gate requires no bias voltage;
d. the transistor is symmetrical, so the functions of S and D are interchangeable;
e. owing to the small dimensions of the depletion zone one transistor can be made on an extremely small crystal area, which is of prime importance for integrated circuits.

Drawbacks are, however, the limited frequency (300 MHz) and the rather great temperature dependency of the characteristic.

Fig. 148e shows the configuration of type BFW10, discussed above. Fig. 148f shows a M.O.S.-tetrode, designed by the Matsushita El. Corp, giving a very high amplification.

## II.J.10. INTEGRATED CIRCUITS

The planar technique has created the possibility of making a transistor or a diode on an area of silicon crystal covering no more than a few tenths of $mm^2$. Fig. 148 shows that usually the connections require more space than the actual junction. It is only logical that the idea was conceived to manufacture the several transistors or diodes occurring in one circuit, on one crystal and to make the interconnections by depositing metal films on the crystal. As the connecting points then do not require proportionally more space, this is a method of placing more transistors on each $mm^2$ of the precious crystal. Vapour depositing a few connecting strips is far cheaper than soldering the terminal wires. An additional advantage is that the deposited connections are more reliable than the soldered joints. Since the resistance of the silicon can be influenced by adding more or less dope, it is possible in principle, to produce resistors of a given value on the same crystal. Finally the reverse direction of a diode represents a capacitor for a (pulsating) direct voltage. By adopting the method used for MOS transistors, an even better capacitor (suited for alternating voltages also) can be obtained by depositing a metal film on an extremely thin layer of $SiO_2$ on the crystal. Thus a circuit is obtained consisting of diodes, transistors, capacitors and resistors, all produced in one common process on a single piece of semiconductor crystal: the INTEGRATED CIRCUIT.

Around 1960 this idea was first put into practice by TEXAS INSTRUMENTS and FAIRCHILD, and in 1962 Philips showed a three-stage amplifier on a germanium crystal. However, the actual break-through of these integrated circuits

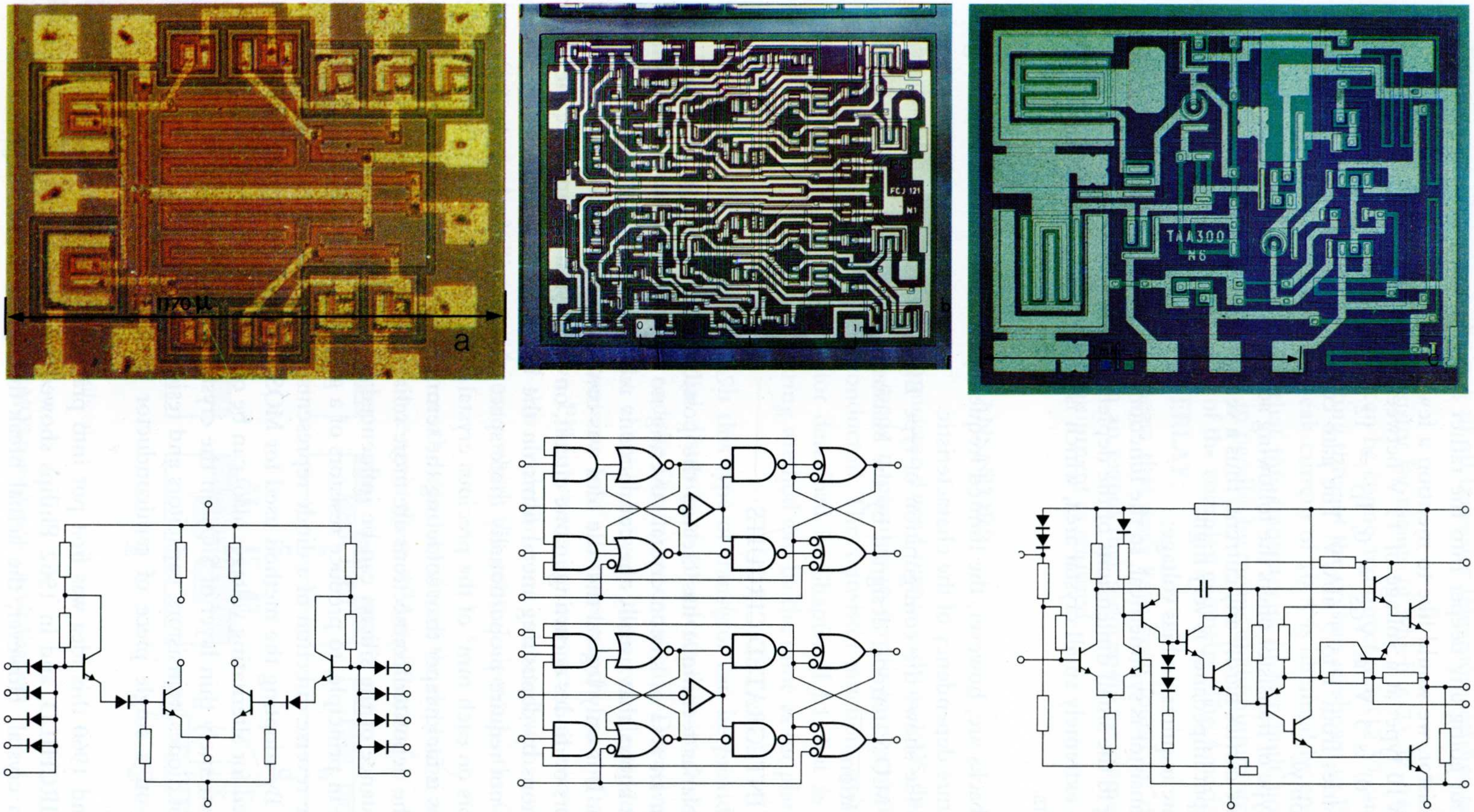


Fig. 152. a, b, c. Some examples of integrated circuits with the corresponding circuits they contain:
a. FFH141: a very simple computer circuit (dual NAND gate), b. FCJ121: a more complex computer circuit (dual JK flip-flop), c. example of a linear circuit. TAA 300, an A.F. amplifier with an output of 1 W at an input of 8.5 mV. (This clearly illustrates how much space the resistors take up).

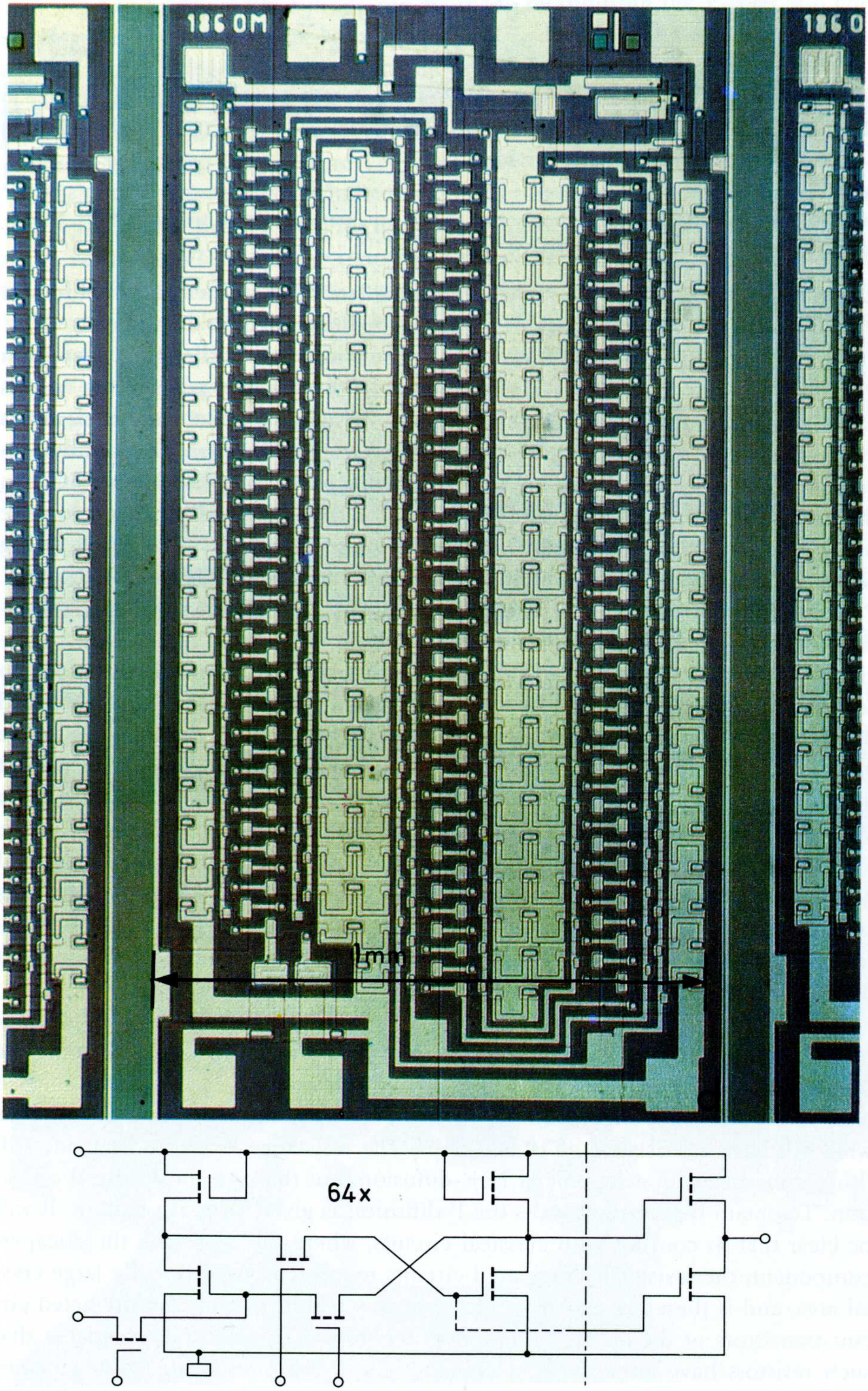


Fig. 152. d. Example of a circuit in MOS technique: FEJ121: 64-bit shift register with the circuit diagram of one "bit".

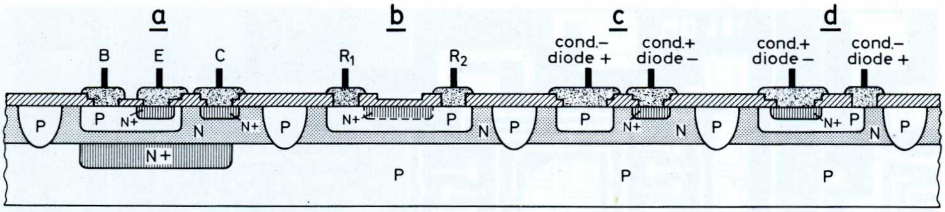


Fig. 150. Various elements made on one crystal by using the same process: a. transistor NPN, b. resistor, c. diode P-N, capacitor N-P, d. diode N-P, capacitor P-N.

had to wait till 1964, after the planar technique had been further perfected so that a large number of integrated circuits could be manufactured on one slice.

Furthermore a few modifications had to be made when changing over from transistors to integrated circuits. In many circuits the negative poles of the various transistors and diodes are not interconnected: in such cases the various elements on their common crystal will have to be isolated from each other. This can be adequately done by etching holes in the crystal which are then lined with $SiO_2$. In the holes thus prepared polycrystalline silicon is then epitaxially applied to serve as the "carrier". Then the original crystal is ground down until the $SiO_2$ lining comes to the surface (DIELECTRIC ISOLATION TECHNIQUE).

A simpler method is to apply the complete integrated circuit in an epitaxial N-layer which is applied to a low conductive P-layer whose sole function is to provide mechanical support and (if necessary) heat transfer to a metal base plate. In accordance with Fig. 150, frames of wedge shaped bars of P-material are diffused into this epitaxial N-layer. Together with the original P-crystal these frames isolate the various elements provided the N-layer is given a positive voltage with respect to the P-crystal (a blocked diode isolates). Now the collector connection can no longer be made at the bottom of the N-layer, but is made via a contact C at the top of the N-layer (Fig. 150a), whilst for better current conduction an additional N+-diffusion (N+ indicates heavily doped N-material) is provided under this contact. The base is formed by a P-diffusion and the emitter by an N+-diffusion (Fig. 150a). Sometimes an N+-diffusion is provided under the N-diffusion (BURIED LAYER) to decrease the collector resistance.

As shown in Fig. 150b a diffused P-layer can be used as a resistor by providing its ends with two contacts $R_1$ and $R_2$. The resistor thus obtained must carry a negative voltage with respect to the N-layer to be isolated from it. If a diffused layer has a resistance of 100 ohm per $mm^2$, a strip 1 mm long and 10 microns wide will have a resistance of 10.000 ohm. The resistance layer can be made still thinner by means of a superficial N+-diffusion, but that is a most critical operation. To obtain higher resistances the P-diffusion is given a zig-zag pattern. It will be clear that in contrast with classical circuits, where the resistor is the cheapest component, the resistor in integrated circuits requires a proportionally large crystal area, and is therefore expensive. It is for this reason that in the integrated circuit transistors or diodes are preferred to resistors. It is also understandable that such resistors have rather wide tolerances. The P-diffusion being homogeneous, the drawback of such wide tolerances is of no importance in a voltage divider (potentiometer) circuit.

According to Fig. 150c a diode can be made by diffusing the P-layer and connecting the N-contact to the N-layer via an N+-diffusion. If a negative voltage is applied to the P-contact, the blocking diode will function as a capacitor.

Fig. 150d shows that an NP-diode can also be obtained by diffusing an N+-layer into the P-diffusion. The P-contact then also serves to short circuit the PN-junction. When the polarity is reversed, we have again a capacitor.

Finally the polarity of the whole assembly as shown in Fig. 150 can be reversed. It is typical of the process used for integrated circuits that the properties of the epitaxial N-layer (collector of the transistor), the diffused P-layer (base of the transistor) plus the N+-layer diffused into it are the same for all the elements. This holds also for the diffusion depth (provided only one diffusion process is used for all the elements). In this connection it is interesting to note that the number of 78 operations are required for a slice of a planar transistor as shown in Fig. 145, whereas an integrated circuit according to Fig. 150 requires no less than 126 operations. An "individual" treatment per element would require an even greater number of operations. The area of the diffusions provide a certain degree of freedom which is limited, however, by the high costs per mm$^2$ of crystal. So an integrated circuit cannot be designed by merely translating discrete components such as transistors, diodes, capacitors and resistors in terms of elements forming the crystal circuits. A compromise will always have to be found between the specified characteristics and what can be realized in the integrated

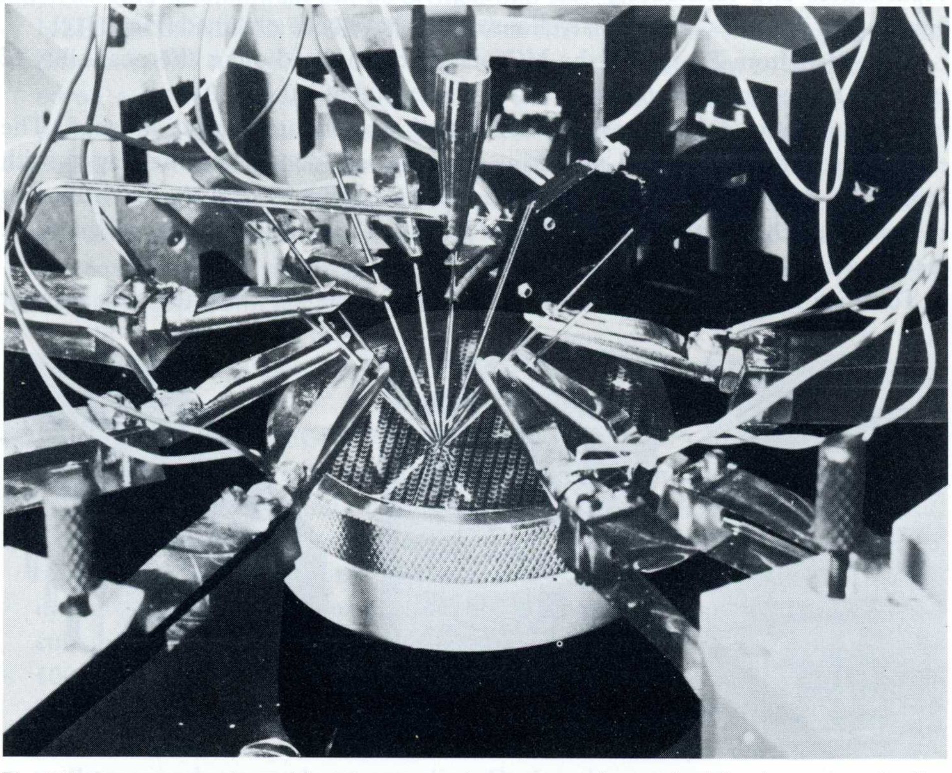


Fig. 151. On an automatic machine 10 contact points are pressed on each of the elements on the slice, and the I.C. is measured. Reject elements are automatically marked.

circuit (usually the requirements for the transistors serve as the starting point). It should also be borne in mind that the various components (particularly the resistors with their rather large area) have a greater capacitance with respect to each other than when mounted on a printed wiring board.

The next problem to be dealt with is that of "reject". As is done with transistors, a precision instrument (in this case one with 14 measuring points Fig. 151) probes each element, marking those that do not function satisfactorily, to be discarded when the slice is broken up. If six transistors are combined into one circuit, each having a yield of 90%, this combination would have a yield of no more than $0.9^6 = 50\%$. For the characteristics the compensation effect, as already mentioned with the resistors, often applies. As regards imperfection on or in the material (dust particles or dislocations), the rule is that on an area five times as large as that of the active parts of one element, the risk of faults is also five times as great. Endeavours are therefore made to keep the area of the parts as small as possible.

Fig. 152 gives some examples of industrially mass produced integrated circuits, and the conclusion is that indeed astounding results have been obtained. The easiest to realize are the relatively slow-switching elements according to Fig. 36 (Fig. 152a). If a greater speed is required, the situation immediately becomes more difficult due to the fact that internal capacitances, for instance, begin to play a part. Fig. 152b shows a much more complex circuit.

Linear circuits (amplifiers etc.) will have to do without self inductances high capacitances and resistances, and although parasitic capacitances play an even greater part in these circuits, marvellous results have been obtained (Fig. 151c).

As appears from Fig. 152d, the MOS technique indeed offers the possibility of producing many more transistors on a single crystal of $2.0 \times 1.0\ \text{mm}^2$.

We must now discuss how the integrated circuits are further finished. The measuring, breaking up and mounting on strips are done the same way as with transistors, only now a maximum number of 14 terminals is involved. The easiest mounting method is to insert the crystal in a metal envelope (or can) with 4 to 12 lead-in wires (Fig. 153a, SOT 5).

To arrive at smaller dimensions, a metal envelope was designed with powder

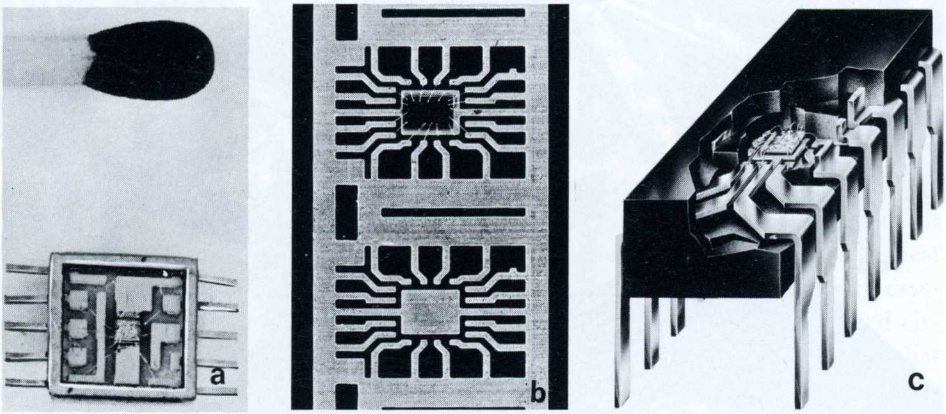


Fig. 153. Some examples of packages for integrated circuits: a. flat pack, b. interior dual in line, c. dual in line XG14.

glass lead-in plates in the side walls (Fig. 153a FLAT-PACKAGE). Finally the DUAL IN LINE PACKAGE was accepted as the standard for general use. Here the crystal is first mounted in a frame with extremely fine strips (Fig. 153b). This frame in turn is welded to sturdy connecting strips, after which the whole assembly is pressed into plastic (Fig. 153c). This structure is strong enough to be pressed into the printed wiring board (similar to the "lock fit" system).

## II.J.11. LARGE SCALE INTEGRATION

Particularly in digital systems (computers, telephone exchanges), the circuits are built up from a multiple repetition of equal circuit elements; examples are the "gate" and "flip flop" circuit in Fig. 36. It is logical that at first one specimen of such circuit was supplied as one integrated circuit in one envelope. This system leads to manufacturing the largest number of identical elements; the integrated circuit is of the simplest design, and the maximum manufacture yield is obtained. However, each connection must be bonded to the aluminium surface on the crystal and to the small frame, then it must be welded to the terminal strips of the "flat package" which must in turn be connected in the printed wiring board. It is attractive to interconnect the elements on the crystal, without breaking it up, by means of an aluminium layer which is deposited in the same operation as the contacts. This development can be divided into two stages:

- MEDIUM SCALE INTEGRATION. With this method the switching units are diminished to such dimensions that a few dozen of them can be accommodated on a crystal area as small as $4.0 \times 4.0$ mm$^2$. The fault risk due to local faults is then the same as if one circuit were formed on the crystal, and several circuits are now obtained for almost the same price. A good example is given in Fig. 152d, where a shift register with 64 bits can be accommodated on one crystal, and no more than six lead-in wires need be connected to the rest of the equipment. In this case the function is not a critical one, and the MOS transistor applied here present but a small reject risk.

  Should the number of interconnections become to large to be made on one plane, cross connections can be made by applying an additional intermediate layer of $SiO_2$.
- LARGE SCALE INTEGRATION. This is the method employed where even larger numbers of similar switching elements, to be interconnected are made on one slice. This method can be used for transistor memories. Here integrated circuits have to compete against magnetic core memories (costing upwards from 5 cents per bit), but they are faster. We speak of "large scale integration" if, for instance, endeavours are made to accommodate about 4000 flip-flop circuits on one single slice. In that case, however, just one reject among these 4000 flip-flops will render the whole slice useless. Therefore this would be an impossible thing to do. A solution to this problem is sought in the method of DISCRETIONARY WIRING. A number of circuits, say, 20% more than actually required, are made on one slice. During the automatic measuring of the slice a computer marks the faulty circuits and stores the information in a

memory. This information is used by a drawing machine, controlled by the same computer, to draw a mask for the interconnections to be made, skipping all reject elements. Thus one slice is obtained with a complete memory in which only the perfect elements are used.

Practice will show whether this system is economically practicable. In most cases where light weight and reliability are so important that economy does not count, there is the possibility of accommodating a complete memory or complete computer on a limited number of slices.

Furthermore this method can be used to manufacture standard slices which, by an appropriate choice of interconnections to be made afterwards, are adapted to the various customer requirements. It is perhaps interesting to know that the human memory can contain $10^{20}$ bits ("1" or "0" information), and that technical science will have quite some ground to cover before it can equal this capacity.

The astounding development over the years 1964 to 1969 justifies the expectation that further developments lie ahead in the direction of "medium scale integration" as described above. Components performing the electronic functions of amplifying, oscillating, mixing, demodulating, switching, counting, storing and finding information, will become cheaper when made in large quantities. This offers possibilities particularly for small computers used for automation, small office machines, and so forth.

# SECTION K. CERAMIC MATERIALS FOR ELECTRONICS

## II.K.1. INSULATING MATERIALS (WHITE CERAMICS)

Whereas the functioning of "semiconductors", described in the preceding section, is based on the internal movement of electrons, it is a requirement of insulating materials that they do not allow the movement of electrons.

Owing to its humidity and temperature resistance, as well as its mechanical strength, PORCELAIN (consisting of aluminium and siliconoxide) found universal application in electrical engineering. Porcelain products are made by pressing, extrusion moulding, or turning the basic material, (a putty of ceramic powder with liquid and a binding agent), into the required shape. The forms thus obtained are fired to a temperature below the melting point where they are SINTERED. In this process new combinations are usually formed, and crystals are transferred into other modifications, (mostly accompanied by a considerable shrinking), with the result that a material is obtained so hard that it can practically be worked only by grinding. In making high-tension insulators it has been found that gas enclosures are to be avoided, for they give rise to locally strong field strengths, in which "dark" discharges might occur introducing complete breakdown. A material better suited for accurate forming, and also little porous, is STEATITE, a magnesium silicate, found in nature as "soap stone" or "French chalk", and widely used in electric switching gear.

Because of its low H.F. losses, great mechanical strength and good heat conduction, pure $Al_2O_3$ is used for the envelopes of electron tubes. By using the molybdenum-manganese process it can be soldered to metal components. For higher frequencies it is better to use the denser MgO. A small percentage of $SiO_2$ improves the mechanical properties of both materials, but the electrical properties deteriorate at the same time.

An insulating material for capacitors is required to have a high dielectric constant (k), which does not vary with temperature, and low dielectric losses. It is not easy for one material to meet all these requirements.

The abovementioned steatite (magnesium silicate), supplied under the name of KERSIMA, has a k of about 7, a positive temperature coefficient (TC = $+100 \times 10^{-6}$/deg C), and a loss angle tan $\delta = 3 \times 10^{-4}$ at 1 MHz. This material is still used for class I capacitors in tuned circuits. Since the self inductance of the coils used in these circuits increases with temperature. (Chapter II.K.2), we wish to compensate this increase with the negative TC of the capacitor. By adding small quantities of titanium oxide, described below, and other ingredients it is possible to obtain a negative TC of $1000 \times 10^{-6}$/degC and a value for k of 100 whilst maintaining the low tan $\delta$.

Around 1948 the behaviour of titanium oxyde (occurring in nature as RUTILE) was studied closer; it allows a k of 100 to be reached. Barium titanate (KERUTILE), investigated afterwards, yielded a k of 2000, so that the value of a capacitor made of this material is 300 times that of a kersima capacitor of equal size. A drawback of this material is, however, that it has a very high temperature coefficient and that the tan $\delta$ is also many times larger. This is where theoretical physics stepped in.

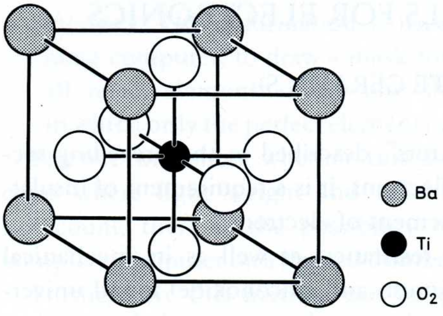


Fig. 154. Crystalline structure of barium titanate.

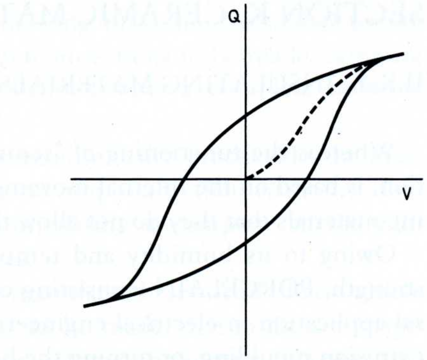


Fig. 155. Charge (Q) and voltage (V) drift caused by inertia (hysteresis) of the charge carriers (dashes = virginal curve).

The pattern of the crystal lattice of $BaTiO_3$ is shown in Fig. 154. The $Ba_2^{+}$ ions are located at the corners of the cube, the titanium ion ($Ti_4^{+}$) in the centre, and the oxygen ions ($O_2^{-}$) in the centres of the six planes. The titanium ion is relatively mobile in this lattice, and will be able to move under influence of an electric field. It is this polarizing effect (charge carriers moving with respect to each other, yet being elastically bound to their positions of equilibrium) that causes the high dielectric constant. The conditions under which this crystal structure is created can be favourably influenced by an appropriate method of firing and cooling.

Orienting the ions in the direction of the field takes time (and energy), and therefore the charge of such a capacitor will vary with the voltage, as shown by the graph in Fig. 155 (note the resemblance with the hysteresis loop of iron). This causes the dielectric losses.

The slightest degree of pollution may have a considerable effect on this complex phenomenon, whilst the addition of certain metals (tin, calcium, etc.) neutralizes the effects of such pollution. Theoretical consideration and systematic experiments have led to the manufacture of a material that at a k of 10 000 and a TC of $1 \times 10^{-3}$/deg C has a tan δ of $250 \times 10^{-4}$ at 1 kHz. It is now possible to make extremely small capacitors as needed for many applications. If a lower TC is required, materials with a lower k (2000-4000) are used.

Finally we must mention the BARRIER LAYER CAPACITORS, consisting of barium titanate made semiconductive by the addition of antimony. The electrodes are made of silver and enamel with contents of manganese oxide. During firing of the silver oxide some manganese diffuses into the barium titanate, as a result of which a diode (barrier layer) is formed under each silver electrode. These "diodes" can stand a voltage from 6 to 20 V. As the barrier layers are extremely thin, a very large capacitance per cm² is obtained.

## II.K.2. FERRITES FOR COIL CORES (FERROXCUBE)

Having discussed the material for capacitors in the preceding chapter, let us now consider the material used for the other part of the resonant circuit; the self inductance and the transformer. Here a core of magnetic material is needed to concentrate the magnetic lines of force, with the result that the self inductance increases considerably and proportionally to the PERMEABILITY (magnetic conductivity, $\mu$, which is 1 for air, and up to 15 000 for iron) of this material. Furthermore, the losses in the core must be kept low. A part of these core losses, which increases with the frequency, is formed by the EDDY CURRENT LOSSES, which are caused by currents perpendicular to the alternating field in the core.

The cores initially used in electrical engineering were made of soft iron, which allows a high degree of magnetization or "inductance", B, up to 15 000. The core itself is divided into thin plates isolated from each other (LAMELLAE) to reduce the area enclosed by the eddy current circuits. By adding silicon to the core metal, the ohmic resistance of the material is increased, and thus the eddy current losses are reduced.

Laminated iron could be used in electronics for audio frequency circuits (up to 20 kHz). For radio frequencies (up to 2 MHz) expensive coil designs (honey comb, toroid coils) without iron core had to be used. For multichannel telephony (60 to 6000 kHz, see Chapter I.23), and video signals (1000 kHz, see Chapter I.11) laminated iron, even when extremely thin, involved too high losses, and air coils were almost impracticable.

At first a solution was sought in making cores of fine iron powder (grain size about 50 microns) which, mixed with an insulating lacquer, was pressed into a rod. Owing to the small diameter of the grains, the eddy current losses were indeed reduced, but the insulating layer between the grains (also in the direction of the field) caused the permeability to be much too low.

Here too, a close study of the crystal structure, in this case of the iron oxide combinations (FERRITES), conducted in the Philips research laboratory around 1943, has led to the ultimate solution.

Ferrites are combined oxides, mostly of the general formula $Me^{++}Fe_2^{+++}O_4$,

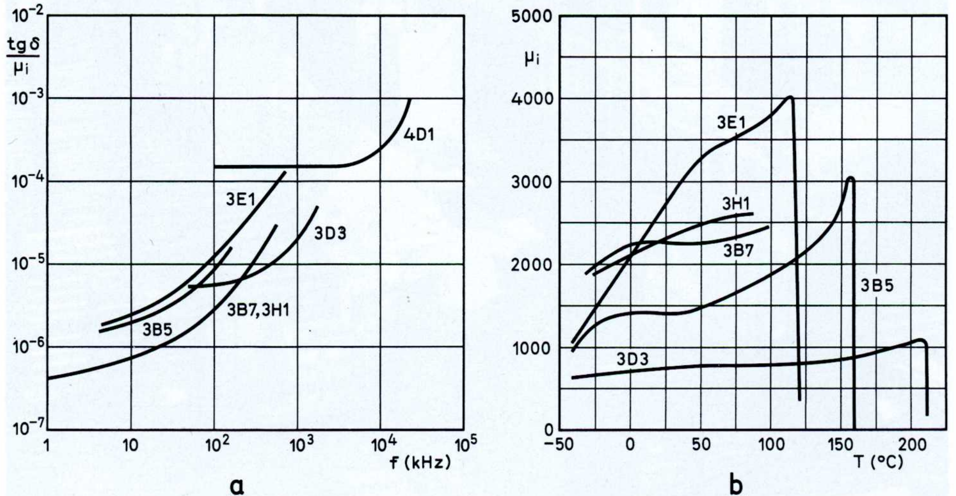


Fig. 156. Properties of different types of ferroxcube: a. loss factor (tg δ/initial permeability) as a function of frequency, b. initial permeability as function of temperature.

in which Me is a bivalent metal ion of about equal ion diameter as iron, like manganese, magnesium, copper and zinc. Since this kind of ferrite has a cubic crystal structure comparable with the mineral spinel (magnesium aluminate) we speak of a SPINEL structure.

Investigation showed that mixed crystals of iron oxide and oxides of other metals, for instance abovementioned ones, have relatively high permeability values (up to 4000). Owing to the high resistance of these oxide combinations the eddy currents are but weak and, the coercive force being small at the same time, the hysteresis losses are low.

And here a new material had been invented, FERROXCUBE, without which, to mention some, it would hardly have been possible to develop multichannel telephony, television receivers, the built-in rod aerial (meaning portable radio sets), the video recorder, high frequency heating with concentrated fields, and the cyclotron.

Over the past years, various types of ferroxcube of widely differing properties have been developed. It would lead us to far to explain what properties were obtained from what ingredients or firing method; be it sufficient to mention the results. Fig. 156b gives the values of $\mu_i$ (initial permeability at $H = 0$) for some types as functions of temperature. Some types are found to be highly temperature dependent; not so the 3D3 material. Fig. 156a gives the loss factor of the various types as functions of frequency, and it can be seen that 4D1 is suitable to reach 10 MHz (with FERROXPLANA, a later development, 200 MHz can be reached).

Fig. 157. A ferroxcube factory.

Making ferroxcube is no simple matter. After the raw materials have been mixed, they undergo a heat treatment (pre-firing) at a temperature considerably lower than at which actual sintering takes place. After a single preparation, granules are then formed by, for instance, spray-drying a slurry. In a mould the granules thus obtained are formed under high pressure, afterwards to be fired in ovens where the material sinters.

Fig. 157 shows sintering furnaces for ferrites in one of the factories.

For the manufacture of rod-shaped objects (e.g. aerial rods) the prefired powder is mixed with water and a binding agent to form a paste which, after de-aeration, is extruded, dried, and fired.

Several types of material require special care in firing; for instance in an electric furnace with accurately controlled conditions of temperature and atmosphere. Shrinkage due to firing amounts to about 20 or 25 %, and it takes quite some skill to obtain accurate dimensional tolerances. In many cases the product is finished by grinding.

Some ferroxcube products are displayed in Fig. 158 which shows that accurate shaping is indeed possible.

Finally it might be mentioned that ferroxcube is used as a UNIDIRECTIONAL ISOLATOR for microwaves, based on the Faraday rotation effect.

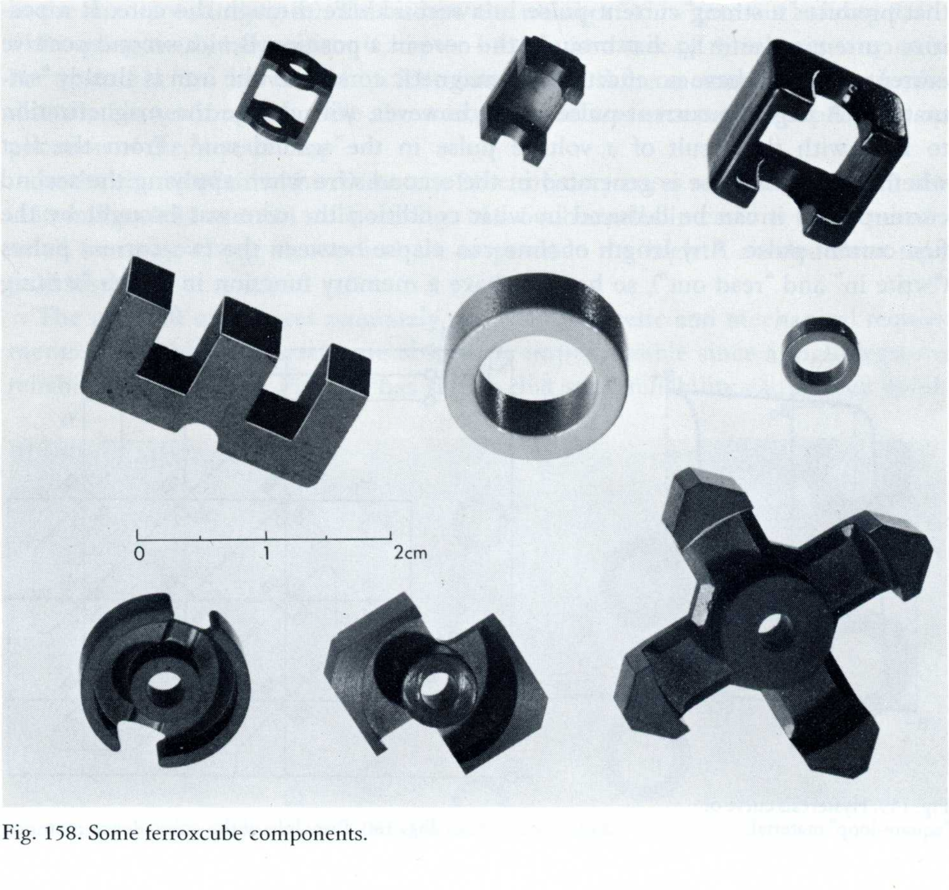


Fig. 158. Some ferroxcube components.

## II.K.3. SQUARE LOOP MATERIALS

In the preceding chapter nothing was said about HYSTERESIS. As drawn in Fig. 155 for an electrically insulating material, the magnetic induction, caused in the magnetic field in the iron core the first time a current is sent through the coil, will also vary according to the dashed line: as the current increases, the point of "saturation" is reached. If then the current drops to zero again, not all the magnetic dipoles, having been oriented in the direction of the magnetic field, will return to their original positions; a certain amount of REMANENT magnetism remains. Only when the current reaches a certain negative value will the magnetism disappear completely. For each cycle of an alternating current the "banana skin" pattern of Fig. 155 is completed. The area of this pattern is a measure of the hysteresis losses, and by a proper choice of material we try to keep this area as small as possible for applications like self inductances and transformers.

In the application discussed below, it is this phenomenon that is turned to particular use. A certain ferrite composition (magnesium-, copper- or lithium-manganese) produces an almost square hysteresis loop as shown in Fig. 159. This loop shows how the magnetic induction (B) varies in a core of this material when a conductor, carrying an alternating current, is led through this core. At a certain current value ($+I_c$ or $-I_c$) the magnetization B changes very rapidly from a very high positive to a very high negative value, and it is particularly this rapid change that produces a strong current pulse in a second wire through the core. If a positive current pulse $+I_m$ has brought the core in a position B+, a second positive current pulse will have no effect on the magnetic condition (the iron is already "saturated"). A negative current pulse $-I_m$, however, will change the magnetization to B−, with the result of a voltage pulse in the second wire. From the fact whether or not a pulse is generated in the second wire when applying the second current pulse it can be deduced in what condition the core was brought by the first current pulse. Any length of time can elapse between the two current pulses ("write in" and "read out"), so here we have a memory function in which "writing

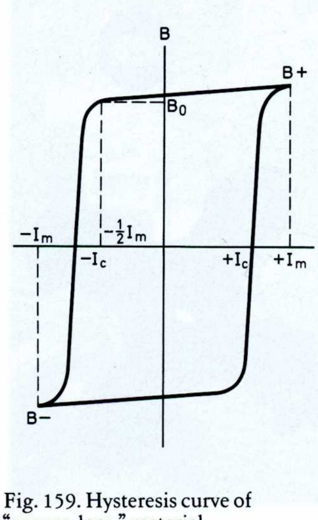


Fig. 159. Hysteresis curve of "square-loop" material.

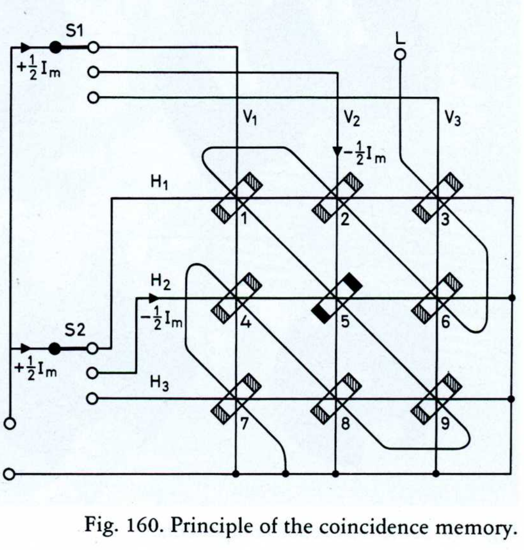


Fig. 160. Principle of the coincidence memory.

in" and "reading out" need not take more than a microsecond. Such a magnetic memory core, made in the form of a small ring can, however, store only one piece of information: "was the previous pulse a positive or negative one?"

If such a magnetic memory is intended for the "write in" and "read out" of more data, the COINCIDENCE principle is used. Starting again from the condition B+ in Fig. 159, a pulse of $-\frac{1}{2}I_m$ will cause but a slight variation of B (the maximum requirement for square loop material is 10%), which will generate but a negligible pulse in the auxiliary winding.

By running horizontal and vertical wires as well as a READ-OUT WIRE through a number of ring cores (Fig. 160), we arrive at the following system. Assuming that all the cores are in the B+ condition (Fig. 159), a current pulse of $-\frac{1}{2}I_m$ through wire $H_2$ will not be able to reverse the cores 4, 5 and 6. If at the same time a current pulse of $-\frac{1}{2}I_m$ is sent through wire $V_2$, and thus through the cores 2, 5 and 8, the cores 2 and 8 will remain in the B+ condition, whereas core 5 is magnetized by $2 \times -\frac{1}{2}I_m$ and changes over. The number "1" has now been written in core 5. If by means of switch $S_1$ pulses of $+\frac{1}{2}I_m$, in the sequential order $V_1, V_2, V_3, V_1, V_2, V_3$, and so on, are sent through the vertical wires while switch $S_2$ is at the same time sending three pulses through $H_1$, then three through $H_2$ etc., all of the rings will successively receive two simultaneous pulses. The fifth pulse, however, will produce a field opposing the existing induction, causing ring 5 to change over, so that a voltage pulse is generated in the auxiliary winding (L). This is how the number "1" is "read out" of ring 5.

By using the binary system (Chapter I.22) in a matrix of nine cores, any number up to 511 can be written in. The function of these memories will be further discussed in section N, but the above explanation was necessary to explain the application of the "square loop".

In the course of time the diameter of the memory cores reduced from 3.75 mm (150 mil) to 0.35 mm (14 mil), see Fig. 161. Such small dimensions are required if small currents (cheap electronics) are to produce the high B needed for a rapid field reversal.

The material must meet accurately specified magnetic and mechanical requirements. Air bubbles or cracks are absolutely impermissible since a high degree of reliability is called for. Practice has shown that such reliability can indeed be ob-

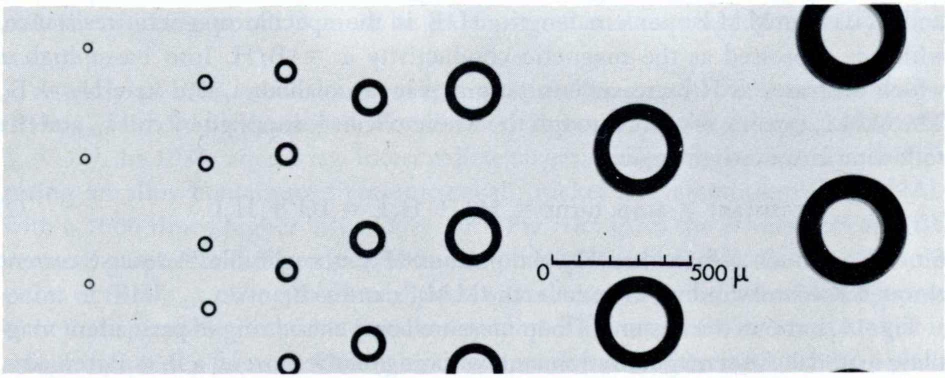


Fig. 161. Some cores of square-loop material as used in memories.

tained. Memories are used containing 5 000 000 cores, not one of which may fail (one failure might well happen to be the 100 000 of the bank balance!). For the mass production of these cores, operations such as moulding, firing, and mechanical and electric measuring are carried out on automatic machines. An interesting piece of equipment is the automatic electrical measuring machine which needs no more than 15 $\mu$s to determine 6 parameters of one core, and passes on only the cores of which all 6 parameters are within the specified tolerances.

### II.K.4. PERMANENT MAGNETS (FERROXDURE)

Certain kinds of stone are PERMANENTLY MAGNETIC. In ancient times such stones were used in the compass (lode stones). Much later it was discovered that soft iron can be magnetized by sending an electric current through a coil surrounding it. Electric motors and generators, as well as loudspeakers, were originally fitted out with such electromagnets. By adding certain materials to the iron, it proved possible to obtain a strong REMANENT magnetism once the iron is magnetized in a coil, thus yielding a PERMANENT magnet. Physical investigations carried out during the years 1925-1950 have led to the development of materials with such favourable permanent magnetic properties that in numerous applications the electromagnets were replaced by permanent magnets.

Before the characteristics of permanent magnets can be discussed, it is necessary first to consider the magnetic circuit. Let us start from the (physically not correct, but practically useful) analogy with the electric current circuit.

Fig. 162a shows the magnetic circuit consisting of a toroid with an air gap, surrounded by a coil. Suppose the magnetic FLUX through this toroid is $\Phi$; the inductance B is then defined as $\Phi$ per cm$^2$. The magnetic flux is the same throughout the circuit, so:

$$B_y S = B_l s. \qquad (1)$$

We now see that the inductance can be increased by reducing the diameter (e.g. at the air gap). To have the flux $\Phi$ flow through the circuit, a certain MAGNETO MOTIVE FORCE (M.M.F.) is needed (in analogy with the electro motive force). The ratio M.M.F./$\Phi$ is the magnetic resistance, and if we define B as flux per cm$^2$, and H as the M.M.F. per cm length, H/B is the specific magnetic resistance, which is expressed as the magnetic conductivity $\mu = B/H$. Iron has a high $\mu$ which decreases as H increases (saturation); $\mu$ in air is about 1, and here $H_l = B_l$. The M.M.F. causing the flux through the whole circuit is supplied by coil $I_1$, and the following formula then applies:

$$\text{constant} \times \text{amp. turns} = H_l l + H_y L = B_l l + H_y L \qquad (2)$$

Since $B_l$ is much greater then $H_y$, the product $H_y L$ is neglibible. A reverse current through a second winding ($I_2$) reduces the M.M.F., and so $B_l$.

Fig. 162b shows the hysteresis loop measured on a closed ring of permanent magnetic material. After magnetization and switching off the current, a $B = B_r$ remains, only to disappear when (for instance with a second coil) a negative

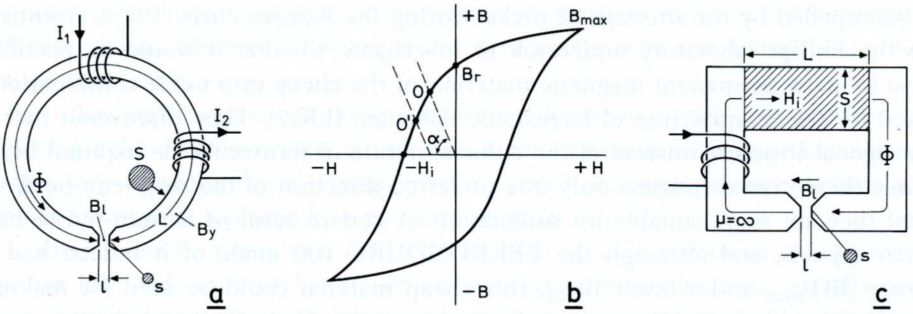


Fig. 162. The magnetic circuit: a. with an electromagnet, b. B-H curve, c. with a permanent magnet.

M.M.F. per cm of $H_i$ is produced in the circuit. For low values of H, B varies according to the upper left part of the graph which is characteristic of the material.

If such a magnet is taken up in the magnetic circuit, this M.M.F. of the remanent magnetism ($H_iL$) will, according to Fig. 162c, have to drive the flux through the circuit. Neglecting the M.M.F. in the pole shoes (they have a very high $\mu$), we find:

$$-H_iL = B_l l. \tag{3}$$

By dividing equation (1) by equation (3) we obtain:

$$B_y/H_i = Ls/lS = \tan\alpha.$$

The angle $\alpha$ can be drawn in the graph of Fig. 162b: the straight dashed line gives the $B_y$ required, the curve the available $B_y$, and the point of intersection (O) is the working point. Tan $\alpha$ and the position of "O" can be influenced to a certain extent by varying the length and diameter of the magnet (at a given dimension s and l of the air gap): O is lowered by using a short, thick magnet. If a current is sent through the coil opposing the direction of the remanent field, the line through O is shifted to O′.

Multiplying equations (1) and (3) gives the magnet power in the air gap:

$$B_l{}^2ls = B_yH_iLS.$$

The magnetic power in the air gap is thus found to be proportional to B × H at point O and the volume of the magnet. To use the least possible magnetic material, the working point O must be so chosen that the energy product B×H is maximum.

As early as 1923, when loudspeakers were on the march, alloys with a higher $(BH)_{max}$ were being developed. For ordinary carbon steel this energy product was $5 \times 10^3$. In 1935, after a few intermediate stages, Philips succeeded in manufacturing an alloy containing titanium, cobalt, nickel and aluminium (TICONAL) with a 1000 times higher value ($5 \times 10^6$). Fig. 163 gives the B versus H and BH versus B graphs for this product. (One can imagine that it is difficult to work at the point of $(BH)_{max}$, owing to the air gap which decreases tan $\alpha$, but which happens to be indispensable in the loudspeaker system). This magnetic steel was melted in a high-frequency furnace. Owing to the expensive raw materials, and the equally expensive installation needed, the price of this steel is rather high.

Compelled by the shortage of nickel during the Korean crisis (1951), scientists in the Philips laboratory undertook to investigate whether it would be possible also to make permanent magnetic materials of the cheap iron oxide combinations used for the manufacture of ferroxcube (Chapter II.K.2). They discovered that a hexagonal structure (instead of the cubic structure in ferroxcube) is required here, since these crystals possess only one preferred direction of the magnetic field, so that they are more suitable for orientation. A certain form of barium-ferrite had such crystals, and although the FERROXDURE 100 made of it indeed had a lower $(BH)_{max}$ and a lower $B_{max}$, this cheap material could be used for making loudspeaker magnets. Because of the high coercive force $H_i$, however, the magnet had to be short, and because of the low B it also had to be wide in order to be able to work near the maximum of the BH curve. For applications with a wide air gap (for instance the two rings in Fig. 30), this high coercive force is of particular advantage. The magnetic rings used for the focussing of television tubes in those days presented an excellent field of application for this material.

However, the physicists in the Philips laboratory were not satisfied with the results obtained; they had found that the magnetized particles, equally oriented during the magnetizing procedure, would afterwards reassume positions pointing in all directions within an angle of 180°, with the result that the remanence was only half the saturation. Then the clever idea was conceived to grind the material down to hexagonal crystals which, mixed with water were pressed into a mould, in which a very powerful magnetic field was generated. Thus the crystals were oriented, and during the firing process they maintained their positions in parallel with the direction of the field. According to Fig. 163 this oriented ferroxdure 330 has a three times higher $(BH)_{max}$ than the unoriented type, and this maximum occurs at higher values of B. Thanks to this material the price of a well dimensioned loudspeaker magnet is already below that of a steel one, and as a result of improved pro-

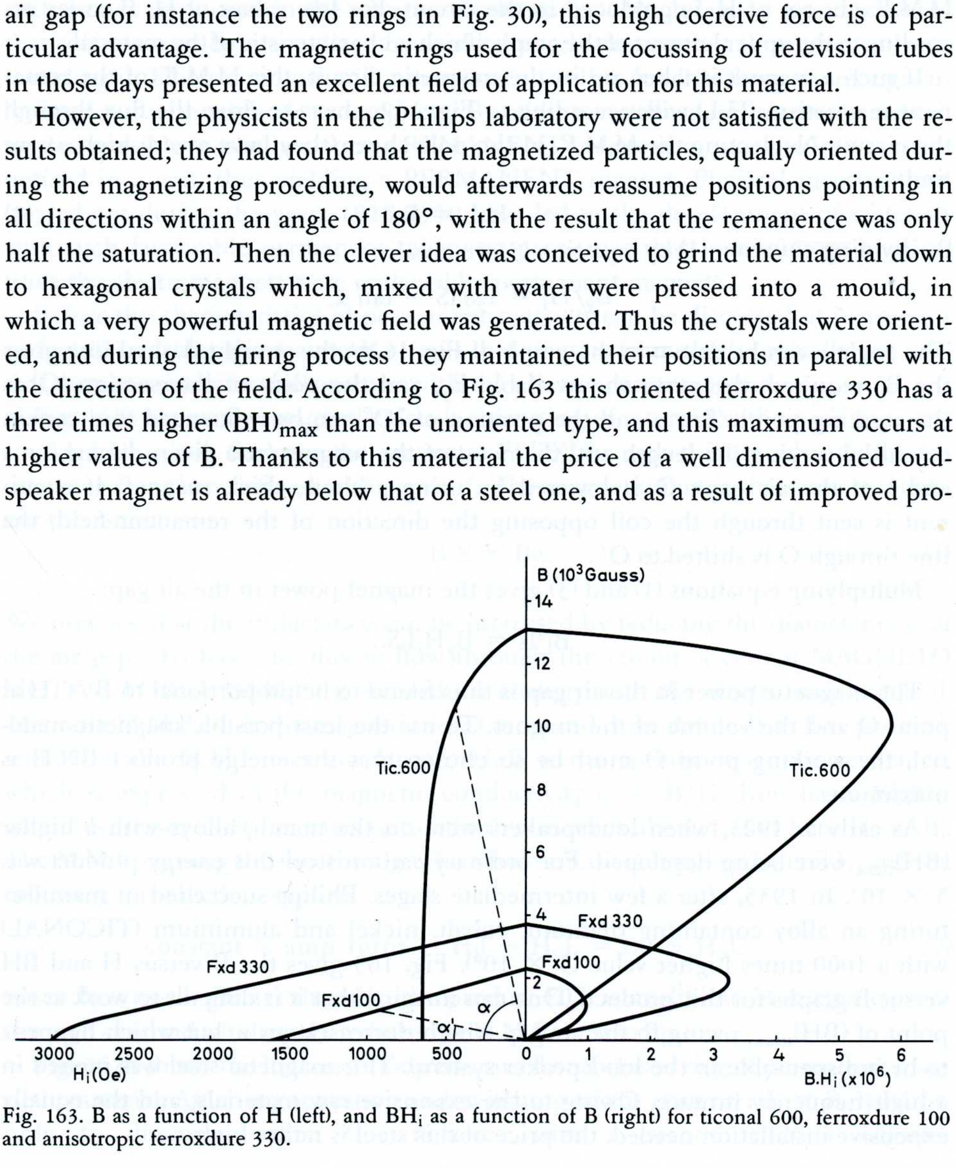


Fig. 163. B as a function of H (left), and $BH_i$ as a function of B (right) for ticonal 600, ferroxdure 100 and anisotropic ferroxdure 330.

cess control and mechanization, it may be expected that the prices of ferroxdure will come down, whereas the prices of the already expensive raw materials used for ticonal are more likely to go up.

For other applications, requiring a wider air gap or in which coils produce an opposing field in the magnetic circuit, the high coercive force is essential. Owing to this improvement the electromagnets in small d.c. motors and synchronous motors can now be replaced by ferroxdure magnets.

It should also be noted that owing to its high specific resistance this material has but low eddy current losses, which is of great importance for applications where a high frequency field is superimposed on the magnetic field, such as in loudspeakers and "transducers".

An entirely different application is that of anisotropic ferroxdure powder pressed into synthetic rubber strip used to lock refrigerators.

Experiments made in later years have led to materials with an even higher $(BH)_{max}$; in 1968 Philips published an article on a samarium-cobalt combination with a $(BH)_{max}$ of $18.5 \times 10^6$.

## II.K.5. CERAMIC (VARIABLE) RESISTORS

In an electronic circuit, a resistor generally serves to establish a voltage to current ratio, or a voltage division. For this application the value of the resistor should remain constant even if the current or temperature varies. Such resistors are made of metal or carbon (Chapters II.L.4 and II.L.5).

Ceramic semiconductor materials are used to manufacture resistors whose values vary characteristically with voltage, temperature, or exposure to light. We shall discuss four types, all of which having found interesting applications in electronics.

VOLTAGE DEPENDENT RESISTORS (V.D.R.) are made of silicon carbide. The voltage as a function of the current varies as shown in Fig. 164a; Fig. 164b shows the same relation plotted logarithmically. The resistance drops steeply as the voltage increases.

The most obvious application of this type of resistor is that of voltage surge ar-

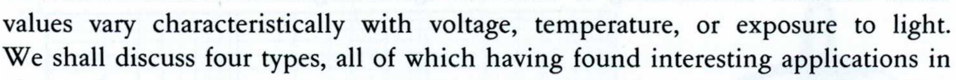


Fig. 164. Characteristics of a V(oltage) D(ependent) R(esistor): a. voltage as a function of current, b. voltage-current curve on a logarithmical scale. Here the formula applies: $V = C.I^{\beta}$. ($\beta = 0.17 - 0.21$).

restor. At normal mains voltage the current through the "V.D.R." is very low; if a voltage surge should occur in the mains (caused by stroke of lightning, for instance), the resistance drops considerably, and the voltage cannot rise excessively. V.D.R. are connected in parallel with relay contacts to avoid "sparking" when the current is switched off, and so these contacts will not burn away. Similarly the sparking of the commutators of direct current motors can be limited. Connected in parallel to a circuit such resistors have a stabilizing effect (the voltage across the circuit does not increase proportionally to supply voltage).

Resistors with a highly NEGATIVE TEMPERATURE COEFFICIENT (N.T.C.) are made of iron, cobalt or nickel oxide with additions of titanium or lithium. The highly negative temperature coefficient is very typical: the resistance decreases almost logarithmically with temperature, as can be seen in Fig. 165. By changing the composition of the material it is possible to obtain widely differing resistance values at room temperature.

The obvious application here is, of course, electrical temperature measuring in which, thanks to the high specific resistance, very small elements (with low heat capacitance) can be used. Now the fever thermometer for "electronic patient monitoring" has become available. A very popular application is that of temperature control of water and oil in motor car engines. For this particular application in liquid, the heat dissipated in the resistor can easily be removed, and a considerable current is permissible (according to the lower curve the current at 12 V and 30 °C is 40 mA, at 60 °C 150 mA, and at 120 °C 600 mA, meaning that a good scale division is obtained). Another application is the one where the time element is involved. If an N.T.C. resistor is connected in series with the heaters of a television receiver, the switching-on current surge through the cold heaters will be

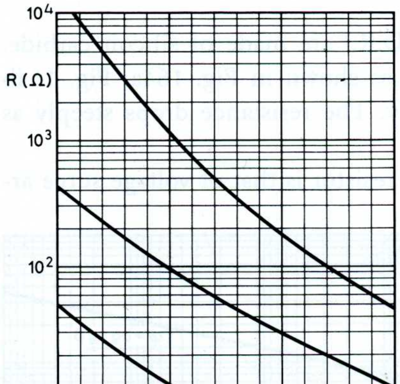


Fig. 165. Resistance versus temperature characteristic for three different compositions of N.T.C. material.

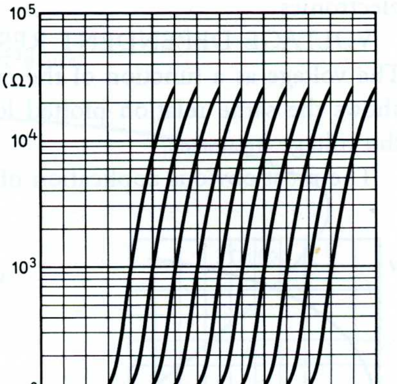


Fig. 166. Resistance versus temperature characteristic for various compositions of P.T.C. material.

limited by this resistor; after some time the N.T.C. resistor heats up, and the resistance drops to a very low value.

In contrast with what was said above, barium or strontium titanates, already mentioned with white ceramic materials, have a very high POSITIVE TEMPERATURE COEFFICIENT (P.T.C.). The steep rise is, however, limited to a certain temperature range that can be influenced by changing the composition of the material used. Fig. 166 gives a family of graphs of an actual series of P.T.C. THERMISTORS. It is shown here that at a temperature increase of 30 deg C (the lower temperature of which can be chosen anywhere between 80 and 170 °C), the resistance of the whole range increases by a factor of 150.

Such type of resistor in series with a current circuit provides an excellent safeguard against overload (in electric motors and transformers). Furthermore they are quite useful for temperature signalling or thermostats. Since heat exchange takes place between a resistor and a liquid in which it is immersed, the resistor can also be used for level control. In automatic washers, certain heater or motor circuits can be switched on without the need for auxiliary relays.

Finally we shall also mention a resistor not made of ceramic material, but of the semiconductor CADMIUM SULPHIDE, whose resistance varies considerably with the intensity of incident light. Fig. 167 shows the characteristic of such LIGHT DEPENDENT RESISTOR (L.D.R.). The tolerances are indeed rather wide, but it is important that in the range from 100 to 1000 lux the resistance varies by a factor of 5. For the accurate measuring of luminous intensity or illumination, the earlier mentioned vacuum, gas-filled, or semiconductor photo cells should be preferred, but thanks to its steep characteristic and the relative-

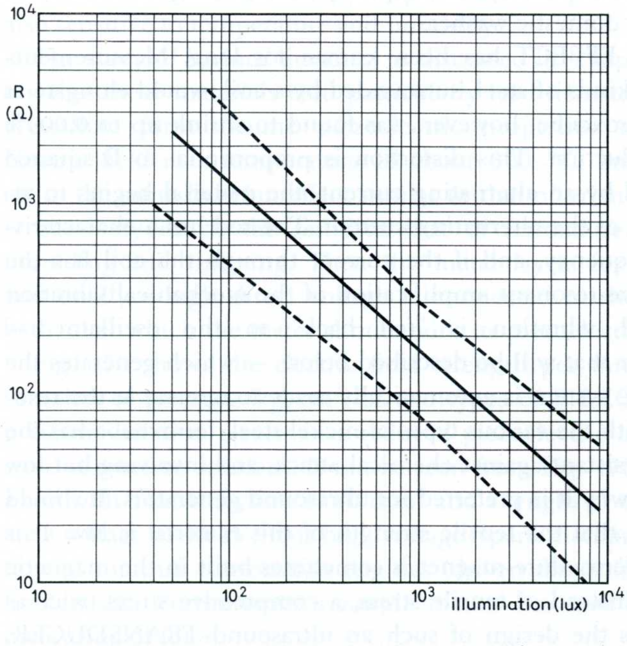


Fig. 167. Resistance versus luminosity characteristic for a cadmium-sulphide L(ight) D(ependent) R(esistor). The dashed line indicates the tolerance.

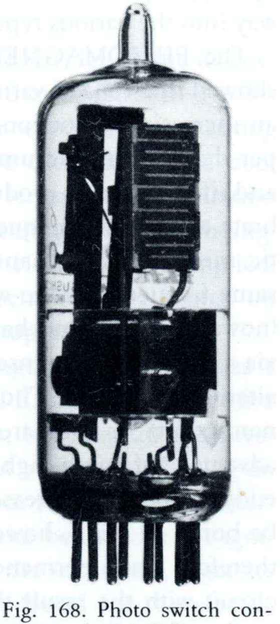


Fig. 168. Photo switch consisting of an L.D.R. resistor, coupled to a bimetal inside the envelope of a radio tube (Matsushita).

ly high permissible current, this L.D.R. is particularly suitable to be used as an OPTICAL RELAY, say to switch on a lamp when it is getting dark.

Fig. 168 shows an L.D.R. and a bimetal relay built together in a radio tube envelope that can be placed in the armature of a street lamp which is then automatically switched on when it gets dark. These cadmium sulphide cells are also widely used for inexpensive lux meters as built into photo cameras.

There are different methods of manufacturing these cadmium sulphide cells; the earliest known is pressing pellets of this material onto which conductive layers are deposited as the electrodes. Better reproducible characteristics are obtained by vapour depositing this material in thin layers. In 1968 Philips published a MONOGRAIN technique, also offering the possibility of this material being used in solar cells. Unfortunately, discussing these new technologies is not within the scope of this book.

## II.K.6. PIEZOMAGNETIC AND PIEZOELECTRIC MATERIALS

Piezomagnetic materials are characterized by the phenomenon that they undergo physical distortion under influence of magnetic fields, and vice versa, that they will produce magnetic fields when under mechanical stress. One of the first applications based on this phenomenon is transferring ultrasound vibrations to a liquid for depth sounding or ULTRASONIC cleaning (the medium high frequency of 30 kHz causes the liquid to move so quickly that the chemical cleaning process reaches a much higher degree of efficiency). The application of this method has up to now been limited to the industry, but expectations are that it will find its way into the various types of domestic washer.

The PIEZOMAGNETIC EFFECT has been known for long. Measurements showed that rods of various kinds of steel, surrounded by a coil, would elongate at an increasing induction; ferroxcube, however, was found to shrink up to 0.003% per deg C at maximum induction. This distortion is proportional to B squared and, if the field is produced by an alternating current, the material begins to vibrate at twice the frequency of the alternating current. The rod has a characteristic mechanical resonant frequency, and if the current through the coil has the same frequency, there will be resonant amplification of the mechanical vibration (nowadays the mechanical vibration is "fed back" to the oscillator — via a piezoelectric transducer that will be described below — which generates the alternating current. Thus the system is automatically made to operate at the resonant point). As compared with the various types of nickel steel, ferroxcube has the advantage of being highly resistant against chemical attack, and involving but low eddy current losses, reasons why it is preferred for ultrasound generators. It should be borne in mind, however, that the tensile strength of this material is low. It is therefore that a permanent ferroxdure magnet is sometimes built in the magnetic circuit with the result that instead of tensile stress, a compressive stress twice as high occurs. Fig. 169 shows the design of such an ultrasound TRANSDUCER. The efficiency (mechanical energy divided by electrical input energy) can rise above 90%, the vibrating output energy can be as much as 3 W/cm$^2$.

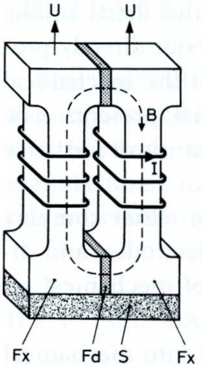


Fig. 169. Piezomagnetic transducer. Fx = ferroxcube, Fd = ferroxdure, S = foam rubber, I = alternating current, B = inductance, U = variation in length.

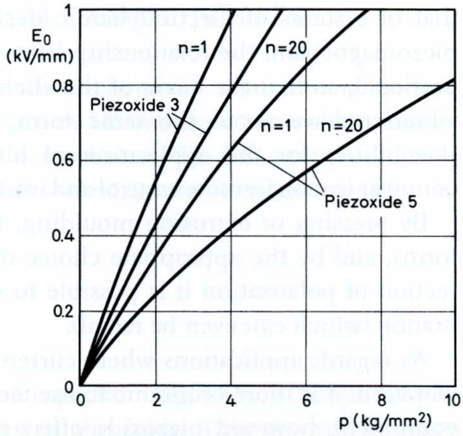


Fig. 170. Voltage versus compression characteristic for two types of piezoxide after 1 and after 20 pressure pulses (after that the material becomes stable).

Piezomagnetism can also be used for converting mechanical vibration into magnetic vibration, and then further into voltage variations, but for such systems it is recommendable to use piezoelectric materials.

Piezoelectricity (generating a voltage by compressing material) was discovered as early as 1883 by the Curie brothers. This piezoelectric effect in QUARTZ crystals was used to obtain a highly stable frequency (Chapter II.L.7). In 1921 it was discovered that the crystals of SEIGNETTE salt are more sensitive in this respect. Crystals of this kind could be produced synthetically (they were "grown"), and were used for gramophone pick-ups, and microphones. Unfortunately they were also sensitive to temperature and humidity.

The study of barium titanates, as discussed in Chapter II.K.1, led to the discovery that these materials (and particularly lead titanates and zirconates) in a polycrystalline form assume piezoelectric properties when exposed to a strong electric field at a temperature of about 300 °C, and are then allowed to cool off in the same field (an effect similar to that of orienting ferroxdure). Preparation is also very much the same as that of ferroxdure: pressing, firing, grinding, moulding and sintering. Then the metal layers for the electrodes are applied, and finally a high direct voltage is connected to the electrodes in a heated oil bath, after which the material maintains its piezoelectric properties.

Fig. 170 shows the electric field strength per mm length (or thickness) as a function of pressure (kg/cm²) of two types of PIEZOXIDE. Initially a repetitive pressure variation results in a reduced piezoelectric effect, but after 20 compressions it remains constant. It can be seen from Fig. 170 that a pressure of 50 kg on a rod with a cross section of 10 mm² and a length of 40 mm is sufficient to generate a voltage of 40 kV. This high voltage (spark discharge) can be used for the ignition of internal combustion engines (driven direct from the camshaft), cigarette lighters etc. Furthermore a simple means has now been found for the electric measuring of force.

In pick-up and microphone applications, vibrations of sound frequency can be converted into voltage variations, the output power being substantially higher than that of systems of electrodynamic design. It should be noted that here, unlike piezomagnetism, the relationship between voltage and force is about directly proportional, so that the curve of the alternating voltage and that of the mechanical vibration have about the same form. Furthermore this material presents new possibilities for the application of ultrasound air vibrations in short distance communication (remote control and intruder alarm).

By pressing or extrusion moulding, this material can be given rather complex forms, and by the appropriate choice of form, position of the electrodes, and direction of polarisation it is possible to determine the direction of mechanical vibration (which can even be radial).

As regards applications where current variations are converted into mechanical vibration, it is more economic to use ferroxcube for ultrasound cleaning. For small equipment, however, piezoxide offers numerous possibilities. Another important application is the generation of the mechanical pulse-train through "delay lines" (Fig. 27, and Chapter II.N.9).

It should be noted that a piece of ceramic material with two electrodes, brought to mechanical vibration (like a tuning fork) by an alternating voltage, will try to vibrate at its mechanical resonance frequency. If the applied alternating voltage has the same frequency, there will be "resonant amplification", causing the current to increase considerably (the reactance becomes lower). By changing the dimensions of the piezoxide disc (by grinding), the resonance frequency can be changed, and the piezoxide element can thus be "tuned". Thanks to the very steep "resonance curve" (the disc has very low dielectric losses), we have here an important means to improve selectivity (RESONATORS, Chapter II.L.8).

## II.K.7. AFTERTHOUGHT

Be it less spectacular than the vacuum and semiconductor techniques, it may be said that the development of these ceramic materials is another good example of how the cooperation between theoretical physicists and electronics engineers has resulted in a range of products which have contributed substantially to the advancement of electronic equipment.

The materials discussed in this section, and whose specifications are compiled in a 1000 page catalogue of Philips (green Data Handbook, part 4 in Fig. 41), are available to any component manufacturer.

# SECTION L. PHYSICAL-CHEMICAL COMPONENTS

The now following subdivision into "physical-chemical" components (Section L) and "electromechanical" components (Section M) may be a somewhat arbitrary one, but has indeed to do with the fact that whereas the first group is mostly manufactured in specializing factories, the second group is made in radio receiver factories (see also Section II.A).

## II.L.1. FOIL CAPACITORS

The amateur of 1915 made his telephone capacitor (Fig. 6) by folding a strip of lacquered paper and inserting silver-paper in the folds. This was the first foil capacitor: the PAPER CAPACITOR.

The capacitance of a capacitor with parallel flat electrodes whose areas are large as compared with the distance between them can be calculated as:

$$C = kO/d$$

O is the area of one electrode which, in foil capacitors, can be made large by winding two strips of metal foil, separated by two strips of an insulating material, into a cylinder which is then usually compressed to form a square package.

k is the dielectric constant of the insulating material, which is 1 for air, 4 for impregnated paper, 6 for mica, 7 for porcelain, and up to 10 000 for barium titanates (Chapter II.K.1).

d is the thickness of the insulating material, and can be made as small as the mechanical strength of the material will allow. At the same time it should be ensured that the field strength at which breakdown occurs is not exceeded. The breakdown voltage for air is about 2.7 kV per mm; that of impregnated paper 50 kV per mm (as we shall see later, the breakdown voltage is not so easy to define). Since paper might contain small holes, at least two layers are wound together.

For a certain capacitance, the losses should be as low as possible, on the one hand to obtain a great tuning selectivity of the relevant resonant circuit, on the other to avoid heat development and the consequent reduction of the breakdown voltage.

The occurring losses are due in the first place to the series resistance of the electrodes, and a parallel resistance presented by the insulating material. Also at a d.c. load, the insulating material draws a leakage current.

When loaded with an alternating voltage, the current through an ideal capacitor (one without losses) shows a phase shift of minus 90° with respect to the voltage. Owing to the self inductance of the wound metal foil, there is a certain voltage loss that lags 90° behind this current. In effect the voltage across the capacitor is reduced. The result is a lower capacitive current thus, a lower capacitance.

The series resistance of the metal foil gives rise to ohmic losses as also occur in the parallel leak resistance of the insulating material. Owing to these losses the phase shift between the terminal voltage and the current will be an angle $\delta$ less than 90°. Tan $\delta$ is defined as the quotient of the losses in watts and the voltamperes of the ideal capacitor. At an increasing frequency the share of losses through the series resistance increases, the share through the parallel resistance diminishes.

More important, however, are the DIELECTRIC LOSSES which we already discussed in Chapter II.K.1: they are the hysteresis losses brought about by the "pole reversal" of the internal structure of the insulating material. In non-homogeneous materials there are then the additional losses caused by the ionization of air or gas enclosures, which we shall now discuss in some further detail.

In 1921 PROOS of the "Nederlandse Kabelfabriek" succeeded in measuring the tan δ of paper cables. He found that above a certain field strength the dielectric losses increase considerably (Fig. 171): the tan δ versus field strength graph clearly shows a knee. Cables loaded to above this value tend to break down after a few hundred hours. Similar phenomena were displayed by suppression capacitors connected across the mains terminals of radio receivers, by capacitors for neon lamps, in ballasts of fluorescent lighting, for the starting of electric motors, and for cos $\varphi$ improvement. Particularly the fact that such a capacitor would stand a test voltage much higher than the operating voltage, to break down after having functioned for a longer period of time, required a thorough study. Also the question why this breakdown did not occur at voltages below 350 V.

The explanation is as follows. Paper is a porous material, and even after impregnation it will contain small air bubbles. In series-connected dielectrics of different materials the voltage is distributed inversely proportional to k. Hence the field strength in the air bubble is k times that of the insulating material. Here, however, no "breakdown" can occur, for the air bubble is separated from the electrodes by the insulating material. At a field strength of 60 - 80 kV/mm in this air bubble the "ionisation by collision" (accelerated electrons knocking more electrons from atoms) becomes so intensive that considerable heat is developed. A similar phenomenon can be observed on a high voltage line, where it is known as the "corona effect". Under influence of the alternating voltage, this ionisation is reversed continuously, and the heat thus produced very gradually attacks the insulating material round the air bubble until, sometimes after hundreds of hours, it breaks down. Fig. 172 shows an unwrapped capacitor, and we can clearly see

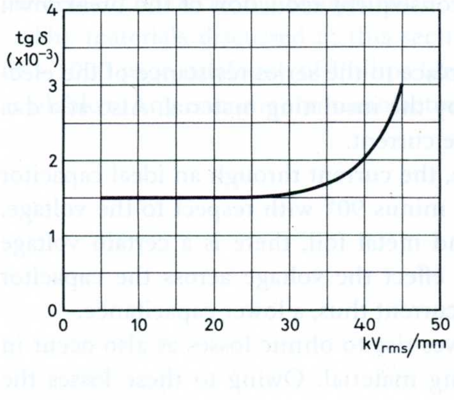


Fig. 171. Tg δ of a paper capacitor as a function of field strength.

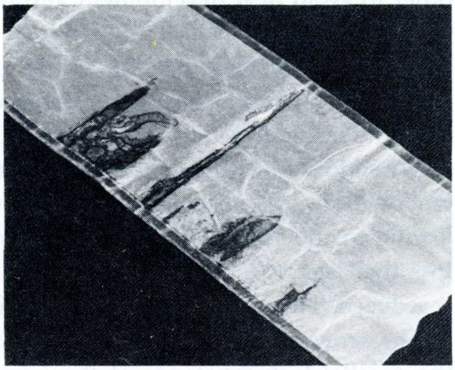


Fig. 172. Paper layer wrapped off a capacitor in which the corona effect has occurred (Philips Technical Review).

where the insulating material has been affected. When trying to find this fault by testing with high voltages, this process is in fact initiated, so that during normal operation the material sooner might be damaged.

According to Paschen's theory the breakdown voltage in air between flat electrodes is practically proportional to the electrode distance. At very short distances, however, the breakdown voltage increases again, for then the electrons have too short a "path" to ionize atoms. Therefore, there is a minimum voltage of about 350 V, below which there is never breakdown in air, and hence no corona.

Electronic measuring methods have provided rather simple means of measuring the tan $\delta$, and so designers and manufacturers of capacitors have found a way to increase the reliability of their products.

The wrapped capacitor package is first evacuated and heated to remove all humidity. Then the capacitor is treated with an impregnating liquid, which has been made very thin by heating, to fill up all pores. Depending on the degree of sealing that can be obtained with the envelope used, oil, vaselin, or wax is used. An impregnating medium used widely nowadays is CLOPHENE, which has a high k (= 4.2) and is but little affected by the corona.

It is found that at a field strength of 18 $kV_{rms}$ per mm, operation is quite safe below the knee point in the tan $\delta$ curve; tan $\delta$ is then less than $60 \times 10^{-4}$ at 1 kHz. The tan $\delta$ of each capacitor is measured to check whether there are any deviations.

The so-called ROLL BLOCK capacitors used in radio receivers were initially mounted in hard paper tubes that were sealed with lacquer at both ends; later they were pressed in wax or plastic. Large paper capacitors were fitted in soldered cans with glass lead-in rings ("compression seal", Chapter II.J.5).

Further development of plastic foils during the years 1950 - 1960 resulted in paper for many applications being replaced by plastic foil. Materials like POLYESTER and POLYCARBONATE can be made in thicknesses upwards from 3 microns; the foil is highly homogeneous and has a breakdown voltage of 200 kV/mm. For polycarbonate the tan $\delta$ is less than $30 \times 10^{-4}$ at 1 kHz, and $100 \times 10^{-4}$ at

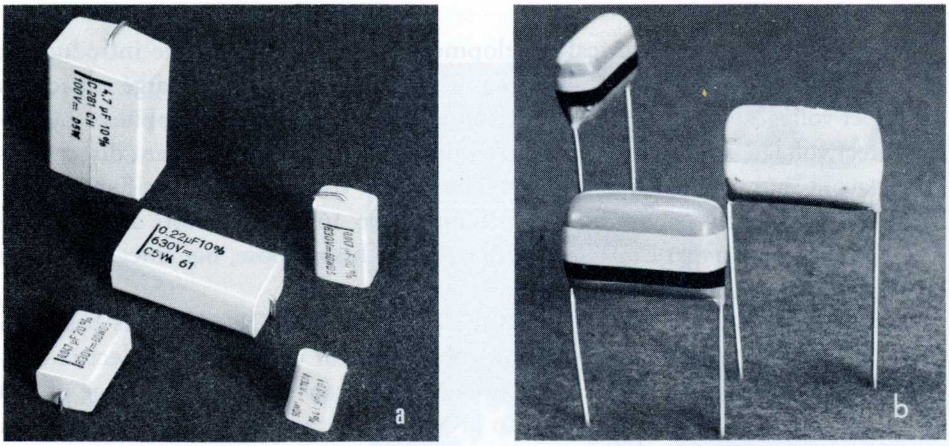


Fig. 173. Some types of polycarbonate foil capacitor: a. pressed into plastic, b. flat version immersed in insulating lacquer.

10 kHz. By depositing the aluminium coatings on the foil in vacuum, a double advantage is gained: room saving and built-in safety. For if the foil should still contain a small hole, the aluminium round it will burn away during the test or in operation (SELF HEALING).

Furthermore a method has been found to make low impedance connections between the terminal wires and the metal coatings. All round, the aluminium layer extends alternatingly to each edge of the foil (EXTENDED FOIL). By metal spraying with zinc, the edge of the wrap is given a cap of zinc which is connected to the coating over its full length. The terminal wire is soldered to the zinc cap.

Foil capacitors too, must be protected against humidity and mechanical damage by means of an envelope. This is done by pressing into plastic (Fig. 173a), and for the smaller types, the FLAT FOIL version, by immersion in an insulating lacquer (Fig. 173b). Thus a whole series of capacitances is obtained up to 6.8 μF, for voltage ratings from 100 - 630 V and frequencies up to 10 kHz.

Owing to the high tan δ, these capacitors are restricted to low frequency applications. For higher frequencies we can use the more expensive material POLYSTYRENE with its tan δ of less than $10 \times 10^{-4}$ at 1 MHz. It can, however, not stand temperatures higher than 85 °C, and hence the coatings cannot be vapour deposited upon it. Since these capacitors are intended for tuned circuits, narrow capacitance tolerances are essential. In this respect a neat solution has been found: the foil is wrapped up under some mechanical stress, and by heating after wrapping, the capacitor is made to shrink. Heating is continued until the required capacitance value has been obtained, and accuracy can be guaranteed to be within plus or minus 1%. The T.C. is slightly negative, so that between 0° and 100 °C the capacitance drops 1% at the most (this compensates for the increase in self induction of the coil). The so-called MINIPOCO capacitors are manufactured for voltages from 63 - 500 V and capacitances up to 0.16 μF. They are particularly used in multi channel telephony.

## II.L.2. ELECTROLYTIC CAPACITORS

When following the historical development, we see that with the introduction of rectifying units in radio sets there was a growing need for very large capacitors for direct voltages up to 500 V. The reason is that a tube rectifier supplies a pulsating direct voltage, and the radio tubes require to be fed with a constant direct voltage to avoid "hum" in the loudspeaker. The output voltage of the rectifier for a radio set is sufficiently "smoothed" by means of two 16 μF capacitors and a choke coil. Initially large paper capacitors were used for this purpose, but since they were large, they were also expensive.

A better solution was the ELECTROLYTIC CAPACITOR, which is based on the principle that aluminium oxide is a good insulator with a high dielectric constant (k = 28). A layer no more than 1 micron thick can withstand a voltage of 500 V, so that a limited area is sufficient to give a large capacitance. If an aluminium electrode is immersed in a suitable liquid (boric acid for instance), a positive voltage is applied to the aluminium, whilst a negative voltage is applied to the liquid, electrolysis will cause a thin film of oxide to be formed on the aluminium. As soon as this film has been formed, it stops all further current conduction, meaning that the oxide film will not become thicker. If, for whatever reason, a small area of insulating film should be lost, a new film is automatically formed there. Small contaminations (chlore) can disturb the process. The oxide film can withstand a small reverse voltage caused by the ripple in the rectified voltage.

Originally these capacitors were made from an aluminium can filled with a liquid into which a rod of aluminium (a drawn profile with the largest possible area) was immersed. The can with its insulated base through which the "anode" was led, had to be mounted vertically in the apparatus. At the top the can was provided with a simple valve to avoid explosion in case gas should be developed. Leaking valves and bases allowed the rather corrosive electrolyte liquid to spread through the apparatus.

As early as in 1935 a method was discovered to manufacture electrolytic capacitors from two aluminium foils (one being oxidized, the other left as it was) which, with an intermediate layer of porous paper were wrapped into a cylinder. The paper was impregnated with the electrolyte after which the whole "wrap" was fitted in an aluminium can. It was not until 1945, however, that it became possible to use this method for making capacitors for higher voltages. In the years thereafter the dimensions could be considerably reduced thanks to a thorough study into the etching process. By etching the surface becomes rough and holes are formed (Fig. 174), so that the active area of the aluminium could be increased up to 30 times. Now the dimensions became so much smaller, and hence the price so low, that even in television receivers the smoothing choke could be replaced by a resistor if two larger capacitors of 100 μF were used.

The manufacture of electrolytic capacitors is rather complicated. The aluminium foil, supplied in rolls, flows continuously through an etching machine, and then through a forming machine (Fig. 175) in which the oxide layer is formed by applying a direct voltage whilst the foil is immersed in the electrolyte. Then the foil is cut into strips which, together with the porous paper, are wrap-

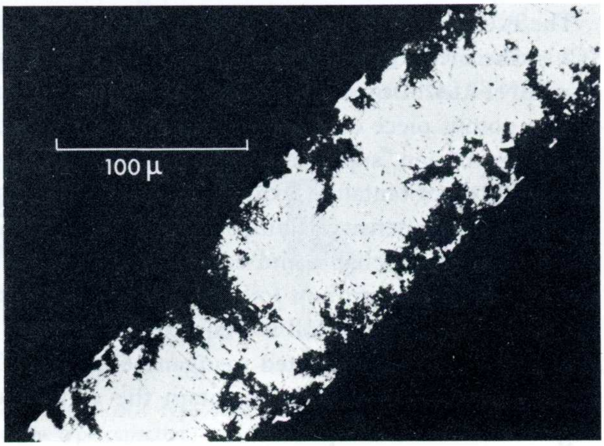


Fig. 174. Microphoto of etched aluminium.

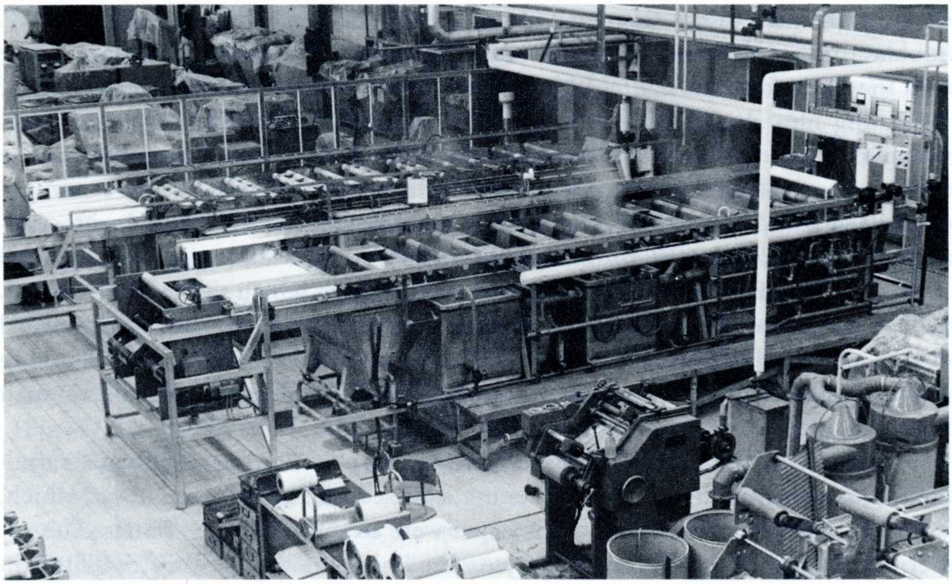


Fig. 175. Machine for forming the cores of electrolytic capacitors

ped round the anode rod which is mounted on a base of insulating material. For some types the "wraps" thus obtained are immersed in the electrolyte in the forming machine, whilst at the same time a positive voltage is applied to the anode (post formation). Afterwards the wraps are placed in an aluminium can into which the base is flanged. Finally the electrolytic capacitors are aged on aging racks.

With the arrival of transistor equipment the use of electrolytic capacitors for "smoothing" direct voltages at various points in the circuitry expanded considerably. Whereas the anode- and screen-grid voltages in tubes are of the order of 200 V, transistors operate at voltages of, say, 9 V. For such low voltages electrolytic capacitors are cheaper than foil capacitors.

The ever present risk of leaks, the limited life and temperature limits (40 to 80 °C), make electrolytic capacitors less suitable for professional equipment. Here dry TANTALUM capacitors are used instead. The anode in these types of capacitor is a porous piece of tantalum which is coated with a layer of oxide, manganese oxide (the oxygen "supplier"), carbon, and silver which forms the cathode. Thanks to the high k of tantalum oxide (35), these capacitors can be very small, but they are also very expensive. Philips have also developed dry electrolytic capacitors which are highly reliable and can be operated from −80 °C. Again the starting point is etched aluminium foil, but now the surface is oxidized by heating in air (pyrolysis). A fibre glass foil is wrapped between the anode and cathode foils, to be impregnated with dehydrated manganese nitrate. This semiconductive layer supplies the oxygen needed to maintain the aluminium oxide. Such capacitors are larger than tantalum capacitors.

## II.L.3. PLATE CAPACITORS

Mica, which is found as a mineral, can be split up into extremely thin plates which have a high breakdown voltage (60 kV/mm), a high k (4 to 8), and low dielectric losses (tan $\delta = 5 \times 10^{-4}$ at 1 MHz). These were the reasons why MICA CAPACITORS were the first to be used in high-frequency oscillating circuits. The coating consisted of vapour deposited silver. The capacitance could be adjusted to the required value by scraping off this silver. A number of such prepared mica plates were stacked and moulded in plastic. However, mica is rather expensive, and the natural sources are too small to satisfy the need of the electronics industry for capacitors.

Therefore they were replaced by polystyrene capacitors (Chapter II.L.1) for high capacitance values, and CERAMIC CAPACITORS for low capacitance values.

As already explained in Chapter II.K.1, pure porcelain (kersima) can be made with a tan $\delta$ of less than $10 \times 10^{-4}$ at 1 MHz and, moreover, the humidity and temperature resistance as well as the stability of a capacitor made of this material are sufficient to meet the so-called class I B requirements. By adding some titanium oxide, the T.C. can be made to match the coil ($-150 \times 10^{-6}$/degC).

The simplest form is the DISC TYPE CAPACITOR (Fig. 176a). The ceramic powder is mixed with liquid and pressed into a disc which is then sintered. The coatings are silver-enamel which, after firing, adhere firmly to the porcelain. Now the terminals are soldered to the coatings, and by immersion into a special lacquer a moisture resistant and electrically insulating layer is applied (this coat of lacquer limits the temperature to 85 °C). To ensure sufficient mechanical rigidity, the porcelain cannot be made much thinner than 0.5 mm, and therefore the capacitance is limited to 18 pF at a diameter of 7 mm.

A tube of ceramic material will be stronger before as well as after firing, and therefore it is permissible to use wall thicknesses upwards from 0.2 mm, so that a capacitance of 270 pF can be obtained with a tube 34 mm long and 4 mm wide. The tubes are freely suspended during firing so that they will not be contaminated from the tray. They are internally silver coated by absorbing the silver slurry. On the outside two silver rings are fitted (Fig. 176b), one of which serves as the second electrode, and the other as the connecting point for the internal coating, and to these rings terminal wires are soldered. This is the TUBULAR CAPACITOR.

With another version a helical wire is soldered to the inside, so that these capacitors can be mounted up-right (Fig. 176c) (the term of the art used for them is

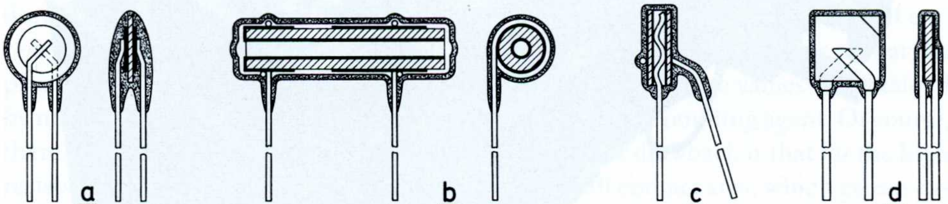


Fig. 176. Various types of ceramic capacitor: a. disc type capacitor, b. tubular capacitor, c. vertical tubular capacitor (pin-up), d. miniature disc type capacitor.

"pin-up", perhaps also because of the beautiful colour code applied by immersion in different colours of lacquer). The outer helix can also be replaced by a flange, and then we have a FEED-THROUGH CAPACITOR.

Since silver is expensive, substantial saving was made when it was found possible to use copper for the coating.

The ever smaller filter coils (Chapter II.N.1) compelled physicists to find ways of making still smaller plate capacitors with a thinner dielectric. They managed to roll the ceramic paste down to a thickness of 0.1 mm in sheets of 8 × 20 cm which were then sintered. Nickel is now deposited on these sheets, which has the advantage that there is no "migration" (movement of metal particles caused by the electric field), whilst moreover it can be removed by sand blasting. The nickel-coated sheets are cut into rectangular pieces of 6 × 7 mm or smaller. Then nickel is removed by sand blasting until the required capacitance value has been obtained. This done, the terminal wires are welded on an automatic machine (Fig. 177) that supplies the capacitors with the terminals on a strip of resin bonded paper. This strip is rolled up, and thus attached, a large number of capacitors can be immersed simultaneously in insulating lacquer. The result is the MINIATURE PLATE CAPACITOR (Fig. 176d), which can be made to have a value up to 150 pF on one single plate.

For lower frequencies (class II), barium titanate compositions can be used (Chapter II.K.1) with which much higher capacitances are obtained, to wit: 1500 pF for disc-type capacitors, 22 000 pF for tubular capacitors and 4700 pF for

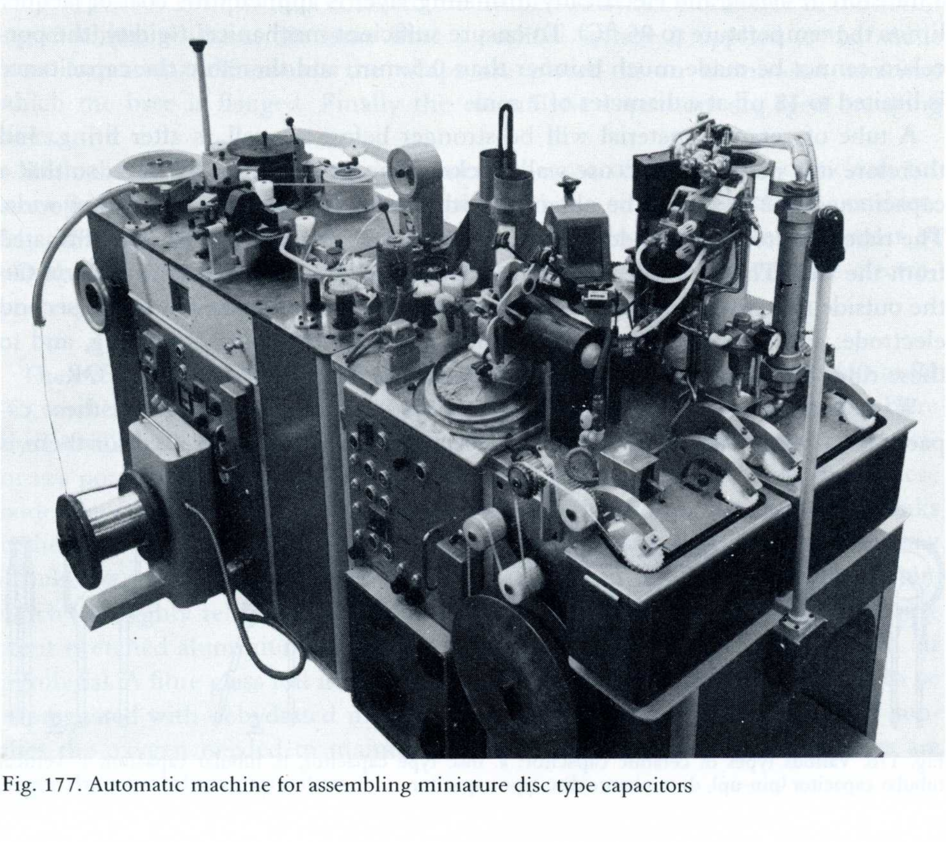


Fig. 177. Automatic machine for assembling miniature disc type capacitors

miniature plate capacitors. The latter can fulfil several functions in the low frequency circuit. And then the BARRIER LAYER capacitors, also described in Chapter II.K.1, can be supplied as "micro plate" types for voltages up to 6 V and with capacitances up to 100 000 pF.

Thus the study of "white" ceramics has led to a series of cheaper capacitors which can be used for several functions.

## II.L.4. WIRE-WOUND RESISTORS

These are the oldest types of resistor used in electrical engineering, which are made by winding a wire round an insulating (usually ceramic) tube. The wire is made of a material with a high specific resistance and a low temperature coefficient. Of the alloys obtainable on the market we must mention two:

- CONSTANTANIUM, a rather precious alloy of copper, nickel and manganese with a specific resistance of 0.49 ohm per $m/mm^2$ and a temperature coefficient of $0.02 \times 10^{-3}$ per deg C, which is used for precision resistors.
- NICHROME, a cheaper composition of chrome nickel, which has a specific resistance of 1.1 ohm per $m/mm^2$ and a temperature coefficient of $0.3 \times 10^{-3}$ per deg C. This material is resistent against high temperatures and is generally used for heating elements, but also in cases where a greater resistance variation due to temperature changes is permitted.

The winding on the ceramic tube must be protected by a moisture-repellent coat. For high temperatures glass enamel or cement is used which is a good electrical insulator, but also a good heat conductor.

The resistance can be measured during winding, so it is possible to work within close limits. Connecting the thin nichrome wire to the end terminals involves some difficulty. Spot welding is too difficult; same as soldering. The solution was found in arc welding in a reducing atmosphere, which gives a corrosion-proof reliable joint.

For the high resistance values as often required in electronic circuits, wire-wound resistors are difficult to produce. For a resistance of 1 MΩ, no less than 170 metres of the thinnest possible nichrome wire (15 microns thick) is required.

## II.L.5. CARBON AND METAL FILM RESISTORS

Carbon has a high specific resistance (40-100 ohm per $m/mm^2$), but cannot be drawn into a wire. One method of making carbon resistors is to extrude a rod of carbon powder bonded with plastic, cut it into pieces and bake these pieces so that the plastic hardens (COMPOSITE RESISTOR). The high resistance values are obtained by making a mixture of little carbon and a large quantity of bonding agent. Of course, there is a very large spread in resistance values. Another drawback is that for the high resistance values the carbon particles have a very small contact area, which causes extra "noise".

A better solution is offered by the CARBON FILM RESISTOR, where a thin film of carbon is applied to a porcelain rod. For higher resistance values a helical groove is cut into this carbon film, so that the resistor proper is a helical strip of carbon along the surface of the rod. There are now two ways of changing the resistance value. Firstly by altering the thickness and composition of the carbon film, and secondly by choosing the appropriate pitch of the helix. By measuring the resistance during cutting, the helix can be terminated as soon as the required value has been obtained.

Of course a number of difficulties must be overcome: the thin carbon film must have a low temperature coefficient, after lacquering it must be resistant against humidity, corrosion and mechanical damage, and it is no easy matter to establish a reliable joint between carbon film and terminal wire. Finally the carbon film must be quite homogeneous and adhere well to the porcelain: if in the cutting operation one particle of carbon comes off, this may well cause a complete interruption of the resistor.

It took, indeed, twenty years (1930-1950) before this procedure could be used to make absolutely reliable resistors. At first the layer was applied by immersing the body of the resistor in an emulsion of carbon in sugar-water. When heated in air the sugar would burn to carbon. This process was hardly reproducible. It stands to reason that a film of pure carbon, applied in thin layers, will yield a more reliable product. In the present manufacturing method we start from rods of kersima, whose surface is first smoothed by "barrelling". Then the rods are placed in a rotating drum in which a hydrocarbon gas is heated to 1050 °C, the temperature at which the gas decomposes (pyrolysis). The carbon deposits on the porcelain rods to form a uniform film which grows thicker as the process is continued. The carbon resistors thus produced have a very low negative temperature coefficient (up to 100 k$\Omega$, $3 \times 10^{-4}$/deg C) and very low noise values (up to 100 k$\Omega$ less than 4 dB). The highest obtainable resistance value is 10 M$\Omega$, but than the film is only a few thousandths of a micron, that is to say a few molecules, thick. As a result the noise value becomes much higher.

Fig. 178 shows in sequence the various production stages of the carbon film resistor. The starting point is the barrelled rod of kersima (a). In a pyrolysis process of gas it is coated with a carbon film (b). Then caps of a resilient material (tombac) are pushed over the ends to ensure good contact with the carbon film. Now $R_0$ is measured and the pitch for the required resistance value is calculated. The helical groove (c) is cut on an automatic machine with an extremely thin grinding wheel which is lifted as soon as the required value has been reached. (For the higher resistance values a much finer groove is now "cut" with a laser beam). This done, the connecting wires are percussion welded to the contact caps (d). Finally, a three-fold, moisture-repellent and electrically insulating layer is applied with on it

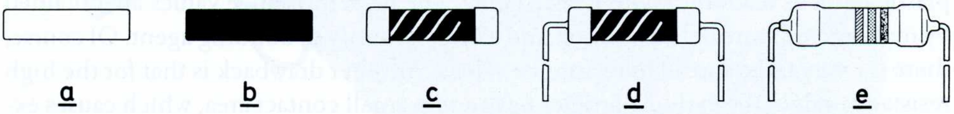


Fig. 178. Various manufacturing stages of a carbon film resistor: a. polished porcelain rod, b. vapour deposited carbon film, c. caps pushed on and grooved, d. wires welded on, e. dipped in lacquer with colour code.

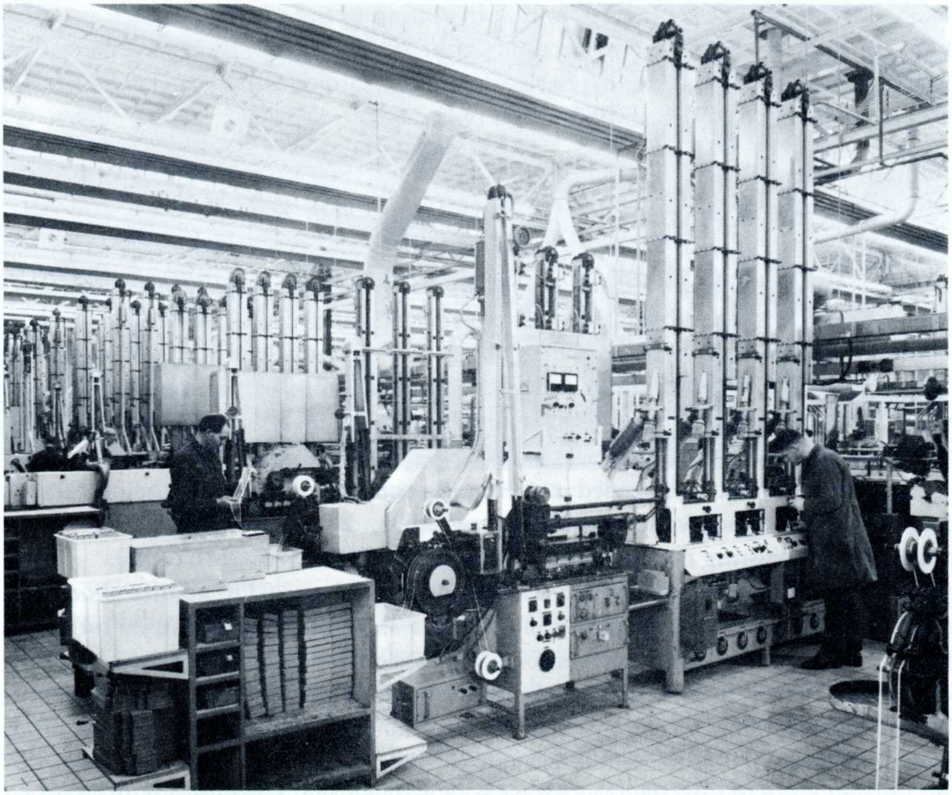


Fig. 179. Machines for the finishing operations in a carbon film resistor factory.

the colour code. This layer of lacquer does not only serve as a protection against humidity, but also against mechanical damage and galvanic contact. Finally the resistors are measured and attached to a paper strip; Fig. 179 shows the machine on which the products are finished. The carbon resistor is probably the electronic component to be manufactured in the greatest numbers, and so a high degree of automation could be developed.

Notwithstanding the important improvements made as regards reliability and stability, there are still a few weak points left. Moisture sensitivity is limited and cannot be improved by using a thicker layer of lacquer, for then the heat removal would be endangered. At higher ambient temperatures and high loads, the resistance value gradually decreases. As a result, when the resistor is connected to a constant voltage, the current will increase with the temperature, which in turn increases the current, and so on.

In order to meet the stringent requirements for certain professional applications, the METAL FILM RESISTORS were developed. Here a film of chrome nickel is vapour deposited in vacuum on the porcelain rod; all further operations are similar to those of the carbon film resistor. The metal film has indeed a lower resistance, so that it is impossible to obtain values higher than 1 M ohm, but the temperature coefficient is lower ($1 \times 10^{-4}$ per deg C), and the temperature and moisture resistance are better. An additional advantage is that there is less resistance drift caused by lacquering, so that in cutting to the accurate value it is possible to guarantee resistance tolerances of about 1%. Of course, owing to the expensive chrome nickel, and the expensive vacuum treatment, these metal film resistors are more expensive than the carbon types.

## II.L.6. POTENTIOMETERS

This term, originally used for a voltage division by branching a resistor (like "volume" control on a radio set) has gradually been accepted as the term for a continuously variable resistor. We shall not give a detailed discussion of WIRE WOUND POTENTIOMETERS where a contact is moved along a resistance wire wound on a cylinder or toroid of insulating material, but we shall restrict ourselves to the CARBON POTENTIOMETERS.

These components are needed in electronic equipment to enable the user to adjust certain quantities (volume, timbre, brightness, contrast etc), or to trim certain values when adjusting the apparatus (TRIMMING POTENTIOMETERS).

With the carbon potentiometers the carbon film is applied to a horse shoe plate of synthetic resin bonded paper, along which a sliding contact is run by turning a spindle. Both ends of the horse shoe are silver coated over the carbon layer for connection of the terminals. The carbon layer must satisfy quite some demands. The resistance values needed in the circuitry range from 1 k ohm to 1 M ohm, with tolerances of ±20%. These tolerances must be maintained at a certain humidity and temperature, whereas no lacquer can be used. Furthermore the contact face must be extremely smooth so as to minimize friction during sliding, and hard wearing to prevent the layer from wearing away too soon in normal use. For volume control it should be borne in mind that the sound impression is directly proportional to the logarithm of the power, so that for this particular application the logarithm of the resistance value must vary directly proportional with the angle of rotation.

The result of tedious experiments is a "composite layer" consisting of carbon powder and resin, carefully applied to the resin bonded plate after which it is baked. The ultimately obtained resistance values can to some extent be influenced by controlling the baking temperature. This has yielded a well functioning and reasonably reliable product. A great problem was, and still is, "crackle". The user, when adjusting the volume, does not wish annoying crackle in his loudspeaker. When the sliding contact moves over the carbon particles, there will unavoidably be resistance variations which will produce all the louder crackle from the loudspeaker as the current through the contact is higher, and the branch voltage is lower. These conditions must be taken into account with the circuit, and it is particularly in transistor sets that the voltage across the contact is lower. Although both the absence of crackle and the reliability have improved considerably over the years, the quality of this carbon potentiometer is inadequate, certainly for professional applications. Physical research is moving in the direction of a conductive glass film on a ceramic plate; this method will certainly lead to a product resistent against tropical conditions.

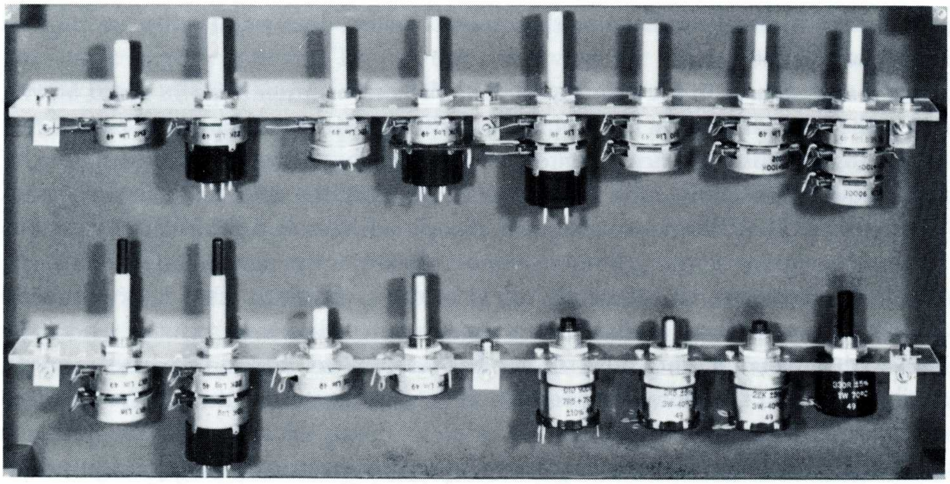


Fig. 180. Different types of carbon potentiometers

The potentiometer plate is fitted in a metal box which also contains the bearing. Two of these boxes assembled together can be fitted with one spindle or two concentrical spindles to form the "tandem" potentiometer.

For many years it has been customary to combine the mains switch with the potentiometer in the sense, that the switch is operated in the "idle" travel of the stroke (where the resistance is still zero). For the direct connection to the mains the switch had to be bipolar and meet the local safety regulations. It was not so simple to operate such a switch by rotating the potentiometer spindle through an angle of no more than 10 or 15°, all the more so when owing to miniaturization the potentiometer dimensions were reduced considerably. Fig. 180 gives a survey of the different types of potentiometer with and without a switch.

## II.L.7. QUARTZ CRYSTALS

Although the quartz crystal technique cannot be fully dealt with here, it is so interesting that at least a short discussion is desirable.

In Chapter II.K.6 the piezoelectric effect was already discussed: generating an electric field under mechanical stress, and being strained by applying an electric field. In the latter case a suitably formed plate will, under influence of a varying electric field strength, begin to vibrate at its natural resonant frequency. A capacitor, formed by the two metal electrodes on either side of this plate, will pass a very high current at this particular resonant frequency. This phenomenon also occurs with a series connection of self inductance and capacitance, whose resonant frequency is $n = 1 : 2\pi\sqrt{LC}$ but here the sharpness of the resonance peak is limited by the losses in the coil. As will be explained in Chapter II.M.1, the sharpness of this resonance peak depends on the quality factor $Q = \omega L/R$, which even for the professional coils cannot rise above 300. In a well designed piezo-electric "RESONATOR" the mechanical and electrical losses are so low that this Q is much higher (for the quartz crystals dealt with below, as high as 30 000 to 300 000).

Such a crystal, connected in the oscillator of a transmitter (Figs 78 a and b) will, when the grid and anode circuits are tuned to about this frequency, ensure that these circuits will oscillate accurately at the resonant frequency of this crystal. If this frequency is stable, and independent of humidity and temperature, it has been ensured that the transmitter will always radiate a certain frequency. If furthermore a similar crystal is fitted in the h.f. circuit of the receiver, we may be certain that this receiver is accurately tuned to the transmitter. This is of particular importance in military applications where mobile units must remain in contact with each other, but where the soldier has no time to "tune in" on the required station. In 1944, therefore, the production of such crystals in the U.S.A. ran up to a total number of 28.5 million.

Quartz, in its natural state, has a fairly high piezoelectric efficiency, and furthermore its frequency is independent of temperature. However, this is only the case when the slices are cut at a certain angle with the X, Y and Z axes of the crystal shown in Figs 181 a and b, and with an accuracy of 15 minutes. In the raw crystal, the crystalline structure is difficult to recognize.

The only way to find the directions of the X, Y, and Z axes is to make test samples and to inspect them according to the X-ray diffraction method. These quartz crystals could not be mass produced until the X-ray equipment needed had been improved (this was done in close cooperation between the research laboratory, the crystal factory and the X-ray equipment factory of North American Philips). Once the correct cutting angle has been found, the slices are cut slightly thicker than actually required for the required frequency. Then the slices are cut down to the required size and are ground between metal plates with an abrasive powder (lap-

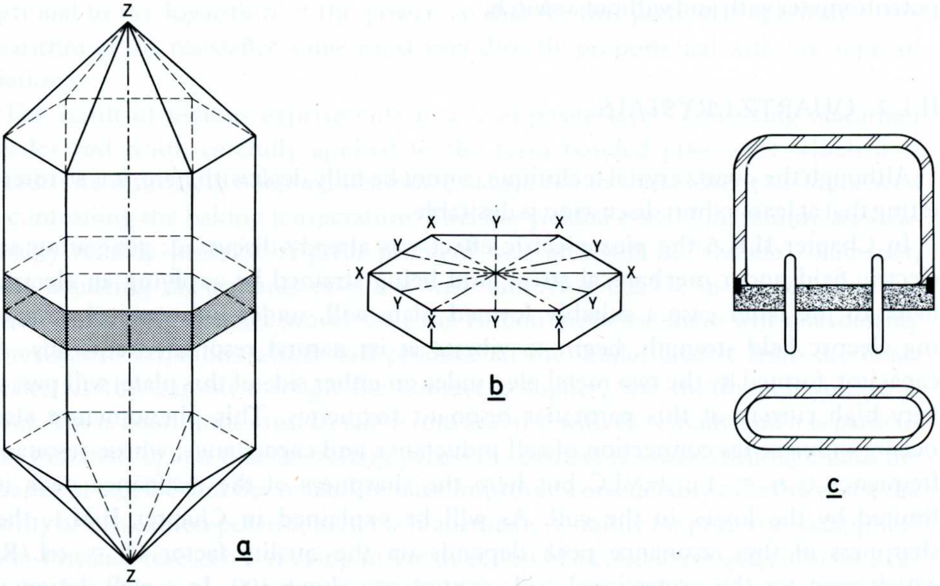


Fig. 181. Quartz crystals: a. hexagonal structure of the crystal, b. X- and Y-axes, c. glass envelope with kovar sealing ring.

ping). During this process a special device on the machine ensures that the slices are made slightly convex. This is done to avoid the occurrence of harmonics of the mechanical vibration. Then the last microns of the thickness are removed by etching, during which process the resonant frequency is constantly checked. With the "shear mode" vibration occurring here, the resonant frequency depends on the thickness of the slice. A crystal for a frequency of 20 MHz has a thickness of 80 microns, and to obtain a frequency tolerance of $\pm 5 \times 10^{-6}$, the thickness tolerance is $4 \times 10^{-4}$ microns.

The crystal thus made is provided with electrodes by depositing metal, or clamping between metal plates, afterwards to be fitted in a moisture-repellent envelope. Since the crystal cannot stand the high sealing temperature of glass, the seal is made in accordance with Fig. 181c (developed by Philips Canada). A kovar ring is placed between the glass envelope and the powder glass base. This ring is then high-frequency heated so that the envelope is sealed off at minimum heat development.

Besides the application in transmitters, quartz-crystals are used for multi channel telephony. Since in the PAL colour television system the subcarrier is not synchronized from the mains, but is contolled by a crystal, every colour television receiver needs a quartz crystal.

## II.L.8. CERAMIC RESONATORS

At the end of Chapter II.K.6 we discussed piezoelectric resonators, which consist of a piece of piezoelectric ceramic which is silver-coated on either side. By grinding, the resonant frequency can be adjusted as required, with a guaranteed accuracy of ±0.2%. At a variable frequency the impedance varies as shown in Fig. 182a; at $f_r$ it is almost zero, at $f_a$ very high. The Q factor is lower than for quartz crystals, but still ranges from 500-1000. These types of resonator find a most important application in INTERMEDIATE FREQUENCY FILTERS.

In a superheterodyne set the received wave lengths are transformed into a fixed intermediate frequency: usually 452 kHz plus or minus the audio frequency (4 kHz). In the i.f. amplifier the quality and selectivity of the receiver are determined: the signal required (452 ± 4 kHz) must be amplified as much as possible, all other frequencies as little as possible to get rid of adjacent transmitters. Ideally the attenuation curve should be rectangular, see the dotted line in Fig. 182b. The common bandfilter is still very remote, for it is impossible to obtain a Q much higher than 100 at reasonable costs. At least two of them must be connected in series to obtain the resonance curve indicated by the dashed line in Fig. 182b. In doing so the attenuation at ±9 kHz is 12 times. If, however, as schematically shown in Fig. 182c, a ceramic filter is connected between the two coils of one band filter, the low attenuation is maintained at ±2 kHz (owing to the losses in the ceramic filter), but the flanks are much steeper, and at ±9 kHz an attenuation of 62 times is obtained (full line in Fig. 182b). So now a band filter is saved, and selectivity has improved considerably. Moreover, it is not necessary to trim the i.f. coils.

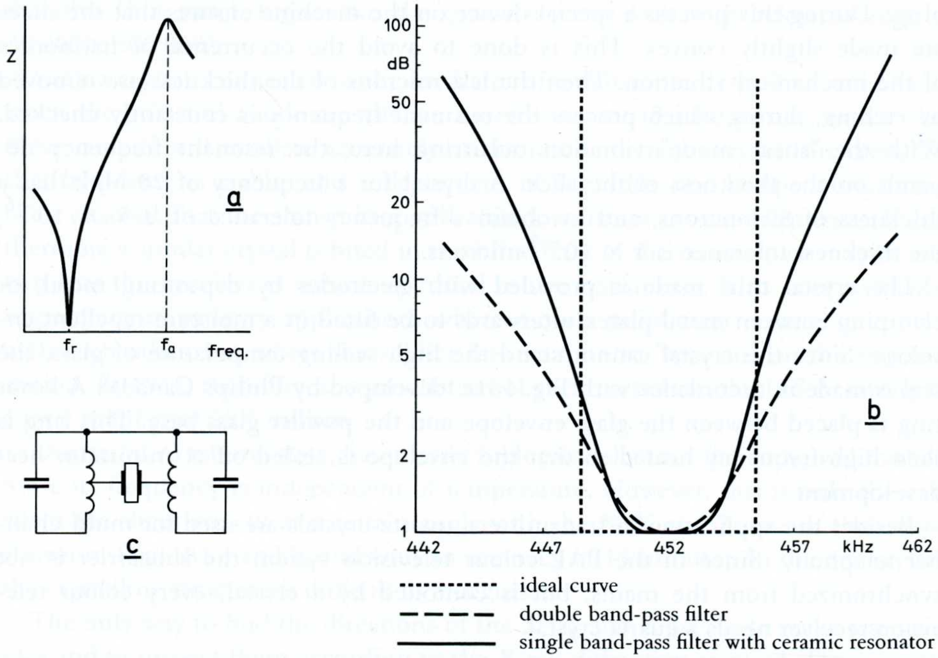


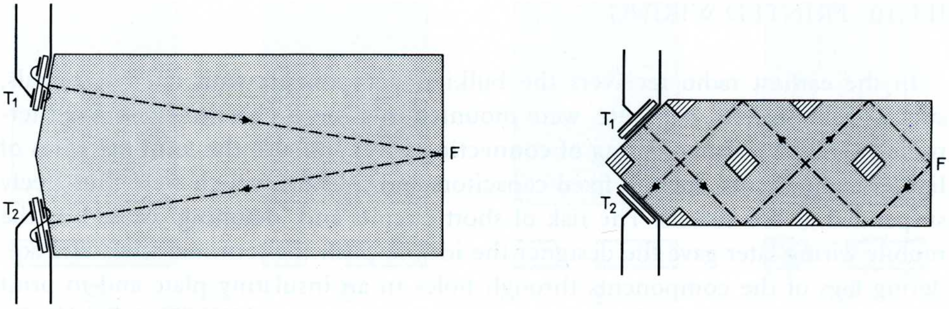
Fig. 182. Ceramic resonators: a. reactance as a function of frequency at a constant voltage of a ceramic resonator (r = point of resonance), b. attenuation as a function of frequency, c. circuit diagram of a band-pass filter with a ceramic resonator.

These band filters are also available for the intermediate frequency of F.M. receivers: 10.7 MHz. A further interesting application is that the steep change from $f_r$ to $f_a$ (Fig. 182a) makes these resonators useful as F.M. DISCRIMINATORS.

For multi channel telephony, where a rectangular tuning curve is desired for the CARRIER SUPPLY FILTERS, quartz crystals are widely used in the U.S.A. and the U.K. In Europe these installations are mostly equipped with rather complex high-quality band filters. Incorparating ceramic resonators in these filters also offers possibilities of making them smaller and improving their quality.

## II.L.9. DELAY LINES

An example of an electronic component which, after a very short development period, had to be available in large numbers is the DELAY LINE. Relatively short before the beginning of colour television broadcast in Europe, it was decided to adopt a system deviating from the American N.T.S.C. system. This was the SECAM system in France, and the PAL system in the rest of Europe. In all these systems an integral (black and white) brightness signal is transmitted for the sake of compatibility. The colour information is added by a signal that indicates the difference between the red, another the difference between the blue and this integral signal. The green is obtained by subtracting red plus blue from the white. To save bandwidth, the SECAM system does not transmit the red signal during the even, and no blue during the uneven lines of the raster. These colour signals are



a. first type (1968) glass thickness 20 mm b. smaller type (1970) glass thickness 0.8 mm

Fig. 183. Delay lines for colour television receivers operating on the SECAM or PAL system.

modulated on a subcarrier of 4.43 MHz. Since for each line of the raster the picture tube requires all three colour informations, the red and blue informations are alternately "stored" for a period equal to one line: $2/50 \times 625 = 64$ μsec. The PAL system also uses a delay line, but its operation is somewhat more complicated.

The delay can be obtained as the propagation time of a longitudinal vibration with a carrier frequency of about 4.43 MHz through a glass body. Fig. 183 shows how the colour signal is fed to a resonator T1. Here the voltage variations are converted into a mechanical vibration which is transferred to the glass body. The wave, following a course as indicated by the dashed line in Fig. 183a, is reflected by plane F and lands on a second resonator T2 where the mechanical vibration is converted back into a voltage. This voltage has the same time function as the one applied to T1, but is slightly delayed. By giving the glass body the correct dimensions it is possible to obtain a delay of 64 μ sec. It is, however, also necessary that the phase is equal, and therefore a tolerance of no more than ±5 n sec is permissible. This can be realized by grinding plane F until equality of phase is obtained, accurate measurements being made during grinding.

The glass thickness of the design shown in Fig. 183a was about 2.5 cm, but closer investigation showed that a thinner glass body is more efficient. Therefore a smaller type was designed in which the glass was only 0.8 mm thick, while owing to the zig-zag pattern of the wave, the glass body can also be made shorter. The hatched square planes are coated with lacquer to damp any waves which would tend to travel from the input to the output taking a way other than the required trajectory.

Delay lines have been used in earlier electronic equipment, but the demand for large numbers of them for colour television receivers has led to an industrial development owing to which the price could be substantially reduced in two years' time.

Such a delay line can also serve as a memory. A train of "1" or "0" pulses with a frequency of 4.0 MHz is fed into it. In 64 microseconds $4 \times 64 = 256$ pulses (and an equal amount of "bits" information) can then be "stored" in the glass body. The pulses from the output (T2) are amplified and again fed to T1. Thus the information keeps circulating and can be "read out" at any required moment.

## II.L.10. PRINTED WIRING

In the earliest radio receivers the bulkiest components with fixed terminals, such as tube sockets, coils, etc. were mounted in a metal chassis. These fixed terminals served for the soldering of connecting wires and also the terminal wires of lighter components, such as fixed capacitors and resistors, which were thus freely suspended in the chassis. The risk of short circuits and "detuning" caused by the mobile wiring later gave the designer the idea to push the terminal wires and soldering tags of the components through holes in an insulating plate and to print thin metal layers on this plate to serve as connecting wires (PRINTED WIRING).

In its simplest form printed wiring consists of a hard paper plate, with on one side the components, and with printed connections on the other. The component terminals protruding from the plate are then soldered to the printed wiring. The starting point is a hard paper plate provided with copper foil on one side. Then a mask is made of fine gauze (SILK SCREEN) on which the areas forming the connection pattern is left open, whereas the rest is covered with lacquer. With a brush lacquer is now applied to the copper foil through the open areas of the silk screen. Then the copper foil is etched away from the places where there is no lacquer, and finally the lacquer is rinsed off.

Over the years this procedure was improved as follows:

— The conductor pattern is applied according to the photoresist method, as discussed with transistors. A transparency is made on which the pattern is left transparent. The copper layer on the resin bonded paper is covered with a photosensitive lacquer which, having been exposed, is not affected by various solvents. After exposure and rinsing the non-exposed areas, the printed wiring remains covered, while the rest of the copper can be etched away. With this system a much finer pattern can be made, that is to say, tracks 0.3 mm wide, and isolating distances of 0.3 mm. As a result it is now possible to include small self inductances in the form of spirals in the printed wiring.

— The thin copper foil is easily mechanically damaged, and soldering involves a risk. Therefore this layer is reinforced by means of a layer of galvanically applied copper, followed by a coat of lead-tin to which the solder adheres firmly. Where contact springs must rest on the "print", or move along its surface, the area involved is coated with gold or nickel-rhodium.

— To make cross connections, the wiring is often applied to both sides of the insulating plate; then there are a few places where the foils on both sides must be electrically connected. At those places holes are drilled in the plate and since soldering to the thin copper layer is a critical job, it is recommendable to metallize the holes internally. This has also the advantage that soldered joints with terminal wires are more reliable.

Thus we arrive at the modern type of printed wiring as shown in Fig. 184. The starting point is a bilaterally copper-clad plate (a) of insulating material for which synthetic resin bonded paper has been replaced by the stronger and more tropical resistent GLASS EPOXY (fibre glass pressed with epoxy resin).

This plate is drilled as shown in b; the minimum hole diameter being 0.8 mm.

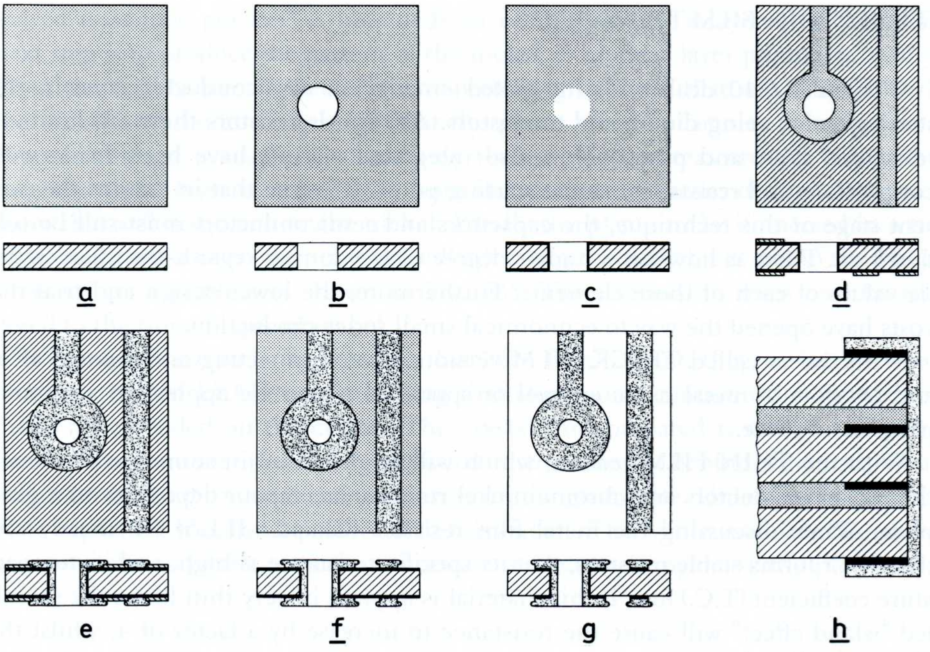


Fig. 184. a-g. Printed circuit panels with silver-lined lead-throughs in various stages of production, h. cross section of a multilayer board.

Internationally the pitch is standardized at one tenth of an inch (2.54 mm).

As illustrated in c, an "electroless" method is used to apply a layer of copper to the copper foil and holes, then a thicker layer of copper is galvanically deposited. Then, according to the photographic method, a protective layer is applied to the areas to be etched away later (d). The open areas are now given a coat of lead-tin (e). This done, the lacquer is rinsed off (f), and finally the copper, where not protected by the lead-tin layer, it etched away (g).

Manually soldering the joints in the holes is a time consuming job, the success of which depends on the attention of the worker. An automatic method has been developed in which the print plate is very gradually and perfectly horizontally immersed in a bath of molten tin which is kept in motion. If all planes are absolutely clean, and a "flux" is added, the surface tension will cause the tin to creep into the holes, surrounding the terminal wires.

More complex circuits often require a larger number of cross connections. In those cases the printed wiring is made in several layers by stacking and compressing a maximum number of twelve of the plates described above with layers of semi polymerized glass epoxy between them. This package is heated on the press, so that the intermediate layers harden to give a compact unit (Fig. 184h). After pressing, the holes are drilled and internally copper-plated (MULTI-LAYER).

This is another example where the manual skill of soldering the terminal wires has developed into an industrial process which requires an extensive physical-chemical research, precision engineering, and pin-point process control.

## II.L.11. THIN FILM TECHNIQUE

Chapter II.J.10 dealt with integrated circuits on semiconductor crystals, the starting point being diodes and transistors. As regards resistors there are limits to resistance value and power. Now also integrated circuits have been made with connections and resistances as the starting point. It is true that in view of the present stage of this technique, the capacitors and semiconductors must still be soldered on. There is however a higher degree of freedom as regards the characteristic values of each of these elements. Furthermore, the lower design and trial-run costs have opened the way to economical small series production.

With the so-called THICK FILM version, silver connecting and contact areas and resistors of metal oxide enamel or sputtered zircon are applied to a ceramic plate or substrate.

With the THIN FILM version, which will be discussed in some further detail, the nickel conductors and chrome nickel resistors are vapour deposited on a glass plate. When discussing the metal film resistors (Chapter II.L.5) we already saw that NiCr forms stable resistors, that its specific resistance is high, and its temperature coefficient (T.C.) low. If this material is applied in very thin films, the so-called "island effect" will cause the resistance to increase by a factor of 4, whilst the T.C. becomes even lower. At a specific resitance of 4 ohm per m/mm$^2$ and a film thickness of 0.01 micron, a strip of deposited NiCr, 10 mm long and 0.1 mm wide, is sufficient to provide a resistance of 40 000 ohms. Higher values can be reached by means of the zig-zag configuration (Fig. 185).

This film is made by heating an NiCr wire in a high vacuum to slightly below the melting point. The glass plate is mounted facing this wire. The NiCr evaporates and settles on the glass as a uniform film.

It is true that initially, owing to the higher vapour pressure, more Cr than Ni evaporates, but as the content of Cr in the wire reduces, this is automatically compensated. The resistance tolerance depends on the width tolerance of the resistor strip, and it is again the photoresist process that brings the solution. First the whole glass plate is vapourized with NiCr, this film is coated with a photo sensitive lacquer and exposed through a mask having the form of the resistor. Afterwards the unexposed areas are etched away.

By applying the connections as a layer of nickel, we arrive at a very simple manufacturing process. First the nichrome film is deposited on the glass (until the re-

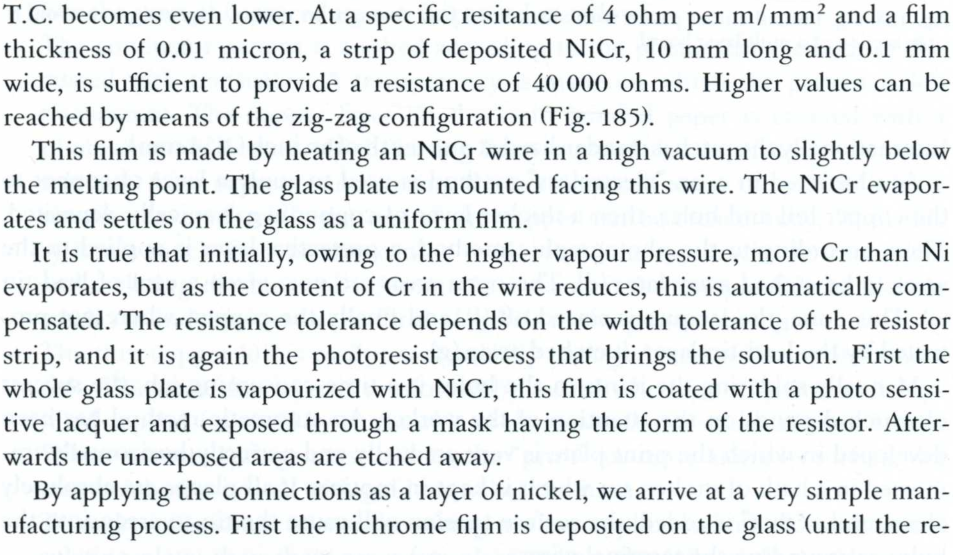


Fig. 185. Thin-film resistance network; black = nickel, grey = chrome nickel.

quired resistance per cm² is obtained), to be followed by a layer of nickel. The next step is to produce the pattern of the nickel + chrome layer photographically. Then the double layer is treated with an etching medium that will dissolve both the nickel and the chrome. After that the photo sensitive lacquer is applied and exposed through a mask in which only the conductive areas are left black, and then the nickel layer is etched away from the resistors by means of an etching medium that will only remove the nickel. Finally the plate can be immersed in a tin bath, making use of the fact that tin will adhere to the nickel, but not to the NiCr. Thus we get a product as shown in Fig. 185. The connecting wires now be soldered to the tinned layer of nickel.

If transistors are to be soldered to this "thin film" circuit, they preferably must be of the BEAM LEAD type, meaning that aluminium contact strips (6 microns thick) are provided on the crystal. The crystal thus prepared is etched along the circumference so that protruding aluminium strips are left behind which are ultrasonically soldered to the aluminium areas on the glass plate. Unlacquered ceramic plate capacitors too, can be soldered to this circuit. The whole assembly is pressed or moulded into plastic to make it moisture resistent and mechanically protected.

## II.L.12. AFTERTHOUGHT, RELIABILITY

The components discussed in Sections II.K and II.L were initially manufactured by Philips' product division ICOMA, established in 1951, which was merged with the product division Electron Tubes and Semiconductors into what is now ELCOMA (Electronic Components and Materials). During the inauguration of the new ICOMA building in 1965 the technical director of this group, DROS, gave a lecture titled "A PALETTE OF APPLIED PHYSICS". The speech was a clear indication that apart from the refined electronic measuring techniques, the modern

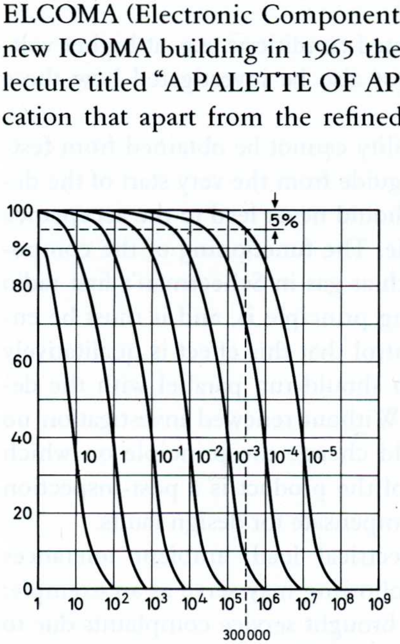


Fig. 186. Reliability of the apparatus (vertical) as a function of the number of components (horizontal) with the percentage of faulty components as a parameter.

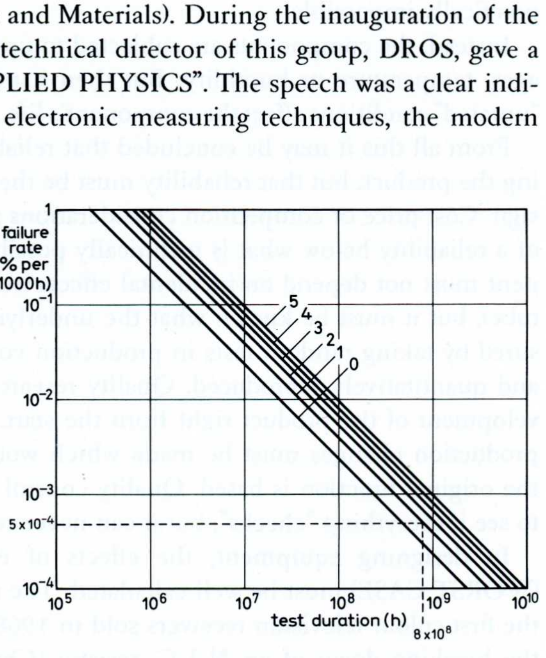


Fig. 187. Failure rate at a 90% confidence level. The percentage of faults as a parameter.

technologist requires knowledge of widely differing subjects belonging to the realm of physics to be able to create all these products.

When discussing radio tubes we already mentioned RELIABILITY. Here there is a distinction between "early failures" (caused by dust, for instance), incidental failures occurring during life, and inherent "end of life" failures (cathode "wear"). This last category of failures no longer occurs in semiconductors and the components discussed here (except perhaps in electrolytic capacitors). This has added substantially to reliability requirements, which became necessary as the circuits were becoming more complex owing to the larger number of components they used.

Let us consider, for example the "early failures" in a missile containing 300 000 components. According to the Gaussean distribution ($F_a = F_1 \times F_2 \times \ldots F_n$) (Fig. 186), one fault in 10 million means that probably 5% of the missiles produced will not function.

A component manufacturer can hardly be expected to prove the reliability of his product by a random test of 10 million samples. A suggestion might be to subject the components to heavier tests (e.g. increased load) to filter out the weak items, but then there is always the risk that "pass" components have passed the test at the cost of a slight damage.

A long-term reliability assessment requires life tests to be carried out. The required "failure rate" for large telephone exchanges amounts to $5 \times 10^{-4}\%$ per 1000 hours, and according to Fig. 187 this means that $8 \times 10^8$ component hours may produce no more than one single fault. (Roughly speaking one year is 10.000 hours, so 80.000 components would then have to be tested for a full year). This is practically impossible.

Instead, the components are subjected to uprated conditions, e.g. at higher voltages, temperature or humidity. But then it must also be investigated how these "uprated" conditions affect the component's life.

From all this it may be concluded that reliability cannot be obtained from testing the product, but that reliability must be the guide from the very start of the design. Cost price or competition considerations should never lead to the acceptance of a reliability below what is technically possible. The functioning of the component must not depend on incidental effects (such as gas in Scheerman's first radio tube), but it must be known what the underlying principle is, and it must be ensured by taking random tests in production control that this effect is qualitatively and quantitatively reproduced. Quality research should run parallel with the development of the product right from the start. Without renewed investigation no production changes must be made which would change the principle on which the original function is based. Quality control of the product is a post-inspection to see if everything "checks", but it can never compensate for design faults.

In designing equipment, the effects of electrical load, ambient tolerances (WORST CASE) must be well calculated. The following may serve as an example: the first colour television receivers sold in 1968 brought service complaints due to the breaking down of an N.T.C. resistor (Chapter II.K.5). Investigation showed that the resistors met the specifications and were not overloaded when the receiver was switched on. At a repeated switching on after a few tens of seconds the resistor was, however, still warm (and therefore had a low resistance), whereas the heaters in the tubes had already cooled off. The result was a high current at repeated switching which caused the resistor to break down (this phenomenon is colloquially referred to as the "father-mother" effect: father switches the television off, mother switches it on again).

Mechanical damage and too high a soldering temperature during assembly can also lead to rejects.

A typical example of how far reliability can go is given by the practical tests to which the components discussed in Chapter II.N.1 are subjected. Here "worst case" calculations of the tolerances, "professional soldering" and a heat and moisture resistent compounding have created favourable conditions. In a total of 88 million component hours there were no more than two rejects.

We may assume that with the application of integrated circuits a number of the abovementioned difficulties will be overcome, but here again "the proof of the pudding is in the eating".

# SECTION M. ELECTROMAGNETIC AND ELECTROMECHANICAL COMPONENTS

## II.M.1. COILS

Before treating the properties of coils, we shall take a closer look at the RESONANT CIRCUIT. In doing so we shall start from a self inductance in series with a resistor, and this in parallel with an ideal capacitor, as shown in Fig. 188. This diagram corresponds to the classical tuning circuit of the radio receiver, for the tuning capacitor with its dielectric of air involves but low losses. The ratio between voltage and current of this circuit is defined as the impedance Z and can be approximated as:

$$E/I = Z = \frac{L/C}{\sqrt{\{(\omega L - 1/\omega C)^2 + R^2\}}}$$

where $\omega$ is the angular frequency. By varying the frequency, a maximum Z will occur at $\omega_o = 1/\sqrt{LC}$, where $Z_o = L/CR$. So at a frequency $\omega_o$ the signal will not pass the bandfilter, but for the greater part it will travel on to the next stage of the circuit (see arrow).

At another value of the frequency, Z will diminish by a factor of "a" which is defined as the attenuation, and which can be calculated as:

$$a = Z_o/Z = \sqrt{\{1 + 4\omega_o{}^2L^2(\omega - \omega_o)^2/\omega^2R^2\}}$$

This formula is presented schematically in Fig. 189.

We now see that Z maximum is equal to $\omega_oL/R$, and also that at a certain frequency deviation the attenuation increases considerably as $\omega_oL/R$ becomes larger. The SELECTIVITY of the circuit is high when Z is high at the resonance wave length, and Z diminishes strongly at a deviating wave length. So this selectivity is determined by the ratio $\omega_oL/R$, which is defined as the quality factor Q of the coil in question.

The self inductance is formed by a coil, often wound on a core of magnetic

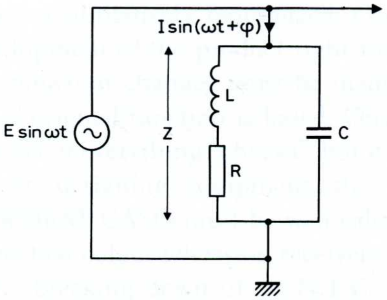


Fig. 188. Resonant circuit with self inductance, series resistor and ideal parallel capacitor.

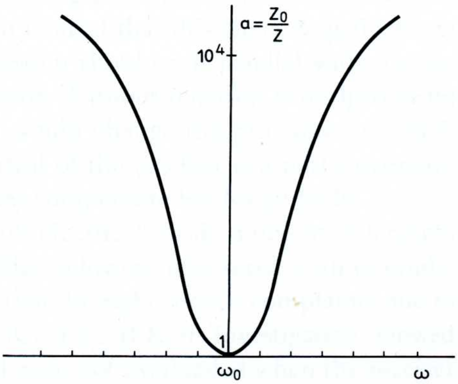


Fig. 189. Resonance response curve.

material. For a "toroid" coil with a constant cross section of the magnetic circuit and a narrow air gap (Fig. 162a) the self inductance can be calculated as:

$$L = 4\pi n^2 O/(d + l/\mu) = 4\pi n^2 O \mu_e / l$$

where $$\mu_e = \mu/(1 + \mu d/l)$$

in which n is the number of turns, O is the cross sectional area, and l the length of the magnetic circuit, $\mu$ the permeability (1000 - 5000 for ferroxcube), and $\mu_e$ the "effective" permeability under influence of the air gap.

It should be noted that the various turns have a certain capacitance with respect to each other, so that the impedance is reduced.

In this coil there are two types of losses:

a. Ohmic losses in the winding. At high frequencies eddy currents will occur in the wire so that the current concentrates along the outer layer of the wire (skin effect). As a remedy the wire is divided into several, parallel connected, insulated strands of wire (STRANDED WIRE) which are sometimes silver coated to reduce the resistance of the outer layers.

b. Eddy current and hysteresis losses in the magnetic core. In Chapter II.K.2 it was already explained how at radio frequencies these losses can only be kept within acceptable limits by using ferroxcube.

The sumtotal of the losses mentioned under a and b can be represented by a resistor R in Fig. 188. It is the coil designer's job to build the required self induc-

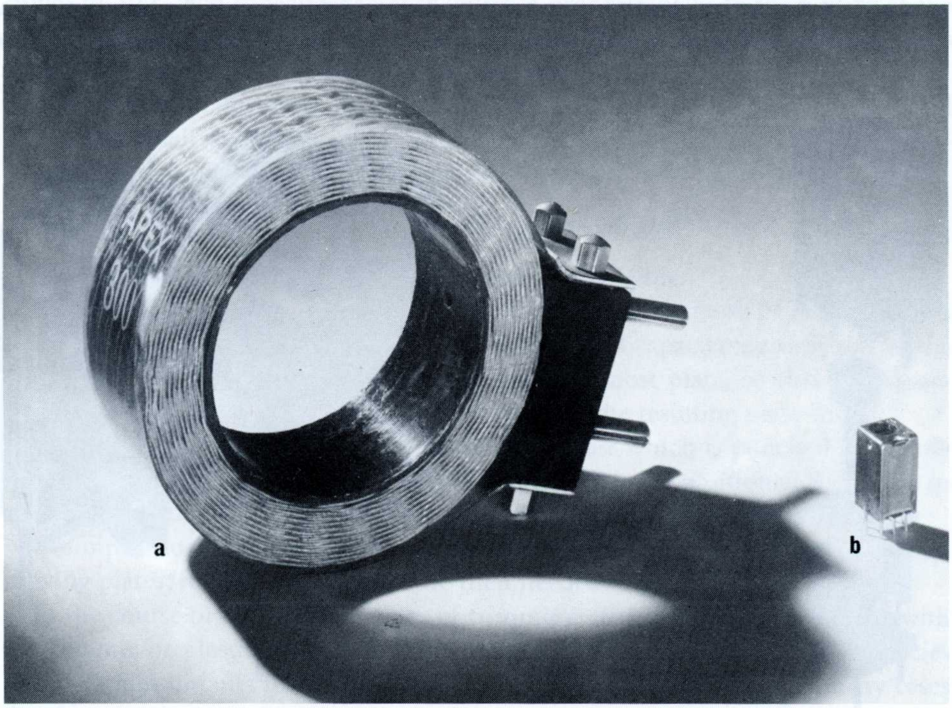


Fig. 190. Development of H.F. coils in 50 years: a. honeycomb coil from 1920, b. I.F. frequency filter with about the same self inductance in 1970.

tance, which must be independent of temperature and frequency, into the smallest possible volume, still ensuring the highest possible Q.

Before the invention of ferroxcube it was impossible to use an iron core for radio frequencies. So air coils were used, and at $\mu = 1$ the area O and the number of turns n had to be increased to obtain the required self inductance. To minimize the parasitic capacitance of the large number of turns, the wire was wound crosswise (HONEYCOMB COIL, Fig. 190a).

With the introduction of ferroxcube it became possible to apply a core material with a high $\mu$ for high frequencies, and so n and O could be reduced considerably. Fig. 190b shows such a coil with a closed ferroxcube core (the various parts are shown in Fig. 191) which has about the same inductance as the honeycomb coil photographed beside it. This clearly illustrates how important the influence of ferroxcube on the design of such coils has actually been. (In Chapter II.K.2 we already discussed powdered iron and its drawbacks.)

In these "miniature" coils an air gap is still used, one reason being that in doing so the influence of the temperature on $\mu$, and thus on the value of L, is minimized. Furthermore the closed magnetic circuit must be formed of various pieces of ferroxcube between which air gaps remain which might present an uncertain factor in the self inductance obtained. Finally the air gap provides an excellent means of adjusting the self inductance by screwing a threaded core up or down (for greater values of $\mu$, L is practically inversely proportional to d).

Fig. 191 shows the construction of the miniature coil of Fig. 190b. By screwing the core in or out, the self inductance can be varied from 80 to 120%. In this format of 7 × 7 mm area and 13 mm height, about 250 turns are sufficient to make a coil of 1180 $\mu$H which, together with a miniature ceramic plate capacitor of 100

Fig. 191. Components forming the I.F. band-pass filter of Fig. 190b.

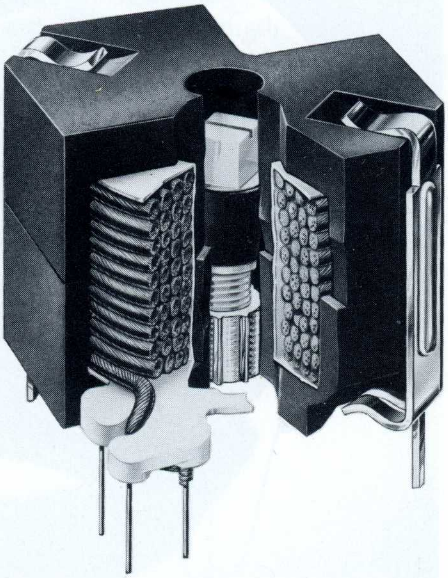


Fig. 192. Pot-core filter coil for multi channel telephony.

pF forms a tuned circuit for 460 kHz (the intermediate frequency of AM radio receivers). The same format can be used for making a coil of 2.08 μH consisting of 18 turns, which in combination with the same capacitor gives a resonant frequency of 10.7 MHz (the intermediate frequency of FM radio receivers). By combining the positive temperature coefficient of the coil with the negative one of the capacitor the tolerance for the resonant frequency remains within 0.3% between −40 and +85 °C. The quality factor Q of the AM coil is 200, that of the FM coil 100.

Fig. 192 shows a coil as used for the lower frequencies in multi-channel telephony, where the magnetic circuit is formed by two ferroxcube POTCORES surrounding the coil. In the centre of these potcores a pin can be moved up and down to adjust the self inductance. These coils have dimensions of 15 × 15 × 15 mm, and a Q of 500 at a temperature coefficient from 1 to $3 \times 10^{-6}$ per deg C.

It is typical of the rapid development in electronics that even the so much improved coils will for a part be replaced by piezoxide resonators (II.L.8).

## II.M.2. VARIABLE CAPACITORS

For "tuning" to a certain wave length it is easier to adjust a variable capacitor throughout the wave length range than a variable self inductance. Such a tuning capacitor consists of a number of fixed plates between which an equal number of plates fitted to a spindle can be moved (Fig. 193). Since the wave length is proportional to $\sqrt{LC}$, the capacitance must increase proportionally to the square of the angle of rotation in order to obtain a linear wave length scale. The moving plates are given a special form, so that this requirement is met.

The moving plates must be positioned about in the middle between the fixed plates, but small deviations are permissible. For at an equal air distance d everywhere, and deviation Δ, the capacitance is:

$$C = O/(d + \Delta) + O/(d - \Delta) = 2dO/(d^2 - \Delta^2)$$

If Δ is small as compared with d, its influence is negligible.

For the simultaneous tuning of several circuits, several stacks of moving plates are fitted on one spindle. To ensure a perfectly equal capacitance variation of the two stacks, slots are sometimes made in the outermost plate, so that the capacitance can be slightly corrected by merely bending the resulting plate strips.

The plates themselves are made of hard sheet brass which is punched absolutely flat and in the required shape. On a jig the fixed plates are positioned accurately in a frame work, the moving plates on a spindle where they are soldered.

A simpler form is obtained by inserting a plastic foil between the fixed and moving plates: fewer moving plates are then needed (Fig. 194).

In the course of time various types of trimming capacitors have been made with air, ceramic or plastic dielectrics. They are used to trim the resonant frequencies of the various circuits in the final production stage or in servicing. In many cases that function is now taken over by the ferroxcube cores which are screwed into the coils.

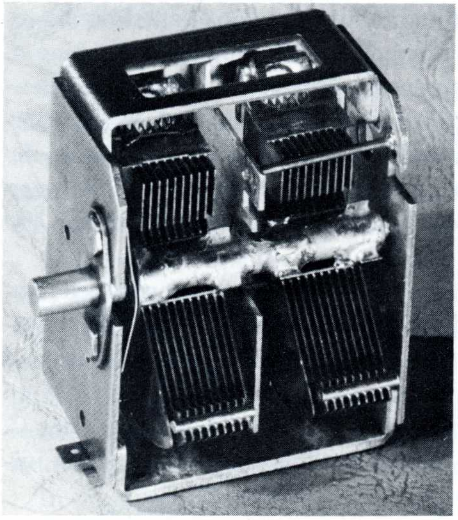


Fig. 193. Double variable capacitor with air insulation.

Fig. 194. Variable capacitor with plastic-foil insulation.

## II.M.3. LOUDSPEAKERS

The design of the ELECTRODYNAMIC loudspeaker has hardly changed since 1925; Fig. 195 shows its components.

The MAGNETIC SYSTEM consists of a cylindrical pot with a core. Between this core and the upper cover of the pot there is an air gap in which a cylindrical coil can move up and down. To ensure high sensitivity the field in the air gap must be as strong as possible. The width of the air gap is determined by the thickness of the coil and the play it requires. The height of the air gap is determined by the height of the coil plus the amplitude. According to Chapter II.K.2 it is now possible to calculate the permanent magnet so that the loudspeaker will operate at about the $BH_{max}$ of the magnetic material used. If ferroxdure is used, the magnet will be short and thick, for ticonal a thinner and longer magnet will be needed.

The COIL must be as thin and as light as possible; it is therefore wound from heat-bonding wire, enamelled copper wire with a thick layer of lacquer by means of which the turns can be "baked" together after winding, and no reel is needed. To have the coil deflect directly proportional to the current, the whole coil must remain inside the air gap even at its greatest deflection. To be connected direct to a tube amplifier, which supplies a low current at a high voltage, the coil would have to consist of a large number of turns. Since this involves a great deal of difficulty, a loudspeaker transformer is used whose transformation ratio is such that the coil needs only a few turns and consequently has a resistance ranging from 4-8 ohm. With a transistor amplifier the voltage is lower and the current higher: now the transformer is not needed if a coil of 15 to 25 ohm is used.

The coil is glued to a CONE which serves to transfer the movements made by the coil to the air. This cone is made of a kind of blotting paper which is obtained by "ladling" a very thin paper slurry with a perforated conical jig. With a corru-

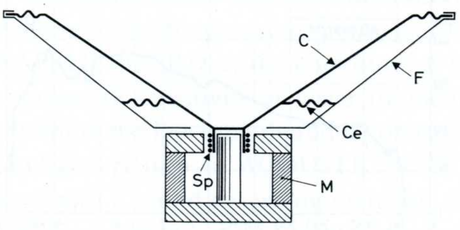


Fig. 195. Electrodynamic loudspeaker. M = magnet, C = cone, Sp = coil, F = frame, Ce = centre ring.

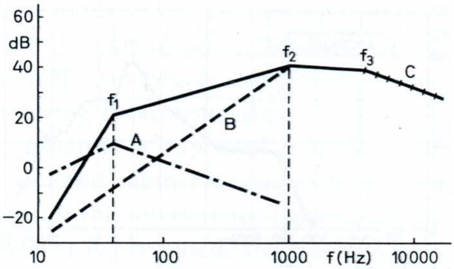


Fig. 196. Analysis of the frequency response curve of a loudspeaker. A = effect of resonance, B = effect of acoustic short circuit, C = effect of the useful cone area.

gated edge the outside of the cone is clamped in the loudspeaker frame. This gives the resilience required to limit the amplitude of the coil. For the low tones the whole cone can follow the coil movements, but the great mass of the cone cannot, however, follow the high frequencies. By lacquering the point and impressing circular ridges it is achieved that at the high frequencies only the point of the cone moves. This lacquering is necessary to make the cone sufficiently rigid so that vibratory patterns are not distributed over the entire cone area. To amplify and direct high frequencies, a small loose cone is sometimes glued into the larger one. The frequency characteristic can be greatly influenced by the choice of material, the thickness and the lacquering of the cone.

With the larger types of loudspeakers the relatively soft cone is not able to keep the coil accurately centred in the air gap. Therefore a CENTRING RING is used which has considerable radial rigidity, yet allowing considerable axial movement.

The final very important item in transferring the vibration from the microphone to sound in the listener's room is the FREQUENCY RESPONSE CURVE of the loudspeaker. This curve can be recorded by applying a constant alternating current of variable frequency to the loudspeaker and measuring the voltage at a microphone which is placed at some distance in front of the loudspeaker. This measurement must be carried out in a so-called "dead" room whose walls, floor and ceiling absorb all sound and reflect nothing.

It has been tried to calculate the curve by means of an "equivalent" circuit diagram as an electrical substitute for all mechanical parameters involved. The power fed to the coil in the form of an electric current is lost as ohmic losses in the coil and friction in the cone. It is used effectively to bring the air into motion: in the electrical equivalent diagram this can be represented by resistors. The resilience of the cone suspension can be replaced by capacitance, and the mass of the cone and the air to be excited by inductance. However, a reliable calculation of the frequency response curve is no simple matter; it must be found from experiment.

By means of Fig. 196 it is possible to mention a few crucial points on the frequency response curve. At a frequency of about 50 Hz ($f_1$) there is a resonance between the mass of the cone and the resilience of the mounting: the effect this has on the sensitivity is indicated with a dash-dot line. At a lower frequency there is also a certain acoustic short circuit: the air pressure variations behind the cone

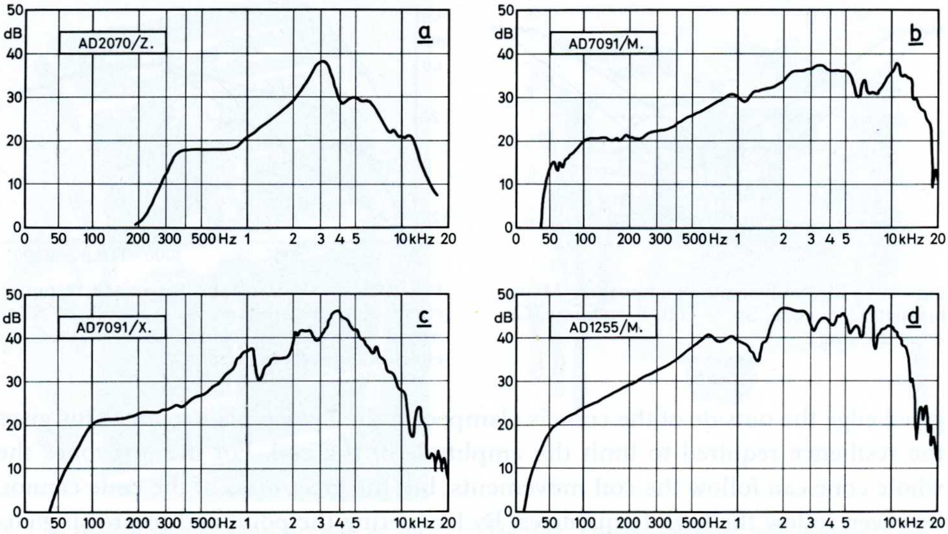


Fig. 197. Some typical examples of frequency response curves of loudspeakers: a. 2½″ standard loudspeaker for a transistor receiver, b. 6½″ loudspeaker for television receivers with F.M. sound, c. 6½″ loudspeaker for television receivers with A.M. sound, d. 12″ HiFi loudspeaker for public address.

compensate the ones in front. This phenomenon becomes less apparent when the frequency is increased, so that the sensitivity increases to $f_2$, the "tilt frequency" at about 1000 Hz, above that this phenomenon no longer plays a part. At frequencies higher than about 5000 Hz the output diminishes since a part of the cone cannot follow the movement, and only the central part will cause the air to vibrate. It should also be noted that the efficiency of a loudspeaker is very low (2-15%).

When choosing loudspeakers, it should be borne in mind that the quality of the loudspeaker need not, and even should not, be better than the quality of the AF-signal received. It is absolutely unnecessary that the speaker can reproduce higher or lower frequencies than the amplifier produces without distortion. Apart from this distortion the undesired signals as hum in the low, and as atmospheric interference in the higher frequencies are reproduced. For A.M. above 4000 kHz there is furthermore interference from adjacent transmitters. Fig. 197a shows the frequency response curve of a 2½ inch 0.5 W loudspeaker of a portable A.M. radio receiver which only reproduces the frequency range from 300-10 000 Hz with a peak at 3000 Hz and a resonant frequency of 360 Hz. Fig. 197b gives the curve for a 3 W loudspeaker used in standard television receivers (F.M. sound) which reproduces frequencies from 100-15 000 Hz reasonably well. For T.V.-receivers with A.M. sound a speaker with a curve according to Fig. 197c is preferred. It gives a higher output in the 4 kHz range, but frequencies over 5 kHz are suppressed.

For F.M. reception, gramophone-, microphone- and magnetophone amplification the speaker has to reproduce the whole frequency spectrum to which the human ear is sensitive (40-18 000 Hz) as faithfully as possible. This does, of course, involve the risk that interference and noise also become audible. Fig. 197d shows the response curve of a high-fidelity loudspeaker with a diameter of 12 inches and a resonant frequency of 50 Hz.

The above curves were measured on the bare loudspeaker (unconnected and unenclosed); the frequency response can however be considerably influenced by the BAFFLE-BOARD or the enclosure. A large baffle can largely compensate for the decrease of sensitivity at lower frequencies owing to acoustic short circuit. Furthermore the low tone reproduction can be substantially improved by building the loudspeaker into an ACOUSTICAL BOX, a closed cabinet whose walls are covered with a sound absorbing material. As a result the interference between the air vibrations behind and in front of the loudspeaker is prevented, and the frequency range can be extended to as low a value as 18 Hz.

With larger installations it is easier to obtain a wider frequency range by using separate loudspeakers for the low and the high tones. Extreme examples are the BOMBARDON, a low-tone loudspeaker with a cone of light, rigid plastic and a resonant frequency of 29 Hz, and then the TWEETER (high-tone loudspeaker) with a resonant frequency of 1000 Hz and distortion-free reproduction up to 22 000 Hz.

It should also be mentioned that the directional effect of the loudspeaker cone increases with the frequency. In the aforementioned "tweeter" measures are sometimes taken to reduce this effect; the result is the DOME TWEETER. The advantage is, however, that the sound perceived within the beam is stronger. In stereophony this directional effect, or directivity, adds considerably to "three-dimensional hearing".

To have the shape of loudspeakers match the cabinet of the apparatus in which they are used, there is also a demand for elliptical loudspeakers. Fig. 198 shows

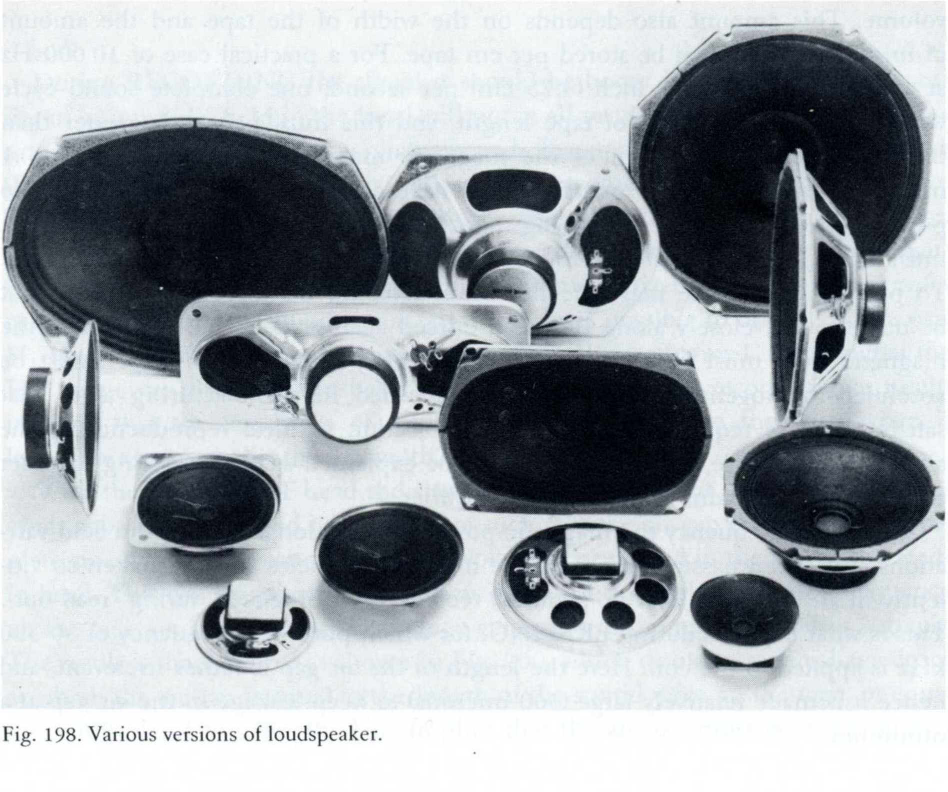


Fig. 198. Various versions of loudspeaker.

some of the different types available. Philips have introduced a certain standardization in this respect. For the sizes between 2½ inches and 12 inches there is now a standard series of 14 different overall dimensions by means of which a series of 65 types of loudspeaker of widely differing electrical and acoustical characteristics can be obtained by using different combinations of magnets, coils and cones.

## II.M.4. MAGNETIC HEADS

When discussing the devices that convert sound vibrations into electric current variations we shall leave the microphone and gramophone pick-up out of consideration since they are not supplied as "building elements". In Chapter I.19 it was, however, explained what an important part the tape recorder plays in electronics, not only for recording and reproducing sound, but also image, and furthermore for the recording and reproduction of numerical data in analogue and digital form. Consequently the magnetic head has become an important component.

As shown in Fig. 29 the tape recorder operates on the principle of magnetizing a layer of magnetic powder which is deposited on a plastic tape, the MAGNETIC TAPE. It is difficult to say what the lines of force on this tape look like, but we may confidently assume that there is no advantage in using a magnetic layer that is thicker than twice the length of the air gap. The plastic tape can be made as thin as its mechanical strength will allow. With a tape thickness of about 20 to 40 microns and a magnetic layer thickness of about 5 to 10 microns a means has been created to store a great amount of information in a small volume. This amount also depends on the width of the tape and the amount of information that can be stored per cm tape. For a practical case of 10 000 Hz at a tape speed of 1 7/8 inch (4.75 cm) per second, one complete sound cycle would equal 4.75 microns of tape length, and this must be much greater than the dimension of one grain of the magnetic material on the tape (about 0.1 micron). Furthermore this material must have a high magnetic conductivity to prevent the flux of the recording head from crossing the air gap, and it must have a high remanence to be able to produce a strong read-out signal. To prevent part of the magnetic flux from crossing the air gap, the tape must be made to run closely along the head (forced against it), which means that the magnetic layer must be hard wearing. And then, of course, this layer must be absolutely homogeneous. Industry has succeeded in manufacturing tapes that satisfy all these requirements. The only objection of ideal reproduction is the always audible noise, which can probably be explained as the clustering together of the individual grains of the magnetic powder.

At a certain frequency the magnetic powder can no longer follow the field variations and, when passing the head, the magnetic particles will be shaken so violently in all directions that no signal is recorded nor produced during "read-out". This is what happens during ERASING, for which purpose a frequency of 30-300 kHz is applied to the coil. Here the length of the air gap is rather irrelevant, and hence it is made relatively large (300 microns) to keep leakage in the air gap at a minimum.

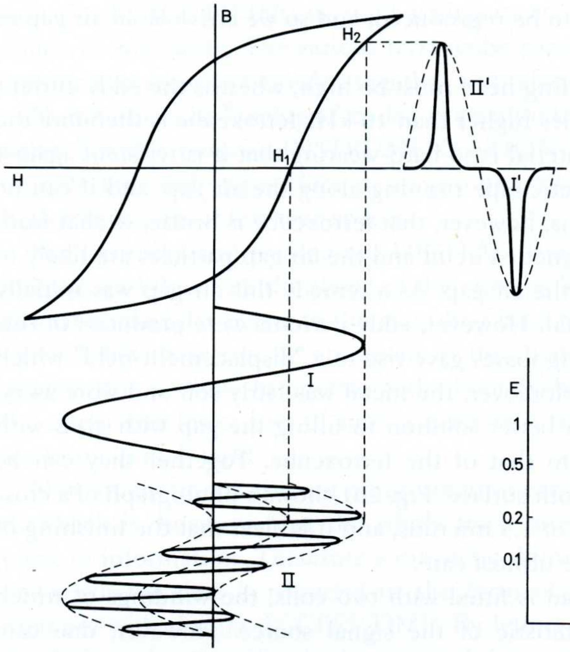


Fig. 199. How a sinusoidal current (I) is registered in the recording coil; magnetization of the tape follows the course indicated by I′. By modulating an H.F. signal on the audio signal (II) the part H1-H2 of the BH curve is used, thus ensuring faithful registration (II′).

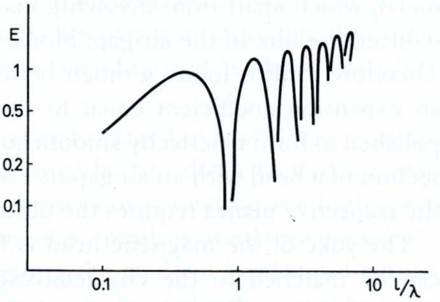


Fig. 200. The read-out voltage (E) as a function of the length (l) of the air gap divided by the wavelength of the audio frequency (λ).

During RECORDING the signal it should be borne in mind that the recording of a sinusoidal field in the head will not at all result in a sinusoidal magnetization of the tape. For due to the relatively wide "banana skin" form of the BH-curve (Fig. 199) a field strength $H_1$ is needed to overcome the remanent magnetism. Therefore a sine form I in Fig. 199 will be recorded on the tape as a magnetization I′, with the result that a distorted signal is read out. By superimposing an alternating current of a higher frequency (which the magnetization of the iron powder cannot follow) on the sound signal (II), it becomes possible to operate in the part $H_1$-$H_2$ of the BH curve and a faithful recording (II′) is obtained. At no signal the H.F. signal in the recording head will also "erase". During recording one might imagine that the last piece of information is passed on to the tape, when it leaves the air gap, so that the gap width is not critical.

With the READ-OUT head the situation is quite different. Provided the values of air gap, frequency, and tape speed are such that the air gap embraces one complete sine wave of the sound signal, no voltage is generated in the coil of the read-in head. This can be ascertained by recording a continuously increasing frequency on the tape, and by plotting the voltage generated in the read-out coil in a graph. The result is then a curve as shown in Fig. 200, which displays steep voltage drops at which the air gap is equal to the length of the sound wave on the tape, or equal to a multiple of this length. This implies that the air gap must be made shorter than the length of a sound wave to be reproduced, and so we arrive at an air gap of 4 microns or less.

The permeability of the recording head must be high, whereas the eddy current losses must be low. For frequencies higher than 10 kHz ferroxcube is therefore the indicated material. This hard material is so hard-wearing that it is resistent against the friction caused by the magnetic tape running along the air gap, and it can be polished quite well. A drawback is, however, that ferroxcube is brittle, so that from the sharp edges between the magnetic circuit and the air gap particles are likely to break off which might lodge in the air gap. As a remedy this air gap was initially filled up with a diamagnetic metal. However, eddy currents were produced in this metal, which apart from involving losses gave rise to a "displacement field" which reduced the flux in the air gap. Moreover, the metal was fairly soft and wore away. Therefore Philips found a much better solution in filling the gap with glass with an expansion coefficient equal to that of the ferroxcube. Together they can be polished to form a perfectly smooth surface. Fig. 201 shows a photograph of a cross section of a head with an air gap of 1.5 microns, and it is clear that the finishing of the respective planes requires the utmost care.

The yoke of the magnetic head is fitted with two coils, the windings of which can be matched to the characteristic of the signal source. However, due care should be taken that the resonant frequency (due to the self inductance of the coil and the parasitic capacitance of the turns on the coil) is amply above the frequency used.

For recording and playing back music and speech, two or four tracks are produced on one tape. The air gap is 4 microns, the distance between the tracks 3.65

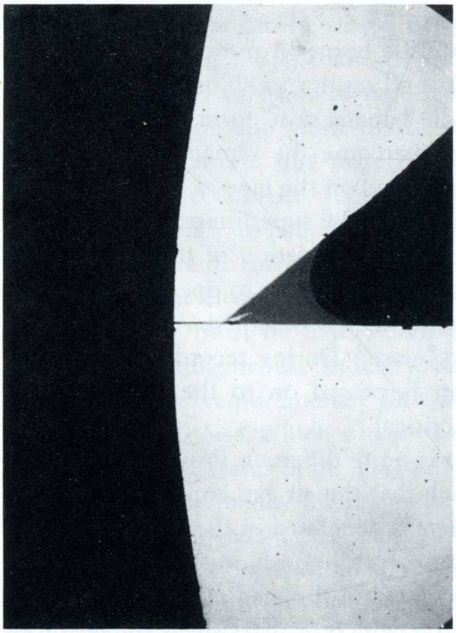


Fig. 201. Photograph showing the cross section of a ferroxcube tape recorder head whose air gap (1.5 micron long) is filled up with glass.

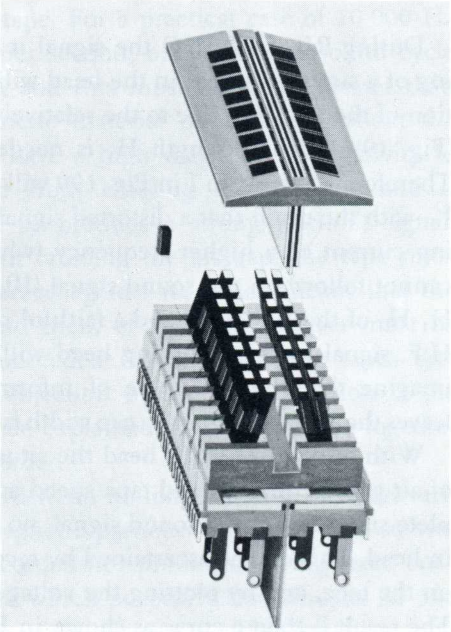


Fig. 202. Components forming the magnetic head for the write-in and read-out of digital data on 9 tracks on a tape ½″ wide.

mm. For VOICE FILING (as used at airports etc.) heads are made for 17 tracks on a tape 25 mm wide. The various ferroxcube parts are separated by ceramic pieces and metal screens to be sealed together with glass enamel.

Numbers can be "analogue" coded as amplitude-, and even as frequency-modulations. For this purpose INSTRUMENTATION heads are designed for 14 tracks on a 25 mm wide tape. Here an air gap of 1.5 to 2.5 microns is used to arrive at high frequencies at low tape speeds.

For recording and reading-out DIGITAL data, only the number of pulses is of interest, and therefore the shape of the curve is less important. Here it is possible to feed rectangular pulses to the recording head to be read out as "spikes" (see Fig. 199 I′), but these can be processed in the digital equipment. According to this process it is possible to operate with about 30 bits per mm tape. Fig. 202 shows the upper and lower part of such a head with recording and read-out units for 9 tracks on ½ inch tape.

Now an enormous amount of information can be stored in such a reel of tape; a drawback is, however, that the whole track must be run down to find a certain piece of information. Therefore a magnetic drum or disc, rotating at a high speed, on which the data is recorded in the form of a spiral, is sometimes used as a memory with shorter ACCESS TIME. By letting the read-out head travel perpendicularly to the sound track, the required data can be located rather quickly.

## II.M.5. DEFLECTION COMPONENTS FOR TELEVISION RECEIVERS

A part of a television receiver consists of circuits for mixing, separating and amplifying the video and the audio signals which are ultimately fed to the grid of the picture tube and the loudspeaker respectively. The electron beam, whose intensity is varied by the video signal, must be deflected in two directions across the screen to describe a complete raster that corresponds with the raster in the camera tube (Fig. 16). The television signal contains current pulses which ensure that the raster in the picture tube runs synchronously with the raster in the camera tube; the receiver must ensure, however, that the deflection is proportional to the time, that the "fly back" is quick enough, the vertical and horizontal lines are straight and run parallel, and that during deflection the electron beam remains bidirectionally focused on the screen, even when owing to the wide deflection angle the distance from the gun to the sides of the screen is much larger than the distance to the centre of the screen.

For colour television it should, furthermore, be ensured that over the entire screen area the beam from the red gun lands accurately on the red dot etc. (colour purity). Moreover, the three electron beams, together forming the colour, must converge in one "spot" (CONVERGENCE).

In Chapter II.C.5 (Fig. 70) the DEFLECTION COIL for black and white television was already described with the adjustments for centring and linearity.

Chapter II.C.10 gave a description of the deflection coil for colour television which is provided with a "colour purity magnet" and fitted with the rings for the static convergence units, the dynamic convergence being obtained from coils in

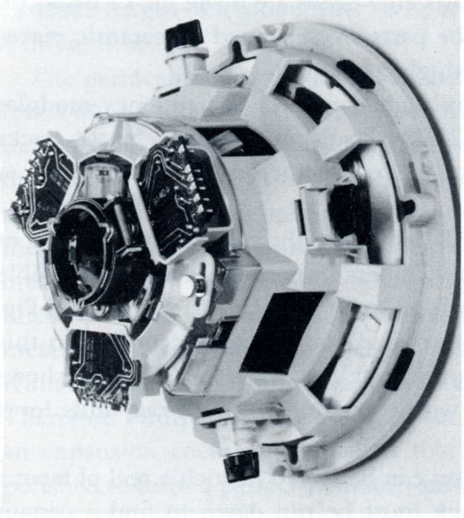


Fig. 203. Deflection coil for a 110° colour picture tube fitted with a convergence unit on top.

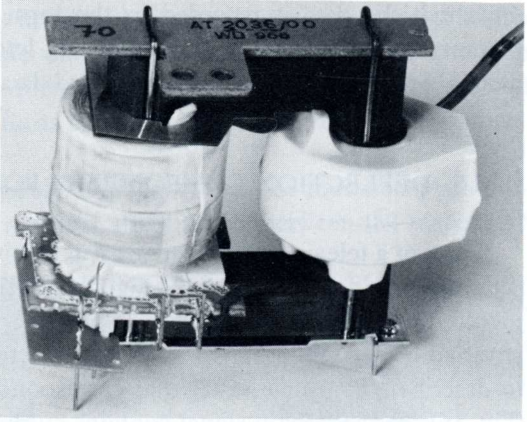


Fig. 204. Line output transformer for colour television (25 000 V)

these units. The 110° colour television tube requires an extra current through the deflection coils to ensure good convergence in the corners as well. Fig. 203 shows a photograph of the colour deflection coil with, mounted on it, the convergence unit. The result is clearly a rather complex piece of precision equipment which can be manufactured in large numbers by special factories only.

Besides these deflection units, coils with adjustable ferroxcube cores are also required for adjusting the picture size and the linearity.

Another important item of the "deflection package" is the LINE OUTPUT TRANSFORMER which supplies both the current for the horizontal deflection coil and the anode voltage for the picture tube. The current needed for horizontal deflection must increase linearly with time from 0 to a certain value to let the spot describe a horizontal line across the screen. Then the current must suddenly drop to 0 again to start a new cycle. When the current suddenly disappears, the high self inductance of the deflection coil releases a large amount of energy which is used to generate the high voltage on the anode of the picture tube. For that purpose the winding which supplies the deflection coil is lengthened to form the high voltage winding. Via a high voltage rectifier the voltage generated in this winding is fed to the anode of the picture tube as a direct voltage. To ensure satisfactory functioning of this complex electromagnetic phenomenon, the transformer requires an iron core which involves but low losses at high frequencies, and again ferroxcube brought the solution. The magnetic circuit is built up from the so-called U-CORES.

A technological difficulty is that the anode voltage of the black and white picture tube is 20 000 V, and for the colour picture tube 25 000 V, and that notwithstanding the influence of moisture and temperature the transformer must be able to stand such a high voltage. This is achieved by compounding the transformer in plastic. Fig. 204 shows a line output transformer for colour T.V.

## II.M.6. SYNCHRONOUS AND STEPPER MOTORS

In Chapter II.K.4 it was already said that ferroxdure is used for small d.c. motors. The fact that with this material it is possible to obtain a sufficiently strong magnetic flux with a relatively wide air gap and a short magnet (in practice an alternately N-S magnetized cylinder) has given impetus to the design of small, handy synchronous motors. The number of revolutions made by a synchronous motor is constant, and equal to the frequency × 60 divided by the number of pole pairs. When connected to a single phase network, the motor will not start automatically, but once it has been brought to rotation it will continue to run synchronously (think of the electric clocks which had to be started manually). By connecting a second phase the motor will start on its own account. This second phase can be provided by introducing a second set of poles carrying a winding which is either short circuited or connected to the mains in series with a capacitor (the direction of rotation can then be changed by switching the capacitor).

Particularly since the mains frequency is kept accurately constant in the power stations, such a synchronous motor is eminently suitable for electric clocks and

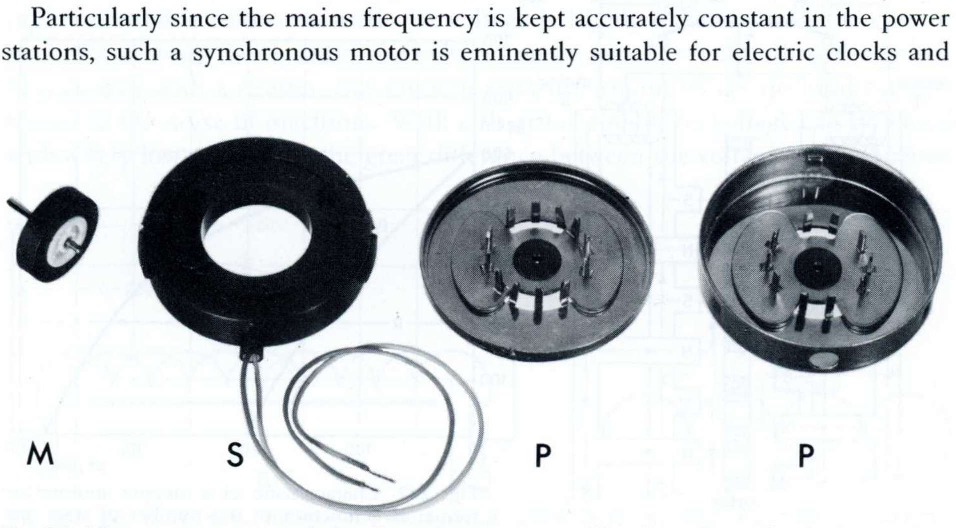


Fig. 205. The constituent parts of a small synchronous motor. M = ferroxdure pole wheel on shaft, S = coil, P = shields with teeth, carrying shorted windings.

time switches (large numbers of the latter are needed in automatic washers). In gramophones, where also a constant speed is required, the synchronous motor did not at first give satisfactory results. The variable friction between stylus and gramophone record is reflected in small speed variations to which the ear is particularly sensitive (the so-called "WOW"). This problem has been solved by using a heavy turn table for the record and a flexible drive between the electric motor and the turn table.

Fig. 205 illustrates how very simple the construction of such a synchronous motor is. The ones Philips make have torques ranging from 15 to 600 gcm and run at a speed of 250 revolutions per minute. Since for application as time switch the speed must be considerably reduced by means of a gear box, there are types on supply with built-in gear reduction (SYNCHRODRIVER), giving an ultimate speed of 8 or 1 revolution per minute.

Such a small motor can also be fed from an electronic INVERTER which changes a d.c. voltage into an a.c. voltage with adjustable frequency so that the speed of the synchronous motor can be controlled.

In modern electrical measuring techniques the measured values are often expressed in a number of current pulses. This is certainly necessary if the measured values are intended for further computation. Automation becomes a fairly simple matter if the motor which controls the machine or process is driven by current pulses. The measured or computed values or the numbers read from a memory or punched tape can then be directly translated into the number of revolutions made by the motor, and hence into a certain "displacement".

The parts of such a synchronous motor can be used to build a simple device that moves one step further at every current pulse fed into it. The magnetic cylinder is then mounted within two stators which are displaced half a pitch with respect to each other. This can be explained with the diagram of Fig. 206. In the drawn position the rotor is held in the position in which the (south) S-poles of

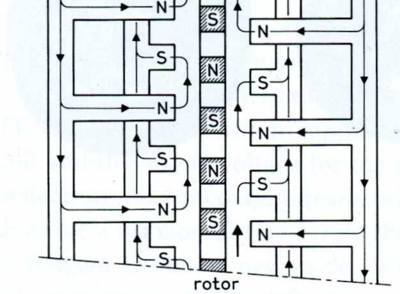


Fig. 206. The principle of the stepper motor.

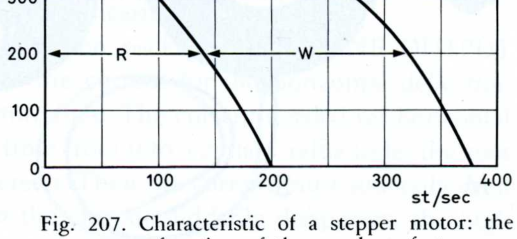


Fig. 207. Characteristic of a stepper motor: the torque as a function of the number of steps per second. A = pull-out, B = pull-in, W = working range, R = starting range ("slew range").

the rotor are between the N-poles of the stator. When $S_2$ is switched, stator II will lose its magnetism during the switching operation, the S-poles of the rotor take up a position facing the N-poles of stator I; in other words, the rotor moves ⅛-pitch upwards (clockwise). If $S_2$-A is closed, stator II changes polarity and the S-poles of the rotor take up a position between the N-poles of the two stators, and thus the rotor moves ⅛-pitch further clockwise. If now $S_1$ is switched from B to A, the rotor moves twice ⅛-pitch further clockwise. However, if we switch $S_1$ in the drawn position from B to A, it is clear that the rotor will move anticlockwise. The switches $S_1$ and $S_2$ can be thought of as replaced by two bistable multivibrators and gate circuits (Fig. 36) and then the motors can be driven by current pulses from a computer or measuring instrument.

Fig. 207 shows a typical characteristic curve of such a "stepper motor". By "pull in" is understood the maximum frequency at which the motor safely starts, "pull out" is the maximum frequency at which the motor keeps running. This frequency reduces as the torque supplied increases. At a friction resistance R the motor must be started at 140 pulses per second and, if required, and if necessary, the frequency can be increased to 350 pulses. The range comprises motors for 48 and 96 steps per revolution, with a pull in speed up to 750 r.p.m. and a maximum torque of 2 kgcm at a power consumption ranging from 2-11 W. Advantages are that such low power is needed to obtain the rather great torque (the efficiency is better than that of a gear transmission) and the "digital" accuracy (each pulse given by a unit is processed).

## II.M.7. VARIABLE TRANSFORMERS

A part not containing electronic components, and yet important for electronic measuring arrangements is the VARIABLE TRANSFORMER. It is not so easy to obtain a continuous control of an alternating voltage as often needed in measuring. If a variable resistor is used, the voltage is highly dependent on the current load. When using the dynamo, the voltage can be varied by controlling the excitation current with a resistor, but modern electrical engineers are no longer accustomed to the noise of machines. With a thyristor circuit the voltage can be varied with a very low power, but the great difference between the voltage curve obtained

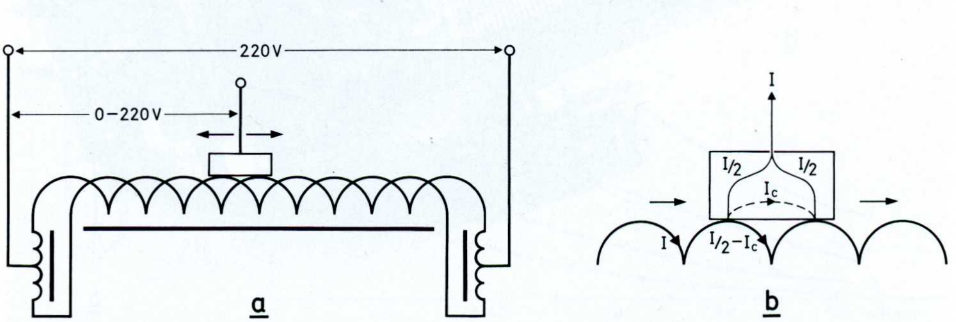


Fig. 208. Two methods for limiting the short circuit current in the winding of a variable transformer that is short circuited by the brush: a. two windings interconnected with choke coils, b. a high-resistance carbon brush.

and the original sine form could give rise to difficulties in a number of cases. An ideal solution is found by mechanically designing the variable transformer as the rotary potentiometer. Winding the transformer coil as a toroid (i.e. on a cylinder of laminated iron) is indeed possible on an ingeneous machine, but the problem of the sliding contact or "cursor" is quite a different story.

In a wire-wound resistor it does not matter if two adjacent turns are short circuited; however, if the sliding contact short circuits two adjacent windings of a transformer, a short circuit current will flow through this contact. A solution is to place two complete windings side by side and to interconnect them with a choke (Fig. 208a). It is simpler to use a brush with a relatively high contact resistance. According to Fig. 208b the short circuited winding will carry a current of $\frac{1}{2}I - I_C$, where I is the current that is to be picked up an $I_C$ the short circuit current which is equal to the voltage of one turn divided by the resistance of the brush. If $I_C$ does not become much greater than $\frac{1}{2}I$, the shorted winding will not assume a higher temperature than its adjacent turn.

So the variable transformer consists of a cylinder, rolled of a strip of transformer lamination steel, provided with a toroid winding. For smaller transformers the contact area is provided at the cylinder head, for larger transformers the cylinder wall offers a larger area. Fig. 209 shows the largest Philips type, variable from 0-260 V at a current of 20 A. The part of the winding along which the brush runs is machined, polished and coated galvanically with a hard, highly conductive material. The brushes slide in a heavy aluminium plate which ensures good removal

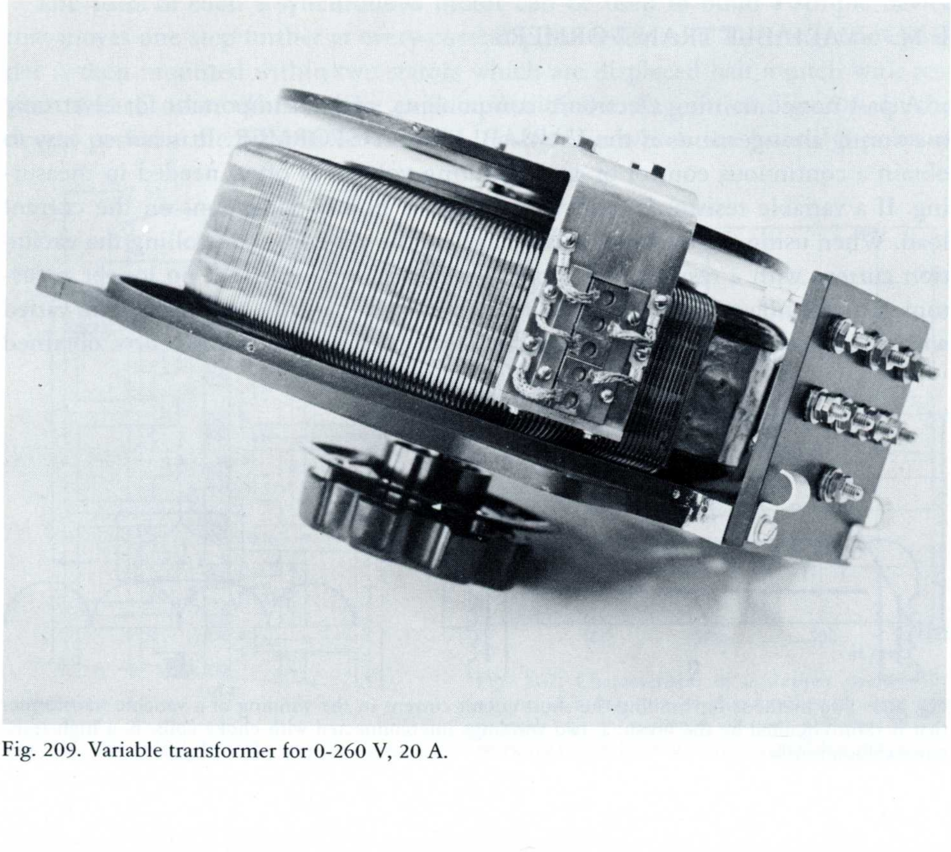


Fig. 209. Variable transformer for 0-260 V, 20 A.

of the heat developed in the brushes. The smallest available type gives 0-240 V at 0.5 A, has a diameter of no more than 85 mm and a weight of 700 grams.

## II.M.8. (REED) RELAYS

In switching, an important part is still played by the RELAY, in which, under influence of a current through a coil, an armature is attracted which, opposing the tension of a spring, opens or closes a contact. Such a relay has three critical values which must be reproducible, to wit: the current above which the contact is closed, above which it remains closed, and below which the contact returns to its initial position. We shall not discuss the well known telephone relay of which millions, particularly in telephone exchanges, are in use and with which a high degree of reliability has been reached. There remains, however, the drawbacks that contacts exposed to the air might assume a higher resistance owing to oxidation or soiling, that as a result of this soiling the friction in the mechanism increases, and that it has a certain mechanical inertia.

In the preceding chapters it has been explained that transistors or thyristors can function as a relay that requires but low operating power, is very rapid and has a long working life. A drawback of such an electronic relay is, however, that in the "closed" condition it still has a certain internal resistance, whilst in the "open" condition the resistance is not infinite either.

As explained in Chapter I.23, it is particularly for telephone exchanges that a magnetic relay, the REED RELAY, was developed which entails the drawbacks of the conventional electromagnetic relays to a much lesser extent, and as compared with transistors and thyristors offers the advantage of a lower resistance in the "closed" condition and a practically infinite resistance in the "open" condition.

The principle of the reed relay is shown in Fig. 210a. Two strips are rolled from 0.55 mm chrome iron wire, which as the result of a mechanical and thermal

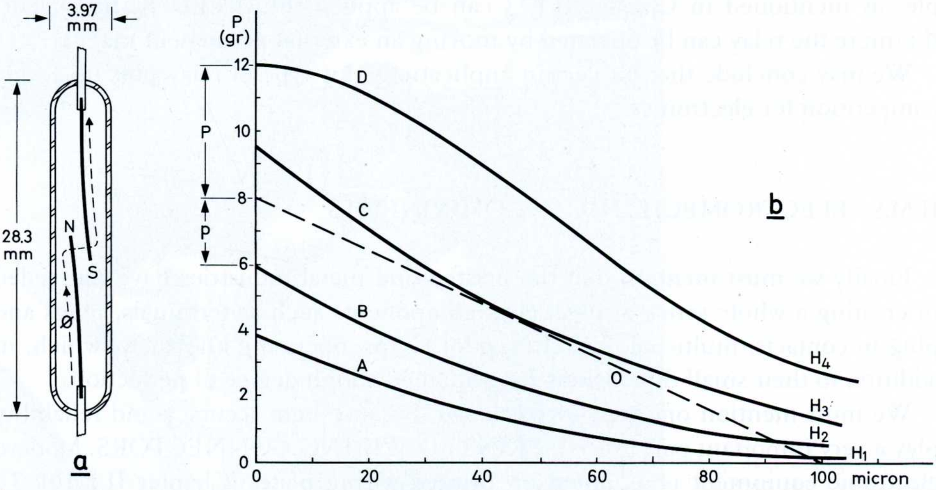


Fig. 210. Reed relay: a. design principle, b. contact pressure as a function of the contact distance (in microns) with the magnetic field strength H as a parameter (H is proportional to the number of ampere-turns).

treatment receive a reproducible resilience and certain curvature. In the open condition the contact distance is 100 microns and according to the Paschen curve mentioned in Chapter II.L.1 the breakdown voltage at this distance in nitrogen of 1 atm is at least 1000 V. To ensure a lower contact resistance, a layer of gold is diffused into the surface. Both strips, carefully positioned, are sealed into a glass tube in an atmosphere of pure nitrogen. So the contact is located in a space that is absolutely free from dust, oxygen and water vapour. The relay is energized by a longitudinal magnetic field. In Fig. 210b the line 0 represents the resilience of the two strips when they are bent towards each other (this resilience is directly proportional to the deflection). The magnetic attraction between the two strips (under influence of the magnetic field) is about inversely proportional to the square of the distance between them.

At a magnetic field strength $H_1$ (curve A) the strips do not move, at $H_3$ (curve C) they will just close. At a field strength $H_4$ (curve D) the contact pressure P will be 4 grams. Furthermore, the strips will then reach saturation, so that variation in field strength will have hardly any effect on the contact pressure. If the field strength is now reduced to $H_2$ (curve B), the contact will open, even after having to overcome a certain sticking force (p). Apparently the three aforementioned critical values are well defined, and are little influenced by friction or mass inertia. The closing time is 0.001 s, the opening time 50 μs, the switching power is 5 W or 100 mA at 65 V. The working life must be more than 5 million switching operations with a failure rate of less than $8.5 \times 10^{-9}$. After that life time the contact resistance should not be more than 2.5 ohm (0.15 ohm initially).

The magnetic field can be supplied by a coil round the relay with matched magnetic circuit; the minimum number of ampere turns (AW) for switching is 58, for non-switching maximum 30.

The strips can also be magnetized by the fields of two coils with magnetic circuits so arranged that the field of one coil alone has no effect, whereas the fields of both coils combined will close the contact. In doing so the "coincidence principle" as mentioned in Chapter II.K.3 can be applied (FERREED RELAY). Furthermore the relay can be operated by moving an external permanent magnet.

We may conclude that for certain applications this type of relay puts up severe competition for electronics.

## II.M.9. ELECTROMECHANICAL COMPONENTS

Finally we must mention that the plastics and metal industries have succeeded in creating a whole series of ingeneous components such as terminals, plugs and plug-in contacts, multi-pole switches, pilot lamps, operating knobs etc. which, in addition to their small dimensions have attained a high degree of perfection.

We must mention one type in particular, because here accuracy and reliability play a very important part indeed: PRINTED WIRING CONNECTORS. Modern electronic equipment is mounted on printed wiring plates (Chapter II.L.10). To facilitate the interchanging of such plates, the circuit on such a "print plate" ends in a number of gold plated contacts along one of the edges, which are inserted

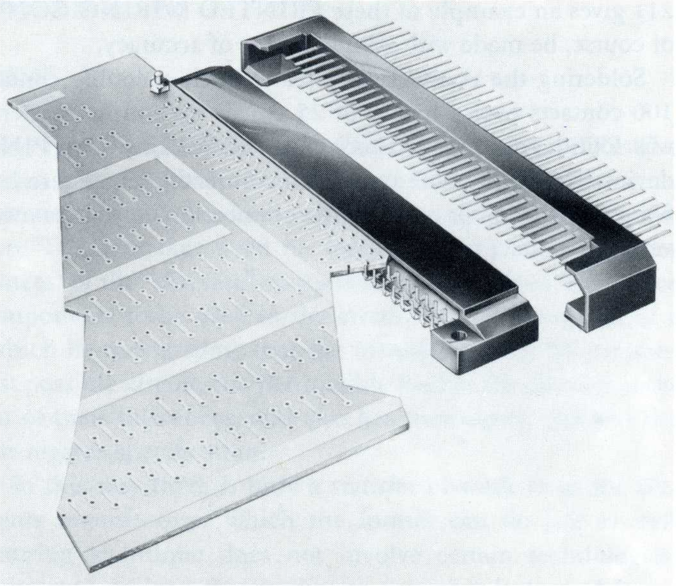


Fig. 211. Printed wiring connector with 2 × 32 contacts at a spacing of 0.1 inch.

into a strip with gold plated spring contacts from where the connecting wires lead to the other parts of the installation.

To obtain two rows of such contacts the method of double spring contacts is sometimes adopted, a double row of plugs being soldered to the print plate. Fig.

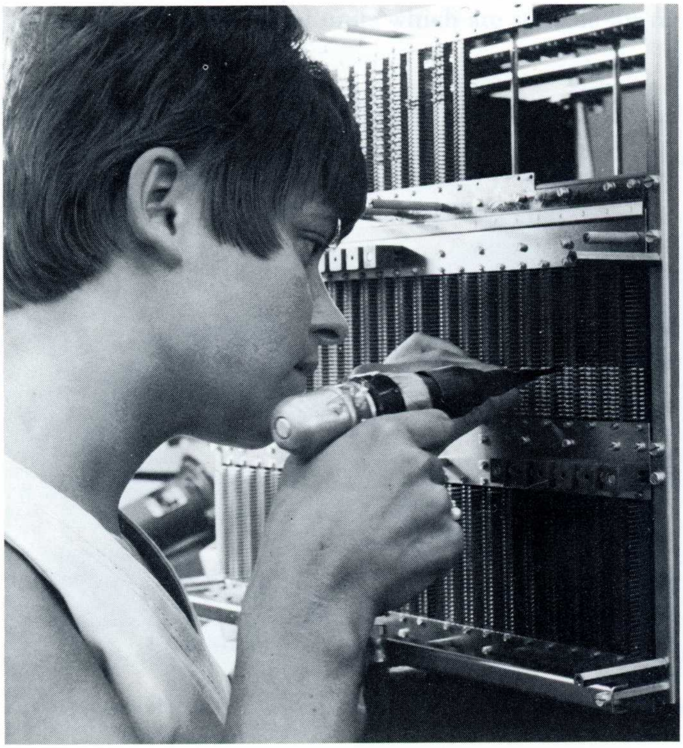


Fig. 212. Connecting the leads to the contacts of the connector according to a process known as "wire wrapping".

211 gives an example of these PRINTED WIRING CONNECTORS which must, of course, be made with a high degree of accuracy.

Soldering the connecting wires to such a double contact row, involving about 100 contacts over a length of 25 cm, is no simple matter. Here a clever solution was found: the so-called technique of WIRE WRAPPING, (Fig. 212). A simple device "wraps" the wire round the connector pin. There is now no more reason to fear dripping solder or resin, the connection combines mechanical strength with a very low contact resistance.

## SECTION N. SUBASSEMBLIES

### II.N.1. INTRODUCTION

In Section II.A we already mentioned the use of FUNCTIONAL UNITS. The complete circuit is than no longer built up from tubes, semiconductors, resistors, capacitors and self inductances, but of standardized switching elements each fulfilling its specific function. The setmaker need no longer consider the properties (specification and tolerances) of the "discrete" components, and neither the choice of the most suitable components to be used in the circuit. He choses a unit of a specified function for which he may assume that the manufacturer of the components guarantees the best possible circuit, the permissible load of the discrete components, the integration of their tolerances, and also has thoroughly checked the complete subassembly as regards specification.

It might be said that in this way there is only a transfer of work from the setmaker to the components manufacturer which the former can do just as well (provided the manufacturing technique does not involve certain technologies, such as integrated circuits etc.). Indeed this applies to the manufacture of larger numbers of radio and television receivers for which it is generally more economic to mount all discrete components (nowadays including integrated circuits) on one print plate, than to build them from separate standard circuits. There are, however, a few considerations which in many cases lead to the application of functional units:

a. Since the components manufacturer supplies to several setmakers and endeavours to make functional units which are suitable for numerous applications, he

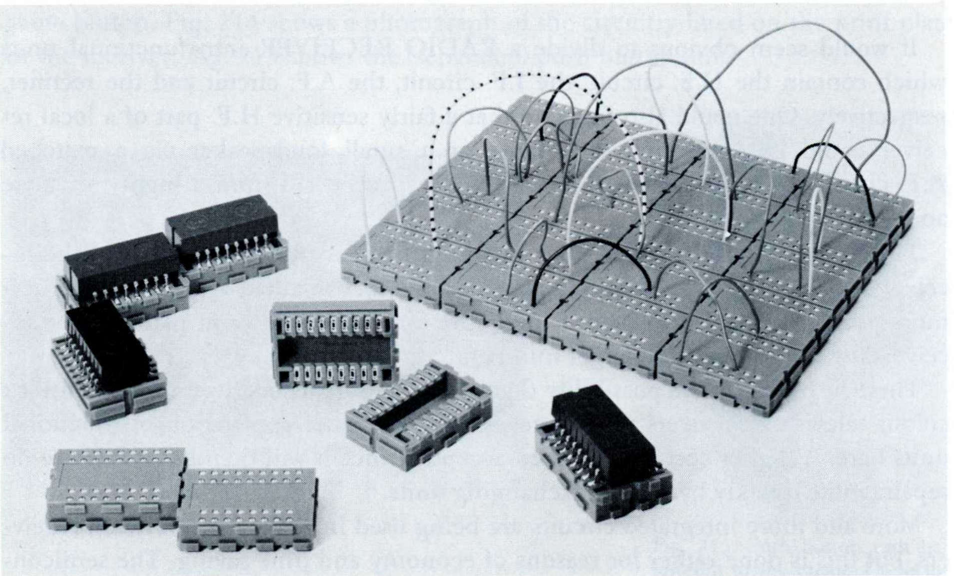


Fig. 213. Some "NORBIT" functional circuit blocks and a plate with resilient contacts for the rapid assembly of experimental circuits.

can produce larger quantities. The cost of design, quality research and expensive (automatic) measuring equipment is spread out over larger quantities.

b. Particularly with the introduction of automation in engineering, the manufacturer of the end-product is often insufficiently skilled in electronics to design the entire circuit.

c. As the circuits are constantly becoming more complex, specialization produces a new generation of electronics engineers who no longer make their first test model (on a piece of wood) with transistors, capacitors, resistors etc, but who start from standardized units of which only the functions they can perform have their interest. In this way they save time and money: they need less time for experiment and they can quickly build small numbers (provided functional units are on supply). As an example, Fig. 213 shows a system of NORbit units for digital circuits. They are pressed in plastic blocks with plugs which can be plugged in resilient contacts into which the connection wires can be inserted from the rear. A "breadboard" of the complete digital circuit is thus completed in a few minutes

One might ask why such functional circuits are still made as print plates in which discrete components are used and not as monolithic integrated circuits. The reason is that power of integrated circuits is limited and that no larger capacitors, resistors and coils can be incorporated. Therefore functional units are still made according to the classical mounting method, to be protected against humidity, dust and mechanical damage by fitting them in a suitable envelope.

We shall now discuss a few applications of these functional units.

## II.N.2. RADIO AND TELEVISION RECEIVERS

It would seem obvious to divide a RADIO RECEIVER into functional units which contain the H.F. circuit, the I.F. circuit, the A.F. circuit and the rectifier, respectively. One could also imagine that a fairly sensitive H.F. part of a local receiver could be connected to a large or a small loudspeaker via a matched A.F. circuit, whilst the same A.F. circuit can also be fed from a highly sensitive and selective H.F. circuit.

This way of reasoning often does not work out, for the various parts of the receiver must be matched to each other. As already explained above, it is furthermore more economical to mount the entire circuit on one print plate for the receivers usually produced in larger numbers.

There is, however, the possibility that service problems involved in the complex colour television receivers will of necessity lead to the application of functional units here. A higher cost price is then accepted since it will then be possible to do repairs more quickly by simply exchanging units.

More and more integrated circuits are being used in radio and television receivers, but this is done rather for reasons of economy and time saving. The semiconductors, small resistors and capacitors are collected in the integrated circuit to which the large capacitors, resistors and self inductances on the print plate of the receiver can be connected. Therefore, these integrated circuits are, in principle, not yet complete "functional units".

An exception to this must be made for the TUNERS, functional units which comprise the H.F. and mixer part of F.M. and T.V. receivers. For F.M. tuning takes place with a variable capacitor or self inductance, for T.V. there are but a few stations to choose from, and the required "channel" is selected with push buttons: the CHANNEL SELECTOR. This is a component which owing to the high frequencies used, is very critical, must be mechanically rigid to avoid detuning caused by conductors moving with respect to each other, and electrically well screened to avoid a mutual influence of the circuits. When choosing different channels in a television receiver other resonant circuits are switched in, and when changing over from V.H.F. to U.H.F. even other tubes or transistors. This switching is done by means of push buttons. By turning the same push buttons a variable capacitor must then be adjusted for accurate tuning which is maintained during the repeated switching operations. Since the push buttons must operate the switch and the tuning capacitor of the H.F. circuit, they must form a close unit with them. These tuners are truly a masterpiece of mechanical and electronics engineering and are made in large numbers in special factories.

In 1969 the design of the channel selectors was simplified in that the circuits could be accurately tuned by means of VARACTOR diodes, whose capacitance changes with the voltage applied. This variable voltage can be supplied via conductors external to the H.F. circuit. Thus it became possible to mount the H.F. circuit in the most favourable position on the print plate of the entire set, the potentiometers for voltage control being fitted in the most suitable position at the front.

The push button unit consists of screw spindle operated linear potentiometers which are switched on by depressing the push button and are set by turning the same button. Fig. 214 shows a photograph of the circuitry fitted on the print plate of the receiver; Fig. 215 shows the (separated) push button unit.

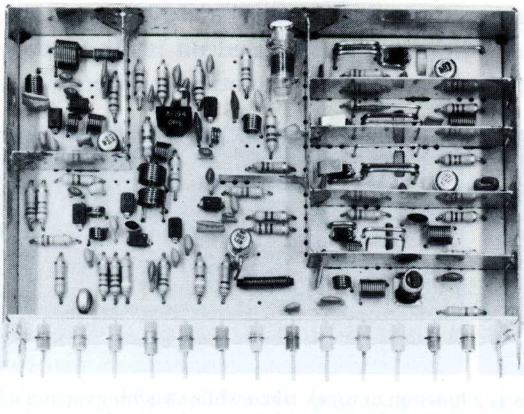


Fig. 214. V.H.F. — U.H.F. circuit of a channel selector with "varactors".

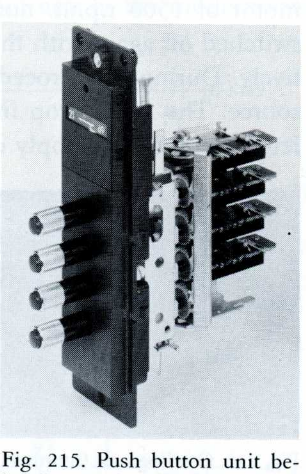


Fig. 215. Push button unit belonging to the channel selector of Fig. 214 with linear potentiometers which are adjusted with a screw spindle.

## II.N.3. MOTOR CONTROL

The field of motor control, as well as the control of electric furnaces and lighting, so in general POWER CONTROL, is particularly suited for the system of functional units. Sometimes the manufacturer does not specialize in electronic circuits. Furthermore the mostly small series and the possibility of making various combinations with standard units are conducive to their application.

For controlling direct current we use THYRISTORS (Chapter II.J.4) which also convert a.c. into d.c. whilst alternating current is controlled with TRIACS. These semiconductors can be joined to form different combinations which supply the required maximum current and voltage, in most cases via a direct connection (without transformer) to the 220 V or 380 V mains. These thyristors or triacs are built into a unit complete with the necessary cooling devices, overcurrent and overvoltage protections and terminals. The current for these units is controlled by merely varying the phase of a current pulse which is fed to the control electrode of the thyristor or triac (Fig. 25). This pulse has but a low power and therefore, one type of pulse generator will suffice for various types of heavy current units. Such pulse generator consists at least of a pulse transformer and a circuit with which the phase of the pulse is adjusted by means of a potentiometer. Various units can be added to this pulse generator. A second unit can measure input voltage fluctuations and correct the pulse phase if deviations should occur. A third unit measures the current and can so correct the pulse to such an extent that the voltage remains constant at a varying load, but if the current should reach an unacceptably high value the unit reduces the voltage. The ideal condition is where a certain potentiometer position corresponds with a certain motor speed. A pulse generator (Chapter II.N.4) is then fitted on the motor shaft, and by means of digital units, as described in the next chapter, the pulse phase is controlled until the motor speed (r.p.m.) corresponds to the potentiometer setting. The oscillogrammes in Fig. 216 show the effect of such an installation, that is to say, how the speed of an 18 hp motor of 1500 r.p.m. nominal, varies when switched on under full load to be switched off again with the potentiometer adjusted to 50 and 1500 r.p.m. respectively. During this procedure the armature winding is connected to a stable d.c. source. The rapid drop in speed when the motor is switched off is obtained by letting the motor supply current back into the mains via another rectifier, which

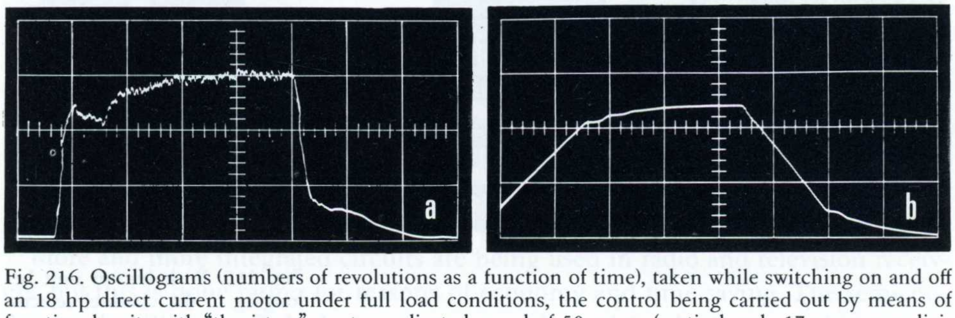


Fig. 216. Oscillograms (numbers of revolutions as a function of time), taken while switching on and off an 18 hp direct current motor under full load conditions, the control being carried out by means of functional units with "thyristors": a. at an adjusted speed of 50 r.p.m. (vertical scale 17 r.p.m. per division, horizontal scale 1 sec per division), b. at an adjusted speed of 1500 r.p.m. (vertical scale 600 r.p.m. per division, horizontal scale 0.5 sec per division).

has a considerable braking effect. Older electrical engineers who in their younger years have had to control a d.c. motor by means of a branch resistor will feel their mouths water at these new control posibilities.

In controlling furnaces the temperature reached can be transferred in an analogue or digital manner to be compared with the temperature setting, the difference causing the pulse phase on the thyristor to shift slightly. For controlling lighting the system is adjusted to a certain luminosity level. Furthermore it is possible to control liquid levels, pressures, tensions, etc.

## II.N.4. CIRCUIT BLOCKS FOR DIGITAL FUNCTIONS

As already explained, digital circuits are built up from a number of standard units as shown in Fig. 36a and 36b. They are therefore particularly suitable to form functional units. Whereas in complex circuitry, such as for computers for instance, integrated circuits are mostly used, simpler subassemblies consisting of "discrete" components on a print plate are mostly used for the automatic control of machines (AUTOMATION). The power can then be somewhat higher, and larger capacitors and resistors as well as coils can be included in the circuit. They are compounded in a metal or plastic encapsulation, which makes them moisture and temperature resistent. Philips supply such functional units under the name of CIRCUIT BLOCKS, MODULAR SUBSYSTEMS or NORBITS.

One of the simplest functions in automation is COUNTING. The signal source can be a photo cell on which a beam of light is incident to be intercepted by moving objects, or an electromagnet with an air gap through which a metal plate moves. In this way it is possible, for instance, to count the number of ready products or the number of revolutions made by a shaft by counting the number of current pulses produced by the signal source.

We already saw in Chapter I.22 that the FLIP-FLOP circuit can count in the binary system. Fig. 217 shows this circuit once again, but now as the modern ver-

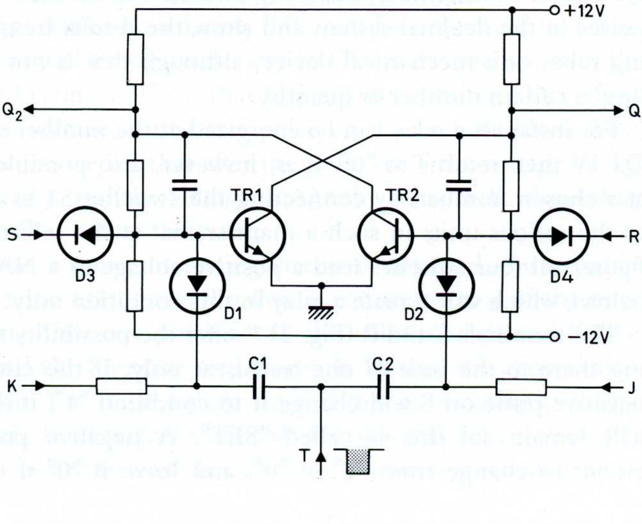


Fig. 217. Flip-flop circuit with various possibilities by making use of the inputs T, K, J, S or R and the outputs Q1 and Q2.

sion with NPN transistors. Some extra terminals have been added offering more switching facilities.

At a correct choice of resistors, capacitors and transistor characteristics this circuit has two stable positions (BISTABLE MULTIVIBRATOR). When TR1 is conducting, the diagonally connected output Q1 is earthed, thus receiving zero potential (this condition of the circuit is defined as "0"). As a result the base of TR2 becomes positive, and so TR2 is blocked. As a result of this again the second output Q2 becomes positive, and with it the base of TR1, so that TR1 is made conductive, bringing us right back to where we started. So in this condition the circuit is stable. A second stable condition is: TR2 conducts, TR1 blocks. Both stable conditions are given below.

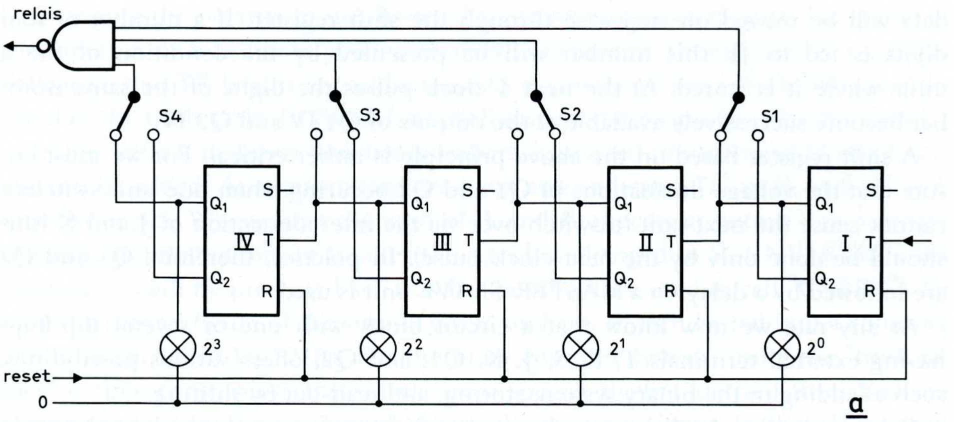
| condition | TR1 | TR2 | base TR1 | base TR2 | output Q1 | output Q2 |
|---|---|---|---|---|---|---|
| "0" | conducts | blocks | positive | negative | zero | positive |
| "1" | blocks | conducts | negative | positive | positive | zero |

The circuit can be brought from one condition to the other by feeding a negative pulse to the base of the conducting transistor which will then block. A negative pulse on trigger contact T will, via one of the capacitors C1 or C2 and the diodes D1 or D2, always find its way to the base of the transistor which is conductive at that moment, for that base is positive, and the diode will pass the pulse in that direction. If a train of negative pulses is fed to TI, each pulse will operate the circuit, and starting from the condition "0", Q1 will become positive at the odd pulses, whilst Q2 will become positive at the even pulses. Hence the circuit will only react to negative pulses on T.

If in accordance with Fig. 218a the trigger terminal T of a second circuit II is connected to output Q1 of the first circuit, it will be clear from what was said above, that at only the odd pulses will cause Q1 to change from positive to "0", thus feeding a negative pulse to T II so that circuit II switches. From the TRUTH TABLE given in Fig. 218b it appears how in 4 units the successive pulses bring these units in different "conditions", the number of pulses fed to TI being written in these units in the binary system. This can be made visible by lamps between Q1 and 0. Furthermore, decoding circuits can be used to translate the number of pulses in the decimal system and show the results by means of numerical counting tubes or a mechanical device, although that is not necessary for just "signalling" a certain number or quantity.

For instance, a relay can be energized at the number of 16 because at that count Q1 IV then returns to "0". It is, however, also possible to have a relay function at a chosen number, by connecting the switches S1 to S4 (Fig. 218) to Q1 or Q2 of the various units in such a manner that at the selected number ("1001" in the figure), all four switches feed a positive voltage to a NAND gate (to be described below), which will actuate a relay in this condition only.

The terminals S and R (Fig. 217) offer the possibility of leading the pulses arriving there to the base of one transistor only. If the circuit is in condition "0", a negative pulse on S will change it to condition "1", if the circuit is already "1", it will remain so: this is called "SET". A negative pulse on R will cause the circuit to change from "1" to "0", and leave it "0" if the first was "0" (RESET).



| number of pulses | pos. of the circuit | | | |
|---|---|---|---|---|
| | IV | III | II | I |
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 |
| 2 | 0 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 | 1 |
| 4 | 0 | 1 | 0 | 0 |
| 5 | 0 | 1 | 0 | 1 |
| 6 | 0 | 1 | 1 | 0 |
| 7 | 0 | 1 | 1 | 1 |
| 8 | 1 | 0 | 0 | 0 |
| 9 | 1 | 0 | 0 | 1 |

b

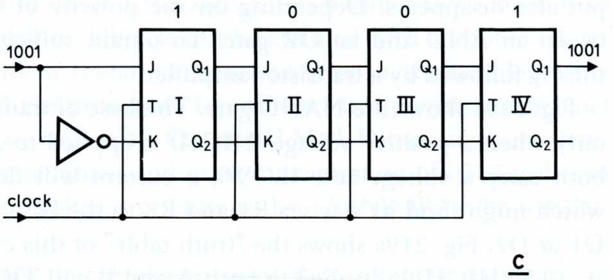


Fig. 218. Counting with flip-flop circuits: a. counter with reset and signalling at a certain count, b. truth table, c. shift register.

Fig. 218a shows how a common reset wire is used to return all units to "0".

Then the terminals J and K enable two kinds of data to be processed in one flip-flop. These terminals "block" the diodes D1 and D2, to which they are connected, when a positive voltage is applied to them. When J and K have a voltage 0, the circuit described will operate as a bistable multivibrator If they both receive a positive voltage, the circuit is blocked for negative pulses on T. If, however, J carries a positive voltage whilst K is 0, only TR1 will be influenced by a negative pulse on T. The circuit will then switch only if TR1 is conductive, so if the circuit is "0". As a result it becomes "1", so that Q1 becomes positive. Vice versa, when K receives a positive voltage, a negative pulse on T can only block TR2, whilst Q2 becomes positive. So we see that after the pulse on T the voltage on Q1 assumes the same polarity as J, and the voltage on Q2 the same as K, depending on the condition of the flip-flop before the pulse.

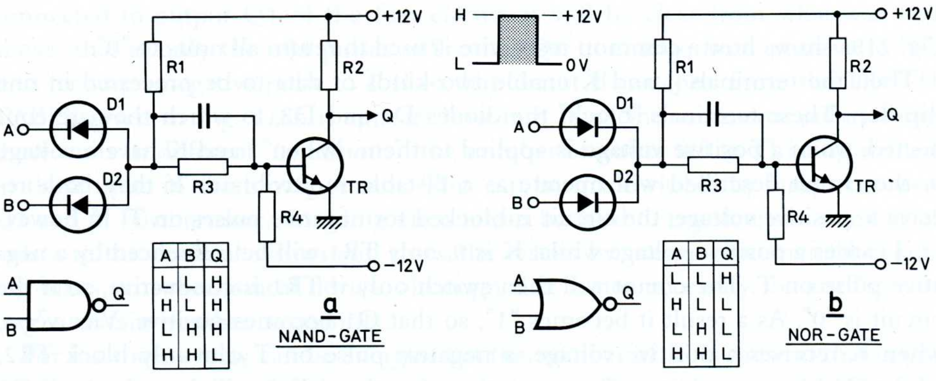
As illustrated in Fig. 218c, a most interesting circuit is obtained, the so-called "SHIFT REGISTER", when the Q1 of one flip-flop is connected to J of the next, and Q2 with K of the next. After one pulse on TI the polarity of JI and KI has been transferred to Q1 I and Q2 I, and so to J II and K II: thus the stored information has been shifted from I to II. At a pulse on T II this information is shifted to J III and K III, and so on. If all trigger terminals are now parallel connected to one "clock-pulse" and the data pulses are fed to JI at the same frequency, these data will be moved on step-wise through the shift register. If a number of four digits is fed to JI, this number will be presented by the condition of the 4 units where it is stored. At the next 4 clock pulses the digits of the same number become successively available at the outputs of Q1 IV and Q2 IV.

A shift register based on the above principle is rather critical. For we must ensure that the voltage fluctuations of Q1 and Q2 occuring when one unit switches, cannot cause the next unit to switch over via the interconnection of J and K (this should be done only by the next clock pulse). In practice, therefore, Q1 and Q2 are followed by a delay, or a MASTER-SLAVE unit is used.

At any rate we now know that a circuit block with one or several flip-flops having external terminals T, R, S, J, K, Q1, and Q2, offers various possibilities, such as adding in the binary system, storing, and read-out by shifting.

Other switching facilities are obtained with "gates", one of which is shown in Fig. 36b. Here there is no memory effect: as soon as the input disappears, the output also disappears. Depending on the polarity of the diodes we distinguish between an AND and an OR gate. To obtain sufficient output power, the gate is mostly followed by a transistor amplifier.

Fig. 219a shows the NAND gate. The base of transistor TR can become positive only when a positive voltage (HIGH) is applied to A and B. If either A or B, or both carry a voltage zero (LOW), a current will flow from the positive voltage, which might find its way via R1 and R3 to the base of the transistor, and away via D1 or D2. Fig. 219a shows the "truth table" of this circuit. Only when a voltage of + 12 V (HIGH) is applied to both A and B will TR1 conduct and will Q be connected to earth (LOW).



| A | B | Q |
|---|---|---|
| L | L | H |
| H | L | H |
| L | H | H |
| H | H | L |

| A | B | Q |
|---|---|---|
| L | L | H |
| H | L | L |
| L | H | L |
| H | H | L |

| A | B | C | D | E | S | S' |
|---|---|---|---|---|---|---|
| L | L | H | H | H | L | L |
| H | L | H | L | H | H | L |
| L | H | H | H | L | H | L |
| H | H | L | H | H | L | H |

Fig. 219. Gate circuits: a. NAND gate, b. NOR gate, c. adder with 4 NAND gates, all with truth table.

Changing the polarity of the diodes (Fig. 219b) gives a NOR gate. For if one of the inputs A or B becomes positive (HIGH), the base of TR will receive a positive voltage so that TR begins to conduct and Q is connected to earth (LOW), as indicated in the truth table. In Part III we shall discuss these circuits in further detail.

In Fig. 219c it is shown how two numbers, expressed in successive H and L levels, can be added. As is customary, we have now indicated "1" as H, and "0" as L (positive logic). We see that the sum of A and B appears at output S as a binary number, and that at output C, after the sign has changed in S′ (INVERTER), the number "2" will be presented as "1" (what in school we used to call "CARRY"). So here the second figure of the sum is given, and can in turn be fed into a counting circuit as shown in Fig. 218a. For the more complex systems using the "truth tables" is time consuming. Instead we use a special algebra (derived from Boolean algebra) and we can let a computer design the required circuit. In Part III we shall discuss in more detail how arithmetical operations are carried out.

For automatic measuring and control the signal (the number to be measured) must be expressed in voltage pulses. Objects can be counted with a photo cell, revolutions can be counted with a metal vane mounted on a shaft and passing through the air gap of a fixed electromagnet (even better is a permanent magnet on the shaft facing a high-speed reed relay, Chapter II.M.8). Dimensions, mechanical pressure, liquid levels, temperature, and even concentration of chemicals can first be converted into analogue voltages or currents by a SENSOR, to be translated into digital data.

The result of the observation or calculation must then often be expressed in mechanical movement to achieve automatic control. This can be done with relays, stepper motors (Chapter II.M.6), thyristor circuits (II.N.3) or electrically operated pressure air valves. Certain data can be supplied from punch cards, punch tapes, magnetic recording tapes or core memories.

Apart from the subassemblies mentioned so far, we also need supply units for the direct voltages for the circuit. And finally a clock pulse generator and a circuit that can restore the "degenerated" pulses to their original rectangular form.

The digital subassemblies discussed in this chapter open the way to automation in numerous fields, since by their use the application of electronics is largely facilitated even for non-specialists.

## II.N.5. MAGNETIC CORE MEMORIES

As we saw in the preceding chapter, a series of flip-flop circuits can be used for storing a series of numbers in the binary system, to be read out again at a later time. In Chapter II.J.11 it was explained that manufacturers of integrated circuits are trying by means of "large scale integration" to find a cheaper solution for making a large memory based on this principle. For the time being the magnetic core memory has the advantage of the lower price.

Data can also be stored in punch cards, punch tapes, magnetic recording tape, drums or discs, but then a whole batch of cards, a great length of tape, or a track on the drum or disc must be searched before the required data can be found.

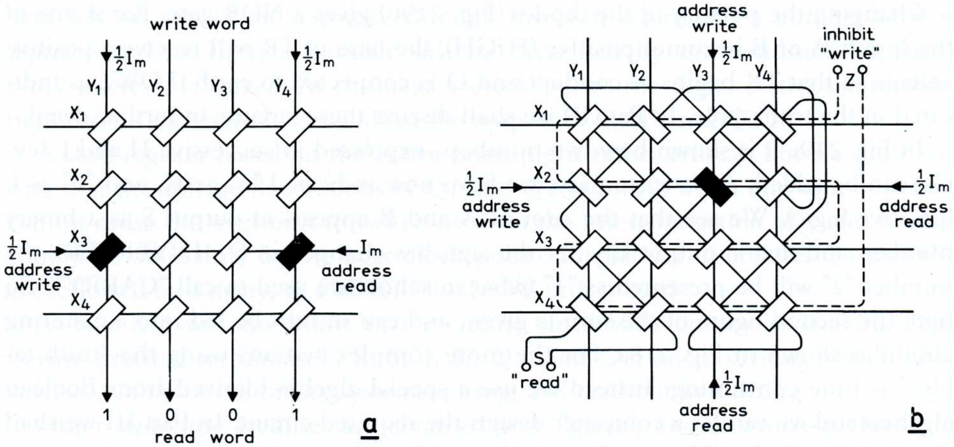


Fig. 220. a. Principle of the "2D" system (address nr 3, number 1001), b. Principle of the "3D" system (1 bit), (address 7, number 1).

The advantage of the magnetic core memory is that it gives immediate access to any data stored in the memory (RANDOM ACCESS).

In this memory too, the data is stored in the form of "BITS". One bit can be stored in one core that is magnetized in the "0" or "1" direction. The number 100 × $10^6$ (a millionair's banking account in cents) requires 27 bits; if letters or other symbols are to be expressed according to the binary system, 5 bits are needed for 32 different symbols. The number stored in a memory is called a WORD.

If several numbers (words) must be stored in a memory, it must be possible to recover a certain word. The "word" concerned then receives a number, an ADDRESS. The capacity of the memory is expressed as follows: 1k8 means that the memory can contain 1000 words of 8 bits each.

The annular cores with the cross wires, the principle of which is shown in Fig. 160, are distributed as logically as possible over flat planes which are known as MATRIX. The abovementioned 1k8 memory contains 8 such matrices of 32 × 32 = 1024 cores each.

The simplest way of "organizing" a memory is using the 2 D(imensional) system as illustrated in Fig. 220a. Here the horizontal X-wires determine the address, whilst the vertical Y-wires determine the "content" of the word. If a pulse of ½ $I_m$ is sent through X3 and simultaneously through Y1 and Y4, the word "1001" has been written at address 3. If we wish to know what address 3 contains, we send a current $I_m$ through the X3 wire, the black cores change polarity and the outputs of Y1 to Y4 produce a combination of pulses forming the word 1001. This system has the advantage that a high current can be used for read out; a drawback is that each Y-wire in series must run through as many cores as there are addresses, which makes the system highly sensitive to interference.

With the 3 D(imensional) system the address is selected according to the coincidence principle, as shown in Fig. 160. The same set-up has been drawn again in Fig. 220b, the cores being so positioned that adjacent currents flow in opposite directions so as to avoid mutual influences. One bit of the word is chosen by coincidence of, say, X2 and Y3 in Fig. 220b. If a current of ½ $I_m$ is sent through these wires, the black core will change polarity. During read out a current of $-\frac{1}{2}\,I_m$ is sent through X2 and Y3, and the switching of the black core produces a voltage pulse in the S-wire.

A separate "matrix" is made for each bit of the word. The X- and Y-wires of all matrices are connected in series. If $X = Y$, the number of addresses is A, and the number of bits per word is B, a total of $B\sqrt{A}$ cores are connected in series.

We have now seen how a "1" is written, but for writing a "0", another piece of information must be added. This is done by means of the "inhibit" wire Z presented as a dashed line in Fig. 220b. This Z-wire runs parallel with the X-wire, but is connected in series in one matrix for all X-wires. In the matrix into which a "0" must be read, a current of $\frac{1}{2}\,I_m$ is sent through the Z-wire during write-in, opposing the current through the X-wire. Fig. 221 shows the special arrangement and current circuits for writing and read out. The word appears on the S-wires of the various matrices (in this figure "1001").

The number of wires through one core has now increased to 4, and as the ring is continually being made smaller (Chapter II.K.3 Fig. 161), this is getting difficult. Therefore an intermediate form between the 2D and 3D, the so called 2½D system, is sometimes used. Here the Y-wires are divided into groups of B bits. When choosing the address, a group of B bits is chosen to which the word data must then be fed in the form of "inhibits". So the circuit required for this must be repeated many times. The number of addresses is now:

$$A = X \times Y/B,$$

from which we can deduce that the number of cores in series has diminished to $\sqrt{A.B}$, or is $\sqrt{B}$ times smaller than with the 3D memory.

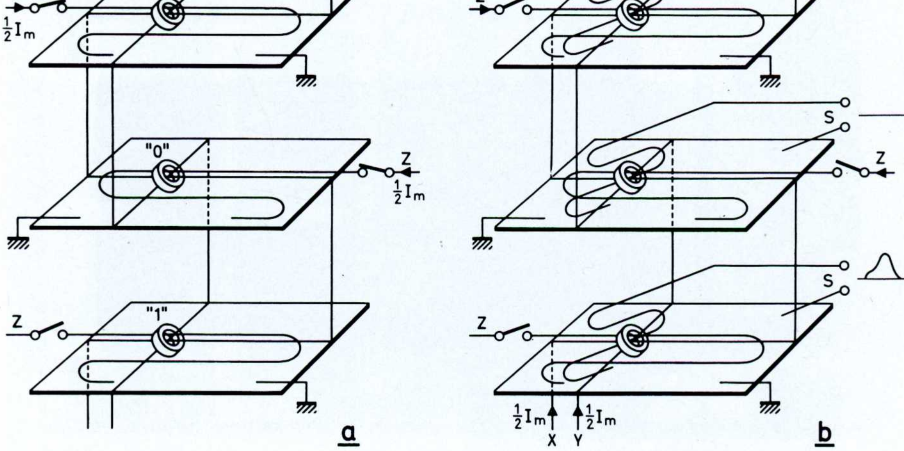


Fig. 221. Survey of the 3D system in several planes: a. writing the "word" 1001, b. reading the "word" 1001.

The electronics involved in connecting the memory to the computer will be discussed in Chapter III.E.5.

It should be noted that all systems are "destructive", meaning that if a read out is to be preserved, it should immediately be written again by means of a positive pulse on X and Y, whilst the signal from S is fed to the "inhibit" wires.

As regards the mechanical design, using very small cores offers the possibility of accommodating the various planes (Fig. 221) on a single panel (Fig. 222). The X-wires can now be drawn through 4, the Y-wires through 2 planes, and the various connections can be made on the print plate on which the matrices are mounted.

From Fig. 222 it appears that the entrances to all X- and Y-wires must be made available on the outside. By mounting a part of the switching system on the print plate the number of spring contacts and soldered connections could be reduced even further (see Fig. 247b).

As said at the beginning of this chapter, the advantage of such a memory is that data can be stored and read, in very short time because the address selection system is electronic. The time elapsing before the input data is fed into the input registers is determined by the electronics concerned and is already very short. The time passing between giving the "read" signal and the data being available at the output is known as ACCESS TIME, which, for a relatively simple 3D 1k8 memory is no more than 0.6 μs. The time needed for the memory to get set for the next operation is the CYCLE TIME which in this case is 4.0 μs. The Philips G(2½D) memory with its capacity of 524288 words of 9 bits (nearly 5 × $10^6$ bits), has an "access time" of 1.2 μs and a "cycle time" of 2.5 μs. This means that about 400 000 words per second can be stored or read. Truly another victory of electronics!

Such a memory has indeed grown into a large and complex subassembly, but one that qualifies as a "functional unit" with the simple direction for use: "These

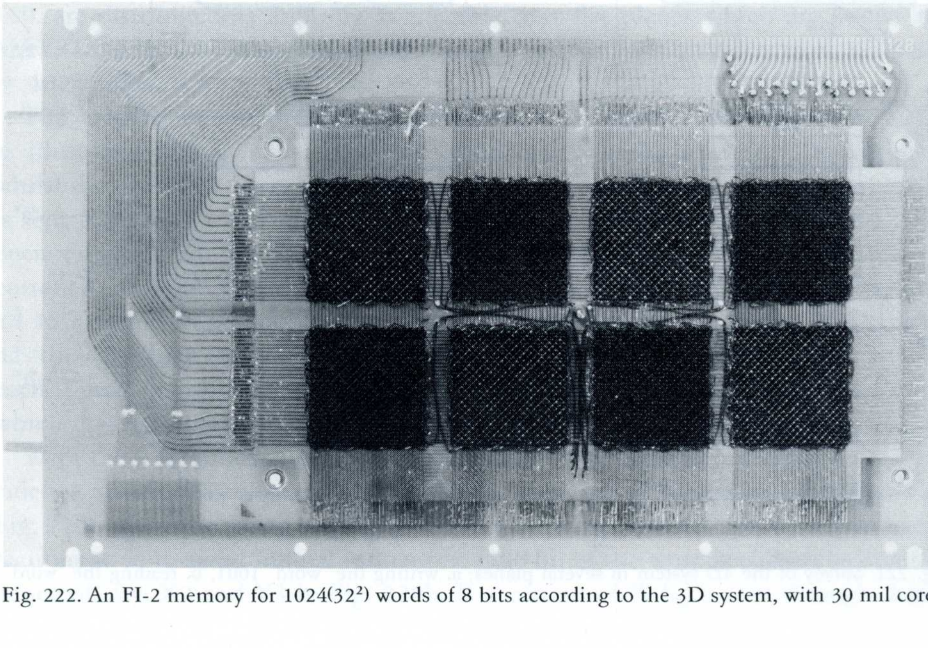


Fig. 222. An FI-2 memory for 1024($32^2$) words of 8 bits according to the 3D system, with 30 mil cores.

are the terminals for the pulses required for the address, these are for the pulses required for the word, and these are for reading out the word".

For certain applications (the "intermediate" memory in computers which operate with switching times of nanoseconds) the memories discussed above are still too slow.

Even steeper pulses can be used provided there is no interference from adjacent wires owing to electromagnetic induction. This can be remedied by replacing each X-wire by two wires which receive simultaneous pulses, one opposing the other (TWO CORE PER BIT SYSTEM, cycle time 250 ns).

In this era of labour saving in industry, the running of 3 or 4 wires through these minute cores is no simple matter even though it can be partly done automatically. Numerous factories farm out this work in Hong Kong or Taiwan where workers to do this job are available. Obviously it has been tried to arrive at a more automated production process by depositing layers of magnetically and electrically conductive metals according to a certain pattern. So far it has been found impossible to obtain results equal to those of the core memory.

Experiments are also made with the "plated wire" system. Here a magnetic layer is applied to copper wire around which the diagonal wire is then woven. Such memories are made with an access time of 125 nanoseconds.

Yet the core memory system keeps the advantage of clearly defined difference between "1" and "0". And so, throughout the world, wires are still being threaded through millions and millions of annular magnetic cores!

# PART III. CIRCUITS

On the following pages it is explained how the parts and sub-assemblies discussed in Part II can be combined into circuits which as such form a complete electronic device. The development of electronics has only been possible by virtue of a continuous interaction between circuit technique and technology of components. This will have to continue if the increasing complexity of this technique leads to specialisation of those engaged in it.

Of course, the space allocated for this part makes it impossible to deal with all possible circuits. The only solution was to make a choice suitable to illustrate how the numerous aspects of electronics could be realized.

In general, the circuits studied by the APPLICATION LABORATORY of the PHILIPS PRODUCT DIVISION ELCOMA over the years, will be discussed.

# SECTION A. BASIC CIRCUITS

## III.A.1. RECTIFIERS, POWER SUPPLY.

As electron tubes and transistors can function as amplifiers only if direct voltage is applied to the various electrodes, it is necessary for the POWER SUPPLY of electronic equipment to convert the alternating voltage of the mains into a di-

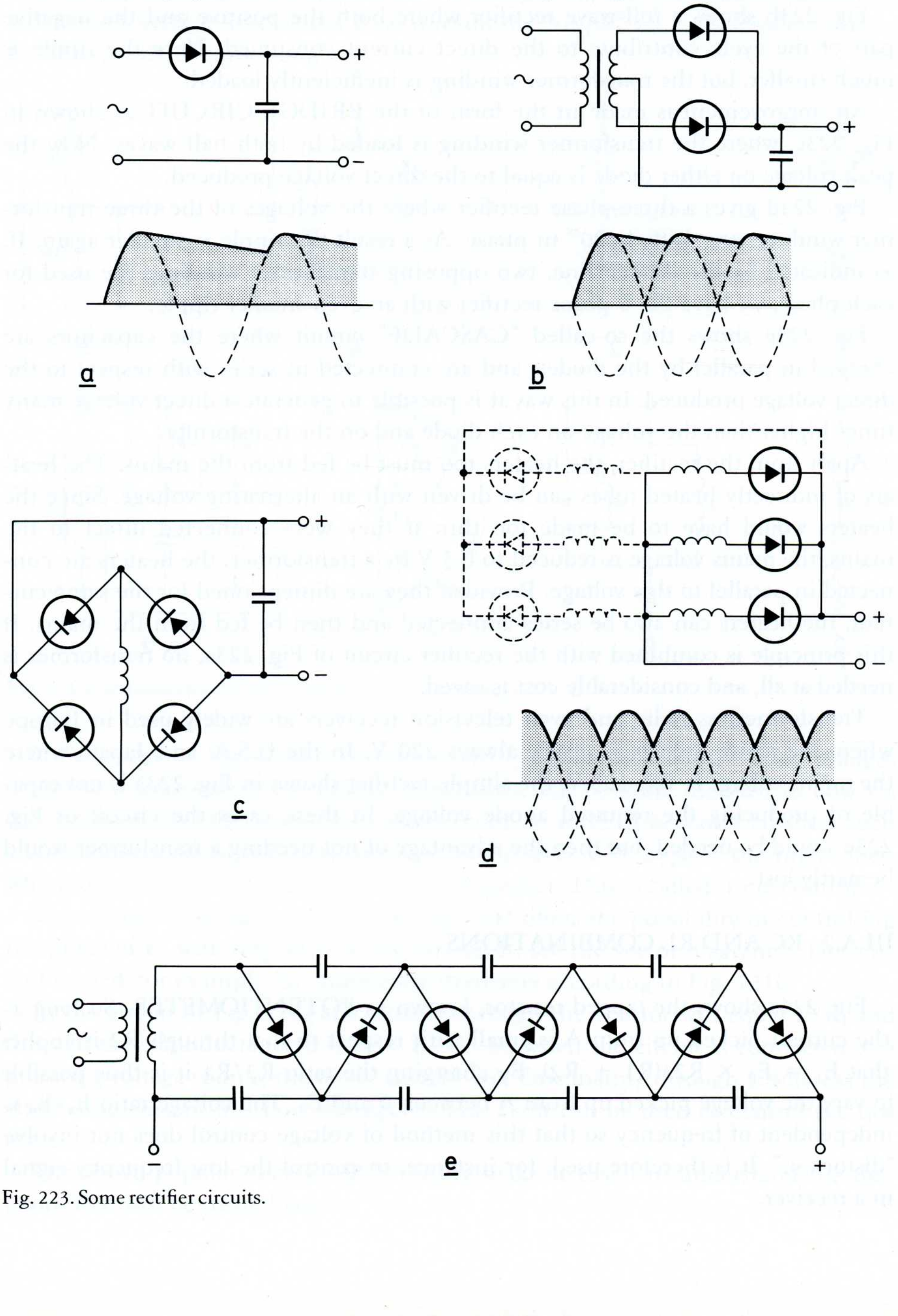


Fig. 223. Some rectifier circuits.

rect voltage. This is done by means of high vacuum or gas filled rectifier tubes or semiconductor diodes. To smooth the "ripple" of the rectified alternating voltage, the output of the rectifier is usually bridged by a capacitor.

Fig. 223a shows a half-wave rectifier (see also Fig. 58) where the capacitor is charged to about the peak value of the mains voltage. Since only the positive half wave of the cycle is used, the ripple is rather large. The peak voltage across the diode is about twice the direct voltage. This circuit is used for AC-DC receivers which can be connected to either d.c. or a.c. mains.

Fig. 223b shows a full-wave rectifier where both the positive and the negative part of the cycle contribute to the direct current consumed. Here the ripple is much smaller, but the transformer winding is inefficiently loaded.

An improvement is made in the form of the BRIDGE CIRCUIT as shown in Fig. 223c, where the transformer winding is loaded by both half waves. Now the peak voltage on either diode is equal to the direct voltage produced.

Fig. 223d gives a three-phase rectifier where the voltages of the three transformer windings are shifted 120° in phase. As a result the ripple is smaller again. If, as indicated by the dashed line, two opposing transformer windings are used for each phase, we have the 6-phase rectifier with an even smaller ripple.

Fig. 223e shows the so-called "CASCADE" circuit where the capacitors are charged in parallel by the diodes, and are connected in series with respect to the direct voltage produced. In this way it is possible to generate a direct voltage many times higher than the voltage on each diode and on the transformer.

Apart from the rectifier, the heaters too must be fed from the mains. The heaters of indirectly heated tubes can be driven with an alternating voltage. Since the heaters would have to be made too thin if they were connected direct to the mains, the mains voltage is reduced to 6.3 V by a transformer; the heaters are connected in parallel to this voltage. Provided they are dimensioned for the same current, the heaters can also be series connected and then be fed from the mains. If this principle is combined with the rectifier circuit of Fig. 223a, no transformer is needed at all, and considerable cost is saved.

Transformerless radio and even television receivers are widely used in Europe where the mains voltage is nearly always 220 V. In the U.S.A. and Japan, where the mains voltage is 105-125 V, the simple rectifier shown in Fig. 223a is not capable of producing the required anode voltage. In these cases the circuit of Fig. 223e would be needed, but then the advantage of not needing a transformer would be partly lost.

## III.A.2. RC AND RL COMBINATIONS

Fig. 224a shows the tapped resistor, known as POTENTIOMETER. So long as the current picked up from A is small with respect to that through R2 it applies that $E_1 = E_0 \times R2/(R1 + R2)$. By changing the ratio R2/R1 it is thus possible to vary the voltage picked up from A between 0 and $E_0$. The voltage ratio $E_1/E_0$ is independent of frequency so that this method of voltage control does not involve "distortion". It is therefore used, for instance, to control the low frequency signal in a receiver.

For a resistor and capacitor in parallel (Fig. 224b) it applies that $I_1 = E/R$, and $I_2 = E\omega C$ ($\omega$ = angular frequency).

If this combination is connected in series with a direct voltage with a ripple, the capacitor will reduce the voltage drop across the resistor caused by the ripple: this is known as decoupling. If one point in a circuit is to be fed with alternating and direct voltages from different circuits, the alternating voltage is supplied via a capacitor, which blocks the direct current.

If in Fig. 224c $E_0$ is a pulsating direct voltage, the a.c. component through C will cause a voltage drop across R; as a result the "ripple" on $E_1$ is smaller than on $E_0$. This is called "smoothing".

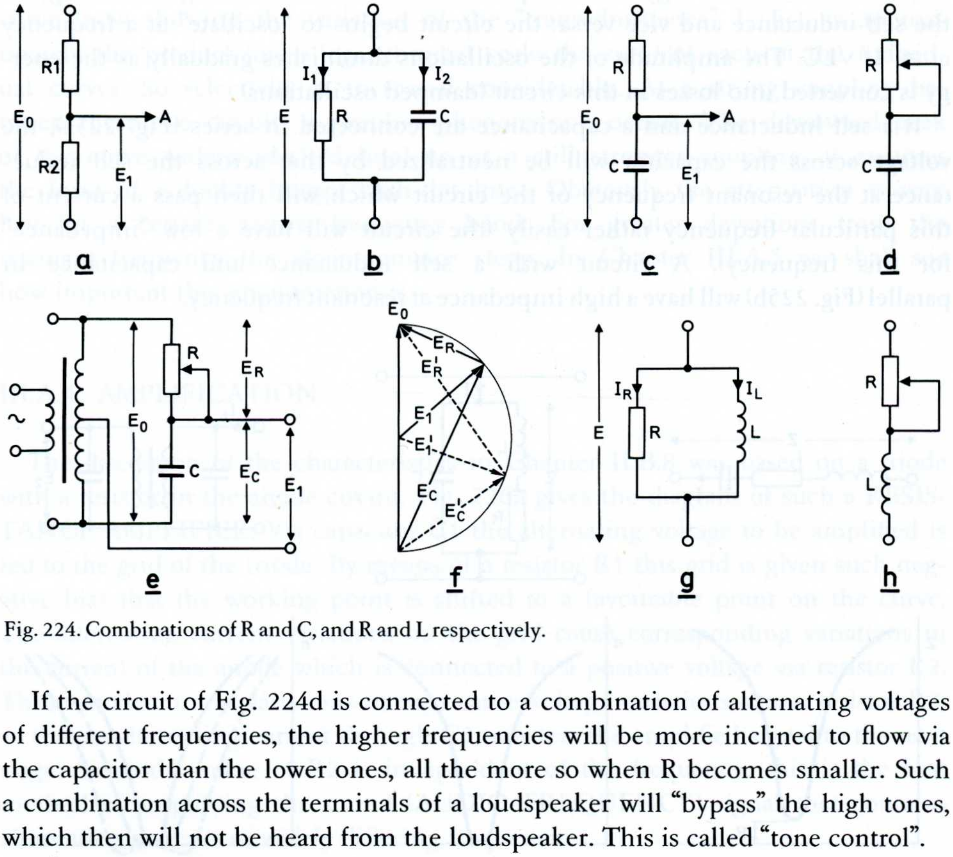


Fig. 224. Combinations of R and C, and R and L respectively.

If the circuit of Fig. 224d is connected to a combination of alternating voltages of different frequencies, the higher frequencies will be more inclined to flow via the capacitor than the lower ones, all the more so when R becomes smaller. Such a combination across the terminals of a loudspeaker will "bypass" the high tones, which then will not be heard from the loudspeaker. This is called "tone control".

The circuit in the form as given in Fig. 224e offers the possibility of controlling the phase of $E_1$ with respect to $E_0$ by varying R, see the vector diagram f. This circuit is used, for example for controlling thyristors according to Fig. 121b.

In the circuit of Fig. 224g the current through the resistor will equal $E/R$, and through the selfinductance it will be $E/\omega L$. So when the circuit is connected to a pulsating voltage source, the direct current will flow mainly through L whereas the alternating current will flow through R. Fig. 224h can be used to control the low tones.

The above popular survey will probably make it easier to understand the diagrams that will be treated later.

## III.A.3. RESONANT CIRCUITS

Using both self inductances and capacitances in circuits gives rise to a new phenomenon. For in the capacitance and the self inductance energy is "stored" to be released again ¼ cycle later. When the voltage reaches its maximum value, the capacitor is fully charged; when the current reaches its maximum the magnetic field in the self inductance is also maximum. The capacitor can supply the current necessary to magnetize the self inductance, whilst the self inductance supplies the voltage to charge the capacitor. If a single current pulse is fed to a circuit containing a capacitance and a self inductance, the result will be an oscillating current while the electrical energy of the capacitor is converted into magnetic energy in the self inductance and vice versa: the circuit begins to "oscillate" at a frequency $\omega = 1/\sqrt{LC}$. The amplitude of the oscillations diminishes gradually as the energy is converted into losses in this circuit (damped oscillations).

If a self inductance and a capacitance are connected in series (Fig. 225a), the voltage across the capacitor will be neutralized by that across the self inductance at the resonant frequency of the circuit which will then pass a current of this particular frequency rather easily (the circuit will have a low "impedance" for this frequency). A circuit with a self inductance and capacitance in parallel (Fig. 225b) will have a high impedance at resonant frequency.

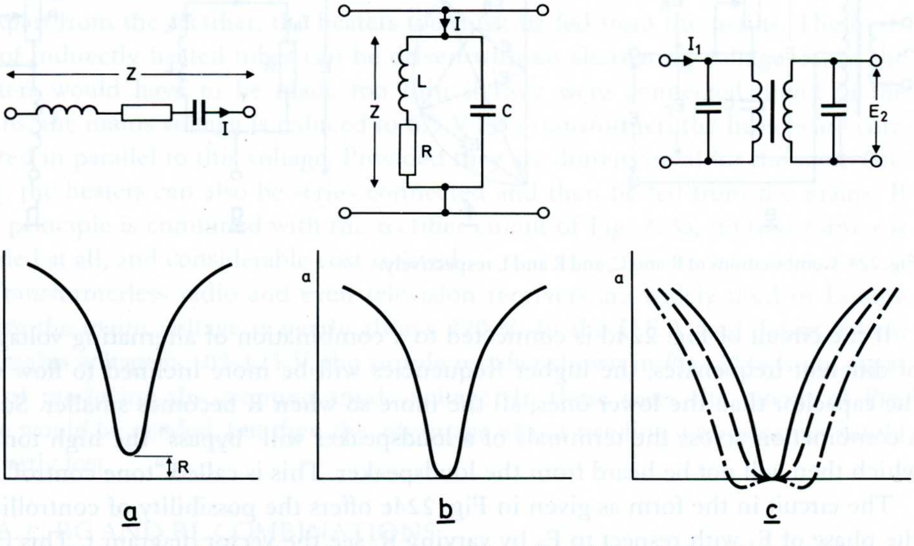


Fig. 225. Resonant circuits: a. L, R and C in series, b. L, R in series and C in parallel, c. band-pass filter.

This can be explained by means of a simple calculation. In case R is small as compared with ωL and 1/ωC, the ratio between current and voltage in the two circuits can be represented in simple formulae:

$$\text{circuit (a)} \qquad I = E/\sqrt{\{R^2 + (\omega L - 1/\omega C)^2\}}$$

$$\text{circuit (b)} \qquad I = EC\sqrt{\{R^2 + (\omega L - 1/\omega C)^2\}}\,/L.$$

At $\omega = 1/\sqrt{LC}$ circuit (a) carries maximum current ($I = E/R$), and circuit (b) the minimum current $ECR/L$. If we try to send an alternating current of resonant frequency through this circuit, a high voltage will occur at the terminals; this phenomenon is known as "resonant magnification". In Chapter II.M.1 it was already shown how the impedance varies as a function of frequency so long as it does not differ too much from the resonant frequency. The function of Z is shown under "a" on a logarithmical scale, "b" shows 1/Z as compared with the condition of resonance, which is defined as the relative "admittance" (a) and is expressed on a logarithmical scale in decibels (dB) (RESONANCE CURVE). As explained earlier: the greater $Q = \omega L/R$ of the circuit, the sharper the peak, the steeper the slope, the greater the "selectivity".

For two magnetically (or capacitively) coupled circuits as shown in Fig. 225c, which are tuned to the same frequency (BAND-PASS FILTER), the resonant curve (now showing the variation of the "transadmittance" $I_1/E_2$), is approximately the product (on a logarithmical scale the sum) of each of the individual curves. So selectivity is increased considerably. At a strong coupling between these two circuits a peculiar phenomenon occurs. The downward peak of the curve widens (dashed line) or, at a still stronger coupling, it assumes the form of a double hump (dash-dot line). Obviously the attenuation is very low for a certain narrow frequency band. For greater deviations from the resonant frequency the slope remains steep. In Chapter III.A.5 we shall see how important this phenomenon is.

## III.A.4. AMPLIFICATION

The discussion of the characteristics in Chapter II.B.8 was based on a triode with a resistor in the anode circuit. Fig. 226a gives the diagram of such a RESISTANCE AMPLIFIER. Via capacitor C1 the alternating voltage to be amplified is fed to the grid of the triode. By means of a resistor R1 this grid is given such negative bias that the working point is shifted to a favourable point on the curve. The alternating current variations on the grid cause corresponding variations in the current of the anode which is connected to a positive voltage via resistor R2. These anode current variations in turn cause voltage variations across resistor R2, so that an alternating current through C2 transfers the amplified signal to the next stage. Since the value of R2 is independent of the frequency, this is the best method of amplifying the sound (AUDIO FREQUENCY) signal that contains alternating currents of widely differing frequencies.

Fig. 226b gives the circuit diagram of the output amplifier where R2 is replaced by the loudspeaker. Hence the latter also carries the direct current component of the anode current, but at a correct dimensioning this does not affect the volume. This diagram indicates an interesting method to apply the appropriate negative voltage to the grid: a resistor R3 is incorporated in the lead between the "0" and the cathode. The direct current component of the anode current causes a voltage drop across this resistor, so that the cathode becomes positive with respect to the "0", and consequently the grid connected to it will become negative with respect to the cathode. Here we have a somewhat automatic control, for should the anode current assume too high a value, the grid becomes more negative, and so the

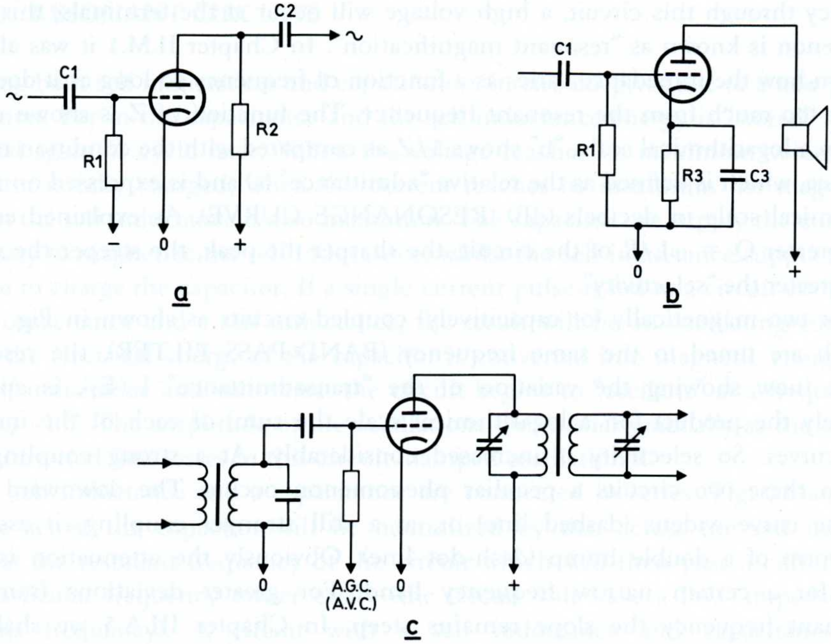


Fig. 226. Resistance and band-pass filter amplification: a. resistance amplification, b. output stage with cathode resistor, c. amplifier stage with band-pass filter.

anode current is reduced again. In order not to have the cathode influenced by the alternating current component of the anode current, a larger capacitor C3 is connected in parallel with R3 to smooth the alternating voltage on the cathode. If this capacitor is omitted, a fluctuation of the anode current will cause a similar fluctuation of the grid voltage. Indeed this means a lower gain, but it can be proved that there is then less distortion. This is known as NEGATIVE FEEDBACK. Often more complex circuits are used for the negative feedback, with which the frequency response can also be influenced.

In Chapter II.D.1 it was already explained that the load resistor in the anode circuit involves losses. To improve efficiency, an element must be found which offers but low resistance for the direct current supply to the anode, but blocks the alternating current, so that a high alternating voltage can be passed on to the next stage. This can be done by means of a choke, but the alternating current flowing through it is frequency dependent. The result is then unequal amplification at different frequencies. If, however, only one frequency (or a narrow frequency band) is to be amplified, it is particularly advantageous to replace the resistor by a resonant circuit as shown in Fig. 225b, whose resonant frequency is tuned to the signal to be received. At this frequency the impedance of the circuit is particularly high, whereas the self inductance with its low resistance will freely pass the direct current from the battery. This amplifier will only pass the "tuned" frequency. Selectivity can be further improved with a band-pass filter as shown in Fig. 225c. With amplification between two band-pass filters the input and output of the signal are isolated (because of the transformers) from the electrodes of the tube. By a proper choice of L and C of each of the band-pass filter circuits within the requirement of $\omega = 1/\sqrt{LC}$ the ratio E/I can be "matched" to the tube or transistor at an equal power E × I.

In Fig. 226c the grid of the triode is connected to the so-called A(utomatic) G(ain) C(ontrol) voltage. Gain control can be introduced by using a tube whose characteristic shows a "tail" which allows the gain to be controlled by varying the grid bias (see Fig. 60). If this grid bias is made more negative when the output signal becomes stronger, we have an automatic gain control (A.G.C.) which prevents overload of the output stage.

## III.A.5. THE MODULATED SIGNAL AND DETECTION

As explained in Chapter I.5, the radio signal at AMPLITUDE MODULATION consists of a sine wave with a radio frequency $\omega$ ($e = E_0 \cos \omega t$), of which a part M (modulation depth) of the amplitude varies sinusoidally with the sound frequency p. The signal can thus be presented as: $e = E_0 (\cos \omega t + M \cos pt \cos \omega t)$ which by means of trigonometry can be elaborated into

$$e = E_0\{\cos \omega t + \tfrac{1}{2}M \cos (\omega + p)t + \tfrac{1}{2}M \cos (\omega - p)t\}.$$

We now see that the A.M. signal consists of the original and two signals of frequencies which are an amount p higher and lower respectively than the basic frequency. The whole sound spectrum is given by two "side bands" and for a faithful reproduction this whole "band width" from $\omega - p$ to $\omega + p$ must be uniformly amplified. At a wavelength of 300 m and a maximum sound frequency of 10 kHz to be transmitted, the band to be amplified lies between 990 and 1010 kHz. And this is exactly what can be done with the abovementioned over-coupled band-pass filter (Fig. 225c). The top of the resonance or response curve can be widened even more by slightly detuning two filters with respect to each other. The same effect can be obtained by reducing Q of the circuit, but then the cut-off at great frequency deviations becomes less satisfactory.

As said in Chapter I.3 (Fig. 6), the modulated H.F. signal can be restored to the original sound signal by a rectifying process, and connecting a capacitor across the telephone (DETECTION). Fig. 227 shows how the modulated H.F. signal, coming from band-pass filter S7-S8, is rectified by diode D1, and "smoothed" by capacitor C9. With potentiometer R15 (compare the telephone in Fig. 6) an adjustable part of the sound signal can be passed on to the A.F. amplifier stage.

A second diode, D2, provides automatic gain control. Via capacitor C13 the H.F. signal is fed to D2; the positive part of the alternating voltage flows away via the diode, the negative part charges capacitor C11. The voltage on C11 against "0" is applied to the first grid of the amplifying tube, so that the gain is controlled. In the meantime the potentiometer R8-R9 has already applied 3 V of the negative to C11. So the voltage on C11 is influenced only when the rectified signal voltage on C11 becomes higher than 3 V, or as soon as a certain threshold voltage of the signal is exceeded (DELAYED A.G.C.).

One might ask to what extent H.F. amplification is used, and where A.F. amplification begins. We saw that the band-pass filter involves less losses than resistance amplification, and that its high impedance results in considerable voltage gain. As a rule the aerial signal is amplified in the H.F. part from a few microvolts in the aerial to a few millivolts at the potentiometer. In the A.F. amplifier the power of a few microwatts is amplified to a few watts into the loudspeaker.

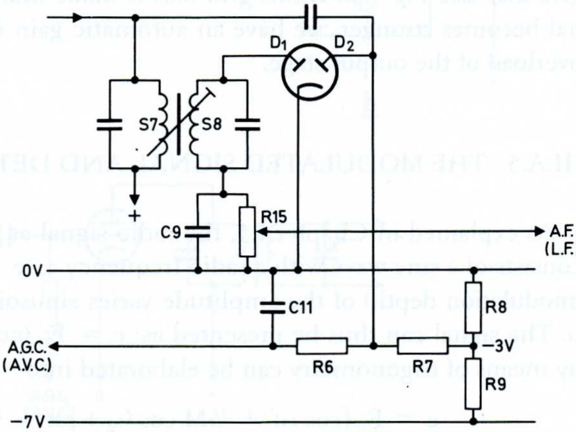


Fig. 227. Detection and A.G.C.

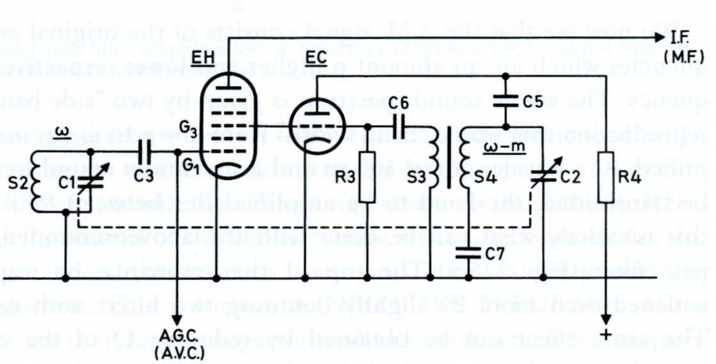


Fig. 228. Mixing.

### III.A.6. MIXING

With a radio set we do not want to receive just one wave length, but a whole series of wave lengths from different stations. To receive all these wave lengths with sufficient "selectivity", the received signal would have to be amplified in a series of band-pass filters, and each band-pass filter would have to be tuned to the corresponding frequency by means of a variable capacitor. However, much better and simpler amplification is offered by the SUPER HETERODYNE principle. Here the received frequency ω is "mixed" with a frequency which is a fixed value "m" lower than all received frequencies. Since this "mixing" is in fact multiplication we can use trigonometry to calculate that:

$$\cos \omega t \cos(\omega - m)t = \tfrac{1}{2}\cos mt + \tfrac{1}{2}\cos(2\omega - m)t.$$

The result is the sumtotal of two sine waves, one with a constant frequency m, which can be further amplified, and the other with a frequency of 2ω − m, which is filtered out in the next amplifying stages. For A.M. receivers the value m is chosen as 452 kHz, which is still 30 times the highest sound frequency.

Although in Fig. 9f the voltages of the respective frequencies $\omega_1$ and $\omega_2$ are added by series connection of the two coils (this method of ADDITIVE MIXING is sometimes used in transistor receivers), it should be noted that mixing is then based on deviations of the linear relationship in the various characteristics with the result that there is indeed multiplication. In a radio tube there is MULTIPLICATIVE mixing. Then the electron current, varied at a frequency "$\omega$" by a first grid, is amplified by a second grid by a factor which is varied with a frequency "$\omega - m$" (Fig. 228). An oscillator with a separate triode serves to generate the frequency $\omega - m$, but here too, we deviate from the diagram given in Fig. 9c. For this oscillator must be able to operate freely, and is therefore "loosely" coupled to the anode and the grid of the oscillator tube by means of capacitors C5 and C6 respectively, whilst the direct voltage is supplied to these electrodes via resistors R3 and R4 independent from the oscillator circuit.

The alternating voltage produced on the grid of EC during oscillation is fed direct to the third grid of the heptode (Chapters II.B.9 and II.B.10). The anode current of this heptode will now vary in accordance with the product of the received signal voltage (frequency $\omega$) and the oscillator voltage ($\omega - m$).

Furthermore the oscillator must generate a frequency $\omega - m$ at all values of $\omega$. This is achieved by tuning the oscillator with a variable capacitor C2 which is mechanically coupled to the one of the aerial tuning circuit (C1). If these two capacitors are to be made identical, the self inductance of S4 must be bigger than that of S2, but since the frequency difference thus obtained is not constant an additional series capacitor C7 (PADDING CAPACITOR) is required to obtain the correct frequency variation.

The super-heterodyne principle has the following advantages:

- the signal is amplified in intermediate frequency circuits, which are tuned to the intermediate frequency and do not require a variable capacitor.
- the frequencies amplified are lower than those of the received transmitters, and therefore the quality, and thus the selectivity, of the circuits can be better.
- if the wave length range to be received is wider than can be covered with one circuit with a tuning capacitor (a factor of 2.5), different coils S2, S3, and S4 can be switched in the circuit without changing the I.F. circuits.

To ensure mixing free from interference and noise, a certain minimum voltage is required on S2. In cases where the received signal is very weak (WORLD RADIO or car radio with a very short aerial), this signal is first amplified in an H.F. pre-amplifier stage. Then three tuning capacitors on one spindle are needed.

Mixing is a critical operation. The resulting I.F. signal must lie exactly in the centre of the frequency response curve of the I.F. band-pass filter (Fig. 225c), otherwise there will be distortion. Interference from adjacent transmitters can give rise to annoying whistles. In view of such interference, the chosen intermediate frequencies differ per country (between 452 and 462 kHz).

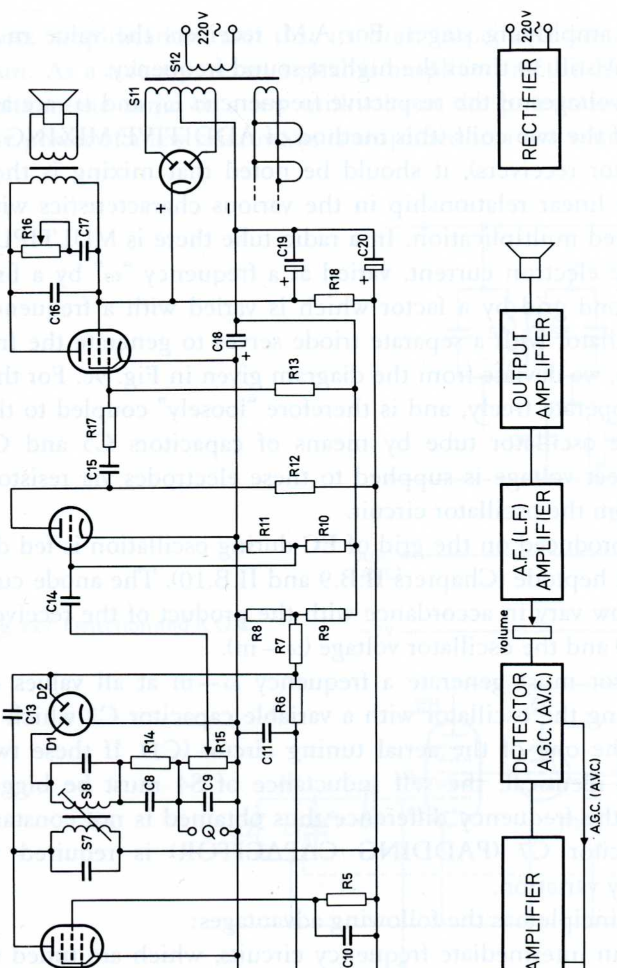


Fig. 229. Complete diagram of a super heterodyne tube receiver for mains supply. Below the block diagram.

## SECTION B. RADIO RECEIVERS

### III.B.1. THE A.M. TUBE RECEIVER

After what was said in the preceding section it will now not be difficult to read the drawing of a receiver as shown in Fig. 229. This diagram has almost become "classical", for since 1935 the principle has not changed and has been reproduced millions of times.

At the right we see the full-wave rectifier (Fig. 223b), whose positive output is smoothed by electrolytic capacitor C19 to be fed to the screen grid, and via the loudspeaker to the anode of output tube EL. For the supply of the other tubes the positive voltage is further smoothed by R18-C20. For the negative grid voltage a special solution has been found here in that the cathodes of all the tubes are smoothed by C18 and are connected to the negative voltage of the rectifier via a common cathode resistor R8-R9. So the negative grid voltage of EL is determined by R8-R9, that for EC is picked up via R10-R11, that for ECH and EF via R8-R9. The advantage of this circuit is that all cathodes are interconnected. Therefore this design can also be used for battery operated receivers (in which the heaters connected to the heater battery are at the same time the cathodes) and offers the free choice of joining the separately indicated tubes into combined tubes. A most economical combination is ECH, ECL and EBF.

The aerial signal of a frequency $\omega \pm p$ reaches the circuit S1 and is fed to the tuned circuit S2-C1 where it is transformed to a higher voltage level. Via C3 the magnified signal is applied to the first grid of the heptode which is connected to the A.G.C. voltage via R1. The screen grids g2 and g4 of EH are connected to "+" via R2 (decoupled by C4). The oscillator is based on the principle shown in Fig. 228. The grid of the triode passes the frequency $\omega - m$ to g3 of the heptode whose anode will then transfer the frequency $m \pm p$ to the first band-pass filter S5-S6. In the combination S5-S6-S7-S8 this signal is selectively transformed to be amplified in EF. The I.F. signal on S8 is rectified by diode D1 and charges C8 and C9. Filtered once again by R14-C8 the frequency p is applied to potentiometer R15 from where, passing output control R15, it is fed to the A.F. amplifier. The pick-up of a record player can be connected in parallel with R15.

The general idea is now that the full "grid space" of EL is used when R15 is in maximum position. This must be achieved at a certain value of the H.F. signal (threshold), and as soon as the signal strength exceeds this threshold, the bias on the grids of EH and EF must be made more negative to reduce the gain in EH and EF to avoid the EC and EL from being overloaded. By means of R8-R9, and of R10-R11 the grid bias of these tubes is adjusted to maximum gain.

To avoid distortion, the A.G.C. circuit is kept fully separated from the circuit of detection diode D1. Via C13 the I.F. signal is branched from S7, and rectified by D2, so that C11 is negatively charged (the positive half cycle of the alternating voltage is passed by the diode). Only when the voltage on D2 is higher than the voltage over R8, the grid bias will become more negative and thus, the A.G.C. is put into operation.

Via C14 the alternating voltage of sound frequency picked up from R15 is fed to the grid of triode EC, whose grid receives the most favourable direct voltage via R10-R11. This triode operates as a resistance amplifier as shown in Fig. 226a, and the alternating voltage across R12 is transferred to the first grid of output pentode EL which, thanks to the common cathode resistance R8-R9, is adjusted to the correct working point. As a result the anode current of EL is varied, and the alternating voltage component (the sound frequency) is fed to the loudspeaker. R16-C17 is the tone control.

Although we cannot discuss all diagrams in detail, this complete diagram of a super heterodyne receiver gives a fair impression of how "electronics" can be manipulated. Fig. 229 also shows the so-called "block diagram" which from now on we shall use for more complex circuitry.

### III.B.2. TRANSISTOR RECEIVERS FOR A.M.

Whereas they were first used in battery receivers only, transistors are now also being used in mains fed receivers. Since the characteristics of the various techniques (germanium alloy and diffusion, silicon planar and M.O.S.) differ widely, there is now a large variety of circuits.

By means of Fig. 230 we shall first discuss the H.F.-I.F. part of a super heterodyne receiver with planar NPN transistors (positive voltage on the collector). On first sight there seems to be resemblance with the diagram shown in Fig. 229 (if the cathode, grid, and anode are replaced by emitter, base and collector, respectively), but still there are great differences.

Usually there is current amplification: the base current is not negligibly small as is the grid current of a tube. The input impedance of a transistor is fairly low and causes considerable damping on the preceding band-pass filter. Therefore the right hand section of the band-pass filters in Fig. 230 have been made aperiodic.

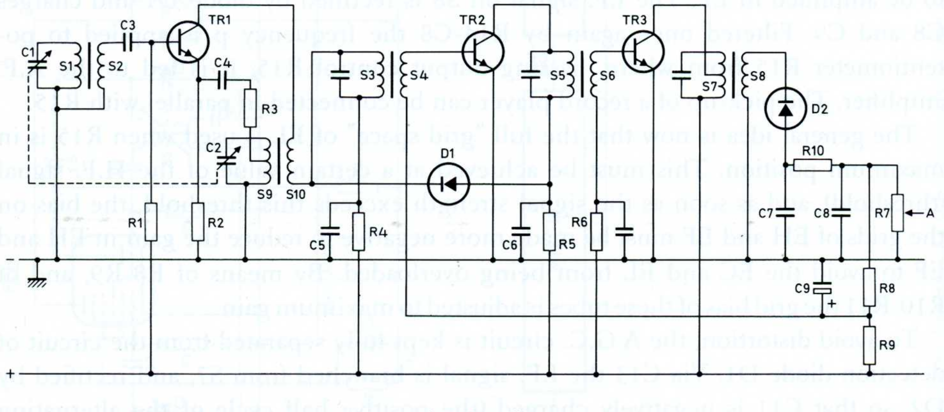
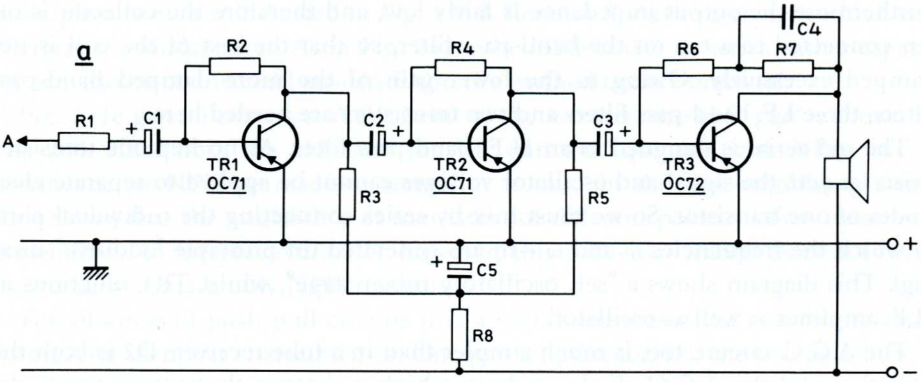
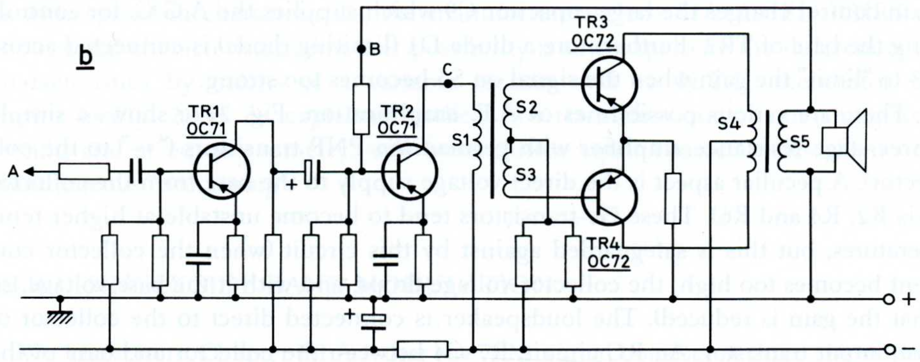
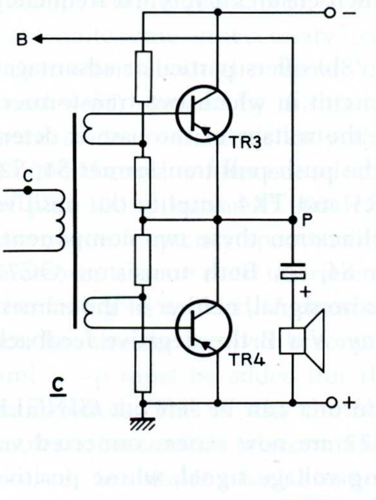


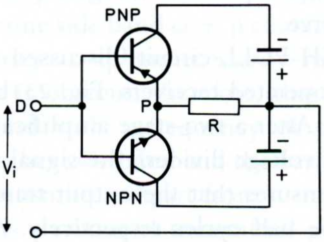
Fig. 230. H.F.-I.F. part of a transistor receiver with NPN silicon transistors.

Fig. 231. Various circuits for low-frequency transistor amplifiers: a. cascade (class A), b. push-pull circuit (class B), c. single-ended push-pull, d. principle of the complementary circuit, e. complete A.F. section of a radio receiver (to be linked up with Fig. 230), in which the principle of "d" is used.

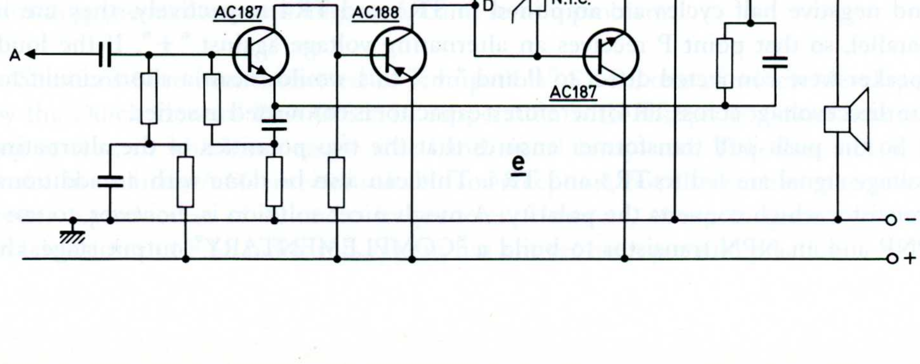

a
A
R1
C1
R2
TR1
OC71
R3
C2
R4
TR2
OC71
C5
R8
R5
C3
R6
C4
R7
TR3
OC72
+
−


b
A
TR1
OC71
B
C
TR2
OC71
S1
S2
S3
TR3
OC72
TR4
OC72
S4
S5
+
−


B
C
TR3
P
TR4
−
+
c


PNP
D
P
R
$V_i$
NPN
−
+

d


A
AC187
AC188
D
N.T.C.
AC188
P
AC187
e
+
−

Furthermore the output impedance is fairly low, and therefore the collector is often connected to a tap on the band-pass filter, so that the rest of the coil is not damped excessively. Owing to the lower gain of the more damped band-pass filters, three I.F. band-pass filters and two transistors are needed here.

The rod aerial is designed as an H.F. band-pass filter. As no heptode transistor exists (as yet), the signal and oscillator voltages cannot be applied to separate electrodes of one transistor. So we must mix by series connecting the individual parts in which the frequencies $\omega$ and $\omega - m$ are generated (in principle "additive" mixing). This diagram shows a "self oscillatory mixer stage", whilst TR1 functions as H.F. amplifier as well as oscillator.

The A.G.C. circuit, too, is much simpler than in a tube receiver: D2 is both the detection and the A.G.C. diode, and via a high resistance the voltage across the gain control charges the large capacitor C9 which supplies the A.G.C. for controlling the base of TR2. Furthermore a diode D1 (limiting diode) is connected across S3 to "limit" the gain when the signal on S5 becomes too strong.

There are various possibilities of A.F. amplification. Fig. 231a shows a simple three-stage resistance amplifier with germanium PNP transistors ("−" to the collector). A peculiar aspect is the direct voltage supply to the base from the collector (via R2, R4 and R6). These Ge-transistors tend to become unstable at higher temperatures, but this is safeguarded against by this circuit (when the collector current becomes too high, the collector voltage drops, and with it the base voltage, so that the gain is reduced). The loudspeaker is connected direct to the collector of the output transistor. An RC circuit, R7-C4 between the collector and base of the output transistor, provides a negative feedback which ensures a very flat frequency response curve.

The PUSH-PULL circuit (discussed with Fig. 78b) offers particular advantages for battery operated receivers. Fig. 231b gives a circuit in which two transformers are needed. After a two-stage amplification (here the voltage of the base is determined by a voltage divider), the signal is fed to the push-pull transformer S1, S2, S3, which ensures that the output transistors TR3 and TR4 amplify the positive and negative half cycles respectively. After amplification these two components are added again in the loudspeaker transformer S4, S5. Both transistors OC72 work in the vicinity of the cut-off point. If there is no signal, neither of the transistors draws current, which means battery economy. Via B the negative feedback goes to TR2.

According to Fig. 231c the loudspeaker transformer can be left out (SINGLE ENDED PUSH-PULL). The two transistors OC72 are now series connected via their collector-emitter circuits. For the alternating voltage signal, whose positive and negative half cycles are amplified in TR3 and TR4 respectively, they are in parallel, so that point P receives an alternating voltage against "+". If the loudspeaker were connected direct to P and "+", this would mean a short circuit for the direct voltage across TR4; therefore a capacitor is connected in series.

So the push-pull transformer ensures that the two polarities of the alternating voltage signal are fed to TR3 and TR4. This can also be done with an additional transistor which converts the polarity. A much nicer solution is, however, to use a PNP and an NPN transistor to build a "COMPLEMENTARY" output stage, the principle of which is indicated in Fig. 231d. Resistor R carries a current which can energize the loudspeaker. As in Fig. 231c, the loudspeaker in series with a capacitor can also be connected between P and the earthed battery terminal.

Fig. 231e gives a complete A.F. circuit of the receiver; the H.F. part is shown in Fig. 230. Since both the H.F. and the I.F. part are equipped with NPN silicon transistors, an NPN germanium transistor is used in the first amplifying stage. This shows that the availability of transistors of different polarity offers numerous possibilities. In integrated circuits, however, this particular feature does not exist; they usually contain only NPN transistors.

For all kinds of push-pull circuits it is a requirement that the two output transistors have identical characteristics. Unequal amplification of the two polarities of the signal would lead to distortion. Since manufactured transistors show a rather wide spread in characteristics, the manufacturer selects MATCHED PAIRS of about equal characteristics. Another possibility is to compensate the differences in characteristics by means of a variable resistor and an N.T.C. thermistor, as shown in Fig. 231e. As explained earlier, two transistors out of one integrated circuit, and at equal geometry, will have almost equal characteristics thanks to simultaneous "processing".

## III.B.3. SINGLE SIDE-BAND MODULATION

We know from Chapter III.A.5 that the A.M. signal consists of three alternating currents with frequencies ω, ω + p, and ω − p respectively. So the signal seems to carry quite some unneccessary "ballast". For one side band ω + p contains sufficient information about the signal, the frequency ω−p unnecessarily increases the band width claimed by the transmitter. It is obvious, therefore, that endeavours were made to transmit only the side band with a frequency of plus 100 to 15 000 Hz, and to filter out selectively the other side band and the basic frequency of plus 100 to minus 15 000 Hz in the transmitter. Apart from a more efficient use of the frequency spectrum the result is a saving of transmitter power, while moreover a source of interference in telecommunication, the so-called "selective fading" is avoided. A drawback is, however, that now not only the transmitter, but also the receiver becomes more complex. In the latter the frequencies ω and ω−p must be added (for that purpose a weak synchronization signal ω is transmitted). This is why the S(ingle) S(ide) B(and) system has been limited to professional radio telecommunication and multi channel telephony, and proved unsuitable for broadcasting, where too many transmitters overcrowding the medium and long wave band form a problem.

Solutions have been indicated of "compatible" S.S.B., one of which (developed by the Dutch General Post Office in cooperation with Philips) is a most interesting system, in which the single side band signal is squared. This does not only make an ordinary A.M. receiver suitable for reception, but also enables it to give a better reproduction of all audio frequencies. However, owing to greater interest in television and F.M. radio, this system has not yet been put into practice.

## III.B.4. FREQUENCY MODULATION

With amplitude modulation at constant frequency, a part M of the amplitude of the H.F. signal, is varied according to the A.F. signal. The F.M. signal can be defined as follows: at a constant amplitude a part K of the frequency of the H.F. signal is varied in accordance with the A.F. signal.

So the frequency is:

$$\omega = \omega_o + K p \cos pt, \tag{1}$$

where K is the modulation index. For the F.M. signal we use the formula:

$$e = E_0 \sin(\omega_0 t + K \sin pt) \tag{2}$$

The amplitude remains constant. Kp is the "frequency sweep". The rate of frequency variation depends on the frequency of the A.F. signal.

This F.M. signal can be produced in a rather simple way, if a capacitor microphone is included in the oscillator circuit of the transmitter. A better method is to connect the anode circuit of a "reactance tube" in the oscillator circuit. The reactance of this tube is then made to vary in accordance with the audio signal fed to its grid.

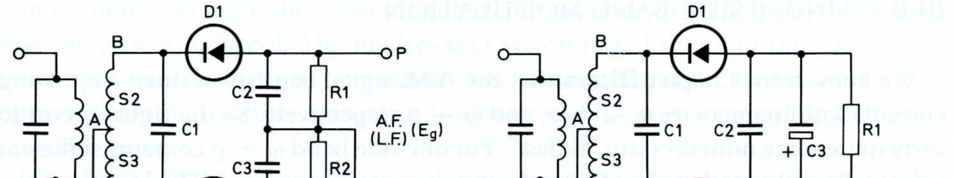


Fig. 232. F.M. detection: a. vector diagram, b. Foster-Seeley discriminator, c. ratio-detector-limiter, d. asymmetrical ratio-detector.

To reproduce the audio signal via the loudspeaker, the F.M. signal according to formula (2) must be converted again into a voltage of the function $E_g$cospt. The circuit used for that purpose is the DISCRIMINATOR.

In accordance with Fig. 232b,c,d, this circuit contains a coil S1 which is magnetically coupled with the input coil of the band-pass filter, and connected to the centre of the coil S2-S3 of the output band-pass filter. If in the vector diagram of Fig. 232a the voltage across S1 is represented as OA, the voltage across S2 and S3 will be shifted 90° in phase with respect to OA when the frequency of the signal equals the resonant frequency of S2, S3, C1. This situation is illustrated by the vectors AB and AC. In that case OB = OC. At a frequency higher or lower than the resonant frequency the phase shift will be accordingly greater or smaller than 90° as is represented by AB′ and AC′. For small frequency deviations the voltage $OB' - OC' = E_g$ is proportional to this deviation (Kpcospt). According to the originally applied Foster-Seeley circuit (Fig. 232b) a voltage will occur across the resistors R1 and R2 equal to OB′ and OC′, to be smoothed by C2-C3. So the difference signal $E_g$, which is proportional to Kpcospt, and hence to the A.F. signal, can be picked up between P and Q.

Fig. 232c shows the RATIO DETECTOR. Here the two diodes are connected in series. The large capacitor C4 will be charged to the r.m.s. (root mean square) value (corresponding to the fundamental frequency ω), and between P and Q the voltage will vary in accordance with the audio signal. Fig. 232d shows the "asymmetrical ratio detector" as generally used in cheaper types of receivers.

In reality the situation is much more complicated than explained here. For during transmission, and in the H.F. and I.F. filters, there occurs a certain amount of amplitude modulation which must be removed by the ratio detector (LIMITER). Furthermore, to reduce the risk of interference, a certain PRE-EMPHASIS is used to increase the frequency sweep for the high tones. The RC filter in the ratio detector corrects this pre-emphasis so that the time constant of this filter must be equal to that in the transmitter (for Europe 50 microseconds is regulation).

In principle the F.M. signal can be amplified and mixed with circuits corresponding to those used for A.M. Since the frequency sweep goes to 75 kHz (so F.M. requires a greater bandwidth than A.M.!) and the ordinated carrier frequency is 87.5 — 108 MHz, the intermediate frequency is selected to be 10.7 MHz. These high frequencies require certain measures to be taken.

For the H.F. amplifier common practice is, using a triode in "grid-base circuit" because then the input and output circuits are better screened from each other. In Fig. 233a a heptode is used as H.F. amplifier since in a combined A.M.-F.M. receiver the heptode of the A.M. mixing tube is used for that purpose. Now the grid-base circuit cannot be used, and extra precautions must be taken to prevent the oscillator voltage reaching the aerial thus causing annoying radiation interference in other receivers. The oscillator circuit L5-C6 which is connected to the grid of the self-oscillating mixer EC must therefore be connected to the output of the H.F. circuit (band-pass filter L4-C5) via a "Wien bridge" (Fig. 233b). A proper choice of the capacitors C7, C8 and C9, as well as the total capacitance to earth via C10, prevents the oscillator voltage from causing a voltage between A and B.

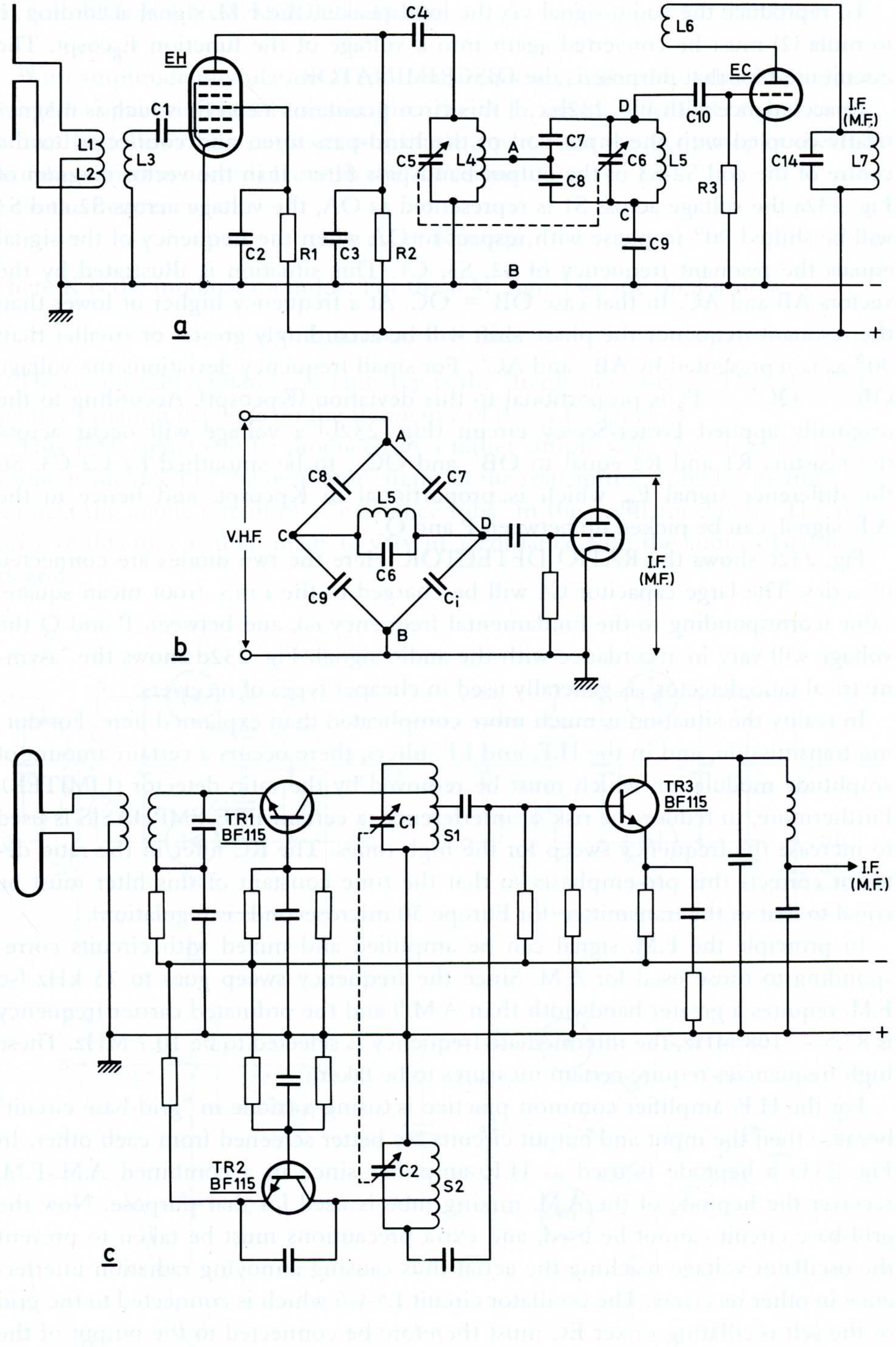


Fig. 233. F.M.-H.F. amplifying and mixing: a. tuner with tubes, b. circuit diagram of Wien bridge (Ci is the total capacitance of D to earth via C10), c. tuner with transistors (TR1 = H.F. amplifier, TR2 = oscillator, both in common base circuit, TR3 = I.F. amplifier in common emitter circuit).

The simplified circuit in Fig. 233a, which can be used to build an A.M.-F.M. receiver with five tubes (ECH, EF, EABC, EL, EZ) has contributed largely to the popularization of F.M. reception.

Fig. 233c gives the circuit of the F.M. FRONT END with transistors. Here TR1 is the H.F. amplifier, and TR2 the oscillator, both in common base circuit. By additive mixing TR3 (with common emitter) amplifies the intermediate frequency.

An interesting possibility presents itself when the H.F. circuit S1-C1 and the oscillator circuit S2-C2 are not tuned with the capacitors C1 and C2, but with the adjustable cores in the coils S1 and S2. Then C2 can be partly replaced by a varactor diode (Chapter II.J.3), whose capacitance changes under influence of a varying direct voltage. If this voltage is obtained from the discriminator, the oscillator frequency will be modified to such an extent that in the vicinity of the resonant frequency of the discriminator, the intermediate frequency becomes equal to this resonant frequency. Thanks to such an automatic frequency control (A.F.C.) it becomes quite easy to tune the receiver accurately to an F.M. station.

It should be noted that in link transmitters operating with frequencies of about 3000 MHz, DOUBLE FREQUENCY MODULATION is used. The audio signal first modulates a sub-carrier, and this F.M. signal in turn is used to modulate the H.F. carrier frequency.

## III.B.5. STEREOPHONY

Frequency modulation with its less interference, wider available frequency band, and greater amplitude variation of the A.F. signal, led to a substantial improvement of faithful reproduction of music via the radio (HiFi). Around 1964, the earlier applied stereophony used in gramophone and tape recorder was added. Hearing with both ears enables us to tell from which direction the sound reaches us. Therefore sound, heard with both ears when it is produced by a loudspeaker on either side of the listener, gives him the impression of actually being in the concert hall.

Stereophony in broadcast radio could easily be realized by transmitting the audio signals heard by the left ear (L) and the right ear (R) respectively (that is to say via two microphones placed in a "dummy" head) via two separate transmitters, to be received via two separate radio sets. However, the system must be made "compatible". A "mono" F.M. receiver must also be able to give a satisfactory reproduction of the stereo signal, and it would never do to have a signal that can be heard with only one ear. Hence the sumtotal (M) of the signals of both ears must be given. Similarly a stereo receiver must also be suitable to receive non-stereo programmes.

The now generally used MULTIPLEX system, the first to be standardized by the F.C.C., operates as follows: modulated on the H.F. carrier, the transmitter transmits the sum signal $M = L + R$ of a frequency 0-15 kHz, which is reproduced by a mono radio receiver. The stereophonic information is given by the difference signal $S = L - R$. This signal is amplitude modulated on a sub-carrier of 38 kHz. The signal thus modulated comprises a frequency of $38 \pm 15 = 23$ to 53 kHz

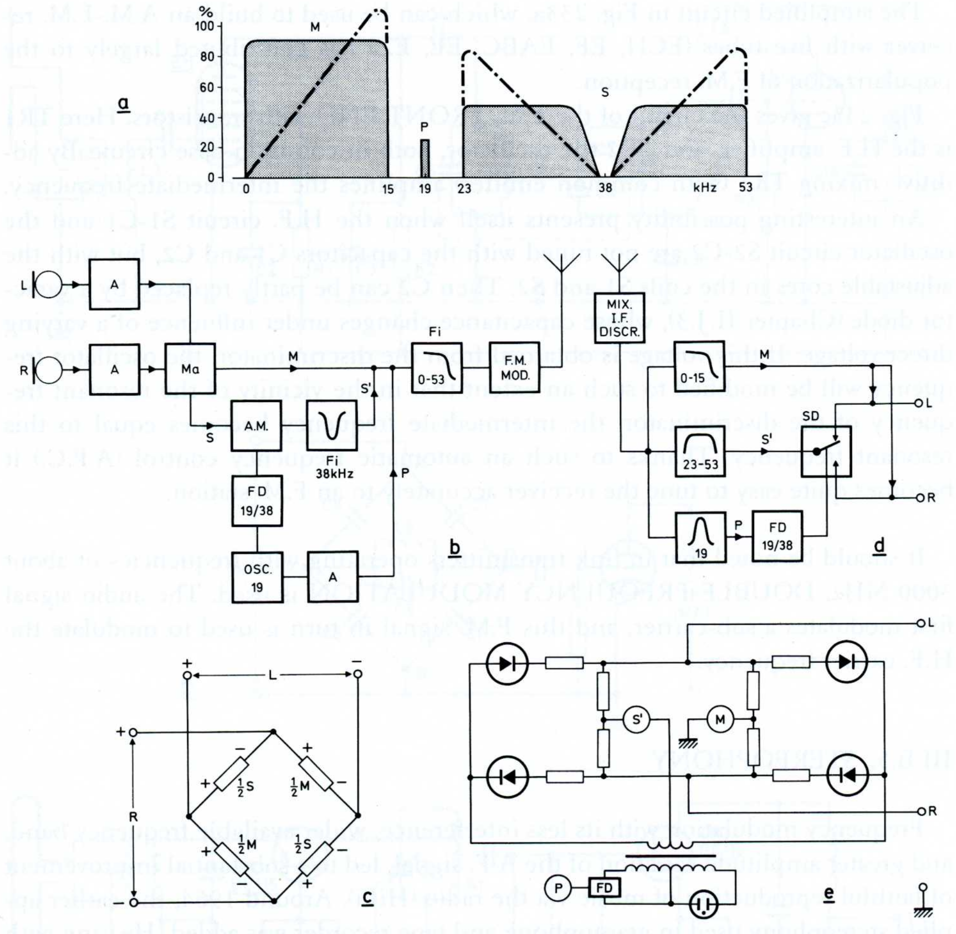


Fig. 234. F.M. multiplex stereophony based on the F.C.C. system: a. frequency spectrum, b. transmitter, c. matrix, d. receiver, e. switching detector, L = left-, R = right-, M = sum-signal, R = pilot tone, A = amplifier, Ma = matrix, F.D. = frequency doubler, SD = switch detector.

and consequently, does not interfere with the M-signal. The fundamental frequency of 38 kHz is removed from this signal (by means of a filter with a steep response curve); and the signal S′ is left. To feed the sub-carrier to the receiver, an additional pilot frequency P of 19 kHz is transmitted which does not interfere with the bands M and S. Fig. 234a shows the resulting frequency spectrum of the modulation. The dash-dot lines indicate the influence of the pre-emphasis.

Fig. 234b gives the block diagram of the transmitter. The L(eft) and R(ight) signals are amplified in A, to be added and subtracted in a matrix (Ma) which is shown as a detail in Fig. 234c (a splendid example of “analog” computation!). In the A(mplitude) M(odulator) the S-signal is modulated on the sub-carrier of 38 kHz which is the result of a doubling (in FD) of the frequency of 19 kHz which is generated in a crystal oscillator. The fundamental frequency of 38 kHz is filtered out of this modulated signal, and what is left is called the S′-signal. The pilot frequency of 19 kHz (P) is picked up from the oscillator, and M, S′, and P together are fed to a filter (Fi) which blocks high interference frequencies. In the frequency modulator M, S′ and P are modulated on the H.F. signal.

Fig. 234d shows the block diagram of the receiver (STEREO DECODER). The received signal is split up in three filters into M, S′ and P. The part M goes direct to the outputs L and R, but S′ is led through an electronic switch (a switch detector) which applies the positive and negative parts of S to the outputs L and R respectively. This switch is governed by the output of the "F(requency) D(oubler)", an oscillator which in turn is synchronized by the "P" signal. Thus the sumtotal $M + S = 2L$ arrives at terminal L, whilst the difference $M - S = 2R$ arrives at terminal R. Fig. 234e gives the diagram of the electronic switch to which the signals M, S′, and FD are supplied. During a mono transmission P, and so FD, receives no voltage and S′ is blocked. Now the switch detector functions as a ratio detector. The stereophonic transmission is characterized by the presence of the P-signal which causes a pilot lamp to light up.

But few listeners will realize all that is happening when the P-lamp lights up and a stereophonic programme can be received.

# SECTION C. TELEVISION RECEIVERS

## III.C.1. THE BLACK AND WHITE SIGNAL

In Chapter I.11 it was explained with Fig. 16 how in a camera tube the cathode ray describes a line raster, the instantaneous value of the current through this tube being a measure for the luminous intensity at each point on the screen hit by the cathode ray. These current variations form the VIDEO SIGNAL which in the picture tube of the receiver will reproduce the image received by the camera tube, provided the cathode rays in both tubes describe the same raster. The receiver itself will be able to generate the "saw tooth" currents which, flowing through the deflection coils, will cause the cathode ray to complete the raster at the same speed as the one in the camera tube. However, the signal must carry information as to the instant in time at which each line should begin from the left and when the raster should be started from the top; this information is the SYNCHRONIZING PULSE for the horizontal deflection (LINE) and vertical deflection (FRAME). The systems used in various countries differ as regards the method of modulation, the number of lines, and the form of the synchronizing pulses (sync pulse). We shall discuss only the C.C.I.R. system used in Europe, west of the Iron Curtain; at first excepting Britain and France. Here the raster consists of 625 horizontal lines which are described 25 times per second. To limit "flickering" of the picture, the method of "interlacing" is adopted: during the first 1/50 second the odd numbered lines are described, to be followed by the even numbered lines during the next 1/50 second and so on.

So each horizontal line is described in $1/25 \times 625 = 1/15625$ second at a constant speed from left to right, and during the last 9/100 part of the time the cathode ray must return from right to left to start the second next line (FLYBACK). During flyback the line must be invisible, and so the video signal must be "blanked". Hence the blanking signal could be used for line synchronization, were it not that the receiver must be able to distinghuish between the "black" of the image and the "black" during flyback. Therefore the black of the picture is started at about 25% of the maximum signal current (which produces "white") whilst the first 25% is used for the sync pulses (blacker than black).

In Fig. 235, left of $t_1$, the video signal and line sync pulse are shown for three horizontal lines. According to the C.C.I.R. system the carrier is negatively modulated: at the "blacker than black" sync pulse the amplitude is 100%; at white about 10% amplitude is left, for without carrier at all the receiver would not be able to function. So the amplitude of the modulated carrier varies as the grey area in Fig. 235 starting from the top.

The frame flyback must take place 50 times per second, once from centre bottom to the top left, and once from the bottom right to the centre top. For each frame flyback we sacrifice about 20 lines during which the video signal is below black level. During that time a train of sync pulses is given as indicated between $t_1$ and $t_4$ in fig. 235. During the same period the line sync pulses must continue (as shown by the arrows), for otherwise the "flywheel" synchronization of the horizon-

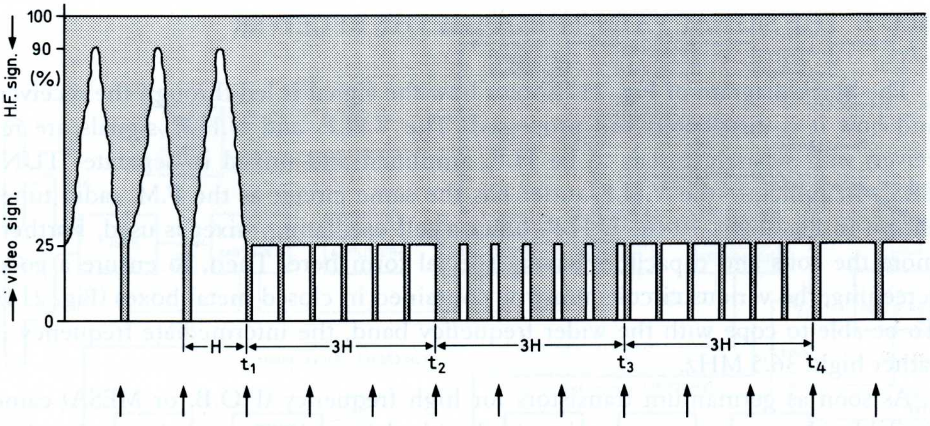


Fig. 235. The video signal reads from bottom to top, the modulated video signal from top to bottom, left of t1 the video signal with sync pulses, t1 to t2 preparation of the vertical synchronization, t2 to t3 vertical sync pulses (the arrows indicate the instances of the horizontal sync pulses).

tal deflection will "fall out of step". During the period $t_1$ to $t_4$ there are, however, sync pulses of double the frequency of the line sync pulses to serve as an "introduction". The narrow sync pulses of double the frequency occurring between $t_2$ and $t_3$ supply the signal for the vertical synchronization in the "blocking oscillator" described below.

Of the H.F. signal that is A.M. modulated by the video and sync pulses, one side band is clipped completely, and the other above 5 MHz (the remaining information is then 2 × 5 000 000/15625 = about 640 dots per line), as shown in Fig. 236. The audio signal is F.M. modulated (with a frequency sweep of about 100 kHz) on a carrier at a distance of 5.5 MHz from the video carrier (Fig. 236), so that the audio and video signal do not interfere with each other. Hence a band width of 7 MHz (one channel) contains all information. Band I, from 41 – 68 MHz can hardly contain 4 channels, and band III from 162 – 230 MHz can contain no more than 7 channels. Therefore U.H.F. bands IV and V from about 470 – 960 MHz, which can in theory contain 70 channels, were introduced later.

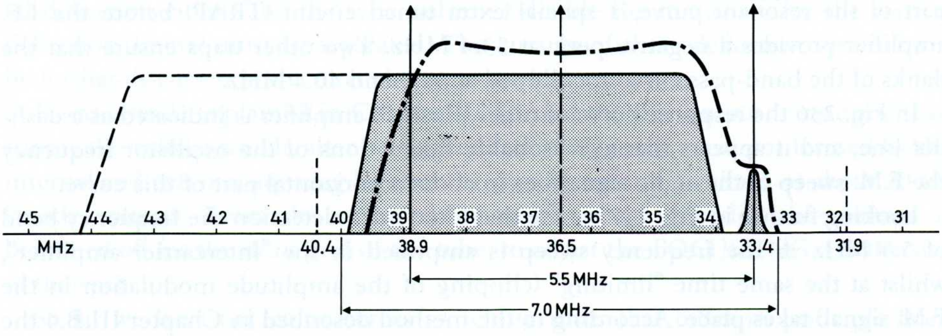


Fig. 236. Frequency spectrum of a television transmitter (full line). The dash-dot line gives the gain from the H.F. circuit to the video detector for a well adjusted tuner. The horizontal scale gives the intermediate frequency (I.F.) reading from right to left because the oscillator frequency is chosen higher than the transmitter frequency.

## III.C.2. THE SIGNAL PATH THROUGH THE RECEIVER

The block diagram of Fig. 237 shows how the signal is led through the receiver, and how it is distributed and processed. The V.H.F. and U.H.F. signals are received with separate aerials to be H.F. amplified and mixed in separate "TUNERS". In principle the V.H.F. tuner has the same circuit as the F.M. radio tuner shown in Fig. 233b. In the U.H.F. tuner a self oscillating mixer is used. Furthermore the coils and capacitors have a special form here. Then, to ensure a good screening, the various circuits must be contained in closed metal boxes (Fig. 214). To be able to cope with the wider frequency band, the intermediate frequency is rather high: 36.5 MHz.

As soon as germanium transistors for high frequency (P.O.B. or MESA) came available, the tuners were transistorized with the result of much less noise than with electron tubes.

The video and audio signal, now modulated on frequencies of 38.9 and 33.4 MHz respectively, are simultaneously amplified in the I(ntermediate) F(requency) amplifier. Thanks to the frame-grid tubes and ferroxcube band-pass coils a gain of 60 dB could be obtained with two stages. The video amplifier can supply the A(utomatic) G(ain) C(ontrol) signal which reacts on the H.F. and the first I.F. amplifier. The output signal of the I.F. amplifier will thus have a certain value which is independent of the strength of the signal received.

It would seem obvious to filter out the audio signal from the combined I.F. video and audio signal. But this entails the difficulty that a minor deviation in the oscillator frequency would cause a deviation of the signal frequency of the F.M. carrier which is great as compared with the frequency sweep. In that case the separated audio signal would lie so far beside the centre of the pass band of the filter, that the result would be serious sound distortion. The solution to this problem is a remarkable one. The A.M. modulated video signal must be "detected" by a simple germanium diode (still according to the principle of Fig. 6). Owing to the typical characteristic of this detector "mixing" occurs with the result of another audio signal that is F.M. modulated on the difference between the carriers of the video and the audio signals: that is 5.5 MHz (the INTERCARRIER). To avoid the audio carrier being situated on the steep part of the resonant curve, a special extra tuned circuit (TRAP) before the I.F. amplifier provides a negative peak at 33.4 MHz. Two other traps ensure that the flanks of the band-pass curve are clipped at 31.9 and 40.4 MHz.

In Fig. 236 the response curve of tuner plus I.F. amplifier is indicated as a dash-dot line, and it appears that at reasonable fluctuations of the oscillator frequency the F.M. sweep of the audio signal lies in about a horizontal part of this curve.

Looking further in Fig. 237 we see that after video detection the frequency band of 5.5 MHz ± the frequency sweep is amplified in the "intercarrier amplifier", whilst at the same time "limiting" (clipping of the amplitude modulation in the F.M. signal) takes place. According to the method described in Chapter III.B.4 the F.M. signal is converted in the ratio detector into the A.F. audio signal which is amplified and fed to the loudspeaker. From the loudspeaker a negative feedback takes place to improve the sound quality.

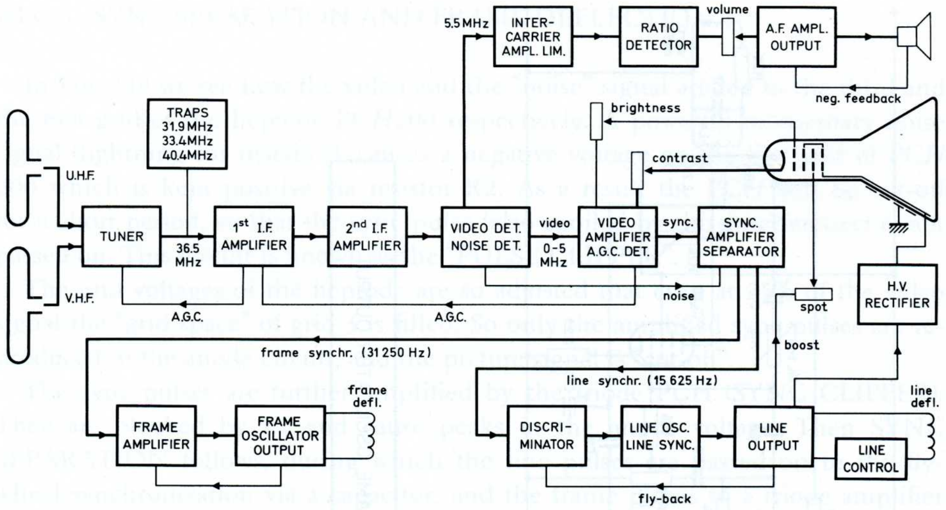


Fig. 237. Block diagram of a complete black and white television receiver.

From the video detector the video signal is now fed to the video amplifier. Via a "CONTRAST CONTROL" the output of this amplifier supplies the cathode and thus the grid voltage for the picture tube. The adjustable zero point setting of the video signal on the grid presents the brightness control.

Behind the video detector an extra signal is branched off by tapping the "no man's land" between 5 and 5.5 MHz in Fig. 235 by means of a special filter. NOISE interference of no specific frequency also occurs in this no man's land and puts the amplifier of the sync pulses momentarily out of operation so that the synchronization cannot possibly be disturbed.

Finally the amplified video signal is processed in a synchronization amplifier. The "grid space" of this amplifier embraces only the 0-25% part of the video signal as shown in Fig. 235 and will thus amplify only the sync pulses (SYNC SEPARATOR). Now the sync pulses of a frequency 25 × 625, which are fed to the horizontal deflection circuit and the sync pulses of double the frequency for the vertical deflection (frame) must be separated. The circuits of these deflection systems will be further discussed in the next chapters.

It was already explained in Chapter II.M.5 that the line-output transformer can supply the voltage which, after being rectified, charges the capacitor formed by the internal and external aquadag layer of the picture tube to a voltage of about 20 kV (anode voltage). From the so-called "booster" voltage an adjustable voltage is branched off for the third grid of the tube, to adjust the FOCUSSING, i.e. the size of the spot.

So far we have not come across switching units that differ in principle from those previously discussed: but that will happen now that we must discuss the frame- and line-output circuit, where the currents or voltages no longer vary sinusoidally, but rectangularly or triangularly with time.

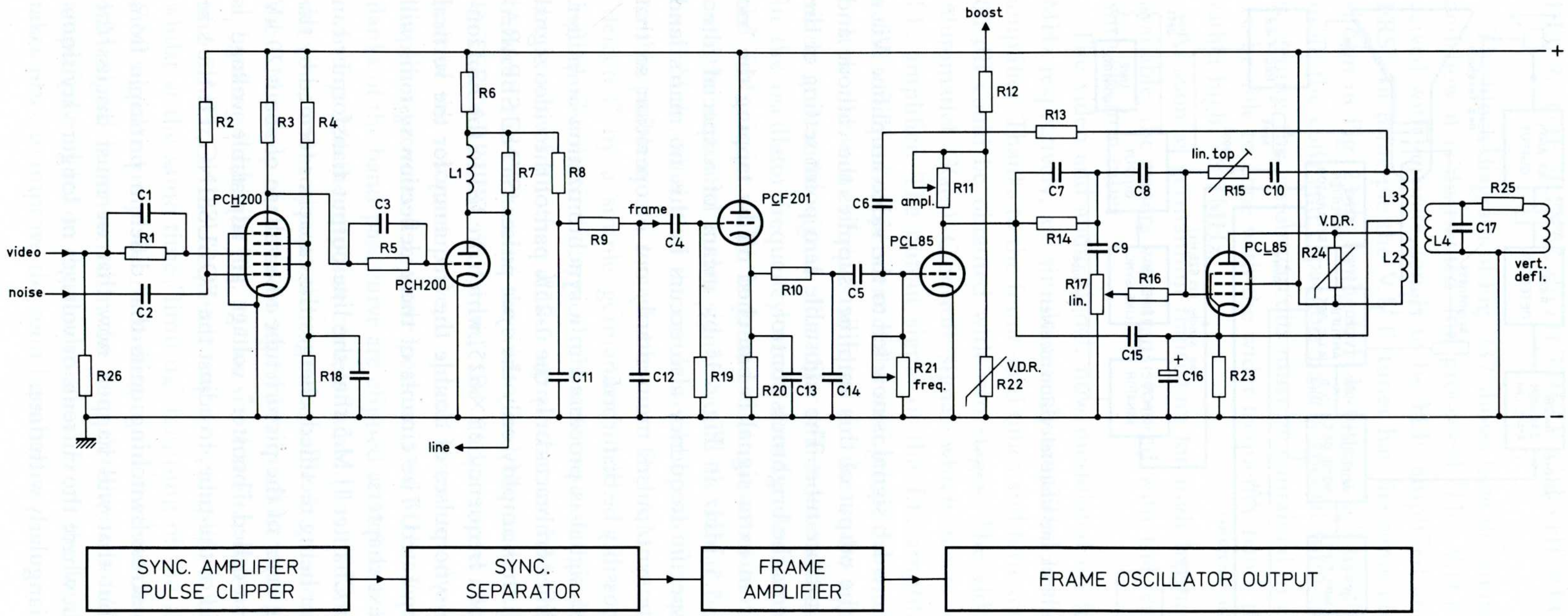


Fig. 238. Synchronization amplifier and vertical deflection.

## III.C.3. SYNC SEPARATION AND FRAME DEFLECTION

In Fig. 238 we see how the video and the "noise" signal are fed to the third and the first grid of the heptode PC*H*200 respectively. A powerful momentary noise signal (lightning for instance), causes a negative voltage on the first grid of PC*H* 200 which is kept positive via resistor R2. As a result the PC*H* will be cut-off for a short period, so that the sync pulse (also spoiled by this interference) is not passed on. This circuit is known as the "PULSE CLIPPER".

The grid voltages of the heptode are so adjusted that even at 25% of the video signal the "grid space" of grid 3 is filled. So only the amplified sync pulses are reproduced in the anode circuit, and the picture signal is "cut-off".

The sync pulses are further amplified by the triode PCH (SYNC CLIPPER). They are blocked by L1 and cause peaks in the anode voltage. Then SYNC SEPARATION follows, during which the line pulses are passed on to the flywheel synchronization via a capacitor, and the frame pulses to a triode amplifier (P*C*F) in the "cathode follower" circuit via an RC network (which blocks the line pulses). By means of this circuit and the RC networks the pulses can be applied to the grid of the P*C*L85 as a rapidly increasing voltage after $t_2$ in Fig. 235.

The vertical deflection coil must be driven with a triangular "saw tooth" current. The current must increase gradually to a certain value to drop steeply to zero again, and so on. A saw tooth voltage could be generated by charging a capacitor via a resistor and short circuiting the capacitor at a certain voltage. The short circuit could be provided by connecting the anode and cathode of a triode in parallel with the capacitor. Since this triode is cut-off until the sync pulse makes the grid positive, the saw tooth generator could be "driven" by the sync pulse. This method has two drawbacks: when the sync pulse disappears, the light spot would still remain, and the screen would "burn in". Furthermore an interfering signal could affect the saw sooth at an unsuitable instant. Therefore preference is given to the "self oscillating" saw tooth generator, which itself generates a saw tooth current at about the right frequency, and the sync pulse need only make minor corrections in that it determines the instant the flyback starts. A suitable method of making such a saw tooth generator is using an amplifying tube with strong negative feedback, which forms part of the circuit P*C*L and PC*L* in Fig. 238. We shall not discuss this circuit in details; let it suffice to mention that this is a self oscillating multivibrator, as will be described in Chapter III.D.2, consisting of two tubes whose anodes and grids are connected crosswise (anode of PC*L* to grid of P*C*L via transformer winding L3). At a correct setting of the tubes and RC filters, L4 (the output of the frame transformer) will produce a saw tooth current of about twice the picture or frame frequency. R11, R17, and R15 serve to adjust the amplitude, linearity, and linearity at the top, respectively. By means of N.T.C. and V.D.R. resistors the saw tooth current is made independent of temperature and supply voltage fluctuations. It is typical of this circuit that a pulse on the grid of P*C*L can only influence the saw tooth generator at about the moment the current is maximum (this can be compared with a petrol engine whose speed could be controlled with the ignition timing). At all other times interfering pulses have no effect on the deflection current. Thus we have "AUTOMATIC SYNCHRONIZATION", at

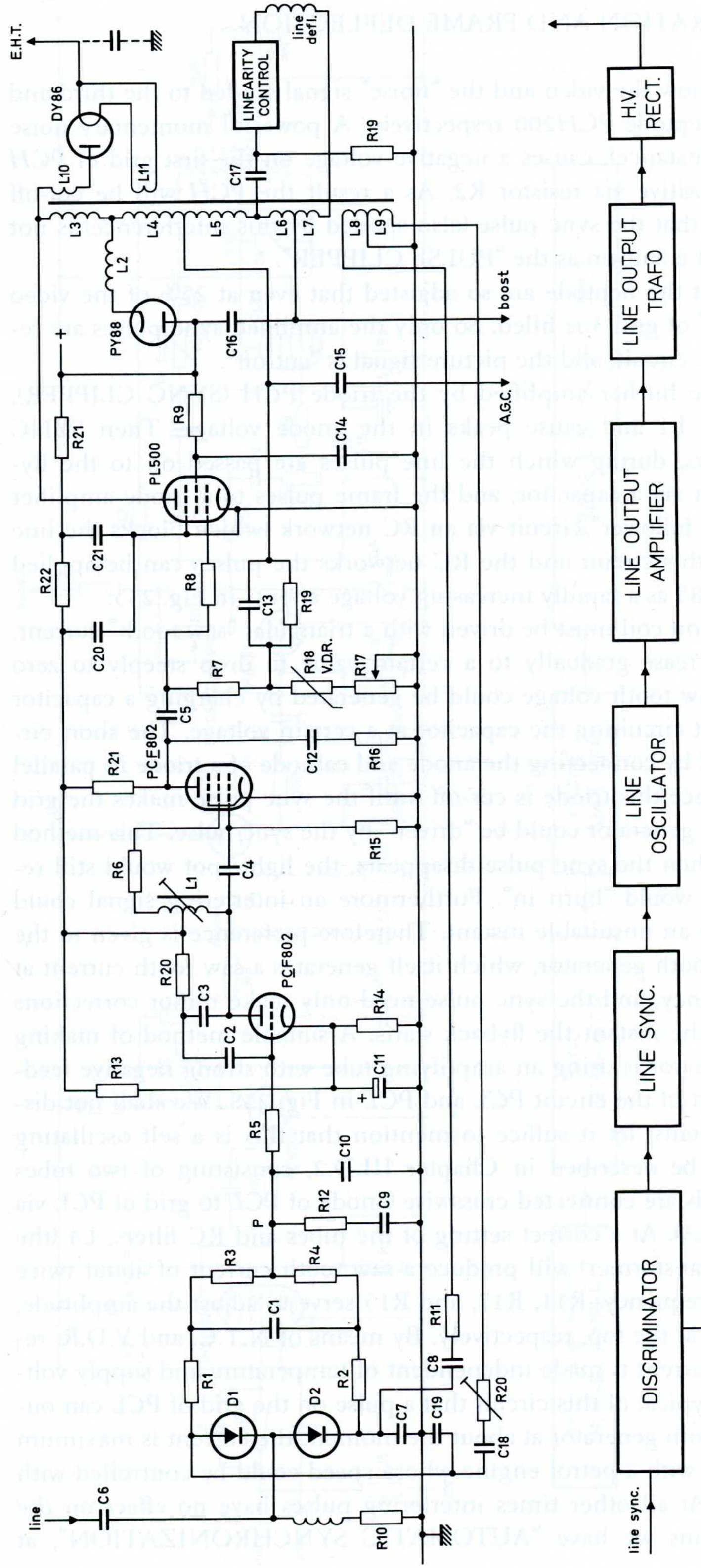


Fig. 239. Stabilized line-output circuit with reactance control and automatic synchronization.

which the correct functioning of the deflection system is to a certain extent independent of interference and of the magnitude of the sync pulse.

## III.C.4. LINE DEFLECTION

Fig. 239 shows the line deflection circuit. Here it is far more difficult to use the system of a self oscillating time base, because the frequency is so much higher and a rapid flyback is impeded by the self inductance of the deflection coil.

An even better method has been found to make the self oscillating saw tooth generator even more independent of interfering effects in the sync pulse, that is to say, by using the FLYWHEEL SYNCHRONIZATION. This means that the oscillator continues to oscillate at its own frequency, and that only a train of sync pulses at the right moment can correct the frequency (all interferences are then eliminated).

The PC*F* functions as the oscillator tube, the oscillating circuit being connected to g1 and g2. Triode P*C*F is connected as a reactance tube across this circuit. At a value of 3300 pF for C3, the triode represents a capacitance of 1000 pF maximum which is dependent on the grid voltage. Thus the characteristic frequency of the oscillator can be varied by a value of 1000 Hz (with respect to the line frequency of 15625 Hz). This grid voltage is supplied by a discriminator with the diodes D1 and D2, in which the incoming pulses of the line sync signal are compared with alternating voltages which are tapped from the output transformer via L8-L9, and indicate the actually occurring oscillator frequency. A phase shift between the sync signal and the oscillator frequency will lead to a voltage variation at point P, which voltage is fed to the grid of the reactance tube and, owing to a variation in reactance parallel to the oscillator, will adjust its frequency. Thus the abovementioned condition is complied with. The oscillator frequency is corrected by regularly repeated sync pulses, and interfering signals have no effect. The oscillator frequency obtained is kept accurately equal to the line synchronization, independent of variations in circuit characteristics brought about by heat, drift, etc, as well as of the values of synchronization and line output voltages. If due to interference an odd sync pulse should be skipped, the oscillation just continues.

As a result of the corrected oscillator voltage on the grid of PC*F* the anode current of this tube displays very steep pulses which are passed on as negative pulses to the grid of line output tube PL500. The line deflection coil is connected to the anode circuit of this tube via the taps L6-L7 on the transformer.

The current fluctuations are so rapid that not the resistance, but only the self inductance of the coil plays a part. If at the completion of the negative pulse, the PL500 becomes conductive, the current in the deflection coil will increase about linearly at a constant voltage because $E = L \times di/dt$. The steep negative peak on the grid of the PL500 will suddenly cut off the current. As an immediate result the deflection coil, and hence the transformer, will generate a voltage surge in the reverse direction. If no proper measures were taken, a damped oscillation would occur which would spoil the saw tooth curve. Therefore a diode and capacitor are connected in parallel with a part of the output transformer. This diode will begin to conduct as soon as the voltage on its cathode becomes negative and the magnetic energy released by the sudden current reduction is led away to BOOSTER capacitor C16, so that the damped oscillation is choked after a half period. During the next saw tooth the capacitor discharges slightly, but the average voltage remains about 250 V. By connecting this BOOSTER voltage in series with the supply rectifier an anode voltage of about 500 V comes available for the line output tube and some other tubes that require a high supply voltage.

Once the flyback has taken place as quickly as possible and the energy from the deflection coil has been recovered (in booster and high tension) the next saw tooth can be started as soon as the negative voltage pulse on the first grid of the PL500 has been completed. The amplitude of the deflection current can be adjusted with R17. In series with the deflection coil there is also a linearity control consisting of a coil with an adjustable saturated ferroxcube core.

Rather great powers are required to obtain high-speed current variations in spite of the large self inductance of the line deflection coil. The line output tube must be capable of dissipating about 20 W if peak currents up to 0.5 A and peak voltages up to 5 kV are to be applied to the anode.

In comparison with this power, the power needed to supply the anode voltage of the picture tube (18 kV at a maximum of 1 mA = 18 W) is fairly low. This power can be supplied by the voltage which is generated by the magnetic energy released in the output transformer during flyback. For that purpose the transformer winding L3-L7 is extended by a high tension coil wound on the second leg of the transformer. This part, loosely coupled with L3-L7 will experience an oscillation of its own with a frequency depending on the capacitance and self inductance of this coil. If we succeed in keeping the capacitance of this coil small enough, its characteristic oscillation will have a frequency five times as high as the flyback frequency. In conjunction with the flyback frequency an almost square voltage form is obtained which can be excellently rectified by vacuum diode DY86 whose heater voltage is supplied by a winding L11 at the top of the coil. Nowadays a cascade circuit with silicon or selenium rectifiers as shown in Fig. 223e is sometimes applied. Then the coil for the extra high tension is not needed.

As explained at the beginning of this section, various functions in a television receiver carried out by electron tubes can be taken over by transistors. However, some of the tubes occurring in Figs. 238 and 239 can hardly be replaced by transistors. Although solutions have been found for battery operated equipment, the HYBRID receivers for mains supply are usually still equipped with tubes for these functions. Most of them have been specially developed for these applications, and contain two systems in one envelope (10-pin DECAL TECHNIQUE).

## III.C.5. THE COLOUR TELEVISION SIGNAL

When dealing with black and white television we were interested only in the brightness of the picture; with colour television the colour must also be taken into consideration. In principle the colour is determined by the frequency of the electromagnetic oscillation that produces the light beams. With a prism we can split up light into beams of different wave lengths. These form a series starting from blue with a wave length of about 380 millimicrons passing via green, yellow, orange to red with a wave length of 780 millimicrons. The colours of the various objects we observe depend on the wave lengths which are absorbed or reflected respectively (and let through by transparant bodies). The colour mixture of various layers of colour on top of each other is obtained by "subtractive" mixing.

In colour television, however, we make use of the fact that any colour of any brightness can be obtained by "additive" mixing of the three "spectral" colours red, green and blue, each with matched brightness. We may assume that the human eye, too, gets an impression of colour from observation via the cones in the retina, which are particularly sensitive to these three colours. Additive mixing occurs when dots or dashes of the three basic colours are placed closely together: they are then seen as the colour mixture. In connection with the colour sensitivity of the eye one lumen of the colour mixture "white" is obtained by adding 0.3 lm red, 0.59 lm green, and 0.11 lm blue (owing to the differences in phosphor sensitivity to the three colours, we find in practice that the ratio of the colour signals are about 0.40 R, 0.34 G and 0.26 B, but this is left out of consideration in the calculation worked out here).

It would be quite simple to lead the three spectral colours through suitable filters to three camera tubes, whose video signal is modulated on three separate carriers to be transmitted from three transmitters. In the receiver these signals could then be led to the three guns in the colour picture tube in which the brightness of the red, green and blue dots would then vary in accordance with what the camera tube observes. For a faithful colour rendition at different brightnesses it would then also be necessary that the camera tube had a linear characteristic.

Unfortunately, matters are more complicated than that, for there must be "COMPATIBILITY". The transmitted colour signal must be reproduced as a black and white picture on black and white receivers. The colour receiver must also be able to reproduce the picture from a black and white transmitter. Therefore a black and white (LUMINANCE) signal must first be produced in the colour transmitter by adding the three colour signals in the abovementioned ratio:

$$Y = 0.30\,R + 0.59\,G + 0.11\,B$$

This signal supplies the black and white receiver with all information required. In the colour receiver the colour information must then be added to the black and white signal in the form of

$$S1 = (R-Y) = (0.70\,R - 0.59\,G - 0.11\,B), \text{ and}$$

$$S2 = (B-Y) = (-0.30\,R - 0.59\,G + 0.89\,B)$$

In the receiver the colour information for green (G−Y) can be obtained from S1 and S2:

$$G-Y = -0.51\,(R-Y) - 0.19\,(B-Y).$$

By transmitting the S1 and S2 signals together with the Y signal the colour receiver can be supplied with full information concerning the colour picture.

To extend compatibility even further, the colour TV signal and the black and white TV signal are "jammed" into the same bands. As explained in Chapter

III.C.1 (Fig. 236) the band width between the carriers of the video signal and the audio signal is only 5.5 MHz wide. The principle is that if a black and white signal has a sufficient band width at its disposal (now reduced to a single side band of 4 MHz), and if we are content with a band width of 1 MHz for the colour signal, there is room for this in the frequency spectrum, earlier referred to as "no-man's land" between the (now somewhat narrowed) black and white signal and the audio signal. Addition of the less sharply defined colour information to the well defined black and white picture, is observed by the human eye as a good colour picture. Now the two signals S1 and S2 must also be accommodated in the frequency range at a distance of about 4.5 MHz from the video carrier. The method of doing this differs from system to system.

According to the American N.T.C.S. system S1 and S2 are amplitude modulated on a carrier of about 4.43 MHz; the modulation phases of S1 and S2 being shifted 90°. Here an additional synchronization signal is required to be able to "recover" the phases of S1 and S2; for that purpose a damped pulse train, the BURST, is transmitted after each line sync pulse in a certain reference phase.

With the French SECAM system (Chapter II.L.9) the information of S1 and S2 is given at alternate lines. In the receiver a delay line ensures that the S1 or S2 signal of the previous line is added to the S2 or S1 signal of the line described. Furthermore, frequency modulation is applied here.

Like the N.T.S.C. system, the PAL system, used in the rest of Europe, operates with amplitude modulation, with a phase shift of 90° between S1 and S2, but now the red signal changes sign at alternate lines. The viewer's eye now "averages out" two successive line signals, so that any phase errors are compensated. An improvement was found in the so-called PALd system, where the "averaging out" is done electrically by means of a delay line. To prevent interference the best carrier for the S1 and S2 signal was found to be: $f_S = (283\frac{3}{4} + 1/625)f_L$, which at $f_L = 25 \times 625$ becomes 4 433 619 Hz

The PAL system, the result of endless conferences between the representatives of the various countries, may be accused of being a European example of "perfectionism" which makes the receiver more complex and hence more expensive, but if the ultimate result is also a better colour picture, it might be expected that the required electronics will continuously become cheaper. A thorough study of the preferred system is of prime importance, for a system once chosen cannot be changed without rendering all existing receivers useless.

## III.C.6. BLOCK DIAGRAM OF THE PAL COLOUR RECEIVER

The block diagram of the colour television receiver shown in Fig. 240 deviates only in principle from that of the black and white receiver (Fig. 237) in that the chrominance signal and the corresponding additional synchronization signal, the BURST, have been added.

The aerial signal, composed as described in the previous chapter, again arrives in the tuner and is led to the I.F. amplifier via the various traps. After that there are three detections:

a. the audio signal, which after detection is modulated on the intercarrier frequency of 5.5 MHz, is amplified and transformed into the A.F. signal in the ratio detector to be further amplified and fed to the loudspeaker;
b. the noise signal which is led to the sync amplifier to interrupt amplification of the sync signal when a powerful interference occurs;
c. the luminance, chrominance, and synchronization signals are detected together. As a result the luminance and synchronization signals are restored to their original forms to be further processed as in the black and white receiver. The chrominance signal is modulated on the sub-carrier of 4.43 MHz (whose fundamental wave is suppressed in the transmitter) and can therefore easily be separated. The "burst" following the line sync pulse consists of a small pulse train with a frequency of 4.43 MHz and can also easily be filtered out. In the PAL decoder, which will be discussed in the next chapter, the B−Y and R−Y signals are restored to their original forms.

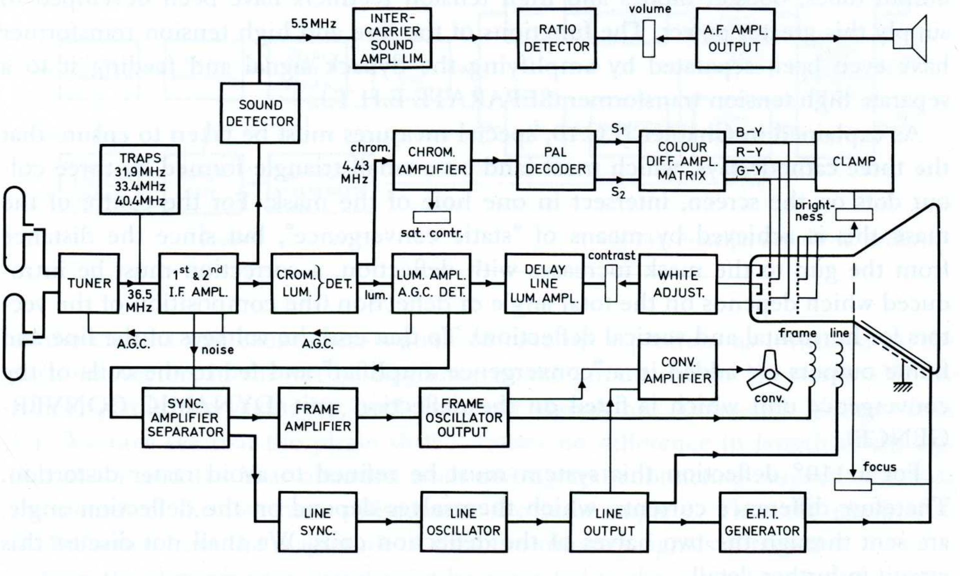


Fig. 240. Block diagram of a colour television receiver.

In the diagram shown in Fig. 240 the luminance signal is fed to the cathodes, and the three chrominance signals are applied to the first grids of the three guns of the colour picture tube. In a black and white tube the luminance signal must produce a black and white picture, and this will also be the case if the signal is applied to the three interconnected cathodes of the colour picture tube whilst the grids are earthed (or vice versa). Only the small difference in characteristics of the guns will then have to be eliminated with the "white adjusters".

After "decoding", the signals S1 = R − Y and S2 = B − Y are obtained as video voltages from the chrominance circuit, and by adding them in a certain ratio in a "matrix", a signal S3 = G − Y is obtained. If the signals S1, S2, and S3 are applied with the correct polarity to the three grids of the three colour guns, the voltage between grid and cathode in each of these guns will be R − Y + Y = R, and similarly for B and G, so that the tube gives satisfactory colour rendition.

Whereas contrast is controlled by the gain of the luminance amplifier, so by varying the cathode voltage, the brightness is controlled by adjusting the voltage on the second grids. Finally the colour "saturation" can be controlled by varying the gain of the chrominance amplifier, so the voltages on the first grids.

The abovementioned circuit, called a SELF MATRIXING circuit, is a logical sequence of the compatibility principle: if a black and white signal is received, the chrominance amplifier is simply put out of operation. Yet, nowadays the R.G.B.-circuit is mostly used, where the chrominance signals and the luminance signal are added in a matrix to be applied to the three first grids of the colour picture tube, whilst the three cathodes are earthed.

In principle the synchronization signals are amplified as in the black and white system, and fed to the self oscillating frame and line circuits. Since the picture tube contains three guns, the diameter of the neck and hence of the deflection coils are much larger, so that the latter require more power. Special line and frame output tubes, booster diodes and high tension rectifiers have been developed to supply this greater power. The functions of the line and high tension transformer have even been separated by amplifying the flyback signal and feeding it to a separate high tension transformer (SEPARATE E.H.T.).

As explained in Chapter II.C.10, special measures must be taken to ensure that the three cathode rays, which must land on a small triangle formed by three colour dots on the screen, intersect in one hole of the mask. For the centre of the mask this is achieved by means of "static convergence", but since the distance from the gun to the mask increases with deflection, a correction must be introduced which depends on the total angle of deflection (the composition of the vectors for horizontal and vertical deflection). To that end the voltages of the line and frame outputs are added in a "convergence amplifier" and fed to the coils of the convergence unit which is fitted on the deflection unit (DYNAMIC CONVERGENCE).

For a 110° deflection this system must be refined to avoid raster distortion. Therefore difference currents, which themselves depend on the deflection angle, are sent through the two halves of the deflection coils. We shall not discuss this circuit in further detail.

## III.C.7. THE PAL DECODER

We must, however, take a closer look at the principle of the circuit which restores the signals S1 and S2 modulated at a 90° phase shift on the 4.43 MHz sub-carrier to their original forms in which they can be applied to the grids of the picture tube.

The principle of the circuit is shown at the top of Fig. 241, below that the vector diagram of the voltage concerned is given. Let us call the incoming signal V1, the signal of the previous line V2, in which the $R-Y$ vector has changed sign. This signal V2 is delayed by one line in the DELAY LINE and appears therefore, at the output of the delay line simultaneously with the incoming V1. Any phase shifts of the combined chrominance signal ($\alpha$) occurring during transmission and

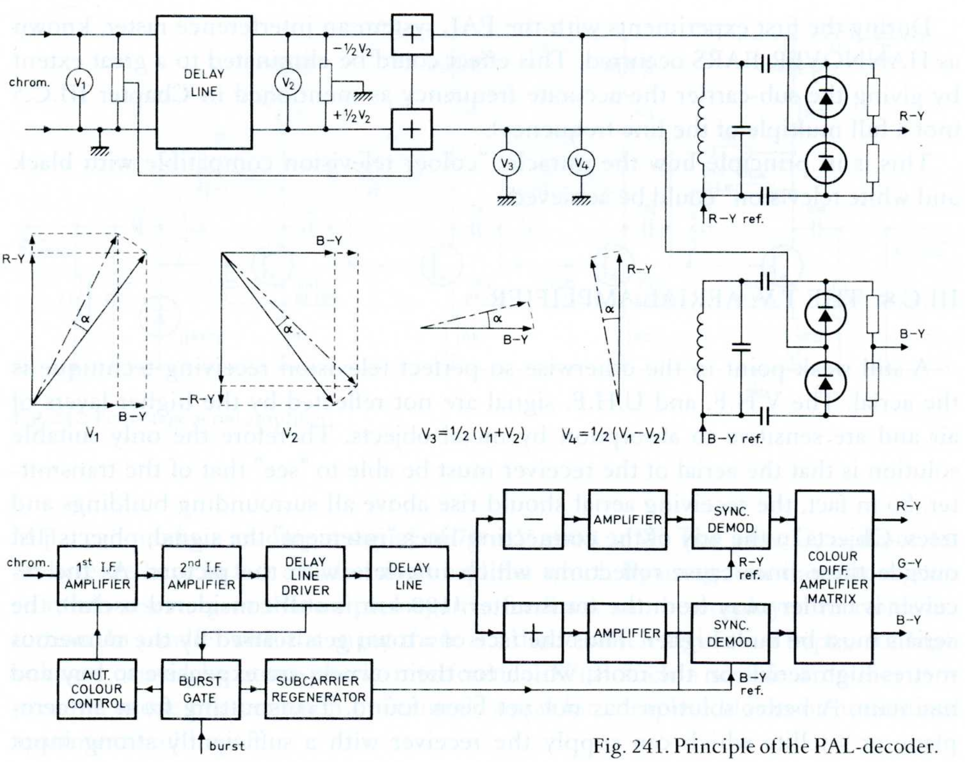


Fig. 241. Principle of the PAL-decoder.

in the circuits will cause these signals to take up a position as indicated by the dash-dot vectors in the vector diagrams. In the circuit shown ½V1 and ½V2 are added, giving the vector V3, and the same signals are subtracted, giving vector V4. We now see that the phase shift α causes no difference in length of the signals R−Y and B−Y as obtained from adding and subtracting respectively; only the phases of the two have shifted over the same angle α. Thus the phase shift is eliminated from the magnitude of the chrominance signals obtained.

Now the chrominance signal must be again fed to the sub-carrier which is suppressed in the transmitter. This sub-carrier is produced by a SUB-CARRIER REGENERATOR with a transmitter crystal. Measures should be taken to ensure that the phase remains accurately equal to that of the transmitter to have the oscillation (the line) start at the same instant. This is done by the BURST signal which, based on the same principle as for the line oscillator, synchronizes the oscillator. As the block diagram at the bottom of Fig. 241 shows, the output of the sub-carrier regenerator is added to the R−Y and B−Y signals which in turn are obtained by adding V1 and V2. In a demodulator (a kind of ratio detector) these signals are restored to their original form of video signal. Finally, the G−Y signal is determined by subtraction in a matrix.

Furthermore it should be noted that the complete decoder causes a phase delay of about 450 nanoseconds, so that the luminance amplifier must be equipped with a simple delay line (Fig. 240), which need not at all be as accurate as the one used in the PAL decoder.

During the first experiments with the PAL system an interference raster, known as HANNOVER BARS occurred. This effect could be eliminated to a great extent by giving the sub-carrier the accurate frequency as mentioned in Chapter III.C.5 (not a full multiple of the line frequency).

This is in principle how the miracle: "colour television compatible with black and white television" could be achieved.

## III.C.8. THE T.V. AERIAL AMPLIFIER

A still weak point in the otherwise so perfect television receiving technique is the aerial. The V.H.F. and U.H.F. signal are not reflected by the higher layers of air and are sensitive to absorption by metal objects. Therefore the only suitable solution is that the aerial of the receiver must be able to "see" that of the transmitter. So in fact, the receiving aerial should rise above all surrounding buildings and trees. Objects in the way of the connecting lines "intercept" the signal, objects just outside these lines cause reflections which interfere with the picture. As the receiver is farther away from the transmitter (100 km is still considered useful), the aerials must be built higher. Thus the face of a town gets scarred by the numerous metres-high aerials on the roofs, which for their owners are expensive to buy and maintain. A better solution has not yet been found. Transmitting from an aeroplane or satellite which can supply the receiver with a sufficiently strong input signal is still wishful thinking, although for countries like India the possibility of one central satellite for the whole country is being considered.

Some sort of improvement is made if several receivers are connected to one aerial. This means that for blocks of houses or large apartment buildings only one aerial need be erected. If municipal permission could be granted to lay ground cables under the pavement, it would even be possible to feed several blocks from one central aerial. Such an aerial, to which a great many receivers could be connected, could then be built very high, receive remote transmitters, and have a more favourable signal-to-noise ratio (in the Netherlands such a C(entral) A(erial) S(ystem) is receiving serious attention).

However, connecting several receivers direct to one aerial does not give satisfactory results, for the signal received by the aerial is so weak that further division would result in a signal which is not powerful enough for noise-free amplification. A good solution is obtained by mounting an amplifier close to the aerial. The noise generated by the lead from the aerial to the receiver is then of little effect on the greater power, and each receiver can be supplied with a signal that is even stronger than can be supplied by one aerial without amplifier.

Until a few years ago it has been impossible to make a WIDE-BAND AMPLIFIER which could cope with the entire frequency range from 40-860 MHz to be received in various countries. It was the arrival of the type BFY90 transistor with its low reactive capacitance, high transition frequency and high power gain (see Table III Chapter II.J.6) that brought the answer. Afterwards a heavier type, the BFW30, was introduced.

Fig. 242 gives the circuit diagram of such amplifier which, between 40 and 860

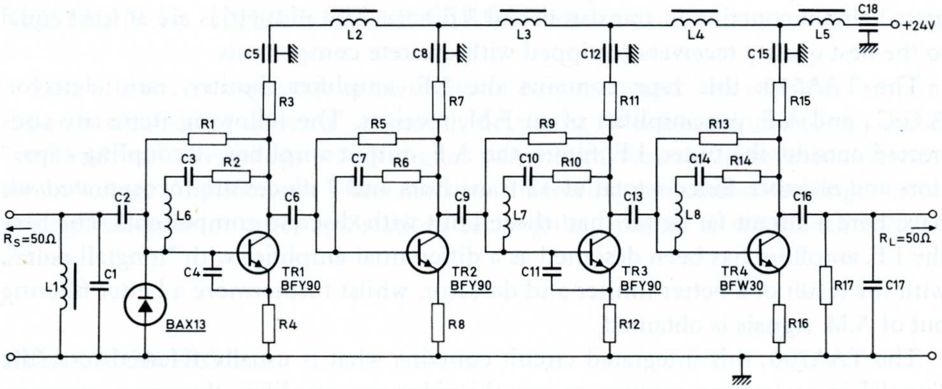


Fig. 242. Four-stage aerial amplifier.

MHz, can produce an output signal of 70 mV at 0.1 mW. The gain, G, is about 28. So here we have in fact a wide-band amplifier which by using four transistors gives so much reserve gain that ample negative feedback can be employed to avoid distortion. In parallel with the input there is a diode which serves as a bypass for high voltages caused by atmospheric interference. With such a simple amplifier it is possible to use one common or "central" aerial for complete blocks of houses and apartment buildings.

## III.C.9. APPLICATION OF LINEAR INTEGRATED CIRCUITS

It would be unjust to end the discussion of the circuits used in radio and television receivers without mentioning the most rapid development which just took place during the year 1969 in the field of integrated circuits used in these receivers. As explained before, a choice is so made that the part of the circuit in which little heat is developed and which contains no large resistors and capacitors is condensed in the form of an integrated circuit. This circuit is equipped with a maximum of 16 pins to which the input, output, supply, coils, and large external resistors and capacitors must be connected. The advantage is that the most complex, and often most critical, part of the circuit is made available as a highly compact and fully tested unit. Furthermore, the possibility of including many more transistors and diodes in the circuit without running into high expenses opens the way to much more complex circuits which have far better properties than those made with conventional discrete components. We shall give three examples of such integrated circuits.

The TAA840: an integrated circuit containing the H.F. amplifier-mixer oscillator, I.F. amplifier, A.G.C. detector and A.F. pre-amplifiers of an A.M. radio receiver operating on a 6 V battery. The following are connected to the integrated circuit: the H.F. and oscillator band-pass filters with variable capacitor, an I.F. filter with a ceramic resonator, a transformerless A.F. output amplifier up to 5 W with loudspeaker, and furthermore some decoupling capacitors and resistors. The integrated circuit contains 11 transistors and 3 diodes. The properties are at least equal to the best quality receivers equipped with discrete components.

The TAA570: this type contains the I.F. amplifier, limiter, ratio detector, A.G.C., and A.F. pre-amplifier of an F.M. receiver. The following items are connected outside: the tuner, I.F. filters, the A.F. output amplifier, decoupling capacitors and resistors. Since a total of 32 transistors and 7 diodes are incorporated, we have here a circuit far better than those built with discrete components. For here the I.F. amplifier has been designed as a differential amplifier with "longtail pairs", with the result of a better limiter and detector, whilst furthermore a better filtering out of A.M. signals is obtained.

The TAA700: this integrated circuit contains what is usually referred to as the "jungle" in a television receiver, to wit: the video pre-amplifier, the noise separator, the A.G.C. with delayed tuner action, the frame pulse separator and amplifier, and the line phase and frequency detector. This circuit incorporates 42 transistors and 11 diodes and is (as yet) the most complex linear integrated circuit, reason why it is shown on the cover of this book.

Obviously, the aerial amplifier, the stereo decoder as well as the PAL decoder qualify for "integration".

With the introduction of integrated circuits a considerable saving in assembly costs, increased reliability and simplified servicing may be expected.

# SECTION D. INDUSTRIAL CIRCUITS

This title covers a number of examples of circuits outside the scope of "amplification of alternating currents" (as mainly occurring in radio and television), and of "yes" and "no" circuits (on which digital switching is based). In industrial circuits direct voltage amplifiers are often used for measuring and control (analogue computing), whereas for other applications rectangular, instead of sinusoidal, voltages are used.

## III.D.1. CONTROL

Fig. 243a represents a circuit for keeping a supply voltage stable. Via a resistor

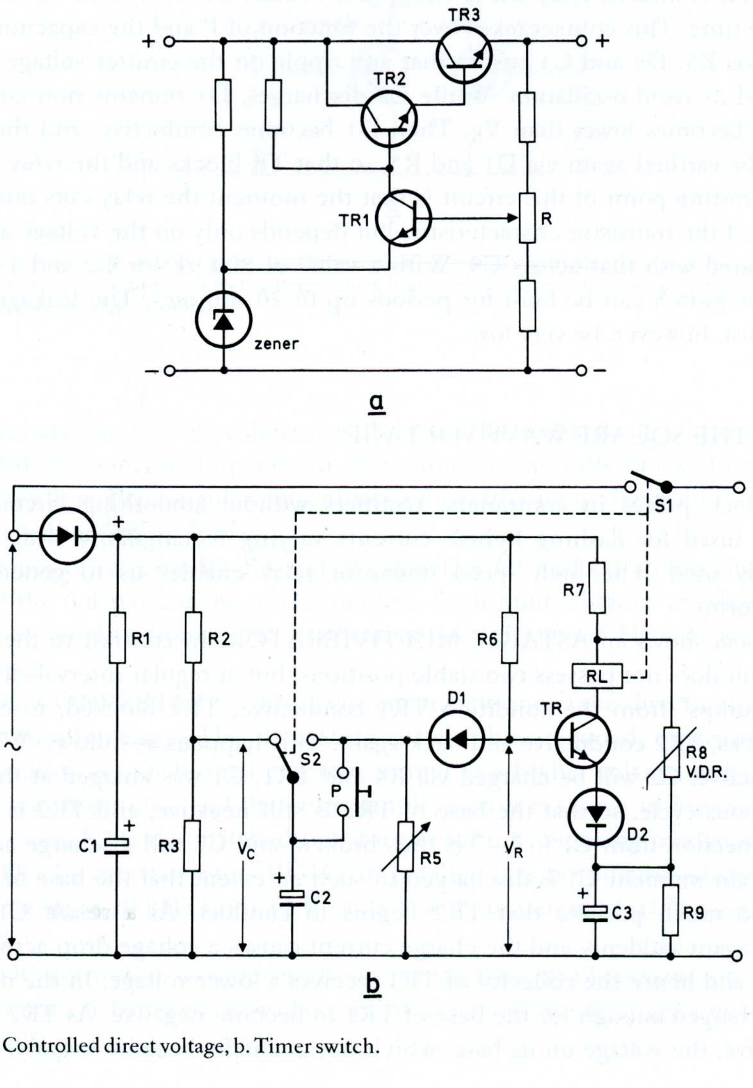


Fig. 243. a. Controlled direct voltage, b. Timer switch.

the input direct voltage is led to a voltage regulator (Zener) diode which supplies a reference voltage. A part of the output voltage, variable with R, is branched off and is compared with the voltage on the voltage regulator diode. The difference between these two voltages is amplified in the transistors TR1, TR2, TR3. The circuit will adjust itself so that the difference voltage between base and emitter of TR1 becomes zero, whilst across emitter and collector of TR3 a voltage drop occurs which compensates the voltage fluctuations at the output. If the voltage regulator diode is replaced by a variable voltage source, the output voltage can be controlled within certain limits in accordance with these voltage variations.

Fig. 243b gives the diagram of a TIME SWITCH. After rectifying, capacitor C1 is charged from the mains, and time defining capacitor C2 is charged with a part of the voltage on C1 that has been adjusted with R2-R3. In the situation as drawn the positive voltage on the base of TR flows away via D1 and R5, and TR is blocked. By closing push button switch P the base of TR is made positive (D1 blocks), TR conducts, relay RL is energized, contact S1 closes and S2 switches at the same time. This contact takes over the function of P and the capacitor C2 discharges via R5. D2 and C3 ensure that any ripple on the emitter voltage of TR is smoothed to avoid oscillation. While C2 discharges, D1 remains non-conductive until $V_C$ becomes lower than $V_R$. Then D1 becomes conductive, and the base of TR will be earthed again via D1 and R5, so that TR blocks and the relay cuts out. The interesting point of this circuit is that the moment the relay cuts out is independent of the transistor characteristic, but depends only on the voltage across C2 as compared with that across C3. With a value of 800 $\mu$F for C2 and 1 M$\Omega$ for R5 a time switch can be built for periods up to 20 minutes. The leakage current of C2 must, however, be very low.

## III.D.2. THE SQUARE WAVE VOLTAGE

For clock pulses in computers, rectifiers without smoothing circuit, on-off switches (used for flashing lights), currents varying rectangularly with time are preferably used. The high speed transistor relay enables us to generate such current forms.

Fig. 244a shows an ASTABLE MULTIVIBRATOR. In contrast to the flip-flop this circuit does not possess two stable positions, but at regular intervals it automatically "jumps" from the condition TR1 conductive, TR2 blocked, to condition TR1 blocks, TR2 conductive and back again. This happens as follows. When TR1 is conductive, C2 will be charged via R4 and TR1. C1 was charged at the end of the previous cycle, so that the base of TR2 is still negative, and TR2 is blocked. The connection from C1 to "—" is thus broken, and C1 will discharge across R3. At a certain moment C1 is discharged to such an extent that the base of TR2 becomes so much positive that TR2 begins to conduct. As a result C1 will be charged again suddenly and the charge current causes a voltage drop across R1, so that Q2, and hence the collector of TR1 receives a lower voltage. In the meantime C2 has charged enough for the base of TR1 to become negative. As TR2 becomes conductive, the voltage on its base, which has gradually become negative, and the

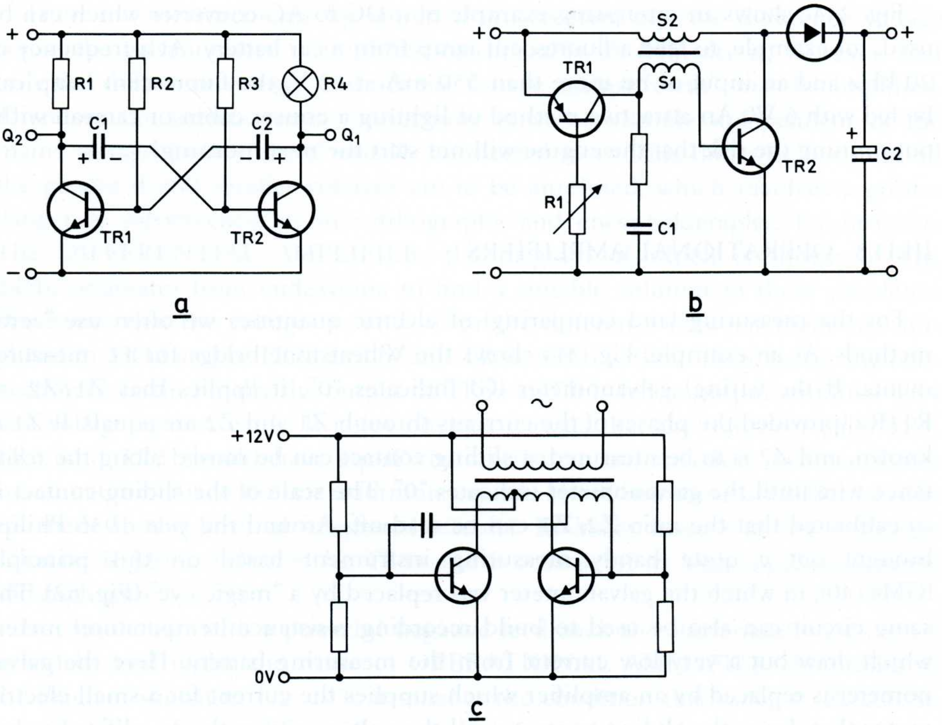


Fig. 244. Generating square-wave pulses and "converters": a. astable multivibrator (e.g. flickering light), b. D.C. converter, c. converting a direct voltage into a higher alternating voltage (supplying a fluorescent lamp from an accumulator battery).

sudden voltage drop on its collector will cause TR1 to block. It is typical of this circuit that the transistors operate in saturation, i.e. are fully blocked or conductive: the result is a rectangular current wave form (the lamp drawn in the place of R4 will be either "on" or "off").

The values of resistors R1-R4 and of capacitors C1 and C2 are decisive for the length of the pulse occurring at Q1 and Q2. Such multivibrators are designed for frequencies ranging from 2 Hz (trafficators) to millions of Hz (clock pulses in computers).

In a DC CONVERTER (a "direct-current transformer") the direct current is interrupted by a transistor. The alternating current thus obtained, consisting of rectangular pulses, is transformed in a transformer to a higher voltage, to be rectified again. A very ingenious, and therefore simple, circuit is shown in Fig. 244b. Here the transistor TR2 and a matched transformer form a "blocking oscillator". By replacing TR2 by two transistors BD121 in parallel, it is possible to achieve voltage transformation at 20 kHz from 12 V to 28 V with an output of 23 V at an efficiency of 84%.

Such DC converters were formerly needed to feed small transmitters (mobilophones) from car batteries; nowadays we have transistors which give enough output at a collector voltage of 12 V.

Fig. 244c shows an interesting example of a DC to AC converter which can be used, for example, to feed a fluorescent lamp from a car battery. At a frequency of 20 kHz and an input of no more than 550 mA at 12 V, the fluorescent lamp can be fed with 6 W. An attractive method of lighting a cruiser cabin or caravan without running the risk that the engine will not start the next morning!

## III.D.3. OPERATIONAL AMPLIFIERS

For the measuring (and comparing) of electric quantities we often use "zero" methods. As an example, Fig. 245 shows the Wheatstone bridge for a.c. measurements. If the (string) galvanometer (G) indicates "0", it applies that Z1/Z2 = R1/R2 (provided the phases of the currents through Z1 and Z2 are equal). If Z1 is known, and Z2 is to be measured, a sliding contact can be moved along the resistance wire until the galvanometer indicates "0". The scale of the sliding contact is so calibrated that the ratio Z2/Z1 can be read off. Around the year 1936 Philips brought out a quite handy measuring instrument based on this principle (GM4140), in which the galvanometer was replaced by a "magic eye" (Fig. 62). The same circuit can also be used to build recording resistance (temperature) meters which draw but a very low current from the measuring battery. Here the galvanometer is replaced by an amplifier which supplies the current for a small electric motor that drives the sliding contact until the voltage across the amplifier has become "0". The sliding contact can be fitted with a recording pen.

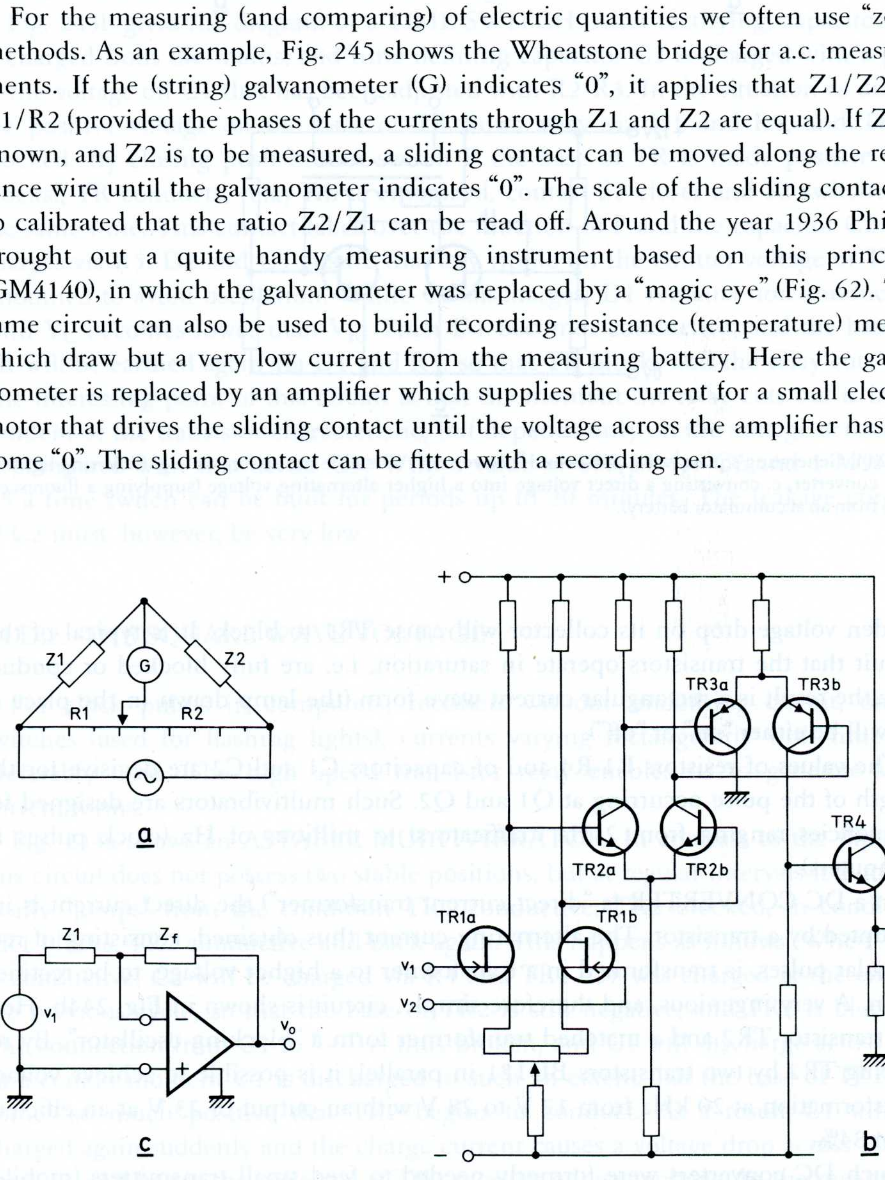


Fig. 245. Differential amplifiers: a. Wheatstone bridge for alternating current, in which G can be replaced by a differential amplifier, b. circuit diagram of a differential amplifier, c. principle of an "operational" amplifier.

There are measurements for which it is necessary that the difference voltage across the galvanometer is amplified considerably. A difficulty is, however, that only this difference voltage, and not the voltage against "earth" must be amplified. Therefore the galvanometer connections must be entirely independent of the mains voltage or the "earth" usually connected to it. This difficulty becomes all the greater if still smaller voltages are to be amplified, which involves a greater danger of interference, as in cardiography and encephalography, for instance. The DIFFERENTIAL AMPLIFIER, the diagram of which is shown in Fig. 245b, originates from endeavours to find a suitable solution to these problems. We notice that the voltages $v_1$ and $v_2$ (or the difference between them) are amplified in TR1a and b, TR2a and b, and TR3a and b without there being any connection to earth. Finally the output of TR3a and b is fed to output amplifier TR4 as a voltage against earth. It is for obvious reasons that the transistor pairs TR1, TR2, and TR3 should have equal characteristics, and therefore two transistors are sometimes made on one common crystal. Nowadays these amplifier circuits are made in the form of integrated circuits. In many cases a M.O.S. transistor (Chapter II.J.9) is indicated for TR1a and b because of its high input impedance and high gain.

In conjunction with a powerful negative feed back the abovementioned circuit is widely used under the name of OPERATIONAL AMPLIFIER. According to the diagram shown in Fig. 245c it satisfies the condition that $v_0/v_1 = Z_f/Z1$. So here we have a d.c. amplifier with a certain gain, and it can be used to integrate, to add, in short: it can be used for ANALOGUE COMPUTING. Furthermore it can be used to build a multivibrator with all circuits derived from it.

# SECTION E. ELECTRONIC DIGITAL COMPUTING

## III.E.1. INTRODUCTION

Chapter II.N.4 was devoted mainly to converting measured quantities into a number of pulses which were counted or stored in a memory system. The function "electronic computing" can be described as follows. From an "input" or a memory, several numbers are fed into the ARITHMETICAL UNIT where they are used to make computations (we shall only consider adding, subtracting, multiplying and dividing). The result of the computation is either produced in visible figures (usually the decimal system), or in the binary system to be stored in a memory.

In electronic computing the numbers in the binary system are presented by a sequence of "high" (H) and "low" (L) voltage levels. Such a number can be fed into a shift register (Chapter II.N.4) to be stored there. From this register the number can be read immediately, if required, in the decimal system. This is done by means of a SELECTOR UNIT which, on the basis of the "H" or "L" information brought into the register, establishes a certain connection, say with a numerical display tube (Fig. 246a). For the performance of arithmetical operations this number can be shifted out of the register by means of clock pulses after which it is combined with other numbers via so-called "gates" (Chapter II.N.4.) In general the outcome arrives in a following shift register from where it can be made available for subsequent arithmetical operations.

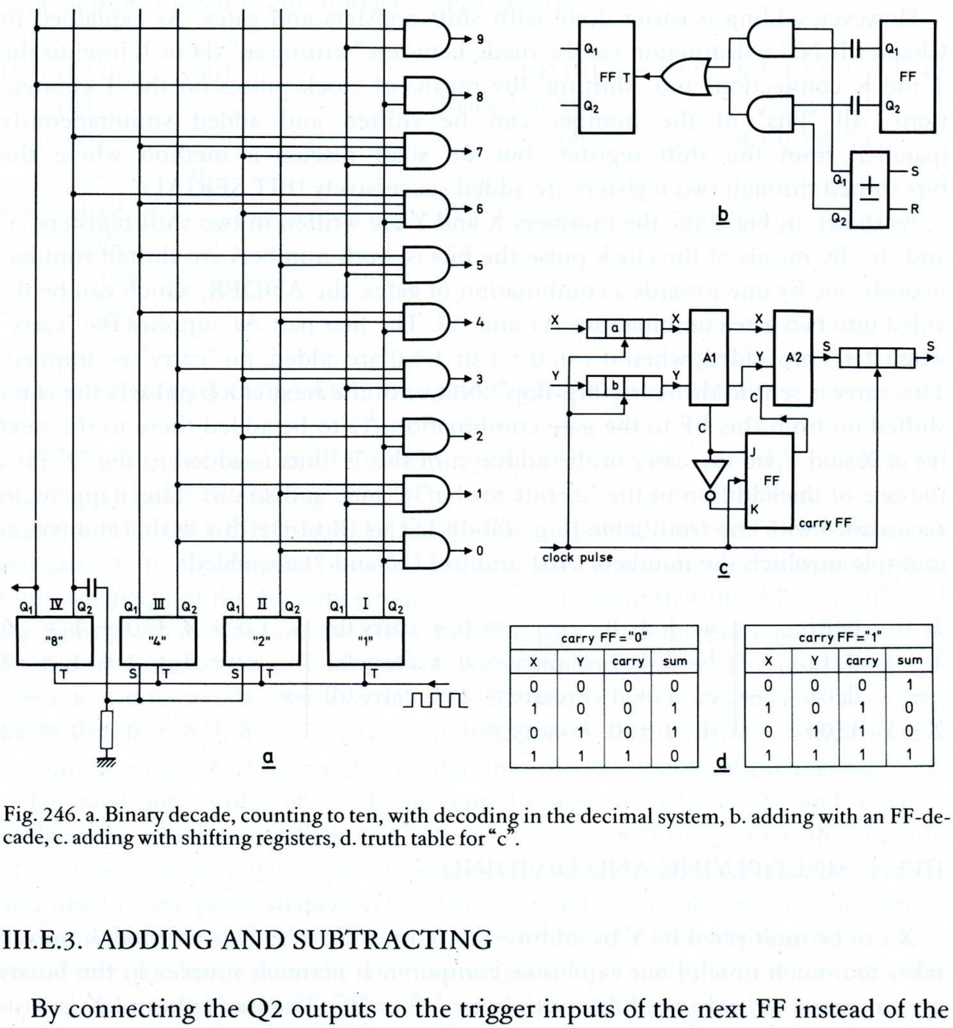
During computation it is possible to "read" numbers out and "write" them into a memory (Chapter II.N.5). The ultimate result is fed to the "output": a row of numerical indicator tubes, a type writer, a high speed printer or a magnetic tape.

Lastly "commands" are needed to have each of the arithmetical operations carried out. This is done by means of a CONTROL UNIT. A simple method is sending the command to the gates concerned in the form of a "high" level. These gates are then "open" for the (positive) clock pulses which are to carry out the required operation (for instance, shifting through the shift register). The succession of commands is known as the PROGRAMME.

## III.E.2. CONVERSION FROM BINARY TO DECIMAL

In Chapter II.N.4 it was explained how a number, in the form of a train of pulses, can be fed into a binary counter (Fig. 218a); written in the binary system. If the number must be read-out of this "register" again, this can be done with the circuit shown in Fig. 246a. The outputs Q1 and Q2 of the decade of flip-flop (FF) circuits are so connected to a series of AND gates that only one gate is available for each number written in; all inputs being "1", so this gate will pass the "1" level. The gates are so arranged that they present an increasing series of numbers. Using such a selector unit for 4FFs offers the choice of one out of maximum 16 output wires whose figure corresponds to the number written into the register

(SELECTOR UNIT). When counting in the decimal system, a "carry-over" pulse must be fed to the counter as soon as a unit of ten has been reached. This is done, as shown in Fig. 246a, by feeding a "set" pulse to the "2" and the "4" counter as soon as the "8" counter changes from position "0" to "1" (after the number 8 has been written in). Now when the counter reaches position "1000", it will automatically jump to "1110". At the ninth pulse it switches to 1111, and at the tenth pulse to 0000, whilst Q1 of the "8" FF feeds a "carry-over" pulse to the decimal counter. There are various other methods of using a binary counter for decimal counting, but we shall not discuss them here.



| carry FF = "0" | | | |
|---|---|---|---|
| X | Y | carry | sum |
| 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 |

| carry FF = "1" | | | |
|---|---|---|---|
| X | Y | carry | sum |
| 0 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 |

Fig. 246. a. Binary decade, counting to ten, with decoding in the decimal system, b. adding with an FF-decade, c. adding with shifting registers, d. truth table for "c".

## III.E.3. ADDING AND SUBTRACTING

By connecting the Q2 outputs to the trigger inputs of the next FF instead of the Q1 outputs (as was done in Fig. 218a), the binary counter will count negatively, that is to say, starting from 0000, to 1111, 1110, 1101, 1100, etc. Fig. 246b gives two elements of a binary counter which can count both "+" and "−". An auxiliary FF, marked "±" applies a positive voltage to one of the two gates which are also connected to Q1 and Q2 respectively of the preceding FF of the binary counter. Now either the voltage of Q1 or of Q2 will be passed for positive or negative counting depending on the position of this "±" FF. The selection can be made by means of a pulse to its S or R input.

Now we write the number X into a binary counter "a" and number Y into counter "b". Then "a" is set for "+" counting, and "b" for "−" counting, whilst a series of clock pulses is sent simultaneously through "a" and "b" to be stopped when "b" returns to 0. Thus the number X + Y has been written into "a".

If the number X is written positively into "a", and Y negatively into "b", and the clock pulses are sent through "a", which is set for negative counting, and also through "b" which is set for positive counting, the difference X−Y will have been written into "a" as soon as "b" returns to 0.

However, adding is easier done with shift registers and gates. As explained in Chapter II.N.4 a distinction can be made between "writing in" (H or L level to the J and K connection) and "shifting" (by means of clock pulses on the T connection). All "bits" of the number can be shifted and added simultaneously (parallel) from the shift register, but we shall discuss a method where the bits shifted through two registers are added successively (BIT SERIAL).

As shown in Fig. 246c the numbers X and Y are written in two shift registers "a" and "b". By means of the clock pulse the bits of both numbers are shifted simultaneously one by one towards a combination of gates, the ADDER, which can be divided into two gate combinations A1 and A2. The first part A1 supplies the "carry" when 1 + 1 are added; when 0 + 0, 0 + 1 or 1 + 0 are added, no "carry" is supplied. This carry is sent to the "carry flip-flop". Not until the next clock pulse is the carry shifted on from this FF to the gate combination A2 to be added there to the next bit of X and Y (so the carry of the addition of the "1" bits is added to the "2" bits, the one of the addition of the "2" bits to the "4" bits, and so on). This happens in accordance with the truth table (Fig. 246d). Let us illustrate this with a numerical example in which the numbers 0101 and 0111 (5 and 7) are added:

| | | | | | |
|---|---|---|---|---|---|
| X | 0101 | | a: 1 + 1 = 0 + carry (b) | | 0 + 4 + 0 + 1 = 5 |
| Y | 0111 | | b: 0 + 1 + carry = 0 + carry (c) | | 0 + 4 + 2 + 1 = 7 |
| —— | dcba | → | c: 1 + 1 + carry = 1 + carry (d) | → | d c b a —— |
| X+Y | 1100 | | d: 0 + 0 + carry = 1 | | 8 + 4 + 0 + 0 = 12 |

## III.E.4. MULTIPLYING AND DIVIDING

X can be multiplied by Y by adding Y−1 times X to X. This method, however, takes too much time of our expensive computer. It is much simpler in the binary system, since we only need to multiply by "1" or "0". The multiplicand X is written into one register, through which it can be shifted bit by bit to the left. After each shift of the multiplicand the contents of the register will be added to a second register, if and only if the multiplier Y contains a "1" at the position corresponding to the shift of X. When X has been shifted through all positions of Y the second register will contain the result of the multiplication. The procedure is illustrated by the example 1101 (13) times 10101 (21):

$$\begin{array}{ll}
1 & \phantom{...}10101 \\
0 & \phantom{..}\mathrm{xxxxx}. \\
1 & \phantom{.}10101.. \\
1 & \underline{10101...} \\
 & 100010001
\end{array}$$

$$\begin{array}{rcrcr}
10101 & = & 16+0+4+0+1 & = & 21 \\
1101 & = & 8+4+0+1 & = & \underline{13} \\
 & & & & 63 \\
 & & & & \underline{210} \\
100010001 & = & 256+0+0+0+16+0+0+0+1 & = & 273
\end{array}$$

When dividing we always try to subtract the denominator, shifted bit by bit, from the dividend. If the difference is positive, there is a "1" in the quotient and the divisor is subtracted; if the difference is negative there is a "0" in the quotient, and the divisor is not subtracted. For example:

$$\begin{array}{ll}
101/101101/ & 1001 \\
\phantom{10}\underline{101} & 1 \\
\phantom{10}001 & \\
\phantom{10}\underline{\mathrm{xxx}} & \phantom{1}0 \\
\phantom{101}010 & \\
\phantom{101}\underline{\mathrm{xxx}} & \phantom{10}0 \\
\phantom{1010}101 & \\
\phantom{1010}\underline{101} & \phantom{100}1
\end{array}$$

$$\begin{array}{rrcr}
101101 = & 32+0+8+4+0+1 & = & 45 \\
\text{divided by } 101 = & 4+0+1 & = & 5 \\
1001 = & 8+0+0+1 & = & 9
\end{array}$$

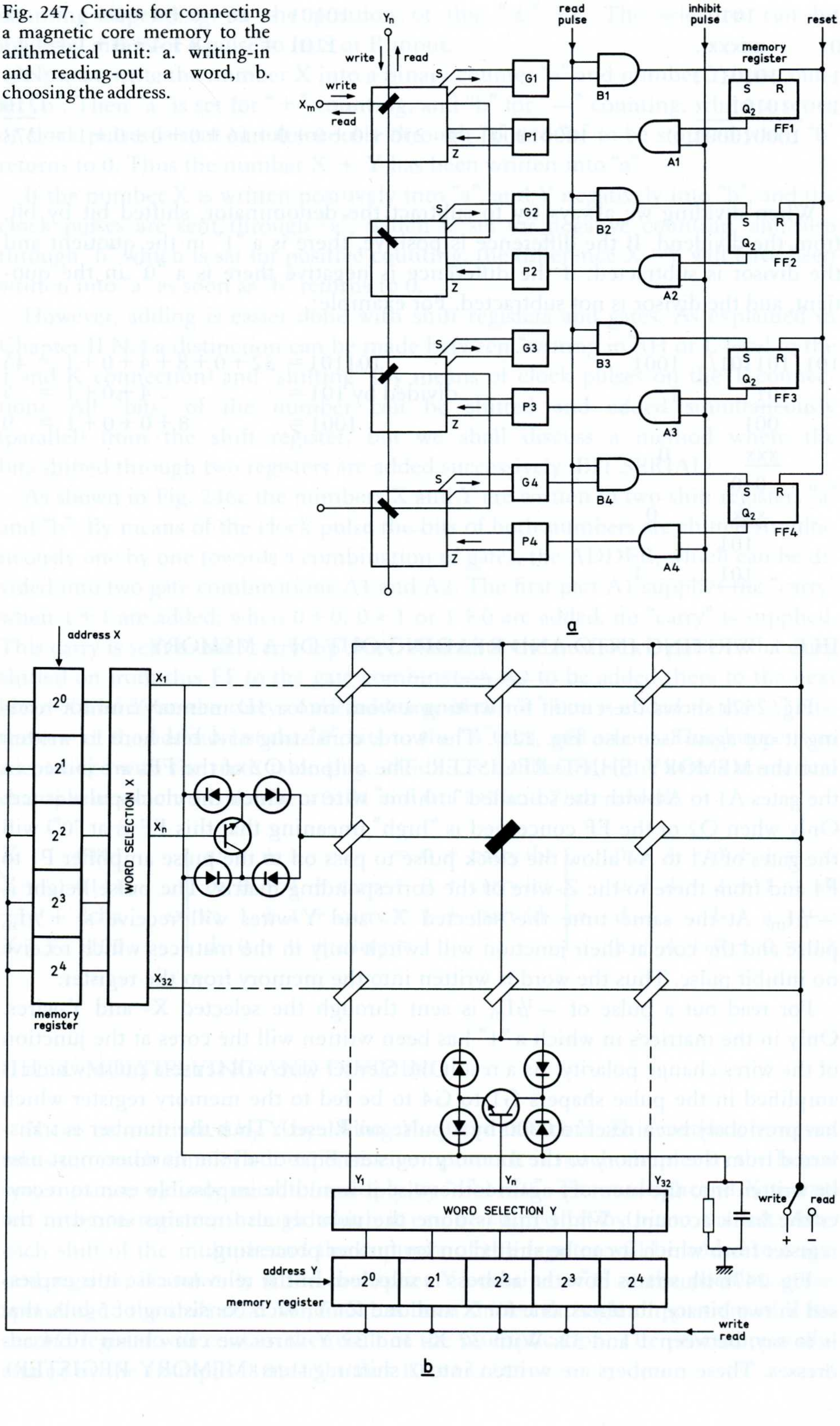
## III.E.5. WRITING INTO AND READING OUT OF A MEMORY

Fig. 247a shows the circuit for writing a word into a 3D memory and for reading it out again (see also Fig. 221). The word, consisting of 4 bits here, is written into the MEMORY (SHIFT) REGISTER. The outputs Q2 of the FFs are joined in the gates A1 to A4 with the so-called "inhibit" wire to which the clock pulse is fed. Only when Q2 of the FF concerned is "high" (meaning that this FF is at "0") will the gates of A1 to A4 allow the clock pulse to pass on to the pulse amplifier P1 to P4 and from there to the Z-wire of the corresponding matrix. The pulse height is $-\frac{1}{2}I_m$. At the same time the selected X- and Y-wires will receive a $+\frac{1}{2}I_m$ pulse and the core at their junction will switch only in the matrices which receive no inhibit pulse. Thus the word is written into the memory from the register.

For read out a pulse of $-\frac{1}{2}I_m$ is sent through the selected X- and Y-wires. Only in the matrices in which a "1" has been written will the cores at the junction of the wires change polarity. As a result the S(ense) wire will sense a pulse which is amplified in the pulse shapers G1 to G4 to be fed to the memory register which has previously been reset to 0000 by a pulse on R(eset). Thus the number is transferred from the memory to the memory register. First of all the number must now be written into the memory again (otherwise it would be impossible ever to recover the bank account). While this is done the number also remains stored in the register from which it can be shifted on for further processing.

Fig. 247b illustrates how the address is selected. In the relevant case it is expressed in two binary numbers, one for X and one for Y, each consisting of 5 bits, that is to say, between 1 and 32. With 32 X- and 32 Y-wires we can obtain 1024 addresses. These numbers are written into 2 shift registers (MEMORY REGISTER).

Fig. 247. Circuits for connecting a magnetic core memory to the arithmetical unit: a. writing-in and reading-out a word, b. choosing the address.



A circuit similar to the one of Fig. 246a, but with 5 bits and 32 gates, ensures that each number to be written-in is led to the proper wire of the memory. This combination of gates is known as the WORD SELECTION UNIT. Once the address has been selected, a WRITE-READ clock pulse will be fed to the selected X- and Y-wire. For each X- and each Y-wire there is a four-diode circuit and a so-called symmetrical transistor. When the base of this transistor becomes positive, it will pass both a positive and a negative voltage. Depending on the position of the "write-read" switch, a pulse of $+\frac{1}{2}I_m$ or $-\frac{1}{2}I_m$ will be sent through the selected X- and Y-wire simultaneously with the inhibit pulse along the Z-wire for write in. In Fig. 247b only one matrix is drawn. In reality the X- and Y-wires run through as many matrices as there are bits per word. We see that each X- and Y-wire requires only one connection from the switching device to the matrix.

For storing more than 32 × 32 addresses the memory register would become too complicated, and therefore numbers are added to the address to select groups of X- and Y-wires.

## III.E.6. PROGRAMMING

The programme of a computer indicates what operations must be carried out successively, that is to say, to which clock pulse inputs (partly shown in Figs 246 and 247) successive pulse trains are to be fed. We saw that for some operations (multiplying, etc) several successive pulses or pulse trains are needed. Therefore a "sub-routine" system is used.

The pulse generator has several outputs which produce different numbers of pulses, usually ended by a current pulse for the change-over to the next operation.

A sub-routine system can be regarded as a collection of switches which establish the necessary connections between the relevant output of the pulse generator and the pulse inputs of the circuits for each of the different operations. By means of a series of FFs, several of such switches can be connected successively. With desk computers a "sub-routine" is started with a push button. If such a computer is equipped with a memory, the address numbers too, must be selected with push buttons.

More complex computers are required to contain a system for the automatic selection of a rapid succession of sub-routines. A programme is then made consisting in its simplest form of a plate with horizontal rows of contacts which are connected to the sub-routine starting switches, and vertical rows of contacts to be connected successively to run down the programme. The programme is fixed by making specific interconnections on this plate. Such plates can be interchanged by means of plug-in contact strips. A similar programme can also be made by means of a matrix with magnetic cores in which a sub-routine is selected by leading a wire through the required cores. Finally the data for the programme can be written in with a punched tape to be processed electronically.

Programming computers, that is finding the simplest sequence of sub-routines to carry out complex operations, has developed into a special science. In the application of computers it has been found that the SOFT WARE (making the programmes) is mostly as important as the HARD WARE (the electronic circuit), but we shall not go into details here.

We do hope, however, that the above short discussion has provided sufficient insight into the principles on which the functioning of the computer is based.

## III.E.7. INCREASING THE SPEED

As explained in Chapter I.22, the very rapid succession of operations required for electronic binary computing can be achieved only if the time required for each of the operations is very short. As the computers became more complex, and consequently more expensive, there was a keen demand for ever shorter switching times, to limit the overall time (and hence the cost per operation) to a minimum.

During the initial period of the computers, diodes were considerably cheaper than transistors, and therefore gates were made with diodes to be followed by only one transistor amplifier. With this D(iode) T(ransistor) L(ogic) (FC family of integrated circuits) it was possible to make circuits for operation times of 30 ns. With integrated circuits in progress, the price of a transistor came down to that of a diode. A transistor has a higher gain and is indeed somewhat faster. And so we come to the T(ransistor) T(ransistor) L(ogic) (F J family of integrated circuits) offering the possibility of obtaining switching times of no more than 13 ns. Even shorter switching times (down to 2.5 ns) could be obtained with C(urrent) M(ode) L(ogic)(GH family of integrated circuits).

We notice that the electronic computer develops in two directions. Firstly towards the large expensive computer which is made efficient by virtue of its enormous speed and the possibility of connecting many users (via telephone lines: DATA TRANSMISSION). For less complicated computations there are the desk computer or bookkeeping machines which, thanks to standardized integrated circuits and core memory circuits can be made much cheaper.

# PART IV. ELECTRONIC APPLICATIONS

## SECTION A. BROADCAST (FROM ONE STATION TO MANY)

### IV.A.1. RADIO

An embittered philosopher once wrote: "God gave man the power of thinking, and what did he use it for?" Likewise one might ask: "Electronics gave man radio, and what did he use it for?"

How right SARNOFF was, when in about 1920 he predicted that radio, in those days used only for telecommunication between ships, would become a source of entertainment for mankind. What tremendous possibilities had now been created because information, education, entertainment and some of the fine arts could be sent into the ether (thanks to advertising usually "gratis"), so that anyone having a receiver costing no more than a few pounds can enjoy them at distances of hundreds of miles.

The spoilt inhabitants of modern countries now have at their disposal an even better means of home entertainment: television, be it that the distance to the transmitter is of necessity shorter, resulting in a limitation of programme selection. Therefore, in these countries radio has mostly been degraded to background music: transistor radios with their small batteries can be taken outdoors, to work, to the seaside, etc. But for the hundreds of millions who still have to do without television the so much cheaper radio receiver is still a highly appreciated means of maintaining contact in any field with the rest of the world.

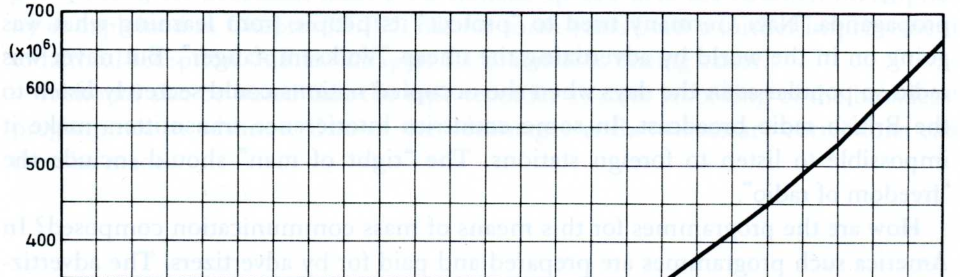


Fig. 248. Growth of the number of radio receivers used in the world (excluding China, Russia and Eastern Europe) (estimate made by Philips C.V. & P.).

Fig. 248 gives an estimate of the number of radio sets used throughout the world, excepting Russia, China, and Eastern Europe. It appears that at the moment there are over 700 million, and that since transistor sets came out around 1957, the annual growth has increased, and still continues to do so. As mentioned before, there are two different reasons for this growth: on the one hand there are the inhabitants of electrified areas who can now carry their "personal set" wherever they go, and then there are the many millions who have to do without electricity, and can now also enjoy the contact with the world.

For the enthusiastic radio listener it is disappointing that in terms of organization and cooperation mankind has failed to draw full profit from this means of many-sided home entertainment. Notwithstanding international consultation in this field, all endeavours to obtain and maintain the most favourable allocation of transmitter wavelengths have failed. With the modern selective super heterodyne receivers there must be a way of getting a reception of reasonable quality with a frequency difference of 20 kHz between neighbouring stations. The so-called medium frequency range from 190-580 metres would then be wide enough to accommodate nearly 50 different stations. In practice, however, the various stations cause mutual interference because their respective wave lengths are too close together. A considerable part of the valuable wave length range is rendered useless because enemy ideologies impede reception of each other's broadcasts by means of interference transmitters.

In Chapter I.9 it was already explained how reflection in higher layers in the air enable radio contact round the earth if short waves are used. Indeed tuning then becomes more difficult, and the quality is affected by interference in the air and by fading, but in principle there is the possibility of broadcast between continents.

Radio is in principle international and could be a means of international exchange of culture. But at the same time radio is feared as a medium for politic propaganda. Nazi Germany tried to "protect" its people from learning what was going on in the world by advertizing the cheap "Volksempfanger". But never was radio so popular as in the days when the occupied nations could secretely listen to the British radio broadcast. In some countries interference transmitters make it impossible to listen to foreign stations. The "right of man" should include the "freedom of radio".

How are the programmes for this means of mass communication composed? In America such programmes are prepared and paid for by advertizers. The advertizing broadcasts will always try to arrange an attractive programme to ensure that the public will tune in and hear a few minute's advertizing. In Europe, the authorities, supported by controlled broadcast companies, have undertaken the responsibility for programming. Money contributed by the public and paid by advertizing sponsors is received by the government. That the listeners are not always pleased with these programmes is evident from the enormous interest for "illegal" radio stations on board ships anchored in the North Sea, which broadcast their programmes to the Netherlands, Belgium and the United Kingdom, and are paid by advertizers.

Considering all abovementioned causes, we must come to the conclusion that mankind has not succeeded in drawing full profit from radio as a means to information, education, entertainment, enjoyment of art and international exchange of culture.

Special mention must be made of the purposeful use of radio as a means of education in developing countries where there is still a high percentage of illiteracy and where there are not sufficient schools and teachers to cover the population spread out over thousands of miles. In this respect an important experiment was initiated in Columbia in 1950. Under the patronage of the church an educational programme was set up for the young and the old, to be broadcast from a few stations. The population of every village is presented with a collective radio receiver which, by means of a crystal, is accurately tuned to the wavelength of the educational station. In several other countries similar activities have been started or are being prepared. The UNO is very active in this field. In 1970 even the European Common Market held a meeting to study cooperation within the community in the field of education per radio and television.

In the early period of radio, when the prices of radio sets were still high and interference suppression inadequate, the system of RADIO RELAY was introduced in many agglomerations. Enterprizing people installed a good radio receiver with a high aerial, an amplifier and a distribution network, so that all a subscriber needed was a loudspeaker, a volume control and a selector switch to be able to listen in on about four different programmes. This system does not enable the listener to select from a large number of stations, but it certainly has its advantages, particularly for those for whom tuning a radio set might be troublesome.

Introduction of FREQUENCY MODULATION meant a considerable improvement of radio reception: more high notes, less distortion and less interference. However, the distance between transmitter and receiver cannot be more than about 100 km. In various countries a part of the programmes, in particular good music, are broadcast from F.M. stations spread over the country. Stereophony as explained in Chapter III.B.5, has for some years been a further improvement to faithful reception. A.F.C. (Chapter III.B.4) makes tuning easier.

It should also be noted that the tape recorder enables recording of the broadcast programmes in advance. One of the advantages is that necessary corrections can then be made.

## IV.A.2. PUBLIC ADDRESS

The amplifier and the loudspeaker provide a means to amplify sound, and particularly the human voice, without restriction. One speaker or singer can be made audible to large crowds, especially if the loudspeakers are adequately spread among the public (Fig. 15). In about 1930 Hitler was one of the first to use this means of influencing the masses. Nowadays we simply cannot imagine a large meeting without a public address installation being used. The technique of speech or singing has changed accordingly: one can whisper into the microphone!

Mechanical means of delay can be used to produce artificial "REVERBERATION", so that sound in the living room will sound as in the church. It has even been tried to use this technique for improving the acoustics in a "dead" concert hall.

Aboard ships, planes, trains, buses and in railway stations travellers can be addressed via loudspeakers (unfortunately the art of using the microphone the right way and adjusting the amplifier correctly is practised far too little). Furthermore background music in factories ("Music while you work"), shops and public houses has really taken wing. The micro-amplifier in the form of a hearing aid was already shown in Fig. 34.

Instead of leading sound to the microphone via the air, the microphone can also be mounted direct on the sound board of the music instrument. The electronic guitar is an important item in dance and "pop" music.

A more interesting instrument is the electronic organ. Here the sound is formed in electric oscillator circuits. Higher harmonics can be added to adjust the timbre to taste. However, here the sound lacks the lively, somewhat irregular timbre of the music as perceived in the mechanical vibration of string, brass and wood-wind instruments. The signals from several oscillating circuits can be fed to the loudspeaker so that it becomes possible to reproduce a complete chord by simply pushing one key. However, the choice of chords remains limited as with an accordeon. Experiments were even made to imitate electronically the typical sound of the felt hammer striking against the piano strings. A great advantage of the electronic music instrument is that the musician can practise by means of headphones so that the other members of the family need not be disturbed.

### IV.A.3. TELEVISION

All the possibilities which Sarnoff predicted with respect to radio apply to a greater extent for television. It is true that a greater financial lay-out is required to get an apparatus with which sound and image can be received via the air. Owing to the higher frequencies used, the distance between transmitter and receiver cannot be much more than 60 km. By means of aerial amplifiers (Chapter II.C.8) the minimum received signal strength can be boosted, but even then the necessity of the aerials being able to "see" each other remains. So a number of transmitters broadcasting the same programme must be distributed over the country; the programmes being transmitted to these relay stations via link transmitters. In areas surrounded by mountains "satellite transmitters" of smaller power are built.

Owing to the high expense of the programmes and the link transmitter networks, only two, at the most three, programmes are broadcast per country (area). Watchers living near the borders can also receive the programmes of neighbouring countries.

The national link transmitter networks of Western Europe can be interconnected to form the "Eurovision Network". It is a pity that insufficient use is made of this splendid opportunity of letting all Europeans enjoy the best that music and dance can offer. For such broadcasts the artists require extra pay, something that many broadcast companies often cannot afford. Since professional football has been strongly commercialized, similar difficulties crop up from that side too.

In the meantime television has become a "mass medium" without equal. With a viewing density of over 80% in Europe, the U.S.A. and Japan, we may confidently assume that a large percentage of the population in these countries spend a few hours per day watching television, more than they would ever spend reading books, periodicals or newspapers. Due to the population moving out to the villa estates outside the large towns, and because it is getting more difficult by the day to have newspapers delivered there, news programmes on television are constantly gaining importance.

However, the television watcher has not the free choice he has as with books, periodicals and newspapers: in the European countries his choice is limited to two or three programmes to which he must pay his subscription. It is particularly for this so very important mass medium that the government refuses to leave the production of programmes in the hands of commercial stations as is done in U.S.A. In the Netherlands, therefore, the money received from subscriptions and advertizing is divided between a number of politically or religiously orientated broadcast companies in proportion to the number of members who subscribe to the corresponding programme time table. Too much dispersion of talent and finance will thus endanger the quality of the programmes.

Since 1963 colour television has first conquered the U.S.A., and then Europe in 1967; a still better source of information, education and entertainment. It is surprising that this so highly complex technical equipment operates with reasonable reliability, and that in spite of a few inherent limitations, the quality of the picture is quite good. The compatibility (see Chapter III.C.5) of colour and black and white television ensures a gradual change over. Transmitters and link transmitters require no special matching; however, in the studio many technical problems had to be solved in connection with camera and lighting.

Fig. 249 shows the growth of the number of television receivers used throughout the world. Since 1965 the increase of the number of black and white receivers is on the decline, a fact which is more than compensated for by the growth in number of colour television receivers. The graph illustrates that the increase in number of colour television receivers in the year 1969/70 exceeds that of the number of black and white receivers when in the year 1954/55 these had reached the same number.

In Europe, U.S.A. and Japan, the link transmitter network (operating with a frequency of about 4000 MHz) has brought an excellent solution of spreading the television programme over the entire country. Even in a great area of that enormous country Brasil a link transmitter network for television and telephony has now been put into operation. Between U.S.A. and nearly the whole earth the transmission of television signals takes place via satellites. For extensive countries like India it is being studied what possibilities there are of broadcasting television programmes via a "national" satellite. Before that can be done, a cheaper solution for the receiving station must be found.

The video tape recorder has made it possible to store television programmes to be reproduced at some other time. This has even led to the development of a new film industry (but then on magnetic tape) which supplies television programmes to other countries who are not in a position to produce their own. Synchronization

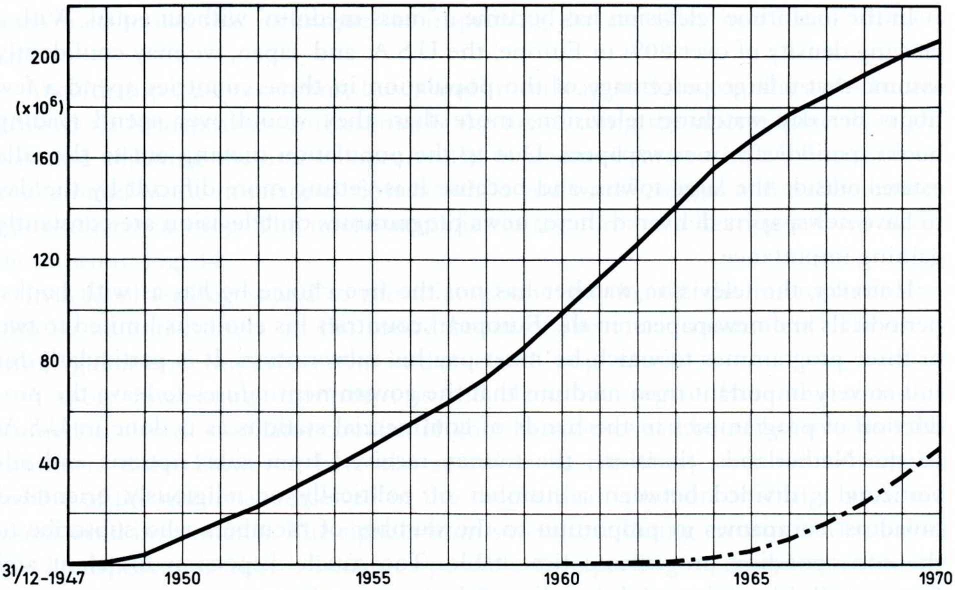


Fig. 249. Growth of the number of "black and white" (solid line) and "colour" (dash-dot line) television receivers used in the world (excluding China, Russia and Eastern Europe) (estimate made by Philips C.V. & P.).

in another language forms no problem. It should be borne in mind, however, that here television relinquishes some of the great advantages it has over cinema, to wit: topicality and originality.

## IV.A.4. CLOSED CIRCUIT TELEVISION

The difficulties involved in the transmission of television signals through the air have led to investigation of the possibilities of distributing television programmes via wires as is done with radio relay. The General Post Office of the Netherlands is studying the C(entral) A(erial) S(ystem) with the idea of receiving the regional and perhaps also the foreign programmes with a high aerial. After amplification these programmes will then be led to the watchers via a coaxial cable. It is hoped that this expensive cable can be used for other purposes (video phone?) too.

Closed circuit television has already found wide use for educational purposes. It can be used to follow a medical operation with a colour camera and to have it watched by a large audience outside the operating theatre. In the dental surgical clinic in Utrecht students can be instructed by letting them follow the treatment per television. With the EIDOPHOR, a large-screen projector of special design, the picture can be displayed to a large audience.

Although it is outside the scope of "broadcast", we would here like to mention the application of closed circuit television for surveillance. The gate keeper can watch the factory yards, the station master the marshalling yard, the police the traffic throughout the town by means of cameras set up in many places. The boiler house guard can monitor the image of the glowing hot fires; in nuclear installations it can be observed what is happening in places where man would be exposed to the danger from radiation.

A most useful application is the transmission of announcements to the public. In railway stations and airports all information concerning times of departure and arrival of trains and planes can be displayed on a TV screen in various places by selecting the corresponding switch. What a simplification as compared with the complex mechanically operated switch boards that are still found in many stations!

A number of firms are working on the VIDEO PHONE, a system where a television image is received simultaneously with a telephone call. By accepting less definition it might be expected that the existing telephone network will be suitable for the new system.

# SECTION B. TELECOMMUNICATION (FROM ONE TRANSMITTER TO ONE RECEIVER)

## IV.B.1. TELEGRAPHY, TELEX

The morse telegraph is practically only used by the railway companies where the staff can read morse. The new form of telegraphy is the high speed telegraph or TELEX, where use is made of a special pulse-code which is converted into letter symbols on an electric typewriter. For radio telegraphy over great distances an engineer of the Dutch General Post Office (PTT), VAN DUUREN, introduced an important improvement: the T(eletype) O(n) R(adio) system. If owing to atmospheric interference or fading a signal is not properly received, the receiver informs the transmitter of the fact, so that the letter symbol is automatically repeated. As a result high speed telegraphy will often work in difficult conditions where telephony fails.

The telephone service has made an important addition to telephone communication by hiring out telex receivers and transmitters to subscribers who can now send telegrams via telephone cables to any other telex connection in the world. The message is recorded in writing and reaches the addressee even if the latter is not present the moment it arrives. Thanks to a refined automatic system the telephone service saves the telegrams for the hours (during the night for instance) when the cables and exchanges are not busy. The bill is made up automatically. It stands to reason that many firms make use of this facility.

On the other hand, the public telegraph is finding less and less use. Delivery of a telegram has become an expensive matter!

Different methods have been worked out to transmit black and white figures or drawings per telegraph (FACSIMILE). Since the arrival of television there has been very little interest for this technique.

## IV.B.2. TELEPHONY

The telephone, supplemented by telex, is undoubtedly the future means of conveying messages. If we consider how much, due to increased wages, the stamping, sorting and delivery of a letter costs, including the selling of the stamp, and added to this the time the sender needs to write, address and post the letter, we must come to the conclusion that the telephone saves a lot of time and money.

Once the connection, cables and exchanges are there, no more wages need be paid to third parties (even the bill is made up automatically).

Fig. 250 gives a survey of the number of telephones used in the world. To this it should be added that in various countries there are long waiting lists for new extensions. In the U.S.A. the number of one telephone apparatus per family has been amply exceeded. In Europe and Japan the number of subscribers is rapidly increasing.

The telephone service will endeavour to turn the existing extensions to the

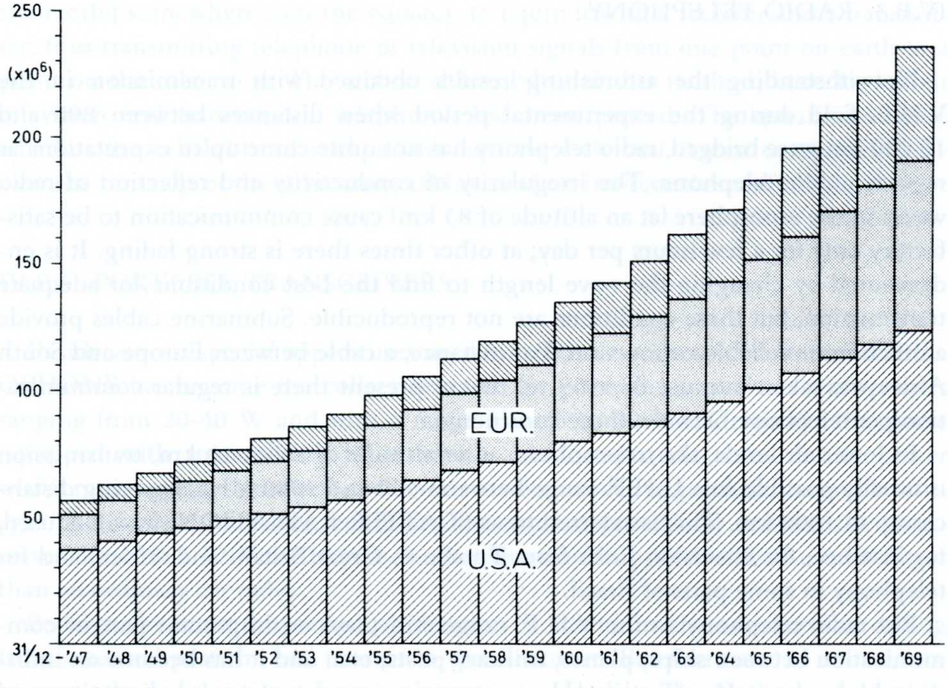


Fig. 250. Growth of the number of telephones used in the world (excluding China, Russia and Eastern Europe) (estimate made by Philips C.V. & P.).

most efficient use. This can be done, for instance, by using the telephone to obtain various kinds of information such as the time of day, weather forecast, news, etc., which are recorded on tape and can be heard through the phone after dialing a certain number. The subscriber can also make use of a tape recorder connected to his telephone apparatus so that during his absence a certain message is spoken, or a call is recorded. It has even been tried to receive radio programmes per telephone, but the sound quality proved insufficient.

Furthermore a subscriber can be connected to a central computer. The Dutch General Post Office (PTT) has announced that within short it will be possible to have "giro"-bookings take place per telephone. A method of meter reading for electricity, water and gas supply via telephone lines is being investigated.

Automation of long-distance telephony has led to a highly intensified long-distance communication. In Chapter I.23 it was already explained how it became possible to lead a large number of calls via one pair of wires in an existing cable. In the U.S.A. the telephone system has been automated throughout the country. This also applies to some great parts of Europe.

Where cables and exchanges tend to get overloaded, the telephone services endeavour to limit the duration of calls made. Therefore the international rate is based on a duration of 1 minute. With such short calls the time needed to dial the number (which can exist of 15 figures), begins to play a part. Electronic dialing and exchanges (Chapter I.23) save a lot of time.

Indeed it will be possible within reasonable time for any subscriber in the world to dial any other subscriber automatically at reasonable rates for short calls.

### IV.B.3. RADIO TELEPHONY

Notwithstanding the astonishing results obtained with transmission in the V.H.F. field during the experimental period when distances between 800 and 10 000 km were bridged, radio telephony has not quite come up to expectations as regards public telephone. The irregularity of conductivity and reflection of radio waves in the ionosphere (at an altitude of 85 km) cause communication to be satisfactory only for a few hours per day; at other times there is strong fading. It is endeavoured by changing the wave length to find the best conditions for adequate transmission, but these conditions are not reproducible. Submarine cables provide a much more reliable connection. For instance, a cable between Europe and South Africa was taken into use in 1969, so that at present there is regular communication per telephone between these continents.

In lower air levels, the troposphere, at an altitude of about 10 km, transmission is usually possible in a U.H.F. range between 500 and 2000 MHz, covering distances up to 1000 km. This is referred to as SCATTER RADIATION, which is used, for instance, for television links from France to Great Britain and Africa, and for telephony in some parts of Brasil.

But radio telephony in the V.H.F. range still plays an important part in communication between ships, planes, military posts, etc., and it has become an indispensable means of navigation. Here a certain amount of sound distortion and fading is taken into the bargain, so that from time to time a message must be repeated. An international security system keeping watch over the wave length allocation must ensure that there is "order" in the ether.

### IV.B.4. LINK TRANSMITTERS

Directed U.H.F. waves are independent of the often irregular reflections of air levels at high altitudes, and therefore they present a reliable means of radio communication. Particularly in development countries where it is almost impossible to lay cables, long-distance telephone communication takes place via link transmitters. But also in the densely populated regions where there is no "room" for cables, a constantly increasing use is made of long-distance telephone communication "through the air". As discussed in Chapter I.23, as many as 900 channels can be used at 4000 MHz; if the frequency is increased to 6000 and 12 000 MHz respectively, many more channels will come available. The maximum range of a link transmitter is about 60 km, but a greater distance can be covered if the transceivers are built on mountain peaks. It should, however, be ensured that the path of the link beam is not obstructed by towers and high voltage masts. Owing to their great band width, these link transmitters are also suitable for the transmission of television programmes.

### IV.B.5. TELECOMMUNICATION VIA SATELLITES

In Chapter I.24 we already saw that it is possible to send a satellite into a geo-static orbit somewhere over the equator, to equip it with a receiver and a transmitter, thus transmitting telephone or television signals from one point on earth to a point on the other side. The technical feasibility of this method, both for colour television and transoceanic telephony, has been proved. The question is only whether the reliability of such a link will be such that it can completely take over the function of the submarine cables used for intercontinental telephony.

## IV.B.6. PORTABLE TRANSCEIVERS

Beside various special apparatuses used for military purposes, the term MOBILOPHONE has become the collective name for portable transceivers with powers ranging from 20-40 W and with a range from 5 to 10 km. Originally they were equipped with transmitting tubes which consumed quite some heater power whilst rotating generators or d.c. converters were needed to produce the anode voltage. The modern transistor technique has made it possible to connect the transceiver direct to a 12 V car battery, so that the whole apparatus need not be larger than an ordinary car radio.

Perhaps the most important application is found in the taxi service, calling from a central post the taxi nearest to the customer. But there are, of course, numerous other applications. In some countries, including The Netherlands, mobilophones can be hired by the public from the General Post Office, and they can be used in conjunction with the public telephone network.

Transistorization has produced a category of small transceivers which are known as WALKIE TALKIES. Their powers range from 0.5-5 W, they cover distances from 0.5 to 2 km, and are fed from dry batteries. It is now possible for the military forces, police force, railway employees working in yards or on building sites to carry on wireless communication with each other. Thanks to the low power and the wave length band reserved for this purpose, the risk of interfering with other transmitters is but small. Such a transmitter can be made so small, that the singer in the show can hide it in her (sometimes diminutive) dress, and whilst dancing unimpeded by the microphone cable she can transmit her singing to a stationary receiver which passes it on via loudspeakers in the hall. Finally we can mention the wireless BABYPHONE.

## IV.B.7. PAGING SYSTEMS

A paging system does not require the wanted person to be spoken to, but only that he receives a signal. The doctor in a hospital, the supervisor in a factory or in the yard carries in his pocket a small piece of equipment that produces a buzzing sound when called up by a code transmitted from the central post. On hearing the signal meant for him he goes to the telephone.

Such a system, but then on a large scale, is supplied by the Dutch General Post Office under the name of SEMAPHONE, a receiver that can be called up as far as 100 km from the central post. A cleverly designed code is used to select a certain apparatus from several hundreds. A simple message can even be conveyed by means of a few signal lamps fitted in the receiver.

## IV.B.8. INTERCOM

By intercom is understood an installation with which, by pushing certain buttons, a chief can communicate in word with his staff elsewhere in the building. In the simplest version the loudspeaker is also used as the microphone.

A variety of this system are the installations where members of a meeting have microphones in front of them which enable them to use the loudspeaker installation. (When Parliament is in session we notice that such an installation is eagerly used for "interruptions").

Still another variety are the SIMULTANEOUS TRANSLATION INSTALLATIONS where interpreters speak the text of the speaker, "simultaneously" translated in another language, into the microphone. By means of a switch the participants at the meeting can select the language of their choice. It is not probable that the "electronic interpreter" will be realized. A dictionary can be written into an electronic memory to be read out again with "random access". For the time being the human mind is indispensable for the interpretation of thoughts expressed in spoken sentences.

## SECTION C. ELECTRONIC RECORDING

### IV.C.1. GRAMOPHONE

The reproduction quality of a gramophone has improved substantially during the development from Edison's wax roll to the modern LP-record. The latter contains about twice thirty minutes of music, with a frequency range from 20 to 18 000 Hz, if required stereophonic, with hardly any noise. This is due to the better material used for the record, the improved method of reproducing from the recorded magnetic tape to the press via the master die, and the light-weight pick-up with a diamond stylus. The latter has a curvature radius of 0.02 mm, and is hardly subject to wear. The record player too, must meet stringent requirements: the speed must be absolutely constant, the stylus pressure constant and rather low, whilst the so-called pick-up must have a linear characteristic. By making use of the vertical and horizontal movement of the stylus it has been made possible to achieve stereophonic sound reproduction. In addition the musical level of the records has improved considerably, with the result that it is now possible to enjoy a "concert in the living room".

During recent years the so-called HiFi installations, consisting of a high quality record player with amplifier and loudspeakers, covering the whole frequency range without distortion, have become increasingly popular. This equipment can also operate in conjunction with an F.M. radio set or a tape recorder. The necessity of having the loudspeakers positioned a few metres apart for stereophonic reproduction has resulted in these installations being formed of separate units that are placed in a book-case or wall cabinet.

Beside these expensive HiFi installations, also the simple electronic record player has been subject to considerable development. Both the motor and the amplifier can be supplied from dry cells, so that these installations have become portable.

Putting on a gramophone record is still a complicated manipulation in which the record can easily sustain damage. A solution is found in using automatic record players.

An advantage of the gramophone is the great production capacity of the press which can produce a record within a few minutes. Consequently a new "hit" can be reproduced by the hundreds of thousands within a few days.

### IV.C.2. MAGNETOPHONE (TAPE RECORDER)

Initially the tape recorder was only used for the recording and reproduction of sound. The first applications took place in the broadcast studio and for "voice filing" (of conversations, for instance, between air traffic control and aircraft). It was not long before the photographic sound tape on the "sound film" was replaced by the magnetic track ("synchronization" in another language was thus greatly facilitated). Then came the applications for speakers and musicians, and the use as a dictation machine. As these apparatuses became cheaper, they came within the reach of amateurs. This led to a new form of amateurism, known as "sound hunting". Finally the tape recorder has also become a useful means in language teaching: "the language laboratory".

The tape recorder, combined with electronic amplification of sound, has brought a new kind of music: ELECTRONIC MUSIC. Here we are not limited by a certain tone scale, and the "timbre" can be selected to taste; it is possible to reproduce extremely low and extremely high notes, and to have the notes follow each other in rapid succession. This application has, however, resulted in mainly extravagant music.

The originally used reels with magnetic tape entailed the drawback of requiring some skill in handling. A solution was found in the form of "music cassettes", in which the tape is automatically wound in both directions. They can simply be slid into the tape recorder. Then it became practicable to supply cassettes complete with recorded music. They can replace the gramophone record and have the advantage of being very small, easy to load, with a better sound quality and less wear. Furthermore it is easier to skip a piece of music if required. Since the tape is played at the maximum speed, at which good sound quality is guaranteed, reproduction of several tapes from the master tape must take place at about normal speed. Experiments have been carried out with magnetic contact methods which allow reproduction at considerably higher speeds. It is a pity that, same as with gramophone records where the 45 and 30 r.p.m. records stand side by side, different types of music cassettes have come into existence, requiring different types of tape recorder.

The part that tape recorders play in the recording of data in the form of figure or letter symbols in conjunction with computers has already been fully discussed (Chapter II.M.4).

## IV.C.3. VIDEO RECORDER

By using a still finer powder, an even smaller magnetic head, a greater tape speed, and making more efficient use of the tape area (parallel oblique lines) it has been found possible to record the video signal on magnetic tape. A video recorder, originally known by its make "Ampex", has been used for years in television studios. At present industry is bringing out a cheaper version which is used in the semi-professional field (sports, theatre, ballet, etc.). The recording and play-back equipment costs 600 dollars and a tape with a playing time of 45 minutes (450 metres long!) 30 dollars. Expectations are that the price of less complex equipment and a simple television camera will soon come within reach of the amateur, thus becoming a competitor of the 8 mm filmcamera.

Furthermore the "home colour video recorder" is waiting to step in. Beside the magnetic tape, experiments are going on with gramophone records and chemical-photographical reproduction methods. Thus ultimately the "magic lantern" in the nursery will be replaced by the video projector.

## SECTION D. DIRECTION FINDING AND LOCATING

### IV.D.1. ULTRASONIC

Introducing an ultrasonic vibration (Chapter II.K.6) in water is one of the simplest methods of "seeing under water". A solid body will reflect this vibration. The time elapsing between transmission and reception of the vibration provides a measure of distance. This method is used, for instance, to sound the depth of water (ECHO SOUNDER), to locate reefs, icebergs, submarines, and even schools of fish. This method is not, however, very accurate for direction finding.

### IV.D.2. MAGNETIC

A magnetic field can be used to find objects that conduct magnetism in a nonmagnetic environment. The magnetic field is generated by an alternating current through a coil, and a magnetic conductor in this field will cause the impedance of the coil to increase, which can be indicated by means of an electric measuring instrument. This is the principle on which the "mine detector" is based, with which mines buried in the ground can be detected at a distance of a few metres. A recent application is "searching"air-travellers for fire arms, bombs, etc. During the sawing or peeling of tree trunks, a metal object hidden in the wood might cause serious damage to saws or blades. A magnetic "metal detector" can be used to indicate the presence of such metal objects; an indispensable means in cutting trees into which shrapnel has been blasted during the war. Of course, there are various other applications.

### IV.D.3. RADIO

If a radio receiver is connected to a frame aerial, the received signal will be strongest when the frame points in the direction of the transmitter. By taking "bearings" on two transmitters it is thus possible to pin-point one's own position. Obviously light beacons, which cannot be used in dense fog, are replaced by radio beacons.

In the beginning of the war from 1940-1945, systems were developed (Decca, Lorran, etc) which laid a cross-pattern of radio beams all over Europe. Sometimes also use was made of phase shift between the transmitted signals. In this way it became possible for aircraft, even over enemy territory, to determine their positions quite accurately.

If the transmitter produces a directed beam, location becomes much more accurate than with a loop aerial. This is the principle of the "approach beacons" on airports which enable the pilot to find the accurate line he must follow to arrive safely at the airport.

In Chapter I.24 we already mentioned the "PROXIMITY FUSE", which by means of a built-in transceiver can cause projectiles to explode when they approach an aircraft or the ground. In modern warfare projectiles finding their own way to the targets (MISSILES) play an important part. The Russians built such installations on Cuba and along the Suez Canal. In the U.S.A. millions of dollars are spent on the A(nti) B(allistic) M(issile) system of which it is hoped that it will put enemy long-range missiles out of action long before they can reach the populated areas.

## IV.D.4. RADAR

Originally RADAR was only used to measure distances. During the battle of Britain in 1940-1941, the British had erected aerials along the Channel coast which radiated across to the French-Belgian coast. The take-off of German bombers was marked by a blip on the screen, and the distance thus determined allowed British fighters to take off in time to engage with the German squadrons before they could reach London.

With the rotatable parabolic aerial for cm-waves it is possible to determine the direction of the target in both the horizontal and the vertical plane. Refined electronic systems can ensure that the aerial, once it has found a target, will continue to follow it, whilst at the same time the anti-aircraft guns remain trained on the target.

Finally the P.P.I (Chapter I.12) enables us to "draw charts" of the environment round the radar station by using a "long persistance" screen of the cathode ray tube in which the luminous line, measuring the distance, rotates synchronously with the aerial. With high powers and longer waves (25 cm) the area displayed can measure tens of miles across; with mm-waves (harbour radar) it is possible to obtain a more accurate picture over distances covering a few hundreds of metres (Fig. 18c).

Another method to obtain a picture which can be used for reconnaissance or "mapping" is SIDEWAYS LOOKING RADAR.

# SECTION E. SAFEGUARDING AND TRAFFIC CONTROL

## IV.E.1. BURGLAR AND FIRE PROTECTION

As a means of fire detection we can use a temperature relay (say an NTC resistor, Chapter II.K.5) to signal the temperature increase or, alternatively, a photo cell might be used to detect smoke development. These elements can put a claxon or spraying installation into operation. The fire insurance companies are justified in reducing their premiums drastically in cases where such systems have been installed.

The increasing number of burglaries call for a burglar protection system slightly more refined than the well-known contacts on windows and in floors. Electronics has failed so far to produce an absolutely fool-proof system. Experiments are going on with photo-cells and installations that are capable of signalling moving persons and objects. This is done by using the "DOPPLER"-effect; in the frequency range from 20-100 kHz with the aid of piezo-oxide transducers, or in the frequency range 1-18 GHz with GUNN-oscillators.

## IV.E.2. ON THE WATER

In preceding chapters it was already explained how much the possibility of radio communication with other ships and the shore has improved safety at sea. The same applies to radio beacons. Radar should prevent collisions between ships at sea sailing busy lines, even during the night and in dense fog. Yet such collisions will occur.

We must not forget to mention some radar safety systems as installed, for instance, along the Nieuwe Waterweg near Rotterdam, the Elbe near Hamburg, and the St Laurens Seaway in Canada. Here a whole chain of radar stations runs along these busy water ways, to register the positions of the various ships. By means of the mobilophone the ship can be instructed what necessary changes in course to make. Better still, the radar picture can be transmitted to the ship which then need only be equipped with a television receiver much cheaper than a complete radar system.

## IV.E.3. RAILWAYS AND TRAMS

The block system in which a certain section or "block" of railway track is indicated as "safe" or "green" only after the previous train has left that particular block, the pointsman, and the gatesman who receives warning when the gates must be closed, are gradually being replaced by automatic electric systems. The approaching train automatically closes the gates of the level crossing, and can automatically be stopped when the signal is red. All points and signals of a railway yard can be operated from a central point, and the positions are displayed on a light panel.

Not until the whole traject has been prepared is it possible for the entrance signal to change to "green". Originally these systems were all of the electromechanical type, but here too, electronics is taking over.

In underground metropolitan railways (METRO) the trains are often controlled fully automatically. The speed of the train along each part of the circuit, including stopping and starting is programmed, same as the switching of points. Only one driver-conductor takes action in case something should go wrong.

## IV.E.4. HIGHWAY TRAFFIC

Freedom on the road with the only restriction of "keep left" (or right as the case may be) has long belonged to the past. At first policemen were put on point duty, but they were later replaced by traffic lights. The duration of the light signals is nowadays often determined in a police post on the basis of monitoring via closed circuit television. This system is, however, too rigid to ensure the maximum flow rate. Therefore the "electronic policeman on point duty" has come into existence. The lengths of the traffic files in both directions are measured, and the computer calculates the most favourable durations for red and green. Furthermore one way traffic can be accelerated by means of the "green wave", where the next traffic lights change to "green" as soon as the file from the previous lights approaches. It is obvious that the definite solution to the problem of ensuring traffic-flow on busy crossings is found in the form of fly-overs, but: electronics is cheaper.

It should be noted in this connection that there is such thing as radar speed control, sometimes operating in conjunction with a system where the offender is photographed.

Technocrats dream of motorways on which cars travel at a prescribed speed and mutual distance along an embedded conductor. But why not travel by train then?

## IV.E.5. AVIATION

For aircraft travelling through the air at speeds of 1000 km per hour or more, electronic communication is of vital importance. The electronics in a modern fighter plane costs more than the propelling installation. Radio transceivers for different frequencies ensure that the aircraft can remain in constant contact with at least one airport. Radio direction finders and radio beacons are of great assistance in determining the course; "air borne radar" gives an orientation as regards obstacles in the vicinity. The gyro compass will be replaced by the electronic compass which can guide the "automatic pilot". For military purposes there are many more instruments.

Large airports are equipped with a radar system capable of following all approaching craft from distances of several hundreds of miles (SATCO). Once the aircraft has been spotted, the computer calculates its position from its speed and course until the outcome is corrected by a next radar sounding. Thus the air traffic control tower has a complete survey of all approaching aircraft and can arrange various successive touch downs.

"World wide" computer systems are used to book reservations on planes. At the airport passenger and luggage handling is dealt with electronically (AIRLORD). Perhaps one day we shall enjoy an automatic system that will take the luggage to the exit simultaneously with the passengers.

# SECTION F. ELECTRONIC MATERIAL WORKING

## IV.F.1. ELECTROLYSIS, ELECTROPHORESIS

Electrolysis is based on the fact that in solutions of acids, bases and salts in water part of the molecules split up into IONS which are positively or negatively charged. The ions of metals or hydrogen carry a positive charge, whereas hydroxyl and acidic ions are negatively charged. If two electrodes connected to a direct current source are placed in a bath containing such a liquid, the positive ions will flow towards the cathode whilst the negative ions move in the direction of the anode. If the liquid contains metal ions, these will settle on the cathode. If the anode contains the same metal as the solution, the negative ions reaching the anode will combine with that particular metal thus maintaining the concentration of the solution. In this way metal can be transferred from the anode to the cathode. This method is generally applied for copper, nickel, and chromium plating etc. of metal objects.

The same process is used for ELECTROLYTIC DEGREASING, where in an alkaline solution hydrogen or oxygen ions move towards the object to be degreased under influence of the electric field, to remove the grease under intensive gas development. If a metal surface is used as an anode in a solution of aggressive acids, we speak of ELECTROLYTIC POLISHING.

Electrolysis of water is a method of producing very pure hydrogen and oxygen. Alloys of (light) metals in molten conditions can be separated in the electrolytic furnace. This process is used, for instance, to prepare aluminium from the oxide.

In Chapter II.B.2 we already mentioned the CATAPHORESIS method of coating the heaters of directly heated cathodes with an emissive layer, and providing the heaters of indirectly heated cathodes with an aluminium oxide coating. Use is then made of the phenomenon that insulating particles in a suspension or emulsion in a liquid are charged in an electric field which causes them to travel to one of the electrodes. In this manner a metal object to which a negative voltage is applied can be provided with an insulating coat. An advantage of this method is that the sharp points and edges, which hold little powder in spraying, are particularly well coated owing to the high field strength occurring there.

For some years ELECTROPHORESIS has also been used for lacquering. The advantage over electrostatic spraying (Chapter IV.F.2) is that a lower voltage is used and that the solvent is withdrawn from the coat of lacquer. Furthermore the already existing layer of lacquer prevents the flow of current in these places, so that paint particles are also carried to less accessible places (crevices etc). With this method it seems even possible to paint car bodies, complete with doors, bonnet, and boot cover (the glossy top layer is applied by spraying on cellulose lacquer).

In the abovementioned process the current can in principle be supplied by d.c. generators. It is obvious that the application of (controlled) electron tube or silicon diode rectifiers, and particularly the possibilities of measuring during the process and using the measured values for automatic current and time control, has led to an extensive use of electronics in this installation.

## IV.F.2. ELECTROSTATIC

In an electrostatic field particles in a gas will be charged to be attracted by the nearest electrode. This forms the basis for electrostatic cleaning of fumes and gases. The fumes or gases are led between two electrodes to which a high d.c. voltage is applied. The dust particles in the gas will settle on the electrodes. If the material is to be recovered, it can be taken off the electrodes in a clean and dry condition. This method can also be used for giving objects (e.g. cigars) a uniform coat of powder. Also XEROGRAPHY, used in certain copying machines is based on the same principle.

In lacquering, the lacquer is often mixed with air when it is sprayed on the work piece. A large quantity of lacquer is then lost because it misses the work piece and must be caught in a water curtain to prevent fire hazard. When the lacquer, before leaving the spray gun, is led through a metal gauze at a voltage of about −70 kV, the particles of lacquer will be charged and attracted by the work piece so that a great deal of lacquer is saved. When depositing metals in a vacuum, metal is evaporated in the vacuum space (usually by heating). The metal vapour will then be attracted by an object in the vacuum which carries a direct voltage. As a result the object is coated with a uniform film of metal. This method is used, for instance, in the manufacture of foil capacitors as described in this book.

## IV.F.3. ARC, PLASM ARC.

The electric arc discharge in air formed the basis of the first electric light source, the ARC LAMP. In Chapter I.4 we already mentioned the application in ARC TRANSMITTERS, based on the oscillatory character of an arc discharge. In an ARC FURNACE alloys of high melting materials can be so heated that the respective components can be recovered in their pure forms.

During the last ten years a new application of the arc discharge has been developed: the PLASM TORCH. By "plasm" is understood here the combination of free electrons and "ions" which are produced in a gas by ionization in an electric discharge. The electric charges in this "plasm" neutralize each other, but the plasm itself stores a great deal of energy. If it can be brought outside the electric field, the electrons and ions will recombine to molecules whilst a large amount of energy is released, and thus very high temperatures occur. By providing one of the electrodes of the arc with a cooled cylindrical aperture, the heat in the space where the arc is formed, will cause a sharply defined "plasm beam" to shoot through this opening. In this beam recombination of the electrons and ions can give rise to temperatures as high as 30 000 °C. By leading a compressed inert or reducing gas (hydrogen, nitrogen or argon) into the space where the arc discharge takes place, a plasm beam is obtained which has much better characteristics than the acetylene burners used for "gas" welding and cutting. It can be used, for instance, for cutting stainless steel and aluminium plates with a thickness of 50 mm, and also for welding different metals together. Furthermore, high melting materials like tungsten, nickel and aluminium oxide can be fed to the plasm beam in the form of powder; these materials will then melt under influence of the high temperature and settle as a well adhering layer on the object brought into the beam. Such a layer offers excellent protection against wear and corrosion, also at high temperatures. The industry which makes parts for jet engines or space craft often uses this process. The high temperatures of the plasm torch can also be used for producing high melting metals from their combinations, and also for chemical research.

## IV.F.4. WELDING

The electric arc provides us with a means to evaporate metal from the electrodes to settle on the work piece. In this way metal components can be joined without loss of strength; if necessary airtight welds can be made. ARC WELDING has found universal application in ship building, boiler building, and assembling metal constructions. The current source (direct current, in smaller equipment also alternating current) must have a certain characteristic to maintain the arc. The electrodes are mostly provided with an insulating coat to facilitate the formation of an arc whilst at the same time it has a favourable effect on the weld in forming "slag" there.

Another method of electric welding consists of melting two pieces of metal together, the heat being produced by a flow of current through the contact resistance of the boundary layers. Apart from a mechanical pressure on the boundary layer an automatic current control is required to compensate variations in the surface resistance. Furthermore, the current must be supplied as a short pulse to prevent the heat developed in the contact resistance from flowing away through the metal. It is clear that current and time control by means of thyratrons, ignitrons and thyristors plus their respective electronic circuitry play an important role here. SPOT WELDING is widely used in the manufacture of sheet metal products such as automobiles, domestic appliances etc. With materials which are themselves good heat conductors (copper), or form an oxide layer of extremely high resistance (aluminium), it is however difficult to use this process. Then we have SEAM WELDING where the two components to be joined are led through two rollers whilst a train of rapid current pulses is sent through the contact area. Finally two rods or wires can be welded head to head by sending a current through them (BUTT WELD).

## IV.F.5. SPARK EROSION

This is a process in which a spark produced from a sharp metal spike is used to burn away metal. This method is used for making very fine dies which serve to punch components out of sheet metal. The process take time, for slow progress must be made. But the machine itself is not expensive because it is not exposed to great forces. An important advantage is that no burrs are formed as is the case with drilling and milling. A similar process is used nowadays for "drilling" small holes through diamonds which are used for drawing very fine tungsten wire. Now the point of a wire which is connected to a high voltage source is forced against the diamond. As a result the part of the diamond directly below the point of the wire is turned into graphite which burns by heating in air. This method yields a perfect conical entrance of the opening, which in turn will give a perfect wire.

### IV.F.6. LASER BEAM

In Chapter II.E.7 we already read about the "laser" which emits an oscillatory wave of a frequency in the infrared range in the form of a very narrow beam. This beam can be used to make holes in hard materials such as tungsten and quartz, and also to cut out plates of a special form. The $CO_2$-laser with its relatively simple construction and high efficiency has paved the way to industrial application.

### IV.F.7. ELECTRIC FURNACES

Due to the limited thermal efficiency of a steam driven power station and the expenses involved in generating and transferring electrical power, electric furnaces are less economic than oil or gas fired furnaces. However, the possibility of working in a clean atmosphere, (if required in a certain gas atmosphere) and of obtaining very high temperatures has resulted in a wide spread use of electric furnaces. Furthermore the temperature distribution can be influenced by appropriate positioning of the heated elements and by circulating the air in the furnace.

Modern industry often requires a highly accurate process control, and for that purpose electronics (particularly the thyristor with its high efficiency) has created the possibility of accurate temperature control in electric furnaces.

### IV.F.8. R.F. HEATING

A good method of melting metal that is supplied in small pieces is using the metal itself as the heater element. To that end the contents of the crucible are made to function as the secondary winding of a transformer whose primary is wound round the crucible. The molten metal is than not contaminated by the heating element and the eddy currents ensure a uniform heating of the metal. In INDUCTION FURNACES, frequencies up to 1000 Hz were originally used, which were produced in rotating generators. In order to obtain a better circulation of the melt and higher efficiency, still higher frequencies were needed. As early as 1936 Philips put into operation an R.F. generator with one transmitting tube and an output power of 200 kW at 10 000 Hz for the production of magnet steel. Similar installations are also used for other kinds of high-quality steel.

R.F. heating can also be used for heating metal components, as is done for degassing (in electron tubes, for instance), hardening, dehardening and soldering. An advantage is that heating can take place in absolutely clean surroundings, and even in an enclosure in vacuum or in a reducing gas.

Radio frequencies are used to heat insulating materials (DIELECTRIC HEATING). This process is used, for instance, in the wood industry (bending and gluing of plywood), in the plastics industry and in restaurants (magnetron cookers).

Apart from R.F. heating, gluing of plastic material has recently been achieved by causing the particles in the material to vibrate (ULTRASONIC).

## SECTION G. NON-DESTRUCTIVE MATERIAL INSPECTION

With the older methods of material inspection the material must be destroyed. This applies, for instance, to compression and tensile tests, chemical and microscopic inspection and breakdown testing of insulators. Electronics has brought new ways of performing tests and making investigations, without material being destroyed. So it has become possible to test every item of a production series.

### IV.G.1. MAGNETIC

In Chapter IV.D.2 it was explained how the presence of parts of magnetic material in objects of non-magnetic material can be determined because the impedance of a coil increases when iron is brought into the field of such a coil. The example given mentioned pieces of iron in tree trunks. Similarly impedance measurements can inform us about homogeneity, porosity, and even the crystal structure of the metal. If a layer of lacquer on iron is regarded as the "air gap" in a magnetic circuit, the impedance measurement will determine precisely the thickness of that layer. The thickness of a galvanically applied coating on metal can be measured in the same manner.

### IV.G.2. HIGH FREQUENCY

Measuring dielectric losses provides an insight into the humidity and porosity of non-metals. In Chapter II.L.1 it was discussed how the "tan δ" measurement can supply useful information; such as about air enclosures in insulators.

### IV.G.3. X-RAY SHADOW

In spite of certain imperfections as compared with the image produced by light (lower brightness, diffraction and less contrast) radiography (Chapter II.G.1 and II.G.2), has become an important means in material inspection. The interior of a composite product can be observed without breaking it open (see Fig. 61, radio tube). This is also of importance where metal components are pressed into plastic. The same holds for components of heavier metal cast into light metal. Furthermore the air bubbles in castings or the porosity of weld seams can be observed on the fluoroscopy screen or can be photographed. The latter method is even regulation for high-pressure boilers and conduits, for large projectiles and critical areas in welded constructions where leaks or fracture could be disastrous. The equipment required ranges from a simple viewing box with a tube up to 100 kV in which assemblies and light metal castings can be inspected, to apparatuses up to 300 kV which are used to make photographs of iron castings and weld seams a few centimetres thick. Since for the latter applications the object is usually unwieldy, the X-ray installation itself must be transportable.

For greater material thickness the penetration power of X-rays (even those generated at a voltage of 300 kV) is insufficient, and the even harder radiation of a COBALT BOMB is used. The intensity of this source of radiation is but low, and therefore it can take a few hours to make a photograph.

## IV.G.4. X-RAY REFLECTION

For another type of X-ray examination use is made of the reflection of X-rays from a crystal surface. Whereas X-rays reflected by an amorphous material produce a uniform blackening on the photograph, reflection from a crystal surface takes place in preferred directions which become visible as a more or less regular pattern of white dots on a photograph made with a so-called LAUE-camera. Hence the pattern of such a photograph tells us something about the form, perfection and size of the crystals, and even about the internal mechanical stress. Furthermore, minute contaminations will be shown in the photograph. So here we have the possibility of analyzing small quantities of powdered material, whilst the photograph of the mixture or alloy can be compared with that of known materials. Thus the X-RAY DIFFRACTION method has become an important means in chemical analysis.

In Chapter II.L.7 we already saw how important this method is for determining the direction of the crystal planes in quartz of which transmitting crystals are made.

X-rays generated at a lower voltage give rise to a different reflection phenomenon, which is known as FLUORESCENCE. Here too, by using equipment suitable for the purpose, certain information concerning the material under investigation is provided.

## IV.G.5. ISOTOPES

Thanks to modern electronics another means of material inspection has been developed making use of "isotopes". By treatment in a neutron tube or a cyclotron, certain atom nuclei can be brought to another energy level in which the electrons follow different orbits. After some hours, days, or years, this ISOTOPE will assume a more stable condition whilst emitting a certain electromagnetic oscillation which can be measured with suitable equipment. If such an isotope is added to a liquid, it is possible by measuring after some hours or days where radiation occurs, to find out where the liquid has gone. By adding isotopes to water that is used to moisten soil, we can measure how quickly and to what extent this water moves through a plant. By adding isotopes to lubricating oil we can check whether the oil spreads out through the whole engine. These are but a few of the many possible applications.

For completeness' sake we shall close this section by mentioning the ELECTRON MICROSCOPE with which a resolution of 5 Angstrom (with a special type 0.6 Angstrom) can be obtained. And then we have the MASS SPECTROSCOPE by means of which a gas mixture can be analysed by observing the deviation of gas ions in a magnetic field.

# SECTION H. MEASURING AND CONTROL

## IV.H.1. THE CATHODE RAY OSCILLOSCOPE

The OSCILLOSCOPE is a most important contribution of electronics to measuring engineering. The cathode ray (electron beam) can be compared with the pointer of a measuring instrument, but a pointer which can follow the variation in measured quantity at almost infinite speed. The oscilloscope can be used as a recording instrument. For the electron beam can be made to deflect proportionally with time in a direction perpendicular to the deflection caused by the measured quantity. But now the "paper speed" can be a few thousands of kilometres per hour. The image the electron beam produces on the screen can also be recorded photographically, but by using long-persistance screens, a phenomenon occurring but once can be observed for a few seconds. To record periodic phenomena, the time base can be synchronized to the corresponding frequency, so that a stationary image can be obtained of current or voltage variations which repeate themselves several million times per second. In this way it is possible to observe, for instance the current and voltage curves in television, radar and computers. Furthermore, the electron beam can be made to deflect in two perpendicular ordinates by means of two quantities varying at about the same frequency. Thus the interrelation between the two is represented graphically.

The oscilloscope has become an indispensable means in research and development, testing and servicing of equipment in all fields of electronics. In its modern form the oscilloscope can handle a maximum frequency of some hundreds of MHz (special versions 2500 MHz). The screen is rectangular and measures 13 cm diagonally across. The beam width is 0.4 mm (up to 0.1 mm), a post-acceleration voltage of 15 to 25 kV and metal backing of the screen still ensuring sufficient luminous intensity. With a "dual-beam tube" two phenomena can be registered simultaneously. At a lower speed an electronic switch can be used to display simultaneously several phenomena by means of dashed lines. By displaying time-shifted parts of the voltage curve representing a periodic phenomenon, it is possible to record still higher frequencies (SAMPLING). The application of the oscilloscope is not restricted to recording electrical quantities. Distance, temperature, pressure, etc. can be converted into electrical quantities and their respective rapid variations can then also be recorded.

## IV.H.2. TELEMEASURING

Electric measuring has the advantage that the instrument indicating the measured values can be placed at a great distance from the place where measuring actually takes place, the connection being a thin wire or even "wireless". There are various methods of converting different quantities into electric currents and voltages. We can measure distance by means of variable resistors, capacitors or self inductances; temperature by means of thermo-couples or temperature dependent resistors; pressure by means of piezoelectric materials; luminous intensity by means of photo cells; chemical composition by means of electrical conductivity etc. The ANALOGUE method is used to transfer a voltage or current that is proportional to the measured value. The accuracy of the measurement then depends on errors in conversion, in transmission, and on the indicating instrument.

With the DIGITAL method the measurement is recorded as a "number" which is transmitted to the recording instrument in the form of pulses, the instrument itself being mostly a "counter". In fact this is not "telemeasuring" but "telecounting". It is an advantage that the accuracy of the measurement is fully maintained during transmission and recording of the measured values. Another advantage is that the measured values in digital form can be "manipulated". Not only can they be compared with other observations or data stored in a magnetic memory or recorded on a punched tape, but they can also be used for computations on the basis of which the machine or the process can be controlled accordingly (Chapter IV.H.4).

To use this method of "telemeasuring" a number of "pulse givers" have been developed which convert the measured values into pulse series (often in the binary system). The measured number is usually reproduced by peripheral equipment as used with computers; numerical indicator tubes and high-speed type writers, for instance.

## IV.H.3. ELECTRIC CONTROL

Thanks to electric drive it is possible to control an entire installation or plant from one central point. To be able to use thin wires, the currents used for control are usually low and are converted into greater powers by means of relays or amplifiers near the machines concerned. Compressed air or hydraulic control also offers the possibility of controlling great powers with low energies. If the control commands are given in a digital form, the STEPPER MOTOR (Chapter II.M.6) can render good services.

## IV.H.4. AUTOMATION

Telemeasuring, electronic computing and electric control combined under the name AUTOMATION have opened the way to fully automatic process control in industry; daily human brainwork in industry being taken over by machines.

A simple example is the switch-board operator in a power plant. He used to take voltmeter readings, calculated the voltage drop to the consumer from the measured current, and adjusted a variable resistor in the exciting circuit of a generator so that the voltage at the consumer maintained the correct value. This control took place automatically with the B.B.C. electric voltage controls which have already been operating for 50 years. In digital control the measured voltage is expressed in a number of pulses, and so is the current which is multiplied by the pre-set cable resistance. the difference is the voltage available to the consumer, which is compared with a pre-set required voltage. The latter difference is decisive for the command for the thyristor which supplies the exciting current for the generator.

Boiler houses, too, can be controlled automatically. From measurements of steam pressure, temperature and composition of the exhaust gases the computer calculates from a pre-set programme to what extent the blowers must be readjusted, and then it passes on the proper command to them. Automation in the chemical industry is carried out in a similar manner.

An example of rather complex automation is the paper mill. For each stage of the process, where the paper is pressed and dried between rollers, a certain amount of tension is permissible in the paper. This tension is measured, and the corresponding motor is controlled accordingly.

We must also mention automation in metal machining. The drawing of the product to be made on the lathe or milling machine is translated into two ordinates, and is then punched into a tape (NUMERICAL CONTROL). The machine measures the dimensions obtained, compares them with the dimensions recorded in the punched tape, and re-adjusts the tool setting. The programme of the automatic controller can include the maximum permissible "feed" as well as the reduction of the "feed" when the required dimensions are approached.

These are but a few examples. A first requirement is, of course, that the installation must be highly reliable. A mechanic costs more than a machine guard, and if we need a man to see that the machine operates satisfactorily, it will cost just as much. Not to speak of spoilt work pieces and waste! Therefore electronics could be introduced into industry only after the components, and hence the complete equipment, had become sufficiently reliable.

# SECTION J. ELECTRONIC COMPUTING

## IV.J.1. LARGE COMPUTERS

In Chapters II.N.4 and II.N.5, and Section III.E it was explained how electronic computing works. It is astounding how during the last decade this new technique has culminated in such a tremendous increase in computer applications in so many aspects of society.

A relatively simple example is the one of WAGE CALCULATION or just "waging". The memory contains the hourly rate per worker plus fixed allowances and deductions, and his weekly income is calculated from the weekly statement of working hours, piece work and special deductions and allowances.

More frequently found is probably the application of computers in BANKING. The balance is stored in a memory and all mutations presented in the form of checks, if required with calculated interest, are fed to the memory and sent to the holder of the account. The treatment of giro accounts plus interest calculation has become possible thanks to the computer. The Dutch G.P.O. giro now pays interest and is trying to keep pace with the increasing number of accounts by further automation. In 1969 the U.S.A. commercial banks handled twenty milliard cheques. An intensive search is going on to find a method of reading handwritten figures automatically (CHARACTER RECOGNITION), which would render the most important manipulations by man in this complete operation redundant. LIFE INSURANCE companies, too, stand in urgent need of computers, particularly now that profit shares are being paid out. The same holds good for the administration of STOCK and BONDS.

Similarly the computer has found its way to STORE CONTROL of a wide variety of parts and products, with automatic mutations, guarding of current selling and buying orders, guarding minimum stocks and placing new orders. A more difficult matter, but already put into practice, is PRE-CALCULATION per product on the basis of the parts lists and the list of operations (of which unit prices are stored in a memory). The final administrative procedure is then formed by post-calculation and the calculation of PROFITS and LOSSES. Here again the introduction of the computer has resulted in such an increase in the factors involved in calculation and acceleration of communication that a human bookkeeper can no longer follow the pace. Like depth investments in mechanization and automation compelled to concentration in industry, the expense involved in computerization now calls for mergers in the service sector.

The computer memory represents a means of recording and columnizing enormous amounts of information; the pushing of a few buttons provides ready access to any piece of information or column. Hence the plans of computerizing the POPULATION REGISTER, complete with medical particulars etc. Even matrimonial agencies try to supply their candidate with a list of the required young lady with blond hair, between 25 and 30 years old, between 5 and 6 feet tall, etc.

The computer calculates at an enormous rate, so that by "trying", it can find a maximum or minimum, or in other words, find the best solution. A well-known demonstration is where the computer produces all "prime" numbers between 0 and 100 000 within a few minutes. And the computer can also solve difficult mathematical problems, such as differential equations etc. It is therefore widely used in research, development and design.

The computer can "reason" to a certain extent. From a number of data it can "consider" the various possibilities to come up with the best solution. For a few years now the Philips exhibition service had on display a "noughts and crosses" game; the computer is always the winner. The Dutch master chess player EUWE studied chess by means of a computer. In the same field we find "management by computer", where a number of existing conditions are programmed, and the computer tries out all possibilities and consequences which would take the manager months of strenuous thought. Older managers claim that the computer lacks intuition and experience, but the young ones believe that this is a great asset, because "prejudice" is then out of the question.

We shall conclude this, by no means complete, summary of possible computer applications by saying that, of course, economy sets certain limits, also to programming.

The demand for computers that can perform highly complex computations with large numbers had led to very large, and consequently very costly computers. As a result, endeavours were made to draw full profit from this expensive piece of equipment by having it operate very quickly. Consequently their was a fierce battle between competitors from 1965 to 1970, during which the time per operation (bit) could be reduced to a few nanoseconds. To draw full benefit from this computer, numerous subscribers were connected to it by telephone lines (see Data Transmission, Chapter IV.J.3).

## IV.J.2. SMALL COMPUTERS

In contrast with the development described above (computers continually growing larger and faster), there is also a tendency towards smaller computers which are installed on the user's premises. This has become possible because, thanks to the introduction of integrated circuits, these types of computer could be made smaller and cheaper. For industrial purposes the computer must always be available, and it is impossible to wait for the central computer. Many offices, too, wish to have immediate access to the calculated results. Thus a series of smaller computers has come into existence, ranging from the desk calculator to the invoicing machine, and even to office computers which have quite a capacity. Thanks to the modern disc memory they are still reasonably compact. Expectations are that this development will continue, so that we shall know types of computer with a more restricted field of application, but which will also be much easier to programme.

## IV.J.3. DATA TRANSMISSION

There are instances where it is required to process data from various places in one computer. An example is booking reservations on a plane. In every travelling agency it is essential to find out on short notice whether there are still vacancies for a certain flight, and if so, to register the reservations booked by the passenger in the central memory. The connection is usually made via telephone wires which are needed but for a few seconds, but the computer must always be ready to receive and return data (ON LINE). The same holds good for the large computers that can carry out certain difficult operations for a number of distributed smaller computers.

### IV.J.4. PAYING AND RECEIVING

Although computerization has greatly facilitated making giro payments, receiving and paying in cash is also being automated. For years now there have been vendors which take coins. The operation of these vendors is based on the dimensions and weight of the coins. If we wish to pay in bank notes, the machine must be capable of identifying the different notes, which has been made possible by a method of electronic scanning and colour analysis. These payment machines are used, for instance, in the automatic petrol pumps recently introduced in the Netherlands (TANKOMAT).

Furthermore, a few banks have installed equipment known as BANKOMAT, from which account holders can draw a certain maximum amount of money at any hour of the day or night. The system has been well thought out. The account holder receives a card that can be inserted in the bankomat built in a safe. Punched holes in that card represent his account number. To avoid unauthorized use of such a card, a second command is required before the money is paid: The account holder must also select a number combination (known only to him) by pressing a series of buttons. Only then will the machine supply a fixed number of banknotes. To avoid repetitive payment, the bankomat blocks the account number concerned for the duration of 24 hours. Payments are recorded on a punch tape which is fed into the bank's computer the next day. With a computer "on line" this system would be better, and the bankomat could be of a simpler design, but it would be more expensive.

# SECTION K. ELECTRONICS IN MEDICINE

## IV.K.1. DIATHERMY

An excellent means of curing certain disorders (aching muscles, for instance) is heating. In this treatment the heat must often be applied in a place somewhere deeper inside the body. With infrared radiation heat development decreases considerably as we penetrate deeper into the tissue: on the one hand there is absorption by the outer layers, on the other there is the divergence of the radiation. Better results are obtained with radio frequency DIATHERMY, where the part of the body concerned is positioned between two flat electrodes, each connected to the output of an R.F. generator. Since the field produced between the electrodes is homogeneous, penetration is quite good. A power of about 300 W at 30 MHz is adequate to provide sufficient heat in any required spot. A drawback is, however, that fat and bone tissues are heated more than muscles. The magnetic method, where an R.F. current is sent through a cable surrounding the patient's body, appears to be better in this respect. Furthermore there is the MICROWAVE diathermy method where the R.F. field is radiated from a single small electrode, at an operating frequency of about 2500 MHz.

Diathermy is often used in revalidation to warm up the muscles before starting the exercises. The treatment is harmless, and the equipment is fairly cheap.

## IV.K.2. X-RAY THERAPY

X-ray therapy is based on the fact that exposure to X-rays can put a stop to the growth of undesirable tissue cells. The wavelength of the radiation seems to have hardly any effect on the results; although the choice is very important in view of the necessity of exposing the sick tissue as much as possible, and the healthy tissue as little as possible to the X-rays.

For the irradiation of peripheral tissues (CUTANEOUS THERAPY) very soft rays are used which are generated at a voltage of about 50 kV in an X-ray tube with a low-absorbing filter. The source of radiation is placed close to the skin, so that even at low power a considerable dose is administered. However, the intensity reduces rapidly with the distance. As a result of this, and the strong absorption of the soft radiation, the deeper tissues are protected. If, however deeper places must be irradiated, the X-ray tube is removed farther away from the patient. Owing to the smaller divergence of the beam, the intensity decrease per cm depth is then less. A higher voltage (up to 200 kV at 30 mA) is used to obtain "harder" rays which penetrate deeper. Soft rays (which would be absorbed by the tissue closer to the surface where no irradiation is required) are filtered out with a metal filter. A sharply defined beam is obtained by a process of diaphragming. By means of cross-fire or rotatory irradiation the dose required in a certain place can be increased whilst the surrounding tissue is less affected.

Even harder radiation, but of a lower intensity, is obtained by using the radio activity of the Co60 isotope of the cobalt that is built in a COBALT BOMB. Another method is performing an operation to bring small "rods" of the material in the immediate vicinity of the object to be irradiated.

## IV.K.3. X-RAY DIAGNOSTICS

The X-ray picture enables the physician to see what is going on inside the human body without the need for an operation. In this instance the curing effect of the X-rays is not used, on the contrary, measures are taken to ensure that the patient receives the smallest dose possible.

Originally X-ray diagnostics was reserved for the osseous system, and even then mainly for the extremities. Owing to the small thickness of the object, satisfactory results could be obtained with simple equipment with a low operating voltage (up to 70 kV) and a low current, for the immobile object allowed longer exposure times. In dental radiography voltages of 50 kV are used.

For pulmonary diagnostics a low voltage (up to 70 kV) was sufficient to send X-rays through the human chest. Photographing this moving object, however, required a shorter exposure time, and consequently a higher current, all the more so because, owing to the greater distance between object and film, the tube had also to be mounted at a greater distance. This method of examination has proved of particular value in fighting pulmonary tuberculosis. This disease could clearly be indicated by means of fluoroscopy followed by a photograph taken if symptoms were found. The method of SCREEN PHOTOGRAPHY, a later development where the image produced on the fluoroscopy screen is photographed with a small camera, has made it possible to examine large groups of the population. In doing so, the occurrence of this disease can be detected at an early stage. The result is that lung tuberculosis hardly ever occurs nowadays.

Examination of stomach and digestive organs set more severe requirements for the equipment used. The thicker and more ray-absorbing tissue required higher voltages (up to 125 kV). Moving rasters had to be used to filter out stray radiation. By inserting contrast media, the shape of the organs could be better observed, but now the physicians required an instantaneous photograph to record a certain phase of the motion of the stomach. This is done with the "serial cassette".

Even more difficulties are encountered in examining the heart, spinal marrow and brains, but here too, suitable solutions have been found; using special contrasting media, for instance. We may confidently state that nowadays any part of the human body is accessible to X-ray diagnostics.

Sometimes X-ray fluoroscopy is needed during an operation. In Chapter II.F.4 it was explained what advantages the image intensifier, if required in conjunction with television, has to offer.

## IV.K.4. CARDIOGRAPHY AND ENCEPHALOGRAPHY

Around the year 1906, EINTHOVEN succeeded in registering the very low voltages that occur between different places on the surface of the human body, by means of a string galvanometer of his own design (for this he received the Nobel prize in 1924). The CARDIOGRAMS thus obtained give an impression of the functioning of the heart muscle. Later the string galvanometer was replaced by a sensitive loop galvanometer, and furthermore the oscilloscope was used, which offers the possibility of observing the interrelationship between the various voltages ("vector cardiogram"). The cardiograph has become an indispensable instrument in treating heart infarcts. By carefully studying the cardiograms recorded for a few hours at an early stage, it is even hoped that an infarct can be prevented.

Even lower voltages (about 30 microvolts) occur between various spots on the surface of the skull. Electronics has provided means of registering these low voltages too. The ENCEPHALOGRAMS thus obtained do not serve only to study the functioning of the brain, but they also enable the physician to record certain disorders in the patient. The effect of say, sleep and light, but particularly of epilepsy is clearly indicated in these curves. Studying the reactions of astronauts has supplied quite some valuable material. However, a diagnosis of any occurring deviation requires an automatic analysis of a number of diagrams that have been recorded in different places on the skull, and for longer periods of time. Here the computer can be of a good service in columnizing the various deviations and changes. Although research must be continued for a considerable time, it might be expected that with this instrument the neurologist will eventually have an important diagnostic means at his disposal.

### IV.K.5. PATIENT MONITORING

In the preceding chapters nothing was said about the electronic amplification of the sound of breathing, and electric temperature measurement, nor of the measuring of infrared radiation and ultrasonic waves (ECHOGRAM). These, and the instruments discussed in the preceding chapter enable "telemetering" the condition of the patient (PATIENT MONITORING), which, particularly in large hospitals, can be of a great value indeed. A patient requiring almost continuous attention can now be electrically connected to the central post where there is always a nurse on duty. The temperature, the cardiogram and what else need be measured, can be displayed by meters and picture tubes mounted in this central post. Thus the nurse can instantaneously survey the condition of the patient, and take immediate action if necessary.

Although we cannot devote a separate chapter to it, we must mention that there are various methods of activating certain nerves (and consequently muscles) by means of voltages. The range of applications includes high voltages (SHOCK THERAPY), low voltages (MYOGRAPHY), and also the minute voltages which are used to regulate and activate the functioning of the heart (PACEMAKER).

This summary, incomplete as it is, might bear evidence of the great part that electronics plays in medicine.

## SECTION L. VARIOUS APPLICATIONS

### IV.L.1. ELECTRIC LIGHTING

The classic method of controlling lighting with incandescent lamps is by means of a variable resistor or transformer (Chapter II.M.7). Nowadays it can be done with thyristors. Controlling gas discharge lamps is somewhat more complicated: here the current (or the time during which the current flows) must be regulated, but the ignition voltage must be maintained. Simple equipment with thyristors has been developed with which, by means of a small variable resistor, the luminous intensity of several fluorescent lamps can be controlled.

Fig. 244c showed the diagram of a part of the "transistor ballast" with which a fluorescent tube can be fed from a car battery. Similar installations are also used for the interior lighting in trains.

As the requirements, which the ballasts of gas discharge lamps must meet, are becoming more severe, electronic circuits are being used here too. The "starter" of the fluorescent tube can be replaced by a thyristor; as yet the price is, however, a bit too high. The possibility is being studied of replacing even the choke coil, which is essential because of the negative characteristic of the gas discharge, by a thyristor circuit.

By using a photometer it is, of course, quite well possible to regulate the operating voltages of lamps by measuring the luminous intensity. This principle is also used for the lighting of traffic tunnels, where the illumination must be made to match the prevailing daylight conditions.

### IV.L.2. PHOTOGRAPHY

The photographer or film producer must adjust his diaphragm aperture in accordance with the illumination of the object. Serious photographers use a lux meter, but that takes time. In automatic cameras a selenium or cadmium sulphide cell (Chapter II.F.1) supplies a current which is proportional to the illumination, and which adjusts the diaphragm. Just mounting a photocell at the front of a camera is basically incorrect; only the illumination of the object should be measured, and not that of the surroundings. In the reflex camera a small photocell can be mounted behind the lens. If correctly positioned, this cell will then measure the average brightness of the entire image.

When making enlargements, the photographer of the old days had to try and find the correct exposure time by means of strips of photographic paper: one negative being darker than the next. Here too, the darkroom photometer brings the solution.

Another application is the "ELECTRONIC FLASH UNIT", in which a capacitor is charged from a battery by means of a converter. When the shutter is operated, this capacitor is discharged across a gas discharge lamp so that a short flash of light of great intensity is produced.

Special equipment has been designed for automatic focussing of projectors. Owing to play in the cassette, and also because the film assumes a slightly convex shape, there are always minor positional errors between film and lens in the now so popular automatic slide projector. This, combined with the wide aperture of the objective used, soon results in unsharpness of the projected image. An electronic device makes the necessary readjustments until maximum sharpness is obtained. Even this rather complex technical novelty has found its way to the amateur market. In most of this equipment the distance between photographic layer and objective is measured, and the objective is adjusted to need. Nevertheless it is basically possible to have the automatic control system influenced by the unsharpness itself.

## IV.L.3. THE HOUSEHOLD

It might be interesting to see how far automation has found its way into the household. The temperature of the electric iron can be adjusted to a value best suited for the material to be ironed. The refrigerator and the cooker assume temperatures that have previously been set. The electric cooker or ring is switched off after a preselected time. The speed of modern mixers can be controlled within wide limits. Toasters can be adjusted to "light or deep" brown. In the future, kitchens in private homes will be equipped with magnetron cookers in which a deep-freeze meal is heated to the right temperature within one minute. The automatic washer runs down its programme without human interference. Temperature, time and motion are so controlled that the maximum effect for each type of laundry is ensured. The same applies to the automatic dish washer. The telephone enables a housewife to order her supplies without having to leave the house, and the babyphone warns her when the baby wants his milk. The postman drops a loose-leaf cash book in her letter-box. Further development of low-cost electronics will bring future improvements. The parts played so far by relays and bimetals will be increasingly taken over by integrated circuits.

## IV.L.4. THE MOTORCAR

The application of electronics in motorcars is progressing but slowly.

There are devices on the market that switch on the parking lights when it gets dark. As yet, no satisfactory solution has been found to the problem of "glaring" and "dipping". Automatic dipping devices based on a photocell did not give satisfactory results in practice.

N.T.C. resistors are used for measuring the temperature of the cooling water, but usually a pilot lamp operated by a bimetal contact is deemed sufficient. The bimetal contact is also cheaper to function as the "thermostat" (reducing the water circulation at low temperatures) and for switching on the "automatic fan" (when the temperature of the cooling water exceeds the permissible maximum). Thyristors are used for screen wipers that give a stroke at set intervals.

The fuel level is measured with a potentiometer contact that is governed by a float. One might wonder whether measuring the capacitance between two electrodes mounted vertically in the tank would not provide a more accurate reading (the dielectric constant of the fuel being higher than that of air). The same principle could be used to indicate on the dash board the oil level in the sump, instead of having to use the conventional rather primitive "dipstick".

Shops are selling electronic revolution counters consisting of a millivoltmeter in series with a diode rectifier and a capacitor for measuring the frequency of the ignition pulses. The speedometer and counter, however, are still driven by means of an inaccessible gear transmission somewhere under the car, and a bowden cable. Would not it be possible to obtain a more reliable and cheaper solution by means of a magnet on the crank shaft and a fixed reed relay? A simple electric wire would then lead the current pulses to the instrument panel with a frequency meter (like the speedometer) to indicate speed, and a pulse counter to indicate the milage done. If all abovementioned measured values were to be transferred in a "digital" form, they could be indicated by one instrument, e.g. numerical indicator tubes, switched to the required reading. The readings can even be "stored" in a memory, and used for calculating, say, the fuel consumption per mile.

Methods are being studied for electronically controlling the braking power as a function of speed, in order to prevent "jamming" of the brakes, and also to select the most suitable "gear".

The most important application of electronics would certainly be the ignition system. A powerful spark is needed for starting a cold engine (low number of revolutions). At a high engine speed the spark voltage should not be so high that the electrodes of the spark plugs burn away, but the spark must be produced at the exact instant to ensure the most complete combustion of the gas mixture, thus drawing maximum power. The accessory trade is offering "transistor ignition systems" which amplify the breaker current. But even then the contact breaker remains a part requiring constant maintenance, whilst it also requires an expensive gear drive. A sensor (as mentioned in Chapter II.N.4) mounted on the crank shaft is a reliable and inexpensive method of generating a number of voltage pulses proportional to the number of revolutions made by the engine. It is then possible to alter the voltage curve of these pulses to such an extent that, in combination with the ignition coil, the most favourable form of the ignition voltage is obtained at any speed. Moreover, the ignition "timing" (advanced or retarded) can then easily be "set" electronically. Thus the basis has been created for "automation" in motorcars.

The most favourable conditions as regards air inlet, fuel supply and timing can be calculated from temperature, engine speed and accelerator position, to be adjusted accordingly. BOSCH have developed such a system for engines with "fuel injection". The importance of reduced fuel consumption and consequently less air pollution (which is becoming a danger to public health), must lead to the conclusion that a greater activity of the entire industry is called for in this field.

The battery-propelled ELECTROMOBILE will certainly be fully automatic, cause no air pollution, and be economic to run.

## IV.L.5. THE PRINTER'S

I am pleased to say that one of the latest electronic applications has played an important part in printing this book: a justifying computer in combination with a character display, known as the DIGISET.

It is customary to give all lines of a printed text the same length (justified columns). Formerly to achieve this, the composer having determined the end of a line, had to distribute the remaining space equally over the whole length of the line by inserting a suitable amount of "white". This is now done automatically by the DIGISET. From the initially endless text punched into a tape the computer determines what parts of the text will fit into the length of one line, by simply adding up the width of characters and spacings. Any of these parts may end in the last word or syllable. Breaking off a word correctly forms a difficult problem for the computer, so that a special programme must be built in. All parts of the text will be slightly shorter than the total length of one line. The computer divides the space left uniformly between the word spacings. The result is a second punched tape which contains the text in numbered lines of equal length.

As the text is put on tape in a "digital" form, it is now possible to use an electronically operated photographic composing machine which is the actual DIGISET. Here each character is stored in a memory as a raster of black or white signals. The digital code of the tape then selects the corresponding character as the "address" from the memory. This character in turn is projected on the screen of a cathode ray tube by a process of scanning as in television. The displayed character is very rapidly photographed after which the film moves one letter space further. When the tape indicates a new line, the film is brought into the required position. Thus the entire text is obtained on a film from which a "galley" proof is copied for correction, and finally the offset plate is prepared.

This machine can display about fifty characters per second. Since a definition of 100 lines is sufficient, the system operates slower than television where 625 lines are scanned in 1/25 second. It is, however, several times faster than the mechanical machine which must first select, or even cast, letter blocks.

Apart from the enormous speed of this method there is the great advantage that making corrections involves no manual labour; this in contrast with the mechanical method where characters or complete lines must be lifted out and replaced. The lines needing correction are punched into a tape together with their numbers. Again the computer is used to produce the corrected tape from the original tape plus the tape containing the corrections. The corrected tape is again processed by the digiset.

This is of great advantage to slovenly or fickle authors who during correction cannot refrain from making new changes in the original manuscript. It is not entirely without embarrassment that I must confess that that is one of the reasons why I have become an advocate of the digiset.

# CLOSING WORD

In Part IV of this book I have endeavoured to give a complete as possible survey of electronic applications. The ultimate conclusion is that this technique has contributed in many respects to making human life easier and more agreeable.

We did not go too deeply into military applications; it stands to reason that not a great deal is published on these matters. It should, however, be realized that electronics can be of vital importance in military defence. On the other hand, it can be used effectively in attack which then will become all the more destructive.

Let us express the hope that man will find the wisdom of using electronics for the benefit instead of the destruction of the world. Only then will the electronics engineer be able to regard his work with justifiable pride.

€ 15,- h